Investigating the utility of SPOT 5 imagery and Artificial Neural Networks, in the identification and mapping of *Acacia mearnsii* within environments of varying complexity.

by

Candice Russell

Submitted in fulfillment of the academic requirements for the degree of Master of Science in the discipline of Geography in the School of Applied Environmental Sciences, Faculty of Science and Agriculture

University of KwaZulu-Natal, Pietermaritzburg

November 2009

Declaration

This study was undertaken in fulfillment of a Geography Masters Degree and represents the original work of the author. Any work taken from other authors or organizations is duly acknowledged within the text and references chapter.

…………………………………..
Candice Russell

…………………………………..
Prof. Onisimo Mutanga
Supervisor

TABLE OF CONTENTS

List of Figures

List of Tables

Abstract

The impacts of invasive species on the environment, human health, and the economy continue to gain interest from public and private agencies, scientists, and the media. This study aimed to investigate the utility of SPOT 5 imagery and Artificial Neural Networks, in the identification and mapping of *Acacia mearnsii* within environments of varying complexity. Results showed that it is possible to identify and map *Acacia mearnsii* using SPOT 5 imagery, depending on the classification algorithm used. It was established that the neural network algorithms performed with greater success when compared to the performance of traditional classifiers, confirming other similar studies. The utility of the various classification algorithms was also investigated in terms of their applicability to environments of varying complexity. The neural networks once again, proved to be more successful and performed well in both the complex and relatively simple environments, indicating the robustness of the neural network algorithm.

Acknowledgements

I would like to thank the following people who, over the course of the past two years, have helped me in some way to accomplish this study.

- Prof. Onisimo Mutanga for his guidance, patience and timely response to any issues encountered relating to this dissertation

- The staff of the cartography department who provided the SPOT 5 image data used

- Mr. Boyd Escott from KZN Wildlife who provided the data which allowed me to select my study sites

- Mom and Dad for making this dissertation possible, for their support, encouragement and generosity, even when it seemed these two years would never end

- Granny and Grandpa for all their support and generosity, always ready and willing to help in any way

- Keagan for always being there through all the ups and downs, for his guidance, encouragement and patience, no matter what!

- Finally, to all my friends in and out of the department

Chapter 1. INTRODUCTION

1.1 BACKGROUND

Acacia mearnsii, commonly known as Black Wattle, is a fast growing shrub or tree native to Australia (IUCN/SSC Invasive Species Specialist Group, 2006). The species was first introduced to South Africa in the 1860's for providing shade and windbreaks. As of the 1960's, a century after its introduction, the species covered an estimated 324 000 ha within South Africa (Nyoka, 2003). Today *Acacia mearnsii* covers an estimated 2.5 million ha within South Africa (de Wit *et.al*, 2001).

Acacia mearnsii is currently used both commercially and on a subsistence level within South Africa. Tannins, resins, thinners, adhesives, timber, charcoal, pulp and woodchips form the bulk of the commercial uses of Black Wattle, while rural communities use the trees as a source of building material and fuel (de Wit *et.al*, 2001; Nyoka, 2003; IUCN/SSC Invasive Species Specialist Group, 2006). The IUCN/SSC Invasive Species Specialist Group (2006) estimate commercial *Acacia mearnsii* plantations cover 130 000 ha of land in the KwaZulu-Natal and Mpumalanga provinces. Both commercial and subsistence uses for *Acacia mearnsii* have been identified as providing pathways through which the species may invade new locations, specifically through:

- Consumption and excretion – the seeds may be distributed by rodents or birds IUCN/SSC Invasive Species Specialist Group, 2006).
- Animals – the seed dispersal is believed to be aided by livestock IUCN/SSC Invasive Species Specialist Group, 2006).
- People – people collecting branches for firewood or building material may spread seeds IUCN/SSC Invasive Species Specialist Group, 2006).
- Transportation of habitat material – seed distribution through the movement of seed-contaminated soil (de Wit *et.al*, 2001).
- Water – seeds are spread readily down water courses (de Wit *et.al*, 2001).

The species is considered particularly invasive due to its ability to produce large quantities of seeds, the germination of which is often triggered by bush fires, as well as its large crown which shades other vegetation. *Acacia mearnsii* therefore out-competes and inevitably replaces indigenous vegetation, the environmental consequences of which are many. It may replace grass communities, reducing the carrying capacity of the land; increase rainfall interception and transpiration, causing a decrease in stream flow; destabilize stream banks and support a lower diversity of species (IUCN/SSC Invasive Species Specialist Group, 2006). KwaZulu-Natal Wildlife have labelled the advance of alien plant species as the most significant past and future threat to conservation in South Africa (de Wit *et.al*, 2001; Goodman, 2003). According to de Wit *et.al*, (2001) commercial plantations and stands of *Acacia mearnsii* in South Africa cause an estimated annual economic loss of $US 2.8 million by reducing surface run-off and water availability.

It has been estimated that within the next fifty years, South Africa will reach the limits of its usable freshwater resources. Clearing alien invasive plants within riparian zones is key to maximizing the water supply in South Africa and it is for this reason the '*Working for Water'* program was created by the Department of Water Affairs and Forestry. In order for this project to plan effectively however, detailed mapping of alien invasive tree distribution needs to be undertaken (Rowlinson *et.al*., 1999).

According to Tsai and Chen (2004), Remote Sensing provides a fast and cost effective means for identifying and mapping invasive plant species as opposed to field based investigations, especially on a large scale. It is for this reason that the application of remote sensing technology has received considerable interest in the field of plant invasion in recent years and is now widely used for collecting and processing data. Many attributes of remote sensing are beneficial in the detection, mapping and monitoring of plant invaders. For example, satellite imagery has been available for most of the world since 1972, meaning the multi-date nature of satellite imagery allows for the monitoring of dynamic landscape features and provides a means to detect land cover changes and therefore calculate rates of change (Joshi, *et.al*, 2004).

Significant volumes of literature exist on the application of Landsat images in the identification of alien invasive plants. One such case study is that of Cobbing (2006) where the use of Landsat ETM imagery was found to be a suitable data capture source of alien *Acacia* (*Acacia mearnsii and Acacia dealbata*) species for the Working for Water program, in terms of their approved mapping techniques for both supervised and unsupervised classification. Both the supervised and unsupervised classification results were compared to control data captured by Working for Water through differential Global Positioning Systems (DGPS). A comparison of the results was done using a ranking system based on variables identified by the Working for Water program, that is, the spatial and positional accuracy, time constraints and cost. Whether the results of the supervised or unsupervised classifications were more accurate varied according to the vegetation type in which the Acacia species were found. It was, however, found that the use of Landsat ETM would be able to enhance the success of the Working for Water program by improving the identification and mapping of the alien species. It was also suggested that the most suitable imagery would be that of the SPOT 5 sensor, as it is able to pan-enhance the image to a 2.5 meter spatial resolution.

SPOT 5 was launched on the 4th of May 2002. It has a sun-synchronous orbit and has a twenty six day orbit around the earth. The spectral bands available from SPOT 5 include two panchromatic bands (5 m) combined to generate a 2.5 m product, three multispectral bands (10 m), and one short-wave infrared band (20 m). The spectral range includes the green, red, near infrared and the short-wave infrared (Images Spot © Cnes , 2005).

Little literature exists on the use of SPOT 5 imagery in the identification of alien invasive vegetation. There has however, been much literature found on the application of SPOT for identifying vegetation in general, an example of which is the study by Pasqualini *et.al* (2005) where SPOT 5 imagery was used for mapping seagrasses. Supervised classifications by depth range were conducted on both the multispectral imagery (with a spatial resolution of 10 m) and the fused imagery (with a spatial resolution of 2.5 m). In comparing the overall accuracy between SPOT 2.5 and SPOT 10, it was found that both were accurate (between 73% and 96%). Although the SPOT 2.5 had a lower overall accuracy than that of SPOT 10, it proved useful in highlighting the patchiness of the seagrass.

The accuracy of a classification is highly dependent on the spatial resolution of the imagery being used (Ju *et.al*, 2005), as well as the nature of the environment being investigated. This is due to the fact that spatial resolution can determine how much generalization can occur within and around a feature: some features may be very detailed at one scale, but may be generalized at another (Lilllesand *et.al*, 2004). As the spatial resolution of an image becomes coarser, so the amount of spectral mixing increases (Ju *et.al*, 2005). A study performed by Markham and Townshend (1981) determined that the accuracy of a classification is governed by two factors; namely the amount of pixels falling on the boundary of features and spectral variation of classes. Mixed pixels, or the number of pixels that contain more than one class, increases as the spatial resolution of the image increases and therefore decreases as the resolution of an image becomes finer. The spectral variation within a class increases with the increase in the resolution of an image. The spectral separability of the classes is therefore reduced creating problems in determining the nature of the class (Allan, 2007). However, the accuracy of classification results depends not only on the resolution of the image but also the classification technique used.

Past experience dictates that traditional statistical techniques of image classification such as minimum distance, parallel piped and maximum likelihood, can fail to detect potential overlaps within image data and so inaccuracies in the classifications can become a problem (Linderman *et.al.*, 2004; Qiu and Jensen, 2004).

Many classification techniques have been suggested in an attempt to counter such issues, one of which is the Artificial Neural Network. According to Qiu and Jensen, (2004) Artificial Neural Networks have been used in the processing of multispectral images with far greater success in terms of accuracy when compared to that of the traditional statistical methods. The reason for this lies in the very nature of the Neural Network. A single neuron is able to simulate a multivariate linear regression model with no *priori* assumptions about the data distribution due to its non-parametric operation (Linderman *et.al*, 2004; Sunar Erbek *et.al*, 2004). These neurons are arranged in layers and perform as non-linear simulators. Neural Networks are also able to learn, making the classification objective. In addition, ancillary data can be used in the classification, shifting the focus to spatial elements within the image, while traditional statistical classifiers focus on the spectral information within the image (Sunar Erbek *et.al*, 2004; Linderman *et.al*, 2004; Qiu and Jensen, 2004; Mutanga and Skidmore, 2004; Mutanga and Skidmore, 2007).

In a study conducted by Mutanga and Skidmore (2004) the ability to integrate several transformed remote sensing data ‘layers’ into the prediction process was made possible through the use of Artificial Neural Networks was key. The study involved the mapping of grass nitrogen concentration in an African savannah environment with the use of integrated image spectroscopy and remote sensing in an attempt to better understand the distribution and feeding patterns of wildlife. Transformed data used in the study included continuum-removed absorption features and the red edge position integrated within a Neural Network. Overall, it was concluded that the Neural Network was able to perform better than a traditional multiple linear regression.

This study will investigate the potential of SPOT 5 imagery in the identification and mapping of an alien invasive species, namely *Acacia mearnsii.* Comparisons will also be made between the accuracies of traditional classifiers and neural networks, as well as between environments of varying complexity based on the spatial resolution of the SPOT 5 imagery (10m). SPOT 5 imagery has been used extensively in field of vegetation identification. Cobbing (2006) suggested SPOT 5 would be the most suitable sensor for such an application due to its high resolution. During this study SPOT (10m) was chosen due to the fact that it has been proven highly accurate in past vegetation based studies (Pasqualini *et.al*, 2005) and SPOT (5m) panchromatic imagery was inaccessible.

1.2 AIM AND OBJECTIVES OF THE STUDY

1.2.1 Aim

The aim of this study was to investigate the utility of SPOT 5 imagery and Artificial Neural Networks, in the identification and mapping of *Acacia mearnsii* within environments of varying complexity.

1.2.2 Objectives

Five objectives were set in order to meet the previously mentioned aim

1. Determine the environmental variables influencing the occurrence and distribution of *Acacia mearnsii* for use within an Artificial Neural Network.

2. Classify and map *Acacia mearnsii* using SPOT 5 images in combination with Artificial Neural Networks and traditional classifiers

3. Compare the accuracy of Artificial Neural Networks with that of traditional classifiers, namely maximum likelihood, minimum distance and parallel piped.

4. Determine whether the methods used in this study could be applied to environments of varying complexity based on the spatial resolution of the SPOT 5 imagery (10m).

5. Investigate the utility of remote sensing for environmental management in terms of identifying alien invasive vegetation on a large scale.

1.3 LOCATION OF THE STUDY SITES

In order to determine whether the methods used in this study could be applied to environments of varying complexity based on the spatial resolution of the SPOT 5 imagery (10m), two study sites were chosen. Study site 1 contains areas where Acacia mearnsii is used in plantation forestry, while Study site 2, according to information received from KZN Wildlife, contains a more natural disjointed distribution. Each study site covers an area of approximately 484 km^2 (see Figure 1.1).

1.3.1 Study Site 1

Study Site 1 occurs approximately 20 km to the north of Pietermaritzburg, in the Albert Falls area (29°29'30.742"S, 30°25'8.412"E). The mean annual precipitation for the area ranges between 747mm and 1389mm while the mean annual temperature ranges between 16.3°c and 17.9°c (GAEA Projects, 2001). The altitude ranges between 601m and 1600m, rising from the south east to the north west. The morphology of the area is considered to be that of undulating hills and lowlands, composed of shallow soils with minimal development (KwaZulu-Natal Data, 2008). According to the Acocks 'veld types of Southern Africa', the natural vegetation in the area is predominantly Ngongoni veld of the Natal mist belt, with small areas of Highland sourveld, Southern tall grassveld, and valley bushveld (Acocks, 1988). The principle land use in Study site 1 is forestry, while large stands of commercial agriculture are also found. Two conservancies are found within the Study area, namely; Albert Falls and Karkloof (KwaZulu-Natal Data, 2008).

1.3.2 Study Site 2

Study site 2 occurs in the Fort Nottingham area (29°24'50.891"S, 29°54'14.897"E). The mean annual precipitation here ranges between 871mm and 1017mm, while the mean annual temperature ranges between 14.2°c and 16.3°c (GAEA Projects, 2001). Altitude ranges between 1101m and 2300m, rising from the east to the west. The morphology of the area includes undulating hills and lowlands to the north east, high mountains to the west and low mountains to the south. The soils in the area have been described as shallow with minimal development and are generally found on hard weathering rock (KwaZulu-Natal Data, 2008). According to the Acocks classification (1988), the natural vegetation in the area is Highland sourveld with small areas of Southern tall grassveld. The principal land use in Study site 2 is commercial agriculture, however small stands of timber plantations are also found. Conservancies in the area include numerous state forests to the west (KwaZulu-Natal Data, 2008).

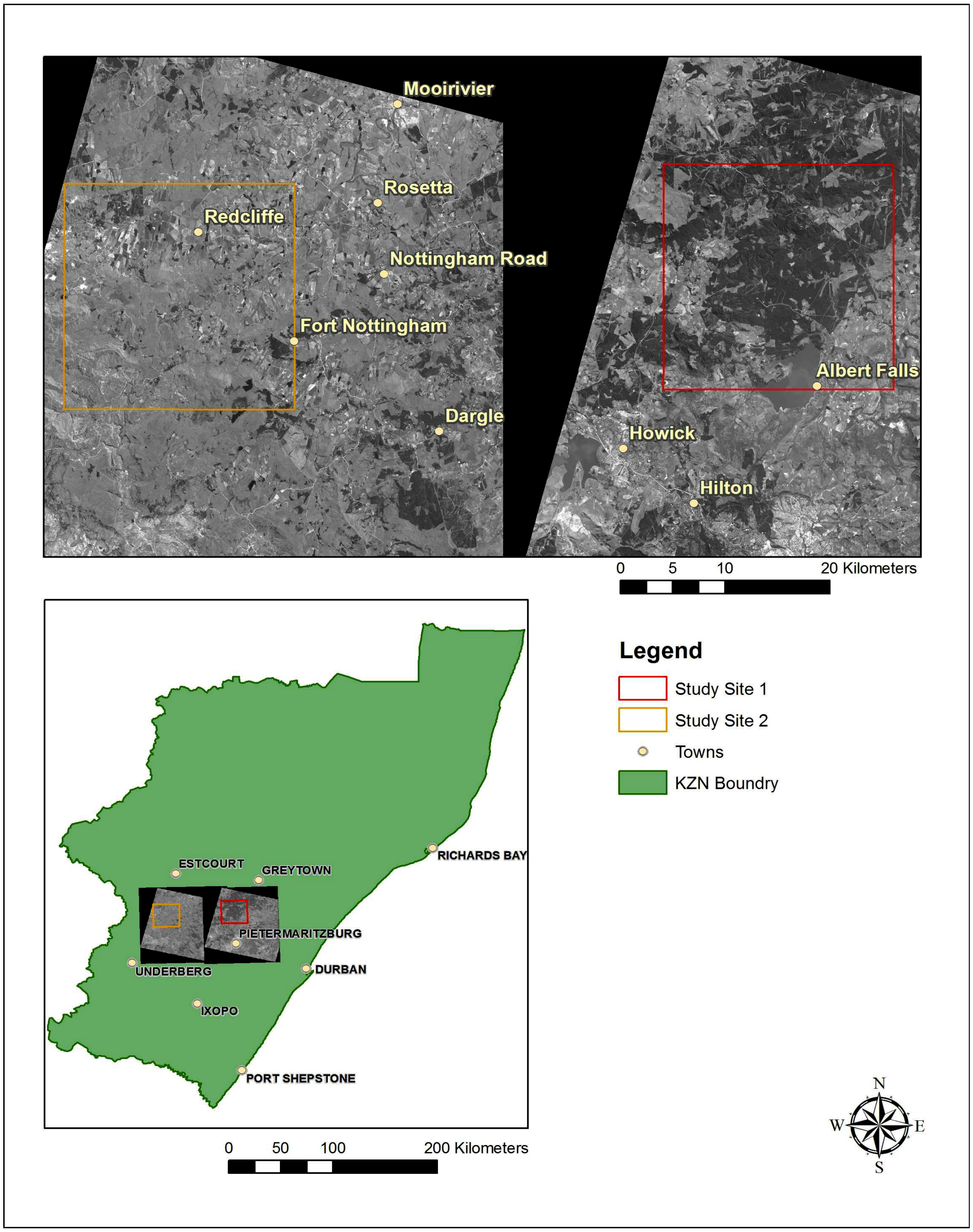

Figure 1.1. Location of the Study Sites in relation to the rest of KwaZulu-Natal

1.3.3 Acquisition of Imagery

The SPOT 5 images were provided by the Cartography Department of the University of KwaZulu-Natal, Pietermaritzburg. Two images were selected based on their proximity to the desired study sites. The image for Study site 1 was captured in April 2007, while the image for Study site 2 was captured in February of the same year. The images can be seen in Figure 1.1.

1.4 OUTLINE OF CHAPTERS

In Chapter 2, the nature and impact of alien invasive species will be discussed, with particular emphasis placed on *Acacia mearnsii*. The theory behind remote sensing and the various classification algorithms will then be explored. The final concept introduced is the application of remote sensing for environmental management in terms of identifying alien invasive vegetation on a large scale.

In Chapter 3, the methods used within this study to achieve the defined aim and objectives will be presented. The chapter is divided into three sections, namely; the baseline data collection and planning, the traditional classification process and finally the neural network classification process.

In Chapters 4 and 5, the results obtained during the study will be presented and discussed. Where possible, the data will be displayed in a tabular format. The data which is not displayed within the text is contained within the appendices.

Chapter 6 concludes the study, with the aim and objectives reviewed to assess the success of the study. Limitations of the study and recommendations will also be discussed.

Chapter 2. LITERATURE REVIEW

2.1 ALIEN INVASIVE SPECIES

Thousands of plant species from across the world have been introduced into South Africa for a variety of reasons including; agriculture, timber and firewood, aesthetics, stabilizing sand dunes and for barricading and hedges (van Wilgen, *et.al*, 2001; Nyoka, 2003). Many of these species have grown to become naturalized and are therefore able to thrive in the South African environment, becoming invasive. Alien invasive species are therefore able to survive, reproduce and spread free of natural predators, at considerable rates. This invasion of 'newly colonized' areas by alien organisms is a global threat of drastic proportions as it has serious implications for the environment (van Wilgen, *et.al*, 2008).

Large parts of South Africa still have natural or semi-natural vegetation. Even in areas where human activities have degraded or transformed habitats, goods and services provided by these ecosystems still play a significant role in human well-being. Through the altered functioning of ecosystems, alien invasive vegetation affects the capacity of the ecosystem to deliver goods and services. An example of an ecosystem service is the provision of water resources. Alien vegetation may affect this resource by invading upper watersheds, thereby reducing stream flow and consequently access to water downstream. Ecosystem goods, on the other hand, include food and fibre for consumption by humans and animals. Some examples of the way in which alien invasive vegetation may directly affect the production of such goods include; the invasion of palatable rangelands by unpalatable alien species which in turn reduce the grazing potential of the land; the displacement of indigenous flora suitable for flower harvesting; or the potential destruction of timber resources supplied by indigenous forests (Richardson and van Wilgen, 2004).

Yet, besides their obvious effect on agriculture, forestry and human health, alien invasive species are widely recognized as the second largest global threat (the first being direct habitat destruction), to biodiversity (Higgins *et.al*, 1999; Richardson and van Wilgen, 2004; Shackleton, *et.al* 2007).

In southern Africa, much of the concern relating to alien invasive species over the years has centred on the consequences for the conservation of biodiversity (van

Wilgen, *et. al*, 2001). According to Wynberg (2002) South Africa is the third most biologically diverse country in the world, containing between 250 thousand and 1 million indigenous species, many of which are endemic. However, alien invasive species pose a great threat to this natural resource (Wynberg, 2002).

South Africa's long colonial history, dating back over 350 years, and it's thriving agricultural and forestry sectors, have been identified by van Wilgen, *et.al* (2001) as having significantly contributed to the introduction, establishment and spread of alien invasive plant species within the country. Approximately 750 tree species and 8 000 shrubs, succulent and herbaceous species have been introduced into South Africa, 161 of which are considered highly invasive (van Wilgen, *et.al*, 2001; Nyoka, 2003). Several attempts have been made to establish the spatial extent of alien vegetation invasions in South Africa. Each study, however, used a different method and was conducted at a different time or focused on particular species or areas. For these reasons no conclusive source of data to produce a national overview of the extent of alien vegetation invasions in South Africa was found (van Wilgen, *et.al*, 2001).

In the past, biological invasions have received much attention from South African scientists and land managers in terms of the history, ecology and management of alien invasive species, yet little research has been conducted on the economic aspects and consequences thereof. The situation is, however, beginning to change as the consequences of these invasions are realized. An example of this realization was the revelation of the current and potential impacts of alien invasive trees and shrubs on South Africa's water resources (van Wilgen, *et.al*, 2001; Nyoka, 2003; Macdonald, 2004).

Numerous studies have been conducted in an attempt to assess the impact of alien invasive vegetation on water resources in South Africa. These studies largely concluded that alien invasive vegetation, including that of plantation forestry, have a significant negative effect on stream flow (Blignaut *et.al*. 2007). A recent nation-wide study attempted to assess the impact of all woody invaders on surface water resources in South Africa. The results of this study suggested that these plants may be using as much as 6.7 % of the total mean annual surface run-off or 9.95 % of the country's usable surface run-off (van Wilgen, *et.al*, 2001; Wynberg, 2002; Nyoka, 2003; van Wilgen *et.al*, 2008).

This break-through demonstrated the importance of intervention and led to the establishment of the Working for Water program (van Wilgen, *et.al*, 2001; Nyoka, 2003). The Working for Water program in South Africa is a Special Public Works Program (SPWP). The program is currently managed by the Department of Water Affairs and Forestry, which aims to clear alien invasive vegetation with the use of manual labour thereby increasing the country's water supply, improving biodiversity and the social and economic empowerment of the most poor and marginalized sectors of South African society (WFW annual report, 1998/ 1999; Wynberg, 2002; Theron *et.al.*, 2004). The Working for Water program initially only worked in watersheds and riparian zones, but now leads alien invasive vegetation management projects in all natural and semi-natural ecosystems (Richardson and van Wilgen, 2004; Turpie, *et.al*, 2008).

Between the Working for Water program's inception in September 1995 and 1 April 2000, the South African government spent over US $ 100 million in support of the initiative (van Wilgen, *et.al*, 2001). While alien invasions have considerable negative environmental and economic impacts, the South African government has used the opportunities generated by the need for such labour-intensive clearing projects to create a range of socio-economic benefits for the employees, including skills development, health care, childcare services, HIV/AIDS awareness and financial savings programs. By linking these social benefits with the clear environmental and economic advantages, the government is able to justify spending such an amount (van Wilgen, *et.al*, 2001; Buch and Dixon, 2008; Turpie, *et.al*, 2008).

Such negative environmental, social and economic impacts by alien invasive vegetation should, however, not be presented in a one sided manner. Nearly all the economically important crops in South Africa are in fact alien species although only a small percentage of these are invasive. In addition, certain alien invasive species hold considerable value, despite their negative impacts. Conflicts of interest therefore arise when significant commercial species become invasive and spread beyond control. An example of this can be seen in the forestry industry where alien plants provide firewood, timber, food, fodder, and where they have aesthetic or utilitarian value (van Wilgen, *et.al*, 2001; Wynberg, 2002; Nyoka, 2003; Shackleton *et.al*, 2007).

Plantation forestry is an important part of South Africa's economy, contributing US $ 300 million (2 %) to the GDP and employing over 100 000 people. Industries linked to forestry produce a further US $ 1.6 billion in the form of timber products, many of which are exported. The economic contribution of plantation forestry is clearly significant, however, a large proportion (38%) of the area invaded by woody alien species (*Pinus* and *Acacia* species in particular) in South Africa, is occupied by species used in commercial forestry. Forestry has therefore been identified as a major pathway through which alien invasive tree species may spread beyond control (van Wilgen, *et.al*, 2001; Wynberg, 2002; Nyoka, 2003).

The same argument has been applied on a small scale in terms of conflicts of interest as they apply to rural livelihoods in South Africa. Shackleton *et.al* (2007) stated that although the negative impacts of alien invasive vegetation on ecosystems are undisputed, the potential positive impacts thereof on rural livelihoods is less understood. This poses a problem as it is their land and water which is most affected by the presence of alien invasive vegetation.

According to Shackleton *et.al* (2007), there are three possible pathways in which alien invasive vegetation may be integrated into the livelihoods of rural communities. The first is communitys' introduction and acceptance of the value of the alien species. The initial introduction of the alien species is usually under controlled farm-like conditions, for example; plantations of tree species used for construction or firewood. Negative consequences occur when the alien species escapes and spreads into the surrounding environment, eventually undermining the livelihoods of those not benefiting from its presence.

The second situation involves the intentional introduction of the alien species into an area, from which it subsequently spreads beyond control. The alien invasive species may not be widely used at first but people in the community may 'switch' from using the indigenous vegetation to that of the alien species. This 'switch' can happen for a number of reasons, for example; the indigenous vegetation may become scarce and the alien vegetation becomes more accessible, or the community may simply want to make the most of the presence of the alien invasive species (Shackleton *et.al* 2007).

In the third situation, the community is forced to live with an alien invasive species which has no apparent use. In the early stages of invasion, the alien species poses little threat. However, as it increases in density and extent, the negative consequences become apparent as ecosystem goods and services, and daily activities are impacted upon, yet the community lacks the capital and, or techniques to control or eradicate it (Shackleton *et.al* 2007).

In each of these situations, rural communities are obligated to weigh up the negative impacts on the ecosystem and the potential positive benefits of using the alien invasive species (Shackleton *et.al* 2007). Such conflicts of interest, both on a rural scale and a national scale, need to be dealt with in a sensitive manner if progress is to be made in the reduction of negative impacts associated with the invasion of alien invasive vegetation (van Wilgen, *et.al*, 2001; Wynberg, 2002).

2.2 *ACACIA MEARNSII*

Acacia mearnsii is a fast growing leguminous tree native to Australia. It is an un-armed evergreen tree generally between 6 and 20m high, growing in dense clusters in riparian zones, grasslands, agricultural, urban, and disturbed areas (see Figure 2.1). The leaves are dark olive green bi-pinnate, finely hairy and crowded with raised glands occurring at and between the junctions of the pinnae pairs (see Figure 2.2). The flowers are pale yellow or cream occurring in globular heads, which appear in late July to October, while the fruits are presented finely hairy, dark brown pods and are usually constricted (Henderson, 2001; de Wit *et.al*, 2001; Nyoka, 2003).

Figure 2.1 Acacia mearnsii trees – photograph taken during the data collection within Study site 2

Figure 2.2 A close-up view of the Acacia mearnsii leaves and flowers (taken from IUCN/SSC Invasive Species Specialist Group, 2006)

In South Africa, *Acacia mearnsii* grows in altitudes between 600 and 1700 m and favours climates from warm temperate to moist tropical with a mean precipitation between 660 and 2280 mm per annum and mean annual temperatures between 14.7℃ and 27.8℃. Acacia mearnsii thrives on soils with a ph between 5.0 and 7.2. It does not, however, do well on very dry or infertile soils (Sherry, 1971; Nyoka, 2003; IUCN/SSC Invasive Species Specialist Group, 2006).

Acacia mearnsii, like many members of the *Acacia* species, is essentially a pioneer species. Unlike the African *Acacia* species, however, which facilitates the succession from grassland to savannah; the Australian *Acacia mearnsii* promotes the succession towards forest climax. *Acacia mearnsii* plays a dual role within a climax forest community. Firstly, as a component of the under-story, contributing to the canopy structure and, due to its leguminous nature, plays an important role in the nutrient cycle by providing nitrogen to the soil (Sherry, 1971).

Secondly, the *Acacia mearnsii* hard coated seeds are able to survive dormant for up to fifty years, when the forest is partially or completely destroyed by biotic factors or fire, ensures the continuation of the species and therefore the forest. Not only is the seed resistant to fire, but the germination of the *Acacia mearnsii* seed is actually stimulated by heat, making the species particularly well equipped to fulfil its primary ecological role as a pioneer species. Outside of Australia, its adaptability to a wide range of environmental conditions and its rapid growth rate makes *Acacia mearnsii* both the perfect agricultural resource and the perfect pest (Sherry, 1971)

The exact date of the first introduction of *Acacia mearnsii* into South Africa is unclear. According to Sherry (1971) the earliest account of *Acacia mearnsii* seed, introduced directly from Australia, was by Mr. Joseph Dicks of the settlement of Howick in 1846, when he purchased an assortment of *Acacia's* and Gums from both the Cape of Good Hope and Australia. Although this fact is little known, it is generally accepted that the species was first introduced to South Africa in the 1860's for providing shade and windbreaks (Nyoka, 2003). The commercial value of the timber and tannin provided by the species soon became clear, and plantations were established across South Africa (Nyoka, 2003; Shackleton *et.al*. 2007). Yet, from the time of its initial introduction into KwaZulu Natal, *Acacia mearnsii* had been more extensively grown in the Province than any other. Sherry (1971) highlighted this point, stating that the greatest concentration of plantation land occurred in the then Natal Midlands, Kranskop, Umvoti, New Hanover, Lions River, Pietermaritzburg, Camperdown, and Richmond areas, and represented an estimated 35 % of the total of the plantation land in South Africa at the time. Today *Acacia mearnsii* covers an estimated 2.5 million ha within South Africa (de Wit *et.al*, 2001).

Acacia mearnsii is currently used both commercially and at a subsistence level within South Africa. Tannins, resins, thinners, adhesives, timber, charcoal, pulp and woodchips form the bulk of the commercial uses of Black Wattle, while rural communities use the trees as a source of building material and fuel (de Wit *et.al*, 2001; Nyoka, 2003; IUCN/SSC Invasive Species Specialist Group, 2006). The IUCN/SSC Invasive Species Specialist Group (2006) estimate that commercial *Acacia mearnsii* plantations cover 130 000 ha of land in the KwaZulu-Natal and Mpumalanga provinces. Both commercial and subsistence uses for *Acacia mearnsii* have been identified as providing pathways through which the species may invade new locations, specifically through:

- Consumption and excretion – the seeds may be distributed by rodents or birds IUCN/SSC Invasive Species Specialist Group, 2006).
- Animals – the seed dispersal is believed to be aided by livestock IUCN/SSC Invasive Species Specialist Group, 2006).

- People – people collecting branches for firewood/ building material may spread seeds IUCN/SSC Invasive Species Specialist Group, 2006).
- Transportation of habitat material – seed distribution through the movement of seed-contaminated soil (de Wit *et.al*, 2001).
- Water – seeds are spread readily down water courses (de Wit *et.al*, 2001).

The species is considered particularly invasive due to its ability to produce large quantities of seeds, the germination of which is often triggered by bush fires, as well as its large crown which shades other vegetation. *Acacia mearnsii* therefore out-competes and inevitably replaces indigenous vegetation, the environmental consequences of which are many. Such environmental consequences include: the loss of biodiversity; the disruption of important ecosystem processes; the threat to conservation areas and agricultural land. Equally important is the invasion of watercourses which result in deep channelling and a reduction in access to the watercourse. This situation is further aggravated by the density of *Acacia mearnsii* stands which could result in an increase in fire intensity and damage (de Wit *et.al*, 2001). It is for this reason; *Acacia mearnsii* was ranked as the world's worst invader in the 'Global Invasive Species Database' funded by La Fondation TOTAL. The invasion of *Acacia mearnsii* also affects the economy and rural livelihoods by reducing the amount of available water by an estimated 577 million cubic meters annually within South Africa, which corresponds to an economic loss of approximately US $ 1.4 billion each year (de Wit *et.al*, 2001; IUCN/SSC Invasive Species Specialist Group, 2006).

With an average precipitation of approximately 500 mm per annum, South Africa falls well below the global average of 860 mm per annum. Factors such as the increasing rainfall irregularity and intensity, as well as the lack of ground water reserves due to the hard underlying rock, have resulted in the classification of South Africa as a country with chronic water scarcity (Blignaut *et.al*. 2007). It has been estimated that within the next fifty years, South Africa will reach the limits of its usable freshwater resources. Clearing alien invasive plants within riparian zones is therefore key to maximizing the water supply in South Africa (Rowlinson *et.al*., 1999).

The invasion of *Acacia mearnsii* however, is not without benefits. In a study conducted by Shackleton *et.al* (2007) on the effect of alien invasive species on rural livelihoods in the Eastern Cape, it was found that almost all rural households collected *Acacia mearnsii* for fire wood and for building or fencing poles. In most cases, *Acacia mearnsii* was collected due to its proximity to the community and the government restrictions on indigenous trees. Interviews conducted during the study confirmed that *Acacia mearnsii* was generally viewed as a year round resource and was readily available. There were however, areas where *Acacia mearnsii* was deemed undesirable such as homesteads, grazing areas, riverbanks, and sacred pools, due to the reduction in the productivity, cultural heritage, or safety. Additional costs to the community included the labour required to continuously remove the new *Acacia mearnsii* saplings from arable land, and the reduction of the carrying capacity of grazing lands (Shackleton *et.al*. 2007).

The effect of *Acacia mearnsii* on rural livelihoods is complex. Some households make extensive use of the species, while others do not; some use it to generate income and others only turn to *Acacia mearnsii* in times of particular need. Yet, the majority of the households interviewed during the study stated that they would welcome greater densities of the species due to its direct use or income potential. It is therefore clear that the difference between *Acacia mearnsii* being a ‘pest’ or a ‘resource’ is simply a matter of perspective and scale (Shackleton *et.al*. 2007).

Yet the problem still exists; on a national scale, the costs of *Acacia mearnsii* outweigh the benefits of the commercial *Acacia mearnsii* industry together with the use by rural communities in South Africa by more than 2.5 times (Shackleton *et.al* 2007). For this reason; the Working for Water program is focused on the active management, control and eradication of *Acacia mearnsii* and other invasive alien species, while at the same time remaining sensitive to rural communities’ reliance there upon (Theron *et.al*., 2004).

In terms of management and control in South Africa, *Acacia mearnsii* is considered to have the invasive status of a ‘transformer’. Transformers are species which are able to dominate or replace any canopy or sub-canopy layer of a natural or semi natural ecosystem, altering its structure, integrity and functioning. This status is reserved for the most detrimental weeds in the country and as such, *Acacia mearnsii* has been declared a category 2 invader (Henderson, 2001). Category 2 invaders are subject to certain regulations defined under the ‘Conservation of Agricultural Resources Act’ (Act No. 43 of 1983) and the amended Act published in the Government Gazette vol. 429 No. 22166 of March 2001. These regulations state the following:

- The species is only allowed in demarcated areas under controlled conditions.
- The importation of the species' propagative material and the trading thereof may only be conducted by permit holders.
- The species must be controlled or eradicated where possible outside demarcated areas.
- The species may not occur within 30m of the 1:50 year flood-line of watercourses and wetlands unless authorization has been granted.

Where *Acacia mearnsii* occurs contrary to these regulations, the land user is obliged to take measures to control or eradicate the species. Control measures vary according to the environmental, social, and economic factors involved in the removal, however three broad categories can be applied, namely; preventative, integrative and biological management.

Integrative management is currently used by the Working for Water program. It is a short term solution involving the combination of physical and chemical eradication techniques (van Wilgen *et.al*. 2001; Milton *et.al*. 2003). Control programs involving physical and, or chemical techniques are, however, labour intensive, expensive and in most cases, are simply short term solutions and need to be applied indefinitely. Biological control agents, on the other hand, have the potential to be sustainable solutions (van Wilgen, *et.al*, 2001). In a study conducted by de Wit *et.al* (2001), a broad cost/benefit analysis associated with Acacia mearnsii in South Africa, found that the most feasible and cost-efficient management plan would be to combine biological control with physical control. Yet no matter the method of control of management, effective planning is crucial, and detailed mapping of alien invasive tree distribution therefore needs to be undertaken (Rowlinson *et.al*., 1999).

2.3 REMOTE SENSING

According to Lillesand *et.al* (2004) remote sensing is the course of action through which information about an object, area or phenomenon can be obtained using data acquired by a device that is not in contact with the subject under investigation. Congalton and Green (1999) emphasize the range of remote sensing techniques be by noting the most basic remote sensing devices as the eyes and ears. However, for the purpose of this research, the term remote sensing will refer solely to satellite remote sensing and will therefore not include other platforms.

According to Tsai and Chen (2004), remote sensing provides a fast and cost effective means for identifying and mapping invasive plant species as opposed to field based investigations, especially on a large scale. The cost of remote sensing is related to economies of scale; the larger the survey area, the cheaper the cost per acre (Everitt et.al, 2005). It is for this reason the application of remote sensing technology has received considerable interest in the field of plant invasion in recent years and is now widely used for collecting and processing data (Tsai and Chen, 2004).

Cobbing (2006), for example, investigated the use of Landsat Enhanced Thematic Mapper (ETM) imagery as a suitable data capture source of alien *Acacia* (*Acacia mearnsii and Acacia dealbata*) species for the Working For Water program, in terms of their approved mapping techniques for both supervised and unsupervised classification. Both the supervised and unsupervised classification results were compared to control data captured by Working For Water through differential Global Positioning Systems (DGPS). The comparison of the results was done using a ranking system based on variables identified by the Working For Water program, that is, the spatial and positional accuracy, time constraints and cost. The results of whether the supervised or unsupervised classifications were more accurate varied according to the vegetation type in which the Acacia species were found. It was however, found that the use of Landsat ETM would be able to progress and enhance the success of the Working For Water program by improving the identification and mapping of the alien species.

Many attributes of remote sensing are beneficial in detecting, mapping and monitoring of plant invaders (Tsai and Chen, 2004). For example, satellite imagery has been available for most of the world since 1972, meaning the multi-date nature of satellite imagery allows for the monitoring of dynamic landscape features and therefore provides a means to detect land cover changes and therefore calculate rates of change (Joshi, *et.al*, 2004).

There are numerous satellite image systems currently available such as Landsat, MODIS, IKONOS and NOAA. However, for the purpose of this research SPOT 5 was chosen given its applicability (in terms of spectral bands and resolution) and accessibility.

2.3.1 SPOT Image Systems

In 1978, the French government developed the 'System Pour L'Observation de la Terre', also known as the 'SPOT' program, and with the aid of Sweden and Belgium, was able to launch the first series of SPOT earth observation satellites. SPOT subsequently developed into a large scale international program with ground receiving stations and data distribution outlets in more than thirty countries around the world. The first satellite, SPOT 1, was launched on board an Ariane launch vehicle in February of 1986, beginning a new era in space remote sensing. This satellite was the system chosen to include a linear array sensor and employ a pushbroom scanning technique. In addition, it boasted pointable optics, enabling side-to-side off-nadir viewing capabilities. This allowed for full-scene stereoscopic imaging from two different satellite tracks permitting coverage of the same area (Lillesand *et.al*, 2004).

SPOT 1 was retired from fulltime service in 1990 and was followed by SPOT 2; 3; 4 and 5 (Lillesand and Kiefer, 2004). SPOT 5 was launched in May 2002, trouncing its predecessors with its increased panchromatic image resolution and multispectral image resolution of 100 %, from 5m to 2.5m, and from 20m to 10m respectively (Cobbing, 2006). It has a sun-synchronous orbit and a twenty six day orbit around the earth (Images Spot © Cnes, 2005). Table 1 provides a summary of the SPOT 5 sensor characteristics.

SPOT 5	
Spectral bands and resolution	2 panchromatic (5m) combined to generate 2.5m product 3 multi-spectral (10 m) 1 short-wave infrared (20m)
Spectral range	P: 0.48 – 0.71 µm B1 (green) 0.5 – 0.59 µm B2 (red) 0.61 – 0.68 µm B3 (near infrared) 0.78 – 0.89 µm B4 (short-wave infrared) 1.58 – 1.75 µm
Imaging swath	60 km by 60 – 80 km

Table 2.1 Summary of SPOT 5 sensor characteristics (adapted from Images Spot © Cnes, 2005)

SPOT 5 imagery has been used extensively in field of vegetation identification. Cobbing (2006) suggested SPOT 5 would be the most suitable sensor for such an application due to its high resolution. An example of which can be seen in a study conducted by Pasqualini *et.al* (2005), where SPOT 5 imagery was used for mapping seagrasses. Supervised classifications by depth range were conducted on both the multispectral imagery (with a spatial resolution of 10 m) and the fused imagery (with a spatial resolution of 2.5 m). In comparing the overall accuracy between SPOT 2.5 and SPOT 10, it was found that both were very accurate (between 73% and 96%). Although the SPOT 2.5 had a lower overall accuracy than that of SPOT 10, it proved useful in highlighting the patchiness of the seagrass.

The accuracy of a classification is highly dependent on the spatial resolution of the imagery being used (Ju *et.al.*, 2005), as well as the nature of the environment and features being investigated (Rocchini, 2007). This is due to the fact that spatial resolution can determine how much generalization can occur within and around a feature: some features may be very detailed at one scale, but may be generalized at another (Lilllesand *et.al*, 2004). As the spatial resolution of an image becomes coarser, so the amount of spectral mixing increases (Ju *et.al.*, 2005). A study conducted by Markham and Townshend, (1981) determined that the accuracy of a classification is governed by two factors; namely the amount of pixels falling on the boundary of features and spectral variation of classes. Everitt *et.al*., (2005) stated the imagery resolution required to map an invasive plant species is dependent on the smallest patch of the species to be mapped and recommend that the pixel size should be at least one forth the area of the smallest patch (see Figure 2.2).

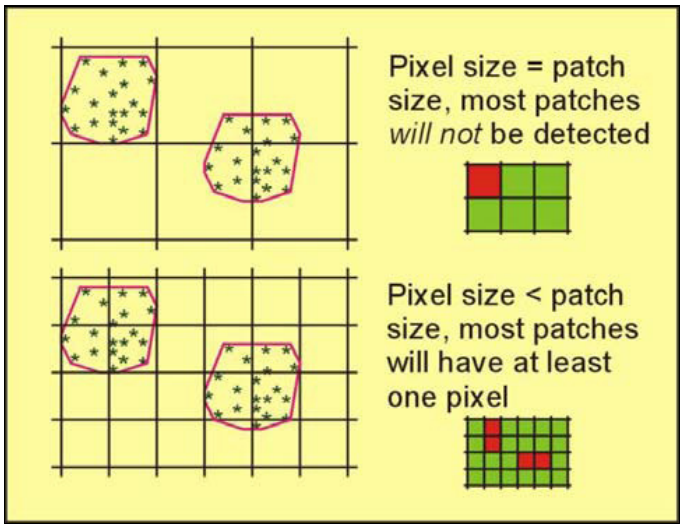

Figure 2.3 Spatial resolution and the area of the features to be mapped (after Everitt et.al, 2005)

Mixed pixels, or the number of pixels that contain more than one class, increases as the spatial resolution of the image increases and therefore decreases as the resolution of an image becomes finer. The spectral variation within a class increases with the increase in the resolution of an image. The spectral separability of the classes is therefore reduced creating problems in determining the nature of the class (Allan, 2007).

In a study conducted by Rocchini (2007) attempts were made to predict plant species richness (spatial heterogeneity) by means of spectral heterogeneity using remotely sensed imagery. Species richness is a fundamental indicator of ecosystem biodiversity at the local and regional scales. The use of spectral variance of satellite images for predicting species richness is known as 'Spectral Variation Hypothesis'. The study focused on the effects of spatial and spectral scale on the spectral and spatial heterogeneity relating to species richness. In order to do so, satellite images of different spatial and spectral resolutions were used, namely; Quickbird, Aster, ortho-Landsat ETM+, and a resampled 60m Landsat ETM+. The spectral variability was found to be both scene and sensor dependent while the coarser resolution data tended to have mixed pixel problems and was therefore less sensitive to spatial complexity. Conclusions were drawn that using a finer dimension would result in a higher level of detail, while the spectral response from different Landcover features in images with higher spectral resolution would exhibit a higher complexity.

Nagendra and Rocchini (2008), agreed, to a certain extent, although clarifying under what conditions such a statement is true. Nagendra and Rocchini (2008) found that hyperspatial data, with its increased pixel resolution is best suited for facilitating the accurate location of features (eg. Tree canopies), but less suited to the identification of aspects such as species identity, particularly when spatial resolution becomes too fine and the pixels are smaller than the size of the object being identified. This relates back to the fact that the cost of remote sensing is based on economies of scale. In order for remote sensing to be a fast and cost effective means for identifying and mapping features, the spatial resolution of the imagery to be used, must be chosen based on the size of the features being studied (Cao and Lam, 1997).

2.4 DIGITAL IMAGE PROCESSING

Lillesand *et.al* (2004), define digital image processing as the manipulation and interpretation of digital images with the aid of a computer. Digital image processing is a very broad subject and has therefore been divided into five broad categories of computer assisted operations, namely; image rectification and restoration, image enhancement, image classification, data merging and GIS integration, and biophysical modelling. These categories are interlinked and follow on from one another, although briefly discussed as separate entities in the following section for ease of reference.

2.4.1 Image rectification and restoration

This is the process through which distorted or degraded image data is corrected to create a true representation of the scene in reality. It typically involves the initial correction of geometric distortions, radiometric calibration and the elimination of noise present within the data. Image rectification and restoration processes usually take place before manipulation and analysis of the data, and are therefore often referred to as “pre-processing” techniques (Lillesand *et.al*, 2004). Often, however, as in this case, image rectification and restoration is conducted prior to the image being made available for public use.

2.4.2 Image Enhancement

Image enhancement techniques are applied to an image and affect how the image is displayed in an attempt to better visual interpretation. These techniques typically involve increasing the visual distinction between features within the image. The objective is to create 'new' images derived from the original image to increase the amount of information which can be visually interpreted. There is no simple set of rules which can be followed in order to produce the best image for any particular application and a 'trial and error' approach is usually required. Several enhancements are therefore often necessary before the desired result can be obtained (Lillesand *et.al*, 2004).

2.4.3 Image Classification

The purpose of image classification is to automatically identify features within an image using quantitative techniques. To put it simply, this typically involves the analysis of multispectral data and the application of statistically based decision rules, resulting in the determination of land cover identities for each pixel within the image (Lillesand *et.al*, 2004).

However, with the imagery now available, containing higher frequency information and a greater number of bands, the risk of additional redundancy is increased, complicating image processing (Qiu and Jensen, 2004). Traditional statistical techniques of image classification such as minimum distance, parallel piped and maximum likelihood, can therefore fail to detect potential overlaps within the data and so inaccuracies in the classifications can become a problem (Linderman *et.al.*, 2004; Qiu and Jensen, 2004). The reason for this can be seen in the way in which these traditional classifiers function:

- Minimum distance to means

The minimum distance to means classification technique calculates the distance between a pixel and the closest class mean, classifying that pixel accordingly (Lillesand *et.al*, 2004).

- Parallel piped

The parallel piped classification technique uses a set of digital number ranges to create "boxes" that define the classes of the classification, classifying those pixels accordingly (Lillesand *et.al*, 2004).

- Maximum likelihood

The maximum likelihood classification technique evaluates the probability of a pixel occurring within a class, if the probability is high, the pixel is classified accordingly, if it is low, the classification process continues until it is classified (Lillesand *et.al*, 2004).

Linderman *et al*., 2004, for example, found that the complex under-story of a forest environment can have a detrimental effect to the reflectance properties of the canopy and so cause inaccuracies in the classification of forest canopies. When using traditional classifiers, problems arose when pixels fell out of the system defined parameters, particularly in areas with high spectral variability or where features had different linear and non-linear contributions to the reflectance. The traditional classification techniques were therefore found to inadequate and other methods of classification had to be explored. Artificial neural networks were subsequently chosen to replace the traditional classification techniques and proved invaluable in the accurate classification of complex and variable features.

Artificial neural networks are essentially mathematical models resembling brain activity (Sunar Erbek *et.al*, 2003). They are made up of many components, known as neurons. A single neuron is able to simulate a multivariate linear regression model with no *priori* assumptions about the data distribution due to its non-parametric operation (Linderman *et.al*, 2004; Sunar Erbek *et.al*, 2004). These neurons are arranged in layers and perform as non-linear simulators. Neural Networks are also able to learn from set parameters for the classification, making it possible for the classification of complex datasets (Linderman *et.al*, 2003 and Qiu and Jensen, 2002). There are two ways in which artificial neural networks are able to learn; supervised and unsupervised. Supervised learning occurs when the final desired output values are known and used during the training of the network, while unsupervised learning occurs when the final output values are not known and therefore not used during the training of the network (Sunar Erbek *et.al*, 2003).

An artificial neural network consists of an input layer and an output layer with nodes and connectors linking the two. These nodes act as connectors within the system, allowing data to be distributed and encoded through the neurons. The connections within a neural network can either be unidirectional or multidirectional. Networks with a unidirectional flow, be it inter-nodal or intra-nodal, are known as 'Feedforward' neural networks, while networks with a multidirectional flow are known as 'Recurrent' neural networks (Allan, 2007; Mutanga and Skidmore, 2003).

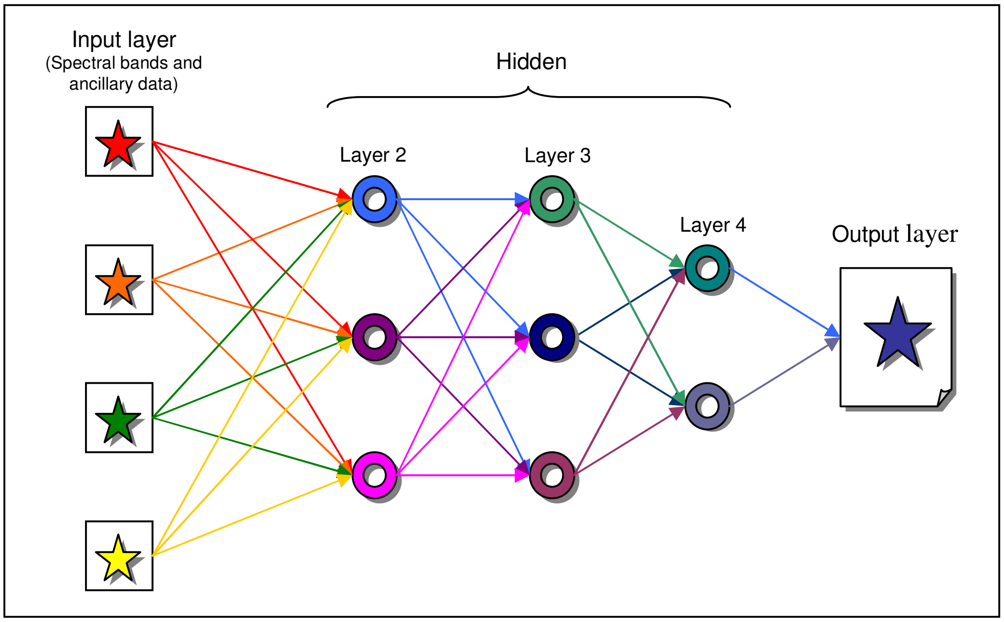

Figure 2.4. Example of a Feedforward neural network (after Allan, 2007; Kazoglu and Mather, 2003)

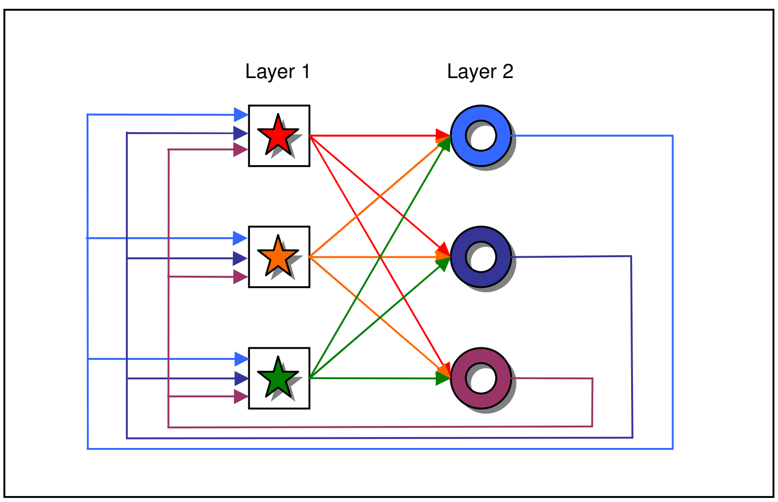

Figure 2.5. Example of a recurrent neural network (after Allan, 2007)

One of the most commonly used Feedforward neural networks is the Multilayer Perceptron (MLP) model (Foody, 2003; Kazoglu and Mather, 2003; Allan, 2007). The MLP is comprised of three layers, namely; input, hidden and output, with each containing nodes. Each node is connected by a user defined weighted function and can not be connected to another node within the same layer. The purpose of a node is to perform a simplified mathematical algorithm (Foody, 2003). Data from spectral bands and ancillary information are fed into the model through the input layer and passed on to the hidden layers. The hidden layers subsequently perform the mathematical analysis on the data, creating the output layer containing the specific classes for classification (Kazoglu and Mather, 2003).

MLP models are considered to be ‘supervised’ models, as information concerning classes and information involved in the model are already known, and so ‘taught’ to the network (Foody, 2003; Kazoglu and Mather, 2003). One of the most popular ‘teaching’ techniques is the backpropagation learning algorithm (Kazoglu and Mather, 2003; Allan, 2007).

The backpropagation learning technique aids in the accuracy of the final result of the model (Foody, 2003; Kazoglu and Mather, 2003; Allan, 2007). Backpropagation works in two phases: phase 1 involves running the model with the initial network weights in place, resulting in estimated output values for each set of input data. Phase 2 commences once the values are estimated and can be compared to known values. The differences between the estimated and known values are calculated and sent backwards through the network, where the weights between the nodes are adjusted. The entire process is then repeated until the error is negligible or non-existent (Foody, 2003; Kazoglu and Mather, 2003, Allan, 2007).

Another type of Feedforward neural network is the Radial Basis Function or RBF. This model shares aspects of the MLP in terms of containing specific layers connected by weighted functions. What differentiates them, however, are the hidden layers. The MLP is able to contain an infinite number of hidden layers, while the RBF is able to contain only one. The single hidden layer of the RBF contains a statistical function which is specifically applied to a small area within a defined input layer. The statistical function calculates the location of a specific point, or designated centre, within the input data. Deviations by input layer vectors from the designated centre are then calculated. The layer is subsequently divided into zones with points closest to the centre of the designated radial point having a maximum value of 1.0 and those furthest, having a value of 0.0 (Foody, 2003; Allan, 2007).

However, whether using MLP or RBF, Feedforward or Recurrent, neural network algorithms are highly complex, making it difficult to document how specific classification decisions are reached. It is for this reason the neural network is often seen to be a 'black box' (Qiu and Jensen, 2002). It has therefore often been suggested that the derivation of many of the values used in a successful classification stems from trial and error (Kazoglu and Mather, 2003). There are, however, several factors which are known to affect the accuracy of a neural network and can therefore be implemented to lessen the amount of 'trial and error', such as the number of connections and layers within a network.

The number of connections and layers within a network plays a role in the accuracy of the classification, as well as generalization and classification of the pixels not within the supervised data set. A greater number of connections positively affect the accuracy in classifying pixels within the supervised data set, while negatively affecting the ability of the network to generalize and classify pixels not within the supervised data set (Foody and Arora, 1997).

Four possible factors affecting the classification accuracy within a neural network have been identified by Froody and Arora (1997) and cited by Allan (2007), they are as follows:

- The structure of the network

The structure of the network refers to the number of layers, including hidden layers and nodes present within the network. It is generally accepted that a more complex network will yield a greater accuracy in classification when compared to smaller less complex networks. Although the structure alone can not be seen as a determining factor in the accuracy of a network, it can be highly influential when combined with other factors (Foody and Arora, 1997).

- The size of the training set

Lillesand *et.al*, (2004) highlight the importance of obtaining a representative sample of the required classes in accurate classifications, be they statistical or neural network. It has, however, been suggested that a neural network can still be accurate without as many training datasets as a statistical classifier due to the fact that a neural network makes no assumptions about the distribution of the samples (Foody and Arora, 1997; Qiu and Jensen, 2002). It has been shown that a greater number of samples can positively affect the accuracy of the classification, although it increases the amount of time required to run the classification.

- Discriminating variables

Discriminating variables allow the system to distinguish between classes based solely on the characteristics of the data. Traditional statistical classifiers use the discriminating variable known as ‘spectral separability’, meaning classes are distinguished based on the spectral bands within an image (Foody and Arora, 1997; Lillesand *et.al*, 2004). Within neural networks, however, the discriminating variables are not limited to spectral separability, but rather incorporate both spectral bands and ancillary data. Examples of ancillary data include; aspect, slope, soil type, and rainfall, to name a few (Foody and Arora, 1997; Lillesand *et.al*, 2004; Mutanga and Skidmore, 2004).

In a study conducted by Mutanga and Skidmore, (2004) the ability to integrate ancillary data into the classification process with the use of Artificial Neural Networks was key. The study involved the mapping of grass nitrogen concentration in an African savannah environment with the use of integrated image spectroscopy and remote sensing in an attempt to better understand the distribution and feeding patterns of wildlife. Ancillary data used in the study included continuum-removed absorption features and the red edge position integrated within a Neural Network. Overall, it was concluded that the Neural Network was able to perform better in terms of classification than a traditional multiple linear regression due to the use of ancillary data.

- Data characteristics

The final factor identified by Foody and Arora (1997) involves the testing of data characteristics, where the final output is compared to real life scenarios in terms of accuracy. A confusion or error matrix is usually used to compare values from the classification to known values from reality, by plotting what was classified correctly versus what was not in the form of a matrix. Ideally, samples representing the statistical representation of each class should be included when determining the accuracy of the classification in such a matrix (Foody and Arora, 1997; Lillesand *et.al*, 2004). According to Lillesand *et.al*, (2004), the image classification process is followed by data merging and GIS.

2.4.4. Data merging and GIS

Data merging and GIS are used to combine image data for a specific area with other geographically referenced data for the same area. The purpose of data merging is to integrate remotely sensed imagery with other sources of information within a GIS. Examples of other sources of information include soil, topography, ownership, zoning and assessment data, to name but a few. Many applications of digital imaging processing are enhanced through the merging of multiple data sets (Lillesand *et.al*, 2004).

2.4.5. Biophysical modelling

The purpose of biophysical modelling is to quantitavely relate the remotely sensed digital data to biophysical features and phenomena in reality. Biophysical models are therefore intended to simulate the functioning of environmental systems in a spatially accurate manner, and to predict their behaviour under user defined 'what if' conditions. Biophysical models have improved the understanding of environmental systems for decades and, with increasing remote sensing and GIS capabilities, have allowed for highly accurate representations of the behaviour of environmental systems. Remote sensing and GIS therefore play an increasingly important role in effective resource management, environmental risk assessment, and predicting and analysing the impacts of global environmental change (Lillesand *et.al*, 2004).

2.5 LESSONS LEARNT

The impacts of invasive species on the environment, human health, and the economy continue to gain interest from public and private agencies, scientists, and the media. Species invasions threaten endangered species, and pose the second largest threat to biodiversity after habitat loss (Higgins *et.al.*, 1999; de Wit, *et.al*, 2001; Goodman, 2003). Although prevention is better than a cure; numerous pathways of invasion make the interception of all species unrealistic. Organizations and individuals attempting to manage invasive species often have limited budgets and insufficient information and tools. According to Barnet *et.al* (2006), the answer to this is efficient management tools which address the multiple phases of plant invaders on a large scale. Early detection and the development of a spatially explicit management strategy is therefore key to improving efforts to prevent the establishment of invasive species (Higgins *et.al.*, 1999; Barnet *et.al*, 2006).

To understand alien plant invasions at local and regional scales, organizations and programs such as Working for Water have developed systems for mapping and compiling non-native plant species information (Cobbing, 2006). Mapping has the potential to record which, how much, and where alien invasive plant species exist on a landscape. Advantages when mapping is implemented over time include the ability to monitor patches of weeds, help predict the spread of species, facilitate the exchange of data between agencies, and increase public and political awareness (Barnet *et.al*, 2006).

Remote sensing provides a fast and cost effective means for identifying and mapping invasive plant species as opposed to field based investigations, especially on a large scale. It is for this reason the application of remote sensing technology has received considerable interest in the field of plant invasion in recent years and is now widely used for collecting and processing data (Tsai and Chen , 2004; Mas and Flores, 2008).

Although remote sensing is an efficient tool for monitoring the Earth at low cost and in a short time, when a strict accuracy assessment is made (based on an unbiased sample and independent classification of verification sites), the results are often disappointing (Mas and Flores, 2008). This is due to the fact that with the imagery now available, containing higher frequency information and a greater number of bands, the risk of redundancy is increased, complicating image processing (Qiu and Jensen, 2004), making any improvement in the methods of classification and analysis crucial (Mas and Flores, 2008). Traditional statistical techniques of image classification such as minimum distance, parallel piped and maximum likelihood, can fail to detect potential overlaps within the data and so inaccuracies in the classifications can become a problem (Linderman *et al*., 2004; Qiu and Jensen, 2004). Since the beginning of the 1990s, artificial neural networks (ANN's), also known as neural networks, have been applied to the analysis of remote sensing images with promising results (Mas and Flores, 2008).

According to Qiu and Jensen, (2004) Artificial Neural Networks have been used in the processing of multispectral images with far greater success in terms of accuracy when compared to that of the traditional statistical methods. The reason for this lies in the very nature of the Neural Network. A single neuron is able to simulate a multivariate linear regression model with no *priori* assumptions about the data distribution due to its non-parametric operation (Linderman *et.al*, 2004; Sunar Erbek *et.al*, 2004). These neurons are arranged in layers and perform as non-linear simulators. Neural Networks are also able to learn, making the classification objective. In addition, ancillary data can be used in the classification, shifting the focus to spatial elements within the image, while traditional statistical classifiers focus on the spectral information within the image (Sunar Erbek *et.al*, 2004; Linderman *et.al*, 2004; Qiu and Jensen, 2004; Mutanga and Skidmore, 2004; Mutanga and Skidmore, 2007).

In a study conducted by Joshi, *et.al*, (2006), remote sensing was successfully applied using an indirect method to map the seed production of *Chromolaena odorata*, one of the world's 100 worst invader species. The study was conducted in the lowland *Shorea robusta* forest in Nepal, where the *Chromolaena* had invaded the understory of the forest. Landsat ETM+ imagery and a neural network algorithm was used to determine the forest canopy density as well as the light intensity reaching the understory. The resulting models were then inverted, allowing the prediction of *Chromolaena* seed productivity. Predicted Chromolaena cover and seed production were then combined, producing a map displaying the total seed production per unit area. Joshi *et.al*., (2006) suggested such a map could be used to significantly reduce the costs associated with the control of alien invasive species by providing information on the spatial segregation of source and sink populations, proving particularly useful when implementing control measures under circumstances of limited capital and manpower, as is the situation within South Africa.

Little literature exists on the application of remote sensing and neural networks in the classification and mapping of *Acacia mearnsii* within South Africa, as well as the affect of the complexity of the environment on the accuracy of the classifications. The accuracy of a classification is highly dependent on the spatial resolution of the imagery being used (Ju *et.al*, 2005), as well as the nature of the environment and features being investigated (Rocchini, 2007). This is due to the fact that spatial resolution can determine how much generalization can occur within and around a feature: some features may be very detailed at one scale, but may be generalized at another (Lilllesand *et.al*, 2004).

This study therefore seeks to investigate the utility of SPOT 5 imagery and Artificial Neural Networks, in the identification and mapping of *Acacia mearnsii* within environments of varying complexity in KwaZulu Natal.

Chapter 3. METHODS

In the past, large scale vegetation mapping and, more specifically, alien vegetation mapping, has been based on what is essentially land cover classification (Carpenter *et.al*, 1999; Rowlinson *et.al*, 1999; Lu *et.al*, 2003; Linderman *et.al*, 2004; Everitt *et.al*, 2005; Cobbing, 2006). This study follows a similar concept, though with the target of identifying and mapping the *Acacia mearnsii* species. In order for training sites to be selected (from which the GPS points would be taken) in a non-bias fashion, it was decided that randomly generated points would be used. What follows is a description of the methods used in the generation of these points, data collection and classification techniques used, the framework for which can be seen in Figure 3.4.

3.1 BASELINE DATA COLLECTION AND PLANNING

An area of interest (AOI) within the original SPOT 5 image using ERDAS 8.4 was created. This AOI was then extracted to produce a subset of the original image, defining Study site 1. The band combination was adjusted using the 'band combinations' tool in ERDAS 8.4. It was found that the use of bands 1, 3 and 4 as red, green and blue respectively, produced the best image for observed vegetation identification. A simple ISODATA unsupervised classification was run in ERDAS 8.4 to determine the classes which would be used in the random generation of training sites. After a number of trials, it was found that the selection of eight classes produced the best result in terms of maintaining a balance between accuracy and simplicity. Simplicity, in the sense of retaining a minimum number of broad classes was identified as being particularly important in this study as the purpose of the study was to identify *Acacia mearnsii* and not to produce a detailed land cover map.

Once complete, the classified image was converted from raster to vector format and opened with ArcGIS 9.2. Each class was then exported as a shape file and re-opened in ArcGIS 9.2. Using Hawth's Tools (Beyer, 2004), 20 random points were created for each class, producing a total of 140 points (see Figure 3.1). The reason behind creating 20 points per class was that; if some of the points prove to be inaccessible, a sufficient number of points would still be collected. The process was then repeated for Study site 2, ensuring the area, band combinations, set parameters, number of classes and points remained identical to those used for Study site 1.

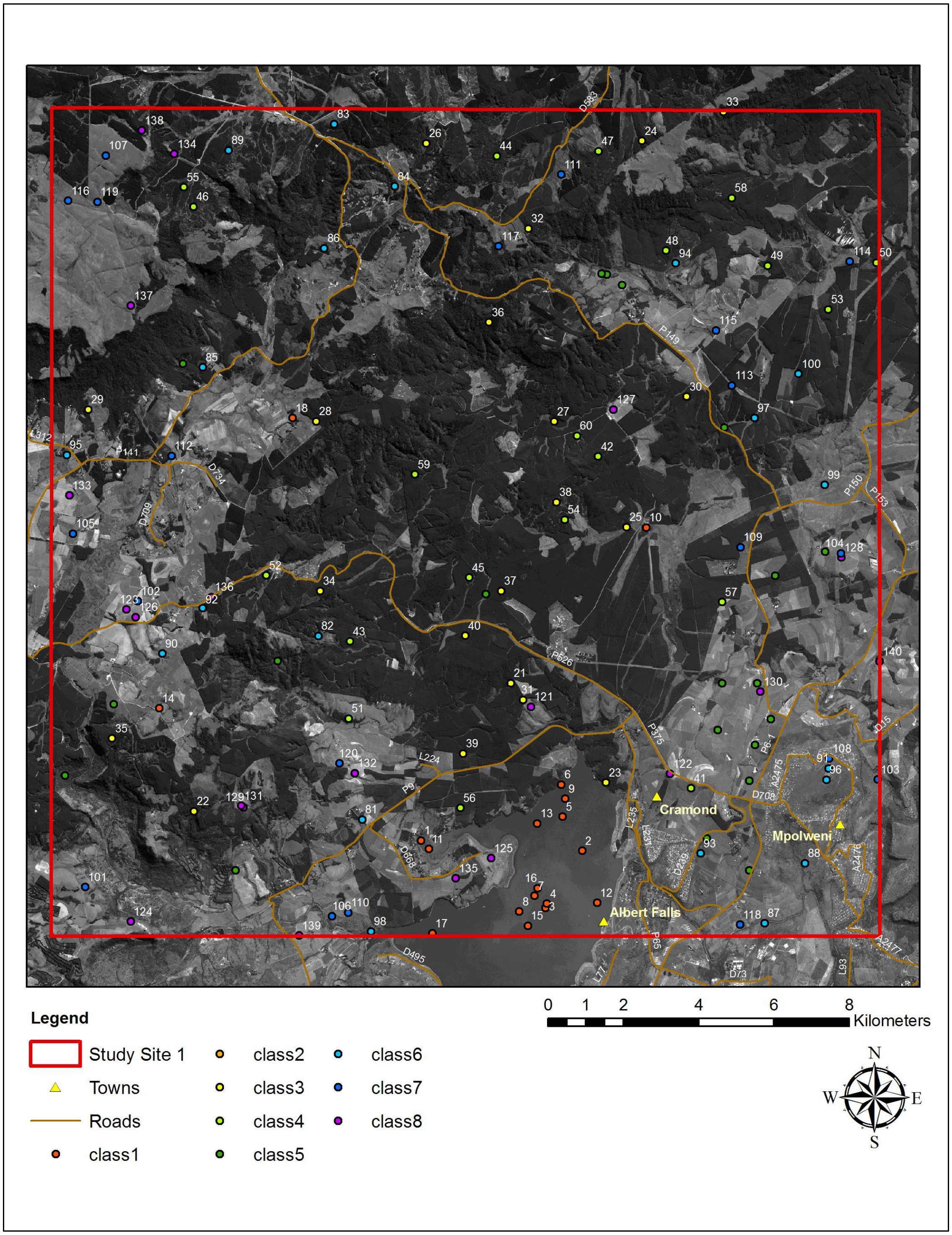

Figure 3.1 Example of the random points generated for Study Site 1

These randomly generated points formed the foundation from which the GPS points for the training sites were collected. Ground truthing took place in February and April 2009, ensuring the season remained the same as when the image was captured. Once collected, the points were opened in conjunction with the SPOT 5 image for Study site 1 and 2, in ArcGIS 9.2, and edited. The editing involved deleting points which were questionable in terms of their class. According to Skidmore (1999), it is better to have fewer highly accurate points, rather than a greater number of less accurate points. In addition, due to the location of several points and their inaccessibility, some of the points initially created during the random point generation process could not be collected. Interpretation and observations made using the SPOT 5 image and 1: 50 000 map data, however, counteracted this and more training points were added. In total, 211 points were collected across Study site 1 and 2.

Once the points were edited, added and found to be accurate, they were divided up into eight classes based on the unsupervised ISODATA classification and observations made in the field according to recommendations made by the National Landcover classification definitions of South Africa (CSIR, 2002). These classes include; water, bare ground/ urban, grassland, woodland, natural bush, agricultural crops, agricultural forestry, and wattle (see Table 3.1). Thus, 50% of the points for each class were then set aside to be used in signature creation, with the remaining 50% being reserved for accuracy assessment (Skidmore, 1999). Using ArcGIS 9.2, shapefiles were created according to the class of the points collected and the corresponding points were exported to the shapefiles. These individual class shapefiles formed the base from which all classifications and accuracy assessments were made.

Table 3.1 Definitions used for each of the classes (after CSIR, 2002)

Class	Class Name	Class Description
1	Water	All open bodies of water including streams and rivers
2	Bare ground/ urban	Includes areas of exposed soil and areas inhabited by humans. The decision was made to combine the classes as the signatures were found to be sufficiently similar
3	Grassland	An area with less than 10% tree or shrub cover, with grass being the dominant species
4	Woodland	Savannah-like natural areas where the cover of tree is between 10% and 70% of the total cover
5	Natural bush	Dense natural vegetation, consisting of shrubbery and natural forest communities
6	Agricultural crops	Areas under crops, including sugar cane and maize
7	Commercial forestry	Commercial plantations not including those of *Acacia mearnsii*
8	Wattle	Any occurrence of *Acacia mearnsii*

3.1.1 Limitations experienced During the Data Collection and Planning

A number of limitations were experienced during the planning and data collection phase of the process. This was generally a result of questionable class definition and inaccessibility.

The first limitation, experienced early on in the data collection process, related to the definition and assignment of classes. During the data collection for Study site 2, for example, it was found that the area had been largely affected by human inhabitants, in that very few areas of natural indigenous bush remained, with commercial timber species being used as windbreaks. These timber species were found in mixed species ‘pockets’. These areas could therefore not be defined as either natural bush or commercial forestry. As a result, a decision was made to combine these two classes to form a class of ‘mixed tree species’.

The second limitation was found to be inaccessibility. Study site 1, for example, is predominantly under commercial forestry and therefore private property. Much of the central region of the site was closed off and access denied. Consequently, attempts had to be made to work around these regions. In Study site 2, on the other hand, mountainous terrain and few roads limited access to much of the western portion of the site. This was however, offset by the fact that such areas were, to a large extent, grassland. Points were later added based on observations made using the SPOT 5 image and 1: 50 000 map data.

3.2 TRADITIONAL IMAGE CLASSIFICATION

3.2.1 Traditional Classification for Study Site 1 and 2

The ground truthed training points were opened in ERDAS 8.4 along with the SPOT 5 image for Study site 1 and signatures were created using the Signature Editor. The signature separability tool within ERDAS 8.4 was used to determine the adequacy of the signatures and after a number of trials, suitable signatures were generated. These signatures were then used in conjunction with bands 1, 3 and 4 of the SPOT 5 image to perform the traditionally used classifications. The reasoning behind using bands 1, 3 and 4 relates to the purpose of the study, that is the identification of *Acacia mearnsii*, a type of vegetation. The reflectance of vegetation is relatively low in the visible range, but is higher for green light than for red or blue. The reflectance pattern of green vegetation in the visible wavelengths is due to selective absorption by chlorophyll, the primary photosynthetic pigment in green plants. The most noticeable feature of the vegetation spectrum is the dramatic rise in reflectance across the visible-near infrared boundary, and the high near infrared reflectance. Infrared radiation penetrates plant leaves, and is intensely scattered by the leaves' complex internal structure, resulting in high reflectance (Lillsand *et.al*, 2004). Band 1 in the SPOT 5 sensor corresponds to the green reflectance ranging from 0.5 – 0.59 µm and band 3 corresponds to the near infrared (NIR) reflectance (0.78 – 0.89 µm), while band 4 corresponds to the short-wave infrared (SWIR) (1.58 – 1.75 µm) (Cobbing, 2006). Due to its structural and morphological characteristics, *Acacia mearnsii* exhibits unique spectral characteristics which can be observed in the NIR and SWIR portions of the electromagnetic spectrum. Bands 1, 3 and 4 of the SPOT 5 image were therefore used to perform the classifications.

Three traditional classification algorithms were used, namely; maximum likelihood, parallel piped and minimum distance to mean. Each produced a classified image comprising of 8 classes (see Table 3.1). This process was repeated several times with minor adjustments in points and classes until an optimum accuracy result was achieved. Based on the results of these trials (Table 3.2), it was decided that the 'Commercial forestry' and 'Natural bush' classes be combined to form the class 'Mixed tree species'. The final classifications were therefore run using 7 classes.

In the interest of retaining a sound basis for comparison between Study site 1 and 2, great efforts were made to ensure the area, band combinations, set parameters, number of classes and points for Study site 2 remained identical to those used for Study site 1. As with Study site 1, three classification algorithms were used, namely; maximum likelihood, parallel piped and minimum distance to mean. Each produced a classified image comprising of only 7 of the initial 8 classes (seen in Table 3.1) due to the fact that Study site 2 did not exhibit the 'Woodland' class as in Study site 1.

The 'Woodland' class is considered to consist of savannah-like natural areas where the cover of trees is between 10% and 70% of the total cover (CSIR, 2002). This is a very unique environment found only in Study site 1 owing to its geographic location and biophysical environment. Due to its unique spectral value it could not be merged with any other class nor eliminated from the classification in the interest of preserving the same number of classes between Study site 1 and 2. Classifications for Study site 2 were therefore performed using the following classes: water, bare ground/ urban, grassland, mixed tree species, agricultural crops, and wattle. The parameters for the classes remained the same as those in Study site 1 and can be seen in Table 3.1. The final classifications were therefore produced using 6 classes.

3.2.2 Traditional Classification for Study Site 1 and 2 Combined

A mosaic was performed in ERDAS 8.4, combining the SPOT 5 images for Study site 1 and 2. For ease of reference, this 'combined' site will, from here on, be referred to as Study site 3. The ground truthed training points for both Study site 1 and 2 were opened in ERDAS 8.4 along with the SPOT 5 mosaic image and signatures were created using the Signature Editor. The signature separability tool was used to determine the adequacy of the signatures and after a number of trials, suitable signatures were generated. These signatures were then used in conjunction with bands 1, 3 and 4 of the SPOT 5 image to perform the traditionally used classifications. Three classification algorithms were used, namely; maximum likelihood, parallel piped and minimum distance to mean. Each produced a classified image comprising of 7 classes, namely water, bare ground/ urban, grassland, woodland, agricultural crops, mixed tree species and wattle (see Table 3.1).

Table 3.2 Traditional classification trials and changes

	Trial	Classification Algorithm	Overall accuracy	Kappa (κ)	Changes
Study site 1	1	maximum likelihood	80.95%	0.7793	Increased number of reference points
	2	maximum likelihood	80.28%	0.7701	Merged classes 'commercial forestry' and 'natural bush'
	3	maximum likelihood	78.87%	0.7496	Final Result (7 classes)
	1	minimum distance	74.60%	0.7025	Increased number of reference points
	2	minimum distance	71.83%	0.6685	Merged classes 'commercial forestry' and 'natural bush'
	3	minimum distance	71.83%	0.663	Final Result (7 classes)
	1	parallel piped	74.60%	0.7025	Increased number of reference points
	2	parallel piped	71.83%	0.6685	Merged classes 'commercial forestry' and 'natural bush'
	3	parallel piped	83.10%	0.7999	Final Result (7 classes)
Study site 2	1	maximum likelihood	60.47%	0.54	Merged classes 'commercial forestry' and 'mixed tree species'
	2	maximum likelihood	79.07%	0.7469	Final Result (6 classes)
	1	minimum distance	53.49%	0.4591	Merged classes 'commercial forestry' and 'mixed tree species'
	2	minimum distance	88.37%	0.853	Final Result (6 classes)
	1	parallel piped	58.14%	0.5129	Merged classes 'commercial forestry' and 'mixed tree species'
	2	parallel piped	76.74%	0.7175	Final Result (6 classes)
Study site 3	1	maximum likelihood	68.42%	0.6209	Final Result (7 classes)
	1	minimum distance	56.14%	0.4849	Final Result (7 classes)
	1	parallel piped	68.42%	0.6208	Final Result (7 classes)

3.2.3 Accuracy Assessment of the Traditional Classifications

Once the classifications were completed, an accuracy assessment was performed on each of the classified images using the 'accuracy assessment' tool provided in ERDAS 8.4. This provided a basis for comparison when later evaluating the performance of the various classification algorithms. The accuracy assessment was performed using half of the collected reference points set aside during the data collection. In total 114 reference points, representing the various classified classes, were used. The system compared the classified image with these known points to determine the accuracy of the user and producer classification. The results were subsequently plotted in an error matrix.

3.2.4 Limitations experienced during the Traditional Classification

Limitations were experienced during the traditional classification phase of the process, particularly in terms of signature creation.

On a large scale, attempts were made during the random generation and collection of training points, to show no bias towards any single class, the same, however, can not be said for the signature creation. Certain classes are obvious and cover large areas, for example; water in Study site 1 and grassland in Study site 2. The signatures created for these classes were therefore much larger than for urban/ bare for example, a factor which can play a major role in the accuracy of the classification.

On a smaller scale, in Study site 1, and to a greater extent Study site 2, the resolution of SPOT 5 (10m by 10m) was in fact found to be a limiting factor. *Acacia mearnsii* was often found in mixed species stands along with a dense understory, both of which affect the reflectance values of a single pixel and therefore the accuracy of the classification.

3.3 ARTIFICIAL NEURAL NETWORK CLASSIFICATION

Artificial Neural Networks (ANN) differ from traditional classification techniques in that ancillary data, or additional data, can be incorporated into the classification process (Linderman *et.al.*, 2004; Qui and Jensen, 2004; Mutanga and Skidmore, 2004; Allan, 2007; Mutanga and Skidmore, 2007). What follows is a description of the testing and training of the ANN as well as the steps and procedures used in the classification of the images, the framework for which can be seen in Figure 3.4.

Image data used within the neural network included the original four SPOT 5 bands ranging from 0.5 – 1.75 µm. In addition, the NDVI (Normalized Differential Vegetation Index) was derived from the original SPOT 5 image in ERDAS 8.4 using equation 3.1 (Lilllesand *et.al*, 2004). The spectral range of the SPOT 5 NIR band falls between 780nm and 890nm, while the Red band falls between 610nm and 680nm. The NDVI, is widely used in vegetation mapping as it aids in improving the distinction between photosynthetically active green biomass (vegetation) types, as well as other classes such as water or bare ground (Lillesand *et.al*, 2004; Ibrahim, 2008).

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad \text{… Equation 3.1}$$

3.3.1 Ancillary data selection and creation

Ancillary data used in an ANN classification shifts the focus of a classification from purely spectral to spatial. The ancillary data used in this study was chosen based on the environmental variables influencing the spatial distribution of *Acacia mearnsii* defined in the literature, as well as data availability. The ancillary data included the elevation, slope, aspect, water sources, and soil type.

The elevation, slope and aspect were derived from relief line data. A TIN was created from the relief line data in ArcGIS 9.2 using the 3D Analyst tool. The TIN was then used to produce the aspect and slope layers, and was subsequently converted to raster format, forming the DTM (digital elevation model), or elevation layer.

The water sources layer was derived from river line and water body data. The river line and water body data was merged in ArcGIS 9.2, resulting in a single vector layer.

The Euclidian distance tool was then used on this layer to produce a raster layer representing distance from water sources.

The soil layer was produced using vector soil data which was converted to raster using ArcGIS 9.2

Before these new layers could be added to the SPOT 5 imagery, it was necessary to standardize them and the SPOT 5 image bands. This was done in three steps, namely; reprojecting, resampling and normalizing.

3.3.1.1 Reprojecting

In order for the ancillary data and satellite image to fit exactly when overlaid, it was crucial that each layer have the same geographical projection (Clark Labs, 2000). This however, was not the case as data from different sources often have different projections. An example of this can be seen in the SPOT 5 image which, according to ArcGIS 9.2, had a WGS_1984_UTM_Zone_35S projection, while the river line data was projected as Hartebeeshoek_1994. It was therefore necessary to match the projection of the ancillary data and images using ArcGIS 9.2.

3.3.1.2 Resampling

Resampling involves the alteration of one pixel grid to fit another pixel grid, in this case the SPOT 5 pixel grid of 10m x 10m (Clark Labs, 2000). Three resampling techniques were available, namely; bilinear interpolation, cubic convolution, and nearest neighbour. The bilinear interpolation technique uses the surrounding pixel values and the inverse distance weighted averages of the nearest four pixels to derive a new pixel value for the pixel in question. The resulting final image tends to be 'smoother', however the grey levels within the image are altered which can lead to problems in terms of spectral pattern recognition. The cubic convolution technique uses 16 pixels surrounding the pixel in question to derive a value for that pixel. The resulting final image tends to be 'smoother' than the other two techniques, however, as with the bilinear interpolation technique, pixel values are altered. The nearest neighbour technique considers neighbouring pixels when altering the position of a given pixel. It is a computationally very simplistic technique and avoids changing the original pixel values (Lilllesand *et.al*, 2004). It is for these reasons the nearest neighbour resampling technique was chosen for this study.

3.3.1.3 Normalizing

In order for the new ancillary data layers to be added to the SPOT 5 image, the data sets were required to be of the same data type (Clark Labs, 2000). The SPOT 5 data was in an unsigned 8-bit format while the elevation, slope, aspect, water sources, and soil type layers were in a continuous format.

The normalization was performed in ERDAS 8.4, where the ancillary data layers were converted to an unsigned 8-bit format. An 8-bit image contains 256 classes, or a numerical range of between 0 and 255 within which the digital values of an image can fall.

Once the data had been standardized, it was possible to perform a layer stack using ERDAS 8.4, merging the ancillary and satellite images to form one image containing 10 bands, namely; SPOT 5 band 1, SPOT 5 band 2, SPOT 5 band 3, SPOT 5 band 4, NDVI, elevation, slope, aspect, water sources, and soil type (see Figure 3.2). This process was then repeated for Study site 2, resulting in a second layer stack.

In terms of Study site 3, the process followed remained the same, but included one additional step; the mosaic of the ancillary data and SPOT 5 images for Study site 1 and 2 once the data had been standardized. As with Study site 1 and 2, the 'mosaicked' ancillary and satellite images were stacked to form a third image containing 10 bands.

These three stacks were subsequently used for the creation of signatures within the neural network

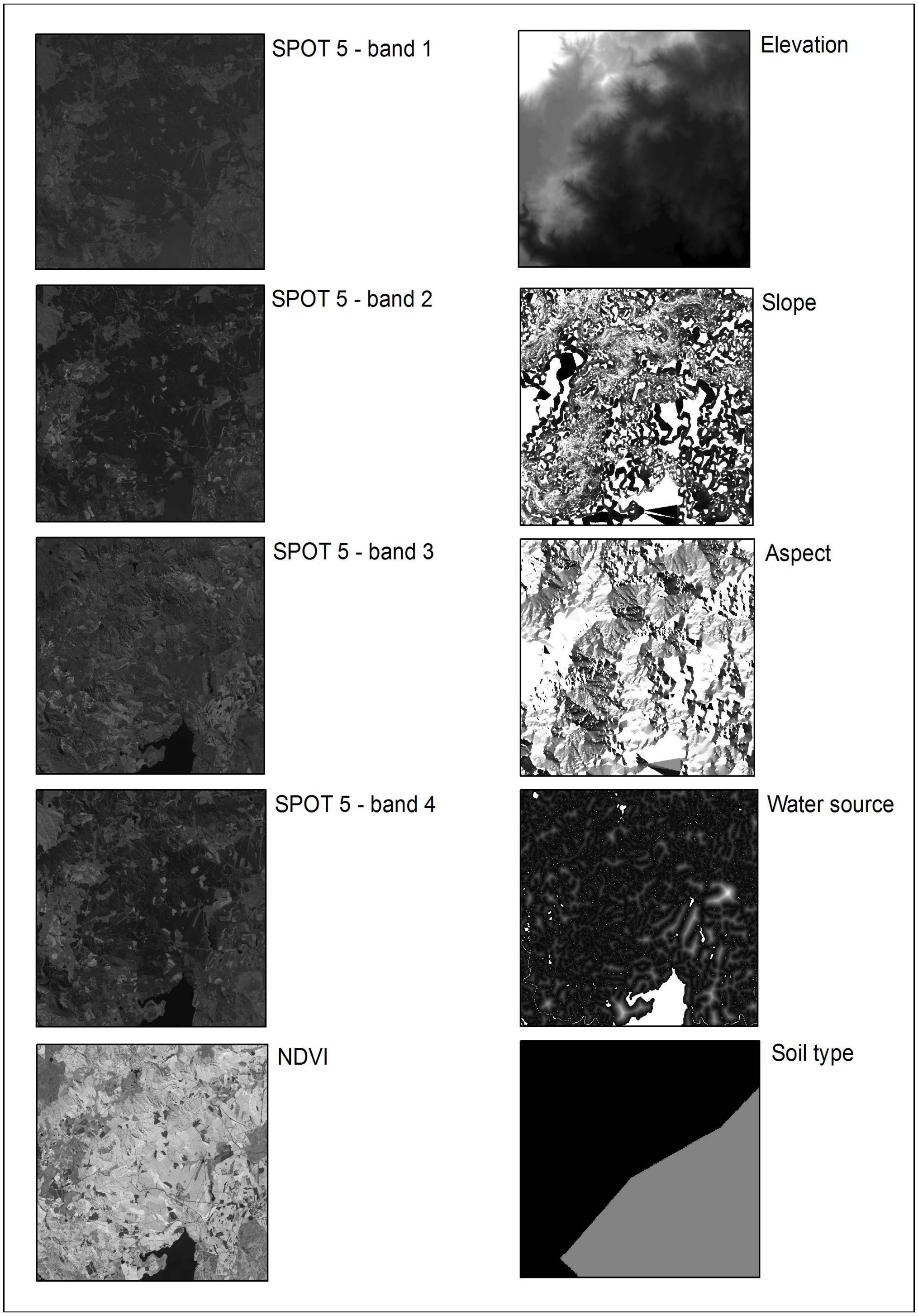

Figure 3.2 Image and ancillary data used in the Study site 1 neural network

3.3.2 Training site selection and signature creation

Once the layer stacks had been created for each of the study sites, they were opened in IDRISI Andes, the software in which the ANN was to be run. Before the process of training and testing the ANN could begin it was necessary to create signatures which would allow the system to discriminate between different spectral classes. Although the principles behind signature creation remained similar to those of the signature creation in ERDAS 8.4 used for the traditional classifications, the techniques used to create the signatures were different.

The signature creation process began with the creation of a vector file within IDRISI Andes. Polygons were then digitized into this vector file based on the classes defined during the traditional classifications, with each class being assigned a code. This code was used to define a set of polygons, each belonging to the same class. Table 3.3 depicts the classes and codes used in the ANN classifications.

Table 3.3 The user defined codes used in the creation of signatures in IDRISI Andes

Code	Study Site 1	Study Site 2	Study Site 3
1	Agric. Crops	Agric. Crops	Agric. Crops
2	Mixed tree sp.	Mixed tree sp.	Mixed tree sp.
3	Grassland	Grassland	Grassland
4	Urban/ bare	Urban/ bare	Urban/ bare
5	Water	Water	Water
6	Wattle	Wattle	Wattle
7	Woodland		Woodland

3.3.3 Neural network design

The neural network used in IDRISI Andes is the Multi-Layered Perception (MLP), which uses a back propagation algorithm. The neural network essentially has two steps; the forward movement of the input data through the nodes within the hidden layers, and backpropagation as the network 'learns' the characteristics of the data, aiding in the accuracy of the final result (Clark Labs, 2000; Foody, 2003 and Kazoglu; Mather, 2003). Backpropagation works in two phases: phase 1 involves running the model with the initial network weights in place, resulting in estimated output values for each set of input data. Phase 2 begins once the values are estimated and can be compared to the known values. The differences between the estimated and known values are calculated and sent backwards through the network, where the weights between the nodes are adjusted. The entire process is then repeated until the error is negligible or non-existent (Foody, 2003; Kazoglu and Mather, 2003, Allan, 2007). The structure of the neural network is presented in Chapter 2, Figure 2.1.

Various parameters within the IDRISI Andes MLP are user defined and can be adjusted, thereby altering the design of the neural network (see Figure 3.3). Before any methods are described within this section, it is best to understand what these parameters are and what options are available to the user.

- Band selection: the number of bands 'inputs' is chosen. This includes the satellite imagery as well as the ancillary data.

- Signature selection: the signature or training site files previously created are inserted into the network.

- Pixels for testing and training input: By default, the testing pixels are set to the number of training pixels. The training pixels are used in the analysis and will be a subset of the total pixels found in the training site or signature file. The testing pixels, on the other hand, are used to validate the results (Clark Labs, 2000). In this study, the number of pixels used for both training and testing in each of the classifications was 40 pixels.

- Hidden layers: the number of hidden layers is chosen

- Nodes per layer: the number of nodes within each hidden layer is chosen

- Learning rate decision: the learning rate is considered to be the most significant factor in the design of a neural network. It is a positive constant that controls the adjustments made to the connecting weights within the neural network, in terms of both quantity and frequency. If the learning rate is too small, the training phase may become overwhelmed and time-consuming, while a rate too large may result in hugely fluctuating adjustments and therefore poor results (Clark Labs, 2000).

- Momentum factor decision: This factor intends to remove the changes to the Root Mean Square Error (RMSE) of the surface classification during the training of the network (Clark Labs, 2000)

- Number of Iterations: the number of times the neural network will run is chosen. The number of iterations is one of the stopping criteria which can be defined by a user. The stopping criteria are parameters which can be set and if met will result in the ending of the network training. Other stopping criteria include accuracy and RMSE.

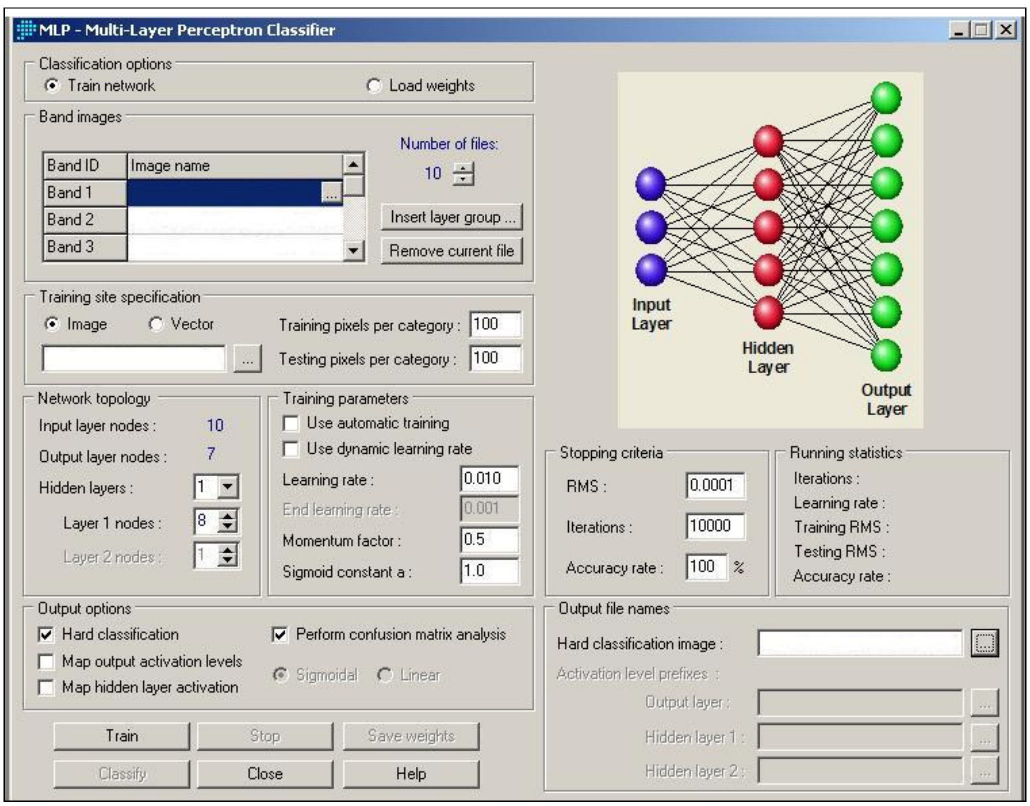

Figure 3.3 The MLP Dialog box of IDRISI Andes with user defined parameters

3.3.4 Testing the neural network

Before classifying an image, the network must first be tested and trained. Testing the network is achieved through the minor adjustment of user defined parameters in the design of the neural network. In order to ensure the highest accuracy and efficiency of the neural network, it was decided that these parameters would be adjusted one at a time. The final parameters were determined based upon the notion that a low RMSE is directly proportional to a high accuracy, and that the increase in accuracy was a result of the previous changes and could therefore be carried forward to the next run. The following section details the testing process of the neural network for each study site.

For the sake of consistency, the changes and intervals remained the same for each of the study sites, as can be seen in Tables 3.4, 3.5 and 3.6. The description of the runs and changes will therefore be dealt with as a whole.

3.3.4.1 Run 1: testing of the ancillary data

As stated in the objectives in Chapter 1, it was necessary to decide which of the ancillary data or environmental variables had the greatest impact on the accuracy of the classification. For this reason, the classification was first run with the four SPOT 5 bands alone. The NDVI, DTM, slope, aspect, water and soil bands were subsequently added, one at a time and the accuracy recorded. It is interesting to note that in each site the neural network performed best when using only the first 9 bands, excluding the soil type band. This was found to be due to the highly generalized nature of the soil type band.

3.3.4.2 Run 2: testing the hidden layers

In this case, the hidden layers were limited to two by the IDRISI Andes MLP. One therefore only had two options; 1 hidden layer or 2 hidden layers.

3.3.4.3 Run 3: testing the nodes per layer

The number of nodes per layer were changed at a set interval until reaching 15. It was found to be unnecessary to continue past 15 since there was a sharp decrease in accuracy as the number of nodes increased passed 10.

3.3.4.4 Run 4: testing the momentum factor

In testing the momentum factor, the rate of increment change was 0.1

3.3.4.5 Run 5: testing the learning rate

Within the IDRISI training manual, the learning rate was identified as being the most important factor in the training of the neural network (Clark Labs, 2000; Hu & Weng, 2009). It is for this reason that this run was more detailed than previous runs, with 16 changes made. The learning rate was changed at a set interval of 0.001 until reaching 0.01 and then 0.01 until reaching 0.1.

3.3.4.6 Run 6: testing the number of iterations

The number of iterations was tested in an attempt to determine if the training and testing of the network could be more efficient, or if letting the process continue for longer period could result in a higher accuracy. Changes were made for this run in intervals of 1000 reaching 10 000. After 10 000 iterations, the increase was changed to 5000 reaching 30 000, after which a final change was made to 50 000 iterations. Excessive iterations can, however, lead to over-training which can result in the poor generalization of the network (Clark Labs, 2000).

3.3.4.7 The final run

Using the best results from runs 1 to 6, the parameters of the neural network were set and the final run was completed. The final outputs for each of the study sites can be seen in Chapter 4.

Table 3.4 Changes made per run for the Study site 1 ANN

Runs	Bands	Hidden layers	Nodes	Momentum factor	Learning rate	Iterations
1	a	1	8	0.5	0.01	10000
2	9	b	8	0.5	0.01	10000
3	9	1	c	0.5	0.01	10000
4	9	1	8	d	0.01	10000
5	9	1	8	0.4	e	10000
6	9	1	8	0.4	0.01	f
final	9	1	8	0.4	0.01	10000

a - 7 changes made: 4,5,6,7,8,9,10
b - 2 changes made: 1,2
c - 12 changes made: 2,3,4,5,…10,11,12,15
d - 7 changes made: 0.3,0.4,0.5,0.6,…0.9
e - 16 changes made: 0.005, 0.006, 0.007…0.01, 0.02,…0.1
f - 10 changes made: 1000, 2000, 3000, 5000, 10000, 15000… 30000, 50000

Table 3.5 Changes made per run for the Study site 2 ANN

Runs	Bands	Hidden layers	Nodes	Momentum factor	Learning rate	Iterations
1	a	1	8	0.5	0.01	10000
2	8	b	8	0.5	0.01	10000
3	8	1	c	0.5	0.01	10000
4	8	1	7	d	0.01	10000
5	8	1	7	0.05	e	10000
6	8	1	7	0.05	0.01	f
final	8	1	7	0.05	0.01	10000

a - 7 changes made: 4,5,6,7,8,9,10
b - 2 changes made: 1,2
c - 12 changes made: 2,3,4,5,…10,11,12,15
d - 7 changes made: 0.3,0.4,0.5,0.6,…0.9
e - 16 changes made: 0.005, 0.006, 0.007…0.01, 0.02,…0.1
f - 10 changes made: 1000, 2000, 3000, 5000, 10000, 15000… 30000, 50000

Table 3.6 Changes made per run for the Study site 3 ANN

Runs	Bands	Hidden layers	Nodes	Momentum factor	Learning rate	Iterations
1	a	1	6	0.5	0.01	10000
2	9	b	6	0.5	0.01	10000
3	9	1	c	0.5	0.01	10000
4	9	1	6	d	0.01	10000
5	9	1	6	0.5	e	10000
6	9	1	6	0.5	0.01	f
final	9	1	6	0.5	0.01	10000

a - 7 changes made: 4,5,6,7,8,9,10
b - 2 changes made: 1,2
c - 12 changes made: 2,3,4,5,…10,11,12,15
d - 7 changes made: 0.3,0.4,0.5,0.6,…0.9
e - 16 changes made: 0.005, 0.006, 0.007…0.01, 0.02,…0.1
f - 10 changes made: 1000, 2000, 3000, 5000, 10000, 15000… 30000, 50000

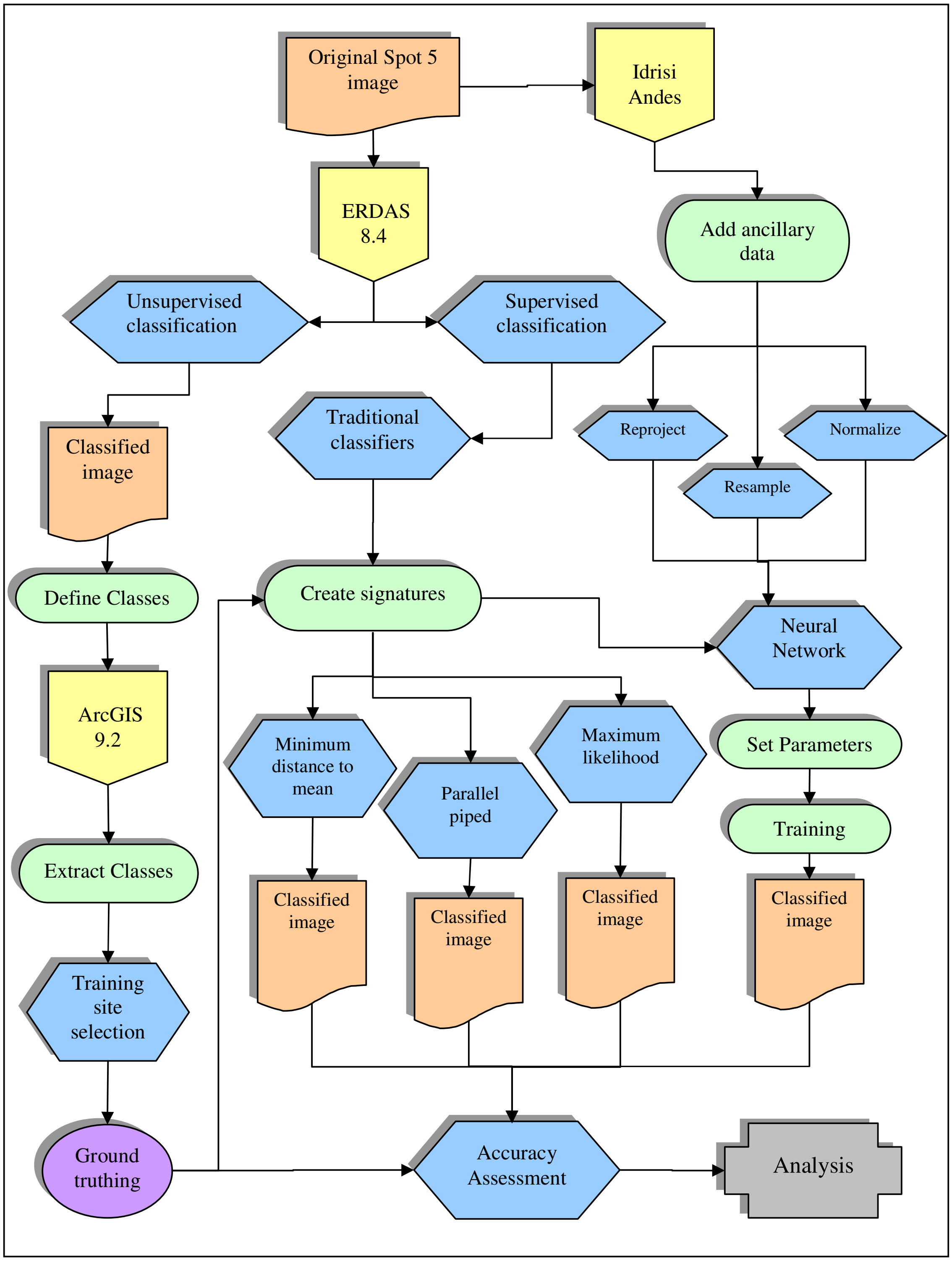

Figure 3.4 Framework for methods used

Chapter 4. RESULTS

This chapter will present the results that were acquired during this study. The structure of this chapter will follow the objectives outlined in Chapter 1. *Acacia mearnsii* was classified and mapped using Spot 5 imagery in combination with Artificial neural networks and traditional classifiers. The accuracy of the artificial neural network and traditional classifications were then compared. Finally, using these accuracy assessment results, the classifications were compared between the two different environmental settings of Study sites 1 and 2.

4.1 TRADITIONAL CLASSIFICATION

This section presents the results for the traditional classification of the SPOT 5 imagery. Chapter 3 outlined the steps taken in obtaining these results.

4.1.1 Traditional Classification for Study Site 1

In order to provide a basis for comparison when evaluating the performance of the various classification algorithms, an accuracy assessment was performed on each of the images and are presented as error matrices. The error matrices were constructed from samples taken from the classified images.

Table 4.1 illustrates the error matrix for the final maximum likelihood classification of Study site 1 after the merge of the 'natural bush' and 'commercial forestry' classes, into the 'mixed tree species' class. In each error matrix that follows, reference, or ground truth points, are represented in the columns and are compared with the classified pixels found along the rows of the matrix. The diagonal values indicate the 'agreement' between the two data sets (Skidmore, 1999). In table 4.1, the most notable confusions during the maximum likelihood classification occurred between the 'grassland' and 'urban/ bare ground' classes and the 'wattle' (*Acacia mearnsii*) and 'mixed tree species' classes. The user and producer accuracies were then calculated based on these diagonal values and were represented in terms of percentages. The user accuracy is the probability that a pixel classified on an image actually represents that category on the ground (Skidmore, 1999). The user accuracy indicates how distinctive a class is, the higher the percentage, the more distinctive a class is (Everitt *et.al.*, 2005). The producer accuracy, on the other hand, is the probability of a referenced pixel being correctly classified (Skidmore, 1999). The user accuracy for 'wattle' was 92.31% while the producer accuracy was 63.16%.

Table 4.1 Error matrix for the accuracy assessment of the classification of Study site 1 using the Maximum Likelihood algorithm

Classified Data	Mixed tree	Woodland	Urban/ bare	Water	Grassland	Wattle	Agric. crops	Row Total	User Accuracy
Mixed tree sp.	8	0	0	0	0	6	0	14	57.14%
Woodland	0	2	0	0	1	0	0	3	66.67%
Urban/ bare	0	0	8	1	4	1	0	14	57.14%
Water	0	0	0	7	0	0	0	7	100.00%
Grassland	0	1	0	0	8	0	0	9	88.89%
Wattle	1	0	0	0	0	12	0	13	92.31%
Agric. crops	0	0	0	0	0	0	11	11	100.00%
Column Total	9	3	8	8	13	19	11	71	
Producer Accuracy	88.89%	66.67%	100.00%	87.50%	61.54%	63.16%	100.00%		

Overall Accuracy = 78.87%

Overall Kappa Statistics = 0.7496

According to Skidmore (1999), a common measure of map accuracy is the overall accuracy. The overall accuracy refers to the number of correctly classified pixels (the sum of the diagonal values), divided by the total number of pixels checked. The overall accuracy of the maximum likelihood classification for Study site 1 was 78.87%.

Another statistical indicator of overall accuracy is Kappa (κ), calculated using equation 4.1. This value rages from 0 to 1, depending on the agreement between the classified image and the reference data (Skidmore, 1999).

$$K = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_i * x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})} \qquad \text{... Equation 4.1}$$

Where: N is the product of marginal totals

r is the number of rows and columns in error matrix

x_{ii} is the number of observations in row 1 and column 1

x_{i+} is the marginal total of row I

x_{+1} is the marginal total of column I

The calculation for the κ statistic for the maximum likelihood classification of Study site 1 therefore proceeded as follows:

$$\sum_{i=1}^{r} x_{ii} = 8+2+8+7+8+12+11= 56$$

$$\sum_{i=1}^{r} (x_{i+} * x_{+i}) = (9*14)+(3*3)+(8*14)+(8*7)+(13*9)+(19*13)+(11*11) = 788$$

$$K = \frac{(71*56) - 788}{71^2 - 788}$$

The κ value is therefore 0.7495, indicating a fairly strong correlation between the classified image and reference data.

Table 4.2 represents the results obtained from the error matrix for the final minimum distance classification of Study site 1 after the merging of the ‘natural bush’ and ‘commercial forestry’ classes, into the ‘mixed tree species’ class. Confusion occurred during the minimum distance classification largely between the ‘wattle’ and ‘mixed tree species’ classes. Consequently, the user accuracy for ‘wattle’ was 64.29% while the producer accuracy fell short at just 47.37%. The overall accuracy however, remained high at 71.83%, with κ value somewhat lower at 0.663.

In the error matrix for the final parallel piped classification of Study site 1, confusion once again occurred between the ‘wattle’ and ‘mixed tree species’ classes, however to a lesser extent than during the minimum distance classification. This can be seen in Table 4.2, in the higher user and producer accuracies for ‘wattle’, which were 100% and 63.16% respectively. The overall accuracy was particularly high in this case at 83.1% with a κ value of 0.7999.

Table 4.2 provides a summary of the results obtained from each of the traditional classifications for Study site 1. From this table it is clear that the parallel piped classification algorithm had the greatest success both in terms of overall accuracy and in the specific classification of *Acacia Mearnsii* (wattle).

Table 4.2 Results summary for the traditional classifications of Study site 1

Classification Algorithm	User Accuracy (Wattle)	Producer Accuracy (Wattle)	Overall Accuracy	Kappa (κ)
Maximum likelihood	92.31%	63.16%	78.87%	0.7496
Minimum Distance	64.29%	47.37%	71.83%	0.663
Parallel Piped	100.00%	63.16%	83.10%	0.7999

4.1.2 Traditional Classification for Study Site 2

Table 4.3 illustrates the error matrix for the final maximum likelihood classification of Study site 2 after the merge of the 'mixed tree species' and 'commercial forestry' classes, into the 'mixed tree species' class. During this classification, significant confusion occurred between the 'mixed tree species' and 'wattle' classes resulting in a very low user accuracy of 36.36% for 'wattle'. The producer accuracy on the other hand, was high at 80%. The overall accuracy (79.07%) and κ value (0.7469), however, indicate that the classification algorithm as a whole performed well.

Table 4.3 Error matrix for the accuracy assessment of the classification of Study site 2 using the Maximum likelihood algorithm

Classified Data	Urban/ bare	Wattle	Mixed tree	Grassland	Agric. crops	Water	Row Total	User Accuracy
Urban/ bare	5	0	0	0	0	0	5	100.00%
Wattle	0	4	7	0	0	0	11	36.36%
Mixed tree sp.	0	1	7	0	0	0	8	87.50%
Grassland	0	0	0	7	0	0	7	100.00%
Agric. Crops	0	0	0	1	7	0	8	87.50%
Water	0	0	0	0	0	4	4	100.00%
Column Total	5	5	14	8	7	4	43	
Producer Accuracy	100.00%	80.00%	50.00%	87.50%	100.00%	100.00%		

Overall Accuracy = 79.07%

Overall Kappa Statistics = 0.7469

In the results obtained from the error matrix the final minimum distance classification of Study site 2, minimal confusion occurred during the minimum distance classification, although some was seen between the ‘mixed tree species’ and ‘wattle’ classes, as well as between the ‘grassland’ and ‘agricultural crops’ classes. The user and producer accuracies for ‘wattle’ were reasonable at 75% and 60% respectively. The overall accuracy and κ value were, however, very high with the overall accuracy reaching 88.37% and κ value of 0.8530.

In the error matrix for the final parallel piped classification of Study site 2, significant confusion occurred between the ‘mixed tree species’ and ‘wattle’ classes during the parallel piped classification resulting in a very low user accuracy of 36.36% for ‘wattle’. The producer accuracy on the other hand, as seen in Table 4.4, was high at 80%. The overall accuracy (76.74%) and κ value (0.7175), however, indicate that the classification algorithm as a whole performed fairly well.

Table 4.4 provides a summary of the results obtained from each of the traditional classifications for Study site 2. From this table it is clear that the minimum distance classification algorithm had the greatest success both in terms of overall accuracy and in the specific classification of *Acacia Mearnsii* (wattle).

Table 4.4 Results summary for the traditional classifications of Study site 2

Classification Algorithm	User Accuracy (Wattle)	Producer Accuracy (Wattle)	Overall Accuracy	Kappa (κ)
Maximum likelihood	36.36%	80.00%	79.07%	0.7469
Minimum Distance	75.00%	60.00%	88.37%	0.853
Parallel Piped	36.36%	80.00%	76.74%	0.7175

4.1.3 Traditional Classification for Study Site 3

Table 4.5 illustrates the error matrix for the final maximum likelihood classification of Study site 3 after the merge of the 'mixed tree species', 'natural bush' and 'commercial forestry' classes, into the 'mixed tree species' class. Significant confusion occurred between the 'mixed tree species' and 'wattle' classes. Consequently, both the user and producer accuracies for the *Acacia mearnsii* (wattle) class are low, with the user accuracy being 46.67% and the producer accuracy being 58.33%. The overall accuracy is indicative of this confusion at 68.42% with the κ value of 0.6209

Table 4.5 Error matrix for the accuracy assessment of the classification of the Study site 3 using the Maximum Likelihood algorithm

Classified Data	Mixed tree	Woodland	Grassland	Water	Urban/ bare	Agric crops	Wattle	Row Total	User Accuracy
Mixed tree sp.	8	0	2	0	0	1	9	20	40.00%
Woodland	0	3	2	0	0	0	0	5	60.00%
Grassland	1	0	13	0	0	1	1	16	81.25%
Water	0	0	0	11	0	0	0	11	100.00%
Urban/ bare	0	0	0	1	13	0	0	14	92.86%
Agric. crops	0	0	2	0	0	16	0	18	88.89%
Wattle	14	0	2	0	0	0	14	30	46.67%
Column Total	23	3	21	12	13	18	24	114	
Producer Accuracy	34.78%	100.00%	61.90%	91.67%	100.00%	88.89%	58.33%		

Overall Accuracy = 68.42%

Overall Kappa Statistics = .6209

Table 4.6 shows results obtained from the error matrix following the final minimum distance classification of Study site 3. Here the greatest confusion occurred between the 'grassland' and 'woodland' classes. There was however, still confusion between the 'wattle' and 'mixed tree species' classes. This is evident in the low 'wattle' user (61.11%) and producer (45.83%) accuracies. The overall accuracy of 56.14% and κ value of 0.4849, suggests that this was not a successful classification.

In the error matrix for the final parallel piped classification of Study site 3, confusion was experienced between the 'wattle' and 'mixed tree species' classes. As a result the user accuracy for 'wattle' was very low at 46.67%, as was the producer accuracy at 58.33%. The overall accuracy was however, reasonable, at 68.42% with a κ value of 0.6208 (see Table 4.6).

Table 4.6 provides a summary of the results obtained from each of the traditional classifications for the Study site 3. As a whole, both in terms of overall accuracy and in the specific classification of *Acacia Mearnsii* (wattle), it is clear that the traditional classifications did not perform as well on the mosaic as on the two separate study sites. These results, although not very high, are in fact more typical in terms of traditional classification accuracy than the results of the two separate study sites. As seen in Table 4.6, the maximum likelihood and parallel piped classification performances were much alike, although the κ value for the maximum likelihood classification was slightly higher. The maximum likelihood classification was therefore considered to be the most successful.

Table 4.6 Results summary for the traditional classifications of Study site 3

Classification Algorithm	User Accuracy (Wattle)	Producer Accuracy (Wattle)	Overall Accuracy	Kappa (κ)
Maximum likelihood	46.67%	58.33%	68.42%	0.6209
Minimum Distance	61.11%	45.83%	56.14%	0.4849
Parallel Piped	46.67%	58.33%	68.42%	0.6208

4.2 ARTIFICIAL NEURAL NETWORK CLASSIFICATION

This section presents the results obtained for the artificial neural network classification of the SPOT 5 imagery and ancillary data. Chapter 3 outlined the steps taken in obtaining these results. The final neural networks were designed to have both high accuracy and high efficiency.

4.2.1 Neural Network Classification for Study Site 1

Table 4.7 represents the error matrix for the neural network classification of Study site 1. Here, confusion occurred between the 'wattle' and 'mixed tree species' classes, however to a lesser extent than during the traditional classifications. Confusion also occurred between the 'urban/bare' and 'water' classes. The overall accuracy was particularly high in this case at 94.46% with a κ value of 0.9254.

Table 4.7 Error matrix for the accuracy assessment of the ANN classification for Study site 1

Classified Data	Agric. Crops	Mixed tree sp.	Grassland	Urban/ bare	Water	Wattle	Woodland	Row Total	User accuracy
Agric Crops	4072	13	2	3	0	102	12	4204	96.86%
Mixed tree sp.	0	9229	0	0	0	473	0	9702	95.12%
Grassland	24	463	5696	13	38	6	33	6273	90.80%
Urban/ bare	50	3	0	584	78	1	32	748	78.07%
Water	0	0	0	0	20856	0	0	20856	100.00%
Wattle	21	1248	0	1	3	3557	9	4839	73.51%
Woodland	1	64	20	14	0	9	2630	2738	96.06%
Column Total	4168	11020	5718	615	20975	4148	2716	49360	
Producer accuracy	97.70%	83.75%	99.62%	94.96%	99.43%	85.75%	96.83%		

Overall Classification Accuracy = 94.46%
Overall Kappa Statistics = 0.9254

4.2.2 Neural Network Classification for Study Site 2

Table 4.8 represents the results obtained from the error matrix for the neural network classification of Study site 2. During this classification, significant confusion occurred between the 'mixed tree species' and 'wattle' classes resulting in a very low user accuracy of 40.17% for 'wattle'. The producer accuracy on the other hand, was higher at 83.45%. The overall accuracy (92.79%) and κ value (0.9019), however, indicate that the classification as a whole performed well.

4.2.3 Neural Network Classification for Study Site 3

In the error matrix for the neural network classification of Study site 3, significant confusion occurred between the 'mixed tree species' and 'wattle' classes. Consequently, both the user and producer accuracies for the *Acacia mearnsii* (wattle) class are low, with the user accuracy being 57.73% and the producer accuracy being 65.39%. The overall accuracy, however remained high at 68.42% with the κ value of 0.6209.

Table 4.8 provides a summary of the results obtained from neural network classifications for each of the study sites. As a whole, both in terms of overall accuracy and in the specific classification of *Acacia Mearnsii* (wattle), it is clear that the neural network performed best in Study site 1.

Table 4.8 Results summary for the neural network classifications of Study site 3

Site	User Accuracy (Wattle)	Producer Accuracy (Wattle)	Overall Accuracy	Kappa (κ)
Study site 1	73.57%	85.75%	94.46%	0.9254
Study site 2	40.17%	83.45%	92.79%	0.9019
Study site 3	57.73%	65.39%	93.45%	0.8959

4.3 COMPARRISON OF CLASSIFICATION RESULTS

Table 4.9 represents the summary of results obtained during this study. In terms of the overall accuracy as well as the κ value, it is clear that the neural networks had the highest accuracy when compared to the traditional classifiers. In addition, classifications run on Study site 1 performed better than those run on the other two sites.

Table 4.9 Results summary all the classifications performed on each of the sites

	Classification Algorithm	User Accuracy (Wattle)	Producer Accuracy (Wattle)	Overall Accuracy	Kappa (κ)	Kappa (κ) Variance
Site 1	Maximum likelihood	92.31%	63.16%	78.87%	0.7496	0.0033309114
	Minimum Distance	64.29%	47.37%	71.83%	0.663	0.0043407455
	Parallel Piped	100.00%	63.16%	83.10%	0.7999	0.0027844094
	Neural network	73.57%	85.75%	94.46%	0.9254	0.0000019091
Site 2	Maximum likelihood	36.36%	80.00%	79.07%	0.7469	0.0058226721
	Minimum Distance	75.00%	60.00%	88.37%	0.853	0.0039144134
	Parallel Piped	36.36%	80.00%	76.74%	0.7175	0.0064906102
	Neural network	40.17%	83.45%	92.79%	0.9019	0.0000025997
Site 3	Maximum likelihood	46.67%	58.33%	68.42%	0.6209	0.0030432112
	Minimum Distance	61.11%	45.83%	56.14%	0.4849	0.0032289930
	Parallel Piped	46.67%	58.33%	68.42%	0.6208	0.0030660357
	Neural network	57.73%	65.39%	93.45%	0.8959	0.0000007451

Checking for significant differences between images produced by different classification techniques is important in determining the utility of one technique over the other (Skidmore, 1999). Skidmore (1999) suggested using the discrete multivariate analysis techniques to statistically determine whether or not two error matrices were significantly different. This is done using the κ values of two independent error matrices and their associated variance by evaluating the normal curve deviate. For obvious reasons, only one image factor can be changed at any time when comparing two matrices, such as the classifier type or date of image collection (Skidmore, 1999).

The asymptomatic variance of Kappa (seen in table 4.9) is calculated by:

$$\sigma^2(K) = \frac{1}{n}\left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4}\right]$$

…Equation 4.2

Where:

$$\theta_1 = \sum_{i=1}^{r}\left(\frac{x_{ii}}{N}\right)$$

$$\theta_2 = \sum_{i=1}^{r}\left(\frac{x_{i+}x_{+i}}{N^2}\right)$$

$$\theta_3 = \sum_{i=1}^{r}\left(\frac{x_{ii}(x_{i+}+x_{+i})}{N^2}\right)$$

$$\theta_4 = \sum_{i,j=1}^{r}\left(\frac{x_{ij}(x_{i+}+x_{+i})^2}{N^3}\right)$$

Where: N is the product of marginal totals

r is the number of rows and columns in error matrix

x_{ii} is the number of observations in row 1 and column 1

x_{i+} is the marginal total of row I

x_{+1} is the marginal total of column I

x_{ij} is the value in the error matrix at the intersection of the *i*th row and the *j*th column.

A null hypothesis can be set up to test whether the κ values of two images differ:

Ho: $\kappa_1 = \kappa_2$ versus Ha: $\kappa_1 \neq \kappa_2$

The null hypothesis is rejected using the 'normal curve deviate statistic' (z) for α=0.05 if z_t > 1.96. z is calculated by :

$$Z = \frac{K_1 - K_2}{\sqrt{\sigma(K_1)+\sigma(K_2)}}$$

... Equation 4.3

Where:

κ is the Kappa value for the image produced

$\sigma(\kappa)$ is the asymptomatic variance of Kappa for the image produced

If the calculated z value exceeds z_t > 1.96 (at α=0.05), one can reject the null hypothesis and conclude that there is a significant difference between the two images.

Table 4.10 The normal curve deviate statistic (Z) calculated for each classification of Site 1

Study site 1	Classification Algorithm	maximum likelihood	minimum distance	parallel piped	neural network
	maximum likelihood				
	minimum distance	0.9887			
	parallel piped	0.6432	1.6218		
	neural network	3.0452	3.9819	2.3775	

Table 4.11 The normal curve deviate statistic (Z) calculated for each classification of Site 2

Study Site 2	Classification Algorithm	maximum likelihood	minimum distance	parallel piped	neural network
	maximum likelihood				
	minimum distance	1.0752			
	parallel piped	0.2649	1.3284		
	neural network	2.0308	0.7813	2.2884	

Table 4.12 The normal curve deviate statistic (Z) calculated for each classification of Site 3

Study Site 3	Classification Algorithm	maximum likelihood	minimum distance	parallel piped	neural network
	maximum likelihood				
	minimum distance	1.7172			
	parallel piped	0.0013	1.7129		
	neural network	4.9844	7.2320	4.9676	

Tables 4.10, 4.11 and 4.12 represent the calculated z values for each site, comparing each classification algorithm. The z values for which the null hypothesis was rejected are shaded. In site 1, as in site 3 the images produced using the neural networks and those produced using the traditional classification techniques were found to be significantly different, although the images produced using the traditional classification techniques were found to be similar to one another. Study site 2 presents a slightly different situation, where the only significant differences found were between the neural network and maximum likelihood images and the neural network and parallel piped images.

Chapter 5 will discuss these and other results obtained and presented within Chapter 4.

Chapter 5. DISCUSSION

This chapter will discuss and expand upon the results obtained in this study, as reported in Chapter 4. The discussion will follow the themes defined in Chapter 1, namely; the identification and mapping of *Acacia mearnsii*, the comparison of traditional and artificial neural network classifiers, and finally, the comparison of environments of varying complexity.

5.1 THE IDENTIFICATION AND MAPPING OF *ACACIA MEARNSII*

According to Tsai and Chen (2004), remote sensing provides a fast and cost effective means for identifying and mapping invasive plant species. This study attempted to test this notion, posing the question; can *Acacia mearnsii* be identified and mapped accurately using spectral reflectance from satellite images?

In Section 4.1, traditional classifiers, namely; maximum likelihood, minimum distance and parallel piped, were performed on each of the three study sites. An error matrix was then produced for each of the classifications, allowing for the calculation of various measurements of accuracy including overall accuracy. Overall accuracy is the ratio of the total number of correctly classified pixels to the total number of pixels in each class. Skidmore (1999), suggests that the minimum level of interpretation accuracy should be at least 85 per cent. Based on this assumption, as seen in table 4.9, the minimum distance classification of Study site 2 (88.37%) was the only traditional classification algorithm which did not fail. It can therefore be deduced that the traditional classifications of Study sites 1 and 3, as well as the maximum likelihood and parallel piped classifications of site 2, failed in the task of identifying and mapping *Acacia mearnsii*.

In Section 4.2 an MLP neural network was run on each of the study sites. In each case the overall accuracy surpassed 85 per cent (see Table 4.9). It can therefore be said that each of the neural networks performed were successful in the identification and mapping of *Acacia mearnsii*, proving that the species can be identified and mapped accurately using spectral reflectance from satellite images in combination with a stochastic neural network model.

5.2 TRADITIONAL CLASSIFICATION *VS* ARTIFICIAL NEURAL NETWORKS

Traditional classification methods, such as the maximum likelihood and minimum distance classifiers, have primarily been used to perform classification of remote-sensing images in the past. However, in recent years, the use of neural networks has grown in popularity (Serpico *et.al*, 1996). According to Linderman *et.al.*, (2004) neural networks have been used in the processing of multispectral images with far greater success in terms of accuracy when compared to that of the traditional statistical methods. The reason for this lies in the very nature of the Neural Network. A single neuron is able to simulate a multivariate linear regression model with no *priori* assumptions about the data distribution due to its non-parametric operation (Linderman *et.al*, 2004; Sunar Erbek *et.al*, 2004). These neurons are arranged in layers and perform as non-linear simulators. Neural Networks are also able to learn, making the classification objective. In addition, ancillary data can be used in the classification, shifting the focus to spatial elements within the image, while traditional statistical classifiers focus on the spectral information within the image (Linderman *et.al*, 2004; Mutanga and Skidmore, 2004; Qiu and Jensen, 2004; Sunar Erbek *et.al*, 2004; Mutanga and Skidmore, 2007).

On the other hand, some difficulties arise when using neural networks, related to the choice of the neural model and of the network architecture upon which classification results depend, for example, the number of hidden layers, the number of neurons per layer etc. (Serpico *et.al*, 1996). The MLP neural network used in this study is considered to be the easiest to understand and the most commonly used. However, it does posses certain weaknesses: it may be very slow to converge properly on an error minimum, it may be trapped in a local minimum, and the parameters that control the training process may be difficult to set (Mas and Flores, 2008). Nevertheless, the neural network applied in this study performed well. The reason for this lies in the very nature of the Neural Network. A single neuron is able to simulate a multivariate linear regression model with no *priori* assumptions about the data distribution due to its non-parametric operation (Linderman *et.al*, 2004; Sunar Erbek *et.al*, 2004). These neurons are arranged in layers and perform as non-linear simulators. Neural Networks are also able to learn, making the classification objective. In addition, ancillary data can be used in the classification, shifting the focus to spatial elements within the image, while traditional statistical classifiers focus on the spectral information within the image (Sunar Erbek *et.al*, 2004; Linderman *et.al*, 2004; Qiu and Jensen, 2004; Mutanga and Skidmore, 2004; Mutanga and Skidmore, 2007).

Section 4.3 deals with the comparison of the traditional classification algorithms with the neural networks for each study site. In Study site 1, the neural network out-performed the traditional classifications in terms of both overall accuracy and Kappa value. Study site 2 and 3 present a similar case, as seen in Table 4.9. In Study sites 1 and 3 the neural networks were the only algorithms which were considered successful as each surpassed the minimum level of interpretation accuracy of 85 per cent.

In order to determine whether or not the classified images were significantly different, the normal curve deviate statistic was calculated, the results of which can be seen in Tables 4.10, 4.11 and 4.12. In Study site 1, as in site 3, the neural networks were found to be significantly different to each of the traditional classifiers. An example of the difference between classified images can be seen in Figure 5.1, where the parallel piped image (the most accurate of the traditional classifications) is compared to the neural network image of site 1. The extent to which the two images differ is statistically verified by a z factor of 2.3775. Study site 2 presented a similar case, although the image produced by the minimum distance algorithm was not found to be significantly different to that produced by the neural network (see Figure 5.2).

The conclusion was therefore reached that within each study site; neural networks classified the SPOT 5 images with far greater success in terms of accuracy when compared to that of the traditional statistical techniques.

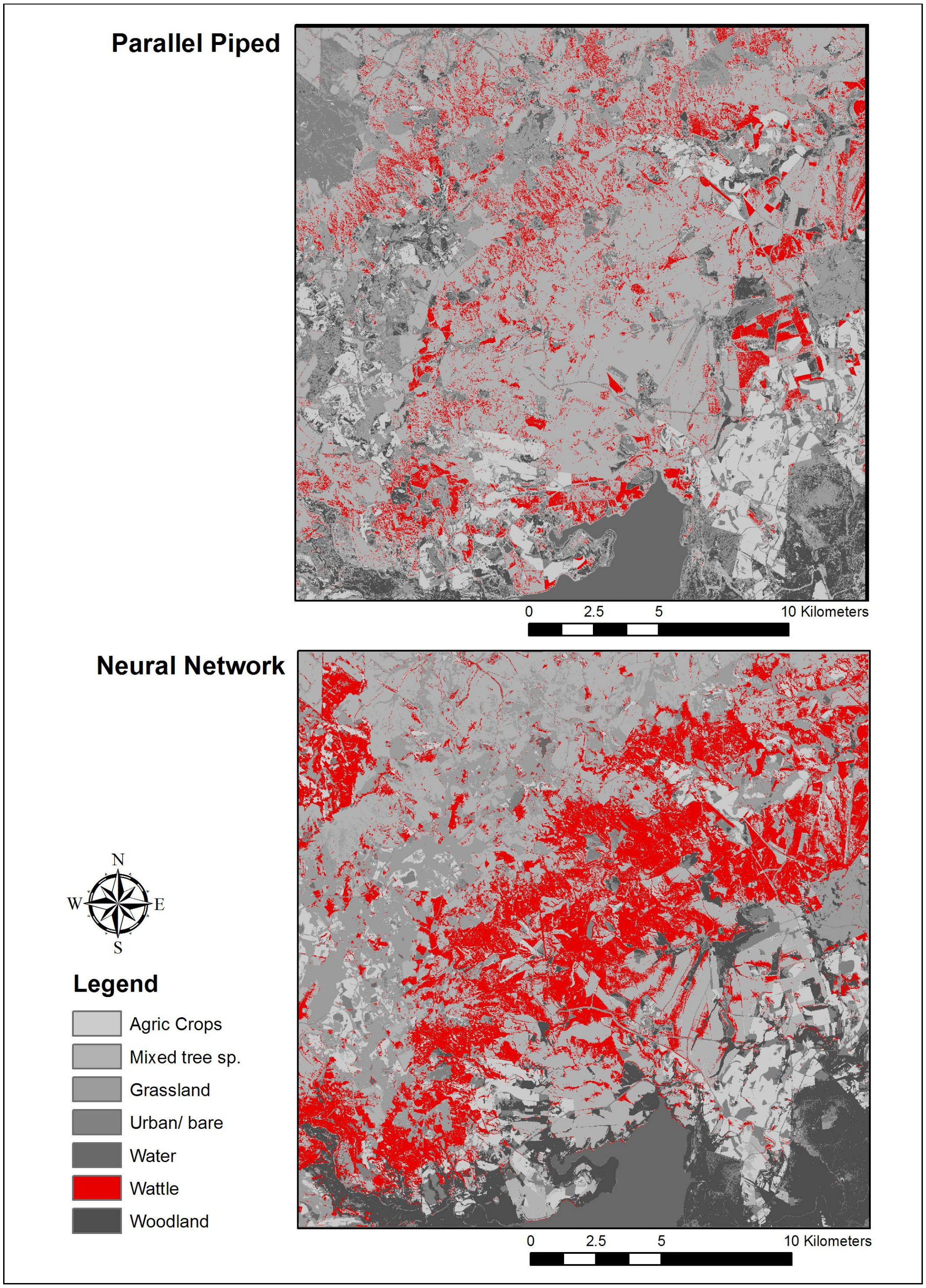

Figure 5.1 Comparison of the images produced using the parallel piped technique and artificial neural network for Study site 1

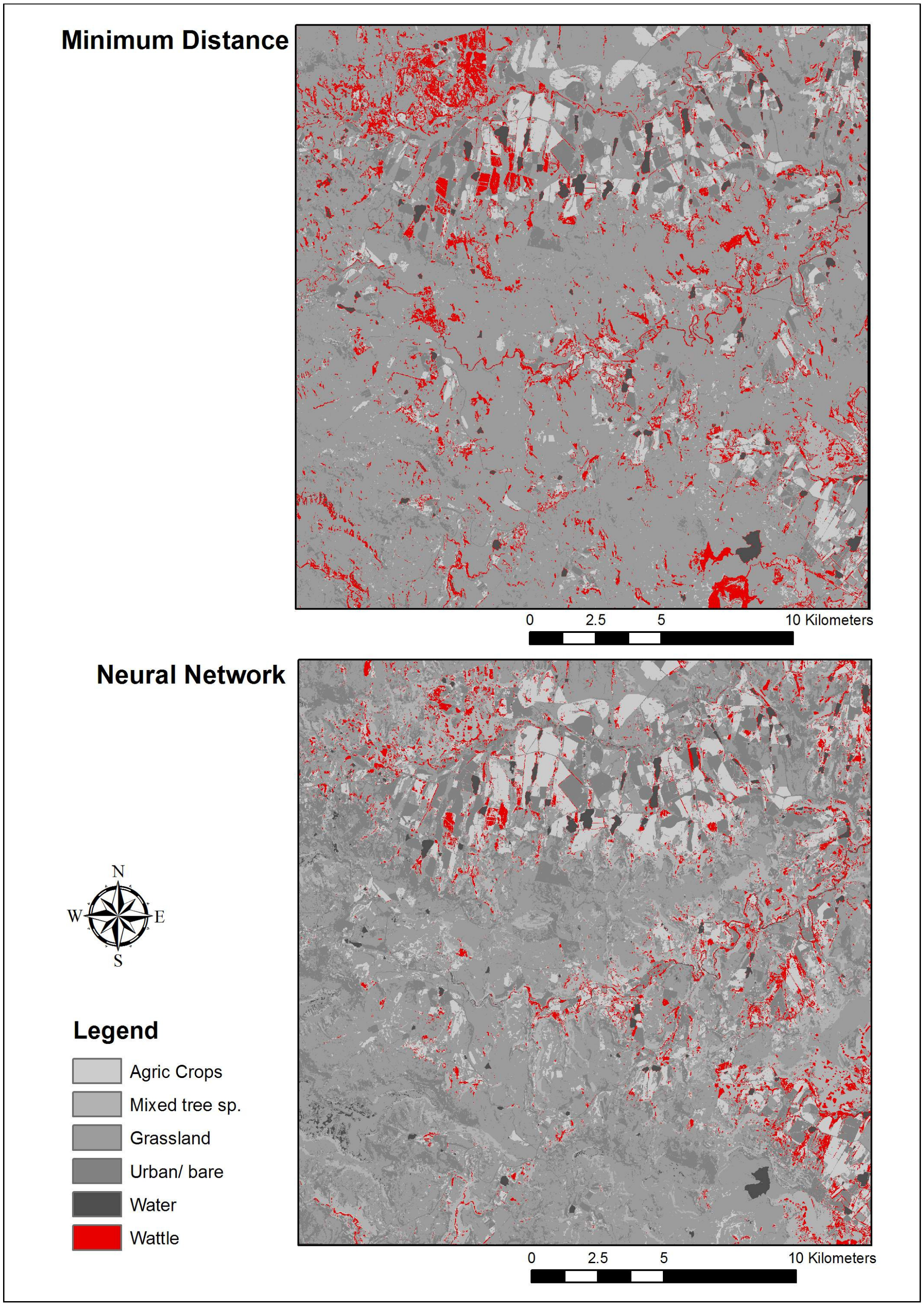

Figure 5.2 Comparison of the images produced using the minimum distance technique and artificial neural network for Study site 2

5.3 COMLEXITY OF THE ENVIRONMENTAL SETTING

The accuracy of a classification is dependent on the spatial resolution of the imagery being used (Ju *et.al*, 2005), as well as the nature of the environment being investigated (Rocchini, 2007). This is due to the fact that spatial resolution can determine how much generalization can occur within and around a feature: some features may be very detailed at one scale, but may be generalized at another (Lilllesand *et.al*, 2004). As the spatial resolution of an image becomes coarser, so the amount of spectral mixing increases (Ju *et.al*, 2005). A study performed by Markham and Townshend (1981) determined that the accuracy of a classification is governed by two factors; namely the amount of pixels falling on the boundary of features and spectral variation of classes. Mixed pixels, or the number of pixels that contain more than one classes, increases as the spatial resolution of the image increases and therefore decreases as the resolution of an image becomes finer. The spectral variation within a class increases with the increase in the resolution of an image. The spectral separability of the classes is therefore reduced creating problems in determining the nature of the class (Allan, 2007). The potential of remotely sensed images for vegetation monitoring is therefore reduced in fragmented landscapes, where most of the pixels are composed of a mixture of different classes (Busetto *et.al*, 2008)

In order to determine whether the methods used in this study could be applied to environments of varying complexity based on the spatial resolution of the SPOT 5 imagery (10m), two study sites were chosen. The first study site (as seen in Figures 5.3 and 5.5), contained large areas of single classes, while site 2 was more complex, containing numerous classes within relatively small areas (see Figures 5.4 and 5.5). During the ground truthing phase of the study, *Acacia mearnsii* was generally found in a plantation situation, where stands of the species alone covered tens of meters squared. In stark contrast to this, only small, fragmented patches of *Acacia mearnsii* mixed with other tree species were found in site 2. Study site 2 therefore presented two possible problems, the first being the selection of training sites and the second being spectral mixing occurring within pixels.

Due to the nature of the complexity of Study site 2, far fewer pixels were able to be incorporated into the signatures generated for each class than for Study site 1. This means that the classifications were performed based on fewer samples, which may lead to a lower accuracy (Skidmore, 1999). This problem with training site selection was exacerbated in many cases by possible spectral mixing in pixels due to the understory and encroachment of other species into a single pixel. *Acacia mearnsii* was often found in mixed species stands along with a dense understory, both of which affect the reflectance values of a single pixel. On the other hand, the effect of spectral mixing can be seen in both sites 1 and 2, as the error matrices of each site depict clear confusion between the 'wattle' and 'mixed tree species' classes in each of the classifications.

These factors account for the lower user accuracies of Study site 2 in each of the classifications. The user accuracy is the probability that a pixel classified on an image actually represents that category on the ground (Skidmore, 1999). The user accuracy therefore indicates how distinctive a class is, the lower the percentage, the less distinctive a class is (Everitt *et.al*, 2005).

The producer accuracies, on the other hand, of Study site 2 are higher than those seen in Study site 1 (see Table 4.9). The producer accuracy is the probability of a referenced pixel being correctly classified (Skidmore, 1999).

In terms of the overall accuracies and Kappa values, the performance of the study sites varied depending on the classification algorithm used, as seen in Table 4.9. In terms of the traditional classifications, the parallel piped algorithm performed best in site 1 (83.10%), while the maximum likelihood algorithm performed best in Study site 2. The reasons for this may be linked to the way in which the classifications function. The parallel piped classification technique uses a set of digital number ranges to create "boxes" that define the classes of the classification, classifying those pixels accordingly (Lillesand *et.al*, 2004). The maximum likelihood classification technique, on the other hand, evaluates the probability of a pixel occurring within a class, if the probability is high, the pixel is classified accordingly, if it is low, the classification process continues until it is classified (Lillesand *et.al*, 2004).

Although the results of these traditional classifications were high, the neural networks out-performed the traditional classifications in each study site. The neural network classification results were therefore compared. The overall accuracy for the Study site 1 neural network (94.46%) was a little higher than that of the Study site 2 neural network (92.79%), indicating the classifications performed slightly better on the less complex site.

Study site 3 was created by performing a mosaic on Study site 1 and 2, as described in Chapter 3, providing a means of running the classifications on the two sites simultaneously. In general the Study site 3 classifications performed worse in terms of overall accuracy than both Study sites 1 and 2. On average the overall accuracy of the traditional classifications of site 3 dropped between 8.32% and 15.09% when compared to those of Study sites 1 and 2, with the exception of the minimum distance classification which dropped significantly by 31.63% when compared to the same classification performed on Study site 2. In terms of the neural networks however, Study site 3 had an overall accuracy of 93.45%, while Study site 1 had 94.46% and Study site 2 had 92.79%, indicating the neural networks performed very well on each of the sites - an interesting result as it shows just how robust the neural network classification algorithms are.

It was therefore determined that the methods used in this study can in fact be applied to environments of varying complexity.

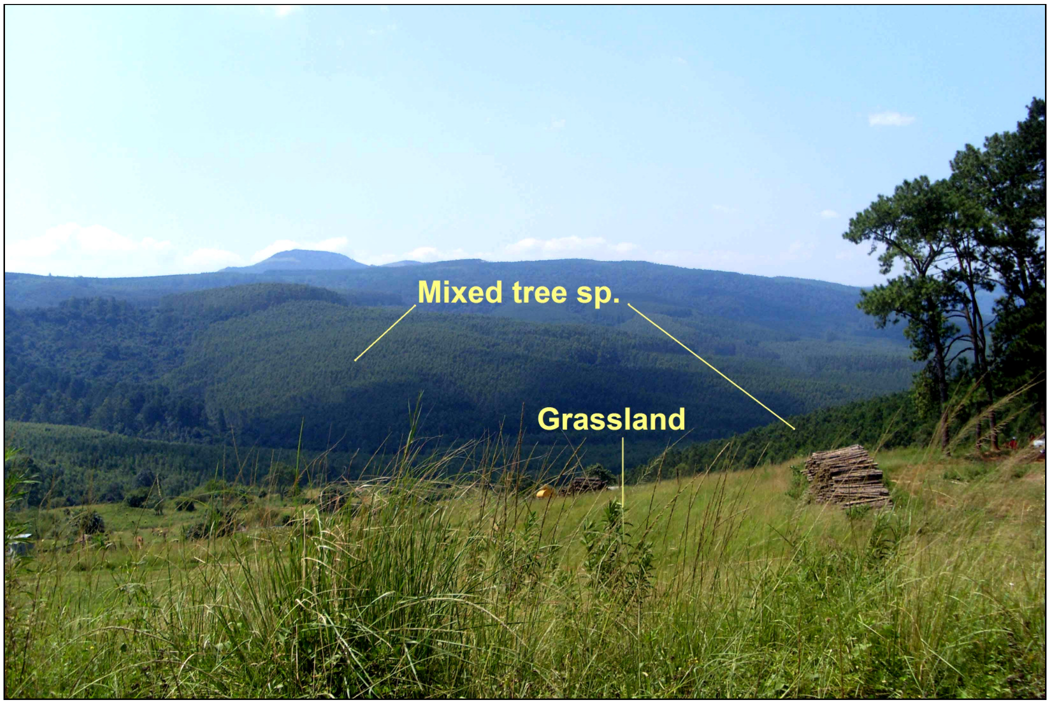

Figure 5.3 Photograph taken during the ground truthing of Study site 1, showing vast expanses of single classes.

Figure 5.4 Photograph taken during the ground truthing of Study site 2, showing a number of classes exhibited within a small area

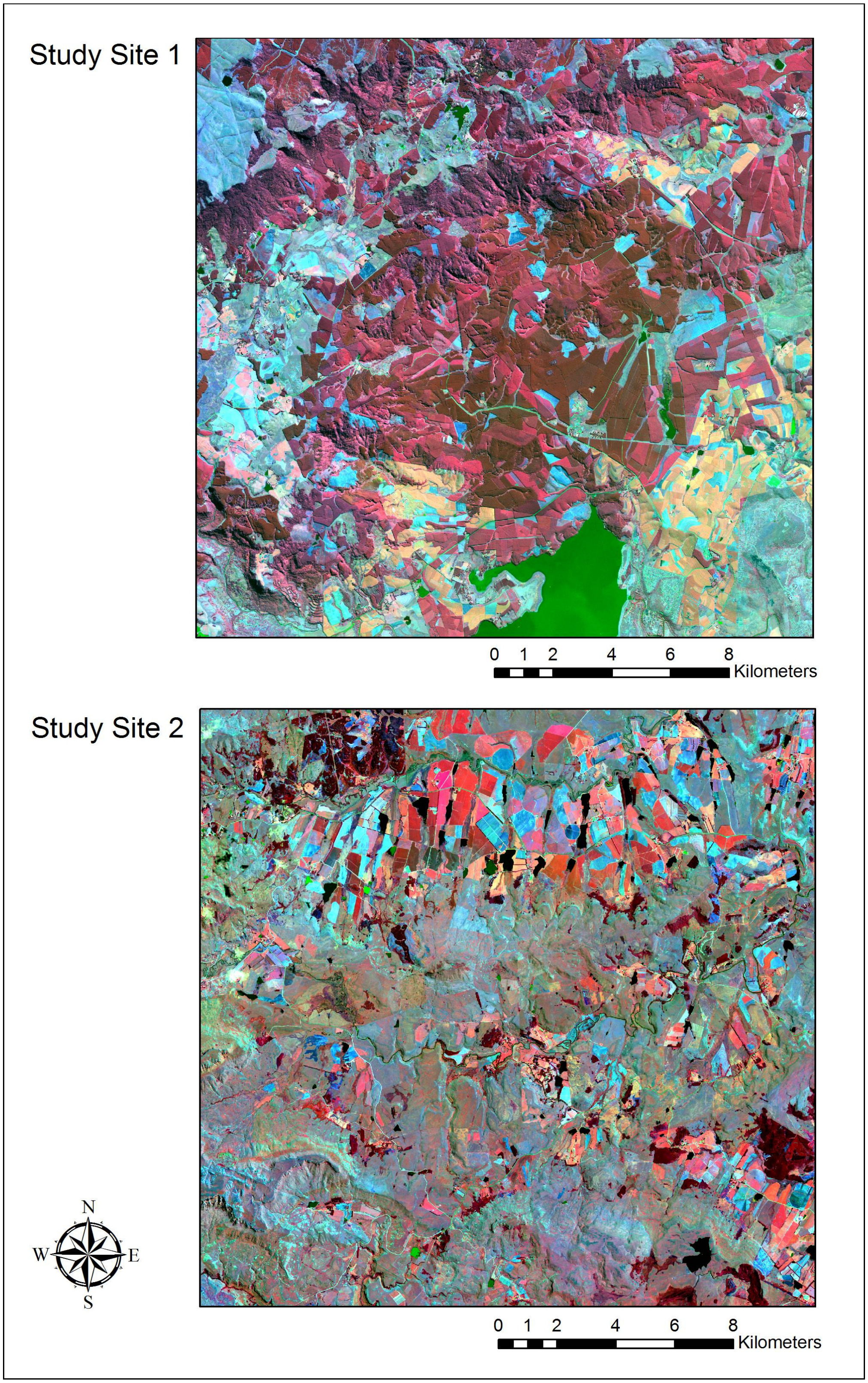

Figure 5.5 Comparison of the complexity between Study site 1 and Study site 2 (displayed as red, green and blue represent bands 3, 1 and 4 respectively).

Chapter 6. CONCLUSION

In this chapter, the aims and objectives outlined in Chapter 1 will be reviewed, examining how close this study came to reaching the set goals. The limitations experienced as well as some recommendations for future studies within this branch of remote sensing will be discussed.

6.1 AIMS AND OBJECTIVES REVIEWED

6.1.1 Aims

The aim of this study was to investigate the utility of SPOT 5 imagery and Artificial Neural Networks, in the identification and mapping of *Acacia mearnsii* within environments of varying complexity.

6.1.2 Objectives

Five objectives were set in order to meet the previously mentioned aim. In this section, how close the study came to meeting the said objectives will be reviewed.

> **1. Determine the environmental variables influencing the occurrence and distribution of *Acacia mearnsii* for use within an Artificial Neural Network.**

The literature found and discussed within Chapter 2, Section 2.2 defined the environmental variables influencing the occurrence and distribution of *Acacia mearnsii.* In brief, these environmental variables were as follows:

- In South Africa, *Acacia mearnsii* grows in altitudes between 600 and 1700 m (Sherry, 1971; IUCN/SSC Invasive Species Specialist Group, 2006).
- It has a tendency to grow in climates from warm temperate to moist tropical with precipitation means between 660 and 2280 mm per annum (Sherry, 1971; IUCN/SSC Invasive Species Specialist Group, 2006).
- *Acacia mearnsii* favours mean annual temperatures of between 14.7℃ and 27.8℃ (Sherry, 1971; IUCN/SSC Invasive Species Specialist Group, 2006).
- *Acacia mearnsii* thrives on soils with a ph between 5.0 and 7.2, however does not do well on very dry or infertile soils (Sherry, 1971; IUCN/SSC Invasive Species Specialist Group, 2006).

- It is generally found in dense clusters in riparian zones, grasslands, agricultural, urban, and disturbed areas (Sherry, 1971)

The ancillary data used in this study was chosen based on these environmental variables influencing the spatial distribution of *Acacia mearnsii*, as well as data availability. The ancillary data therefore included the elevation, slope, aspect, water sources and soil type.

2. Classify and map *Acacia mearnsii* using SPOT 5 images in combination with Artificial Neural Networks and traditional classifiers

Traditional classifiers, namely; maximum likelihood, minimum distance and parallel piped, were performed on each of the three study sites. An error matrix was then produced for each of the classifications, allowing for the calculation of various measurements of accuracy including overall accuracy. Skidmore (1999) suggests that the minimum level of interpretation accuracy should be at least 85 per cent. Based on this assumption, the minimum distance classification of Study site 2 (88.37%) was the only traditional classification algorithm which did not fail, with the maximum likelihood (79.07%) and parallel piped (76.74%) classifications falling below the suggested minimum level of interpretation accuracy. It was therefore deduced that the traditional classifications of Study sites 1 and 3, as well as the maximum likelihood and parallel piped classifications of Study site 2, failed in the task of identifying and mapping *Acacia mearnsii*.

In Section 4.2 an MLP neural network was run on each of the study sites. In each case the overall accuracy surpassed 85 per cent. Each of the neural networks performed were therefore successful in the identification and mapping of *Acacia mearnsii*, proving that the species can be classified and mapped accurately using spectral reflectance from SPOT 5 imagery in combination with a stochastic neural network model.

3. Compare the accuracy of Artificial Neural Networks with that of traditional classifiers, i.e. Maximum Likelihood

Section 4.3 dealt with the comparison of the traditional classification algorithms with the neural networks for each study site. In Study site 1, the neural network out-performed the traditional classifications in terms of both overall accuracy and Kappa value. Study sites 2 and 3 presented a similar case. In Study sites 1 and 3 the neural networks were the only algorithms which were considered successful as each surpassed the minimum level of interpretation accuracy of 85 per cent.

In order to determine whether or not the classified images were significantly different, the normal curve deviate statistic was calculated. In Study site 1, as in Study site 3, the neural networks were found to be significantly different to each of the traditional classifiers. The extent to which the two images differ is statistically verified by a z factor of 2.3775. Study site 2 presented a similar case, although the image produced by the minimum distance algorithm was not found to be significantly different to that produced by the neural network.

The conclusion was therefore reached that within each study site; neural networks classified the SPOT 5 images with far greater success in terms of accuracy when compared to that of the traditional statistical techniques.

4. Determine whether the methods used in this study could be applied to environments of varying complexity based on the spatial resolution of the SPOT 5 imagery (10m).

In order to determine whether the methods used in this study could be applied to environments of varying complexity based on the spatial resolution of the SPOT 5 imagery (10m), two study sites were chosen. The first study site, contained large areas of single classes, while site 2 was more complex.

In terms of the overall accuracies and Kappa values, the performance of the study sites varied depending on the classification algorithm used. In the traditional classifications, the parallel piped algorithm performed best in site 1 (83.10%), while the maximum likelihood algorithm performed best in site 2.

Although the results of these traditional classifications were high, the neural networks out-performed the traditional classifications in each study site, and the neural network classification results were therefore compared. The overall accuracy for the Study site 1 neural network (94.46%) was a little higher than that of the Study site 2 neural network (92.79%), indicating the classifications performed slightly better on the less complex site.

Study site 3 was created by performing a mosaic on site 1 and 2, as described in Chapter 3, providing a means of running the classifications on the two sites simultaneously. Site 3 had an overall accuracy of 93.45%, while site 1 had 94.46% and site 2 had 92.79%, indicating the neural networks performed very well on each of the sites, an interesting result as it shows just how robust the neural network classification algorithms are.

It was therefore determined that the methods used in this study can in fact be applied to environments of varying complexity.

> **5. Investigate the utility of remote sensing for environmental management in terms of identifying alien invasive vegetation on a large scale.**

Species invasions threaten endangered species, and, after habitat loss, pose the second largest threat to biodiversity (Higgins *et.al.*, 1999; de Wit *et.al*, 2001; Goodman, 2003). Prevention is better than cure; however numerous pathways of invasion make the interception of all species unrealistic. Organizations and individuals attempting to manage invasive species often have limited budgets and insufficient information and tools. According to Barnet *et.al* (2006), the answer to this is efficient management tools which address the multiple phases of plant invaders on a large scale. Early detection and the development of a spatially explicit management strategy is therefore key to improving efforts to prevent the establishment of invasive species (Higgins *et.al.*, 1999; Barnet *et.al*, 2006).

To understand alien plant invasions at local and regional scales, organizations and programs such as Working for Water have developed systems for mapping and compiling non-native plant species information (Cobbing, 2006). Mapping has the potential to record which, how much, and where alien invasive plant species exist on a landscape, and, when implemented over time and space, monitor patches of weeds, help predict the spread of species, facilitate the exchange of data between agencies, and increase public and political awareness (Barnet *et.al*, 2006).

Remote sensing provides a fast and cost effective means for identifying and mapping invasive plant species as opposed to field based investigations, especially on a large scale. The cost of remote sensing is related to economies of scale; the larger the survey area, the cheaper the cost per acre (Everitt *et.al*, 2005). It is for this reason the application of remote sensing technology has received considerable interest in the field of plant invasion in recent years and is now widely used around the world for collecting and processing data (Tsai and Chen , 2004; Mas and Flores, 2008).

6.2 LIMITATIONS AND RECOMMENDATIONS

This section will outline some of the limitations experienced during this study. Where necessary, recommendations have been made in order to correct these identified problem areas. It must be noted that this study was undertaken as close to accepted practice as possible, however some improvements can be made to the methodology.

6.2.1 Resolution

SPOT 5 imagery has been used extensively in field of vegetation identification. In a study conducted by Cobbing (2006), where the use of Landsat ETM imagery was investigated as a suitable data capture source of alien *Acacia* (*Acacia mearnsii and Acacia dealbata*), SPOT 5 was recommended. In this study the 'medium resolution' of SPOT 5 was in fact found to have disadvantages in terms of problems with spectral mixing and the definition of boundaries, especially in the more complex site 2. *Acacia mearnsii* was often found in mixed species stands along with a dense understory, both of which affect the reflectance values of a single pixel and therefore the accuracy of the classification. The use of the pan-enhanced 2.5m SPOT 5 imagery has been recommended (Pasqualini *et.al*, 2005) to solve these problems, and was considered for use within this study, however the limited budget did not allow for this. If the pan-enhanced 2.5m SPOT 5 imagery was used, it is likely that the results from Study site 2 would improve reinforcing the findings of this study, namely; that the neural network is a particularly robust algorithm and can perform well in both relatively simple and complex environments.

6.2.2 Data collection

A number of limitations were experienced during the data collection phase of the process. This was generally a result of questionable class definition and inaccessibility.

The first limitation, experienced early on in the data collection process, related to the definition and assignment of classes. During the data collection for Study site 2, for example, it was found that the area had been largely affected by human inhabitants, in that very few areas of natural indigenous bush remained, with commercial timber species being used as windbreaks. These were found in mixed species 'pockets'. These areas could therefore not be defined as natural bush nor commercial forestry. As a result, a decision was made to combine these two classes to form a class 'mixed tree species'.

Inaccessibility was also found to be a problem. Study site 1, for example, is predominantly under commercial forestry and therefore private property. Much of the central region of the site was closed off and access denied. Consequently, attempts had to be made to work around these regions. In Study site 2, on the other hand, mountainous terrain and few roads limited access to much of the western portion of the site. This was however, offset by the fact that such areas were, to a large extent, grassland and points were later added based on observations made using the SPOT 5 image and 1: 50 000 topographical map data.

6.2.3 Signature creation and classification

Limitations were experienced during the classification phase of the process, particularly in terms of signature creation. Attempts were made during the random generation and collection of training points, to show no bias towards any single class. The same, however, cannot be said for the signature creation. Certain classes are obvious and cover large areas, for example; water in Study site 1 and grassland in Study site 2. The signatures created for these classes were therefore much larger than for urban/ bare for example, a factor which can play a major role in the accuracy of the classification. It is therefore recommended that a greater number of ground control points be used to allow for more training sites and points for accuracy assessment.

6.3 CONCLUDING REMARKS

The impacts of invasive species on the environment, human health, and the economy continue to gain interest from public and private agencies, scientists, and the media. Species invasions threaten endangered species, and pose the second largest threat to biodiversity after habitat loss (Higgins *et.al.*, 1999; de Wit *et.al*, 2001; Goodman, 2003). Outside of Australia, its adaptability to a wide range of environmental conditions and its rapid growth rate makes *Acacia mearnsii* both the perfect agricultural resource and the perfect pest (Sherry, 1971). It is for this reason; *Acacia mearnsii* was ranked as the world's worst invader in the Global Invasive Species Database' funded by La Fondation TOTAL.

Researchers whose literary work has been reviewed have suggested that remote sensing has the potential to help curb plant invasions in terms of mapping and monitoring population distributions, prioritizing areas and species, allowing for the development of spatially explicit control and management strategies (Higgins *et.al.*, 1999; Joshi, *et.al.* 2004; Tsai & Chen, 2004; Barnet *et.al*, 2006; Mas and Flores, 2008).

This study aimed to investigate the utility of SPOT 5 imagery and Artificial Neural Networks, in the identification and mapping of *Acacia mearnsii* within environments of varying complexity. The aim was achieved and it was found that it is possible to identify and map *Acacia mearnsii* using SPOT 5 imagery, depending on the classification algorithm used. It was established that the neural network algorithms performed with greater success when compared to the performance of traditional classifiers, a reality that has been advocated by numerous researchers (Sunar Erbek *et.al*, 2004; Linderman *et.al*, 2004; Qiu and Jensen, 2004; Mutanga and Skidmore, 2004; Mutanga and Skidmore, 2007).

The utility of the various classification algorithms was also investigated in terms of their applicability to environments of varying complexity. The neural networks once again, proved to be more successful and performed well in both the complex and relatively simple environments, indicating just how robust the neural network algorithm can be.

Although the neural network has proved to be a highly accurate and robust technique, the testing and training of the network can become a time consuming and therefore costly process. Traditional classifiers can therefore provide a 'quick and dirty' solution where time constraints are a problem and budgets are limited.

Chapter 7. REFERENCES

Acocks, J.H.P. (1988) *Veld Types of South Africa, 3rd Edition*. Botanical Research Institute, South Africa.

Allan, K. (2007) *Landcover Classification in a Heterogeneous Savannah Environment: Investigating the Performance of an Artificial Neural Network and the Effect of Image Resolution*. Masters Thesis. University of KwaZulu Natal, Pietermaritzburg.

Barnet, D.I., Stohlgren, T.J., Jarnevich, C.S., Chong, G.W., Ericson, J.A., Davern, T.R. & Simpson, S.E. (2007) The Art and Science of Weed Mapping. *Environmental Monitoring and Assessment*, 132 (1-3), pg. 235 – 252

Beyer, H.L. (2004) *Hawth's Analysis Tools for ArcGIS*. Available at http://www.spatialecology.com/htools. Date Accessed: 26/01/2009

Blignaut, J.N., Marais, C. & Turpie, J.K. (2007) Determining a Charge for the Clearing of Invasive Alien Plant Species (IAPs) to Augment Water Supply in South Africa. *Water SA,* 33 (1) pg. 27 – 34

Buch, A. & Dixon, A.B. (2008) South Africa's Working for Water Programme: Searching for Win-Win Outcomes for People and the Environment. *Sustainable Development,* 17 (3) pg. 129 – 141

Busetto, L., Meroni, M. & Colombo, R. (2008) Combining the Medium and Coarse Spatial Resolution Satellite Data to Improve the Estimation of Sub-Pixel NDVI Time Series. *Remote Sensing of Environment*, 3 (112) pg. 118 – 131

Cao, C. & Lam, N.S-L (1997) Understanding the Scale and Resolution Effects in Remote Sensing and GIS. In Quattrochi, D.A. & Goodchild, M.F. (Eds.) *Scale in Remote Sensing and GIS*. CRC Press, Inc. New York

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., Woodcock, C.E. & Franklin, J. (1999). A Neural Network Method for Efficient Vegetation Mapping. *Remote Sensing of Environment,* 70 pg. 326 – 338

Cobbing, B.L. (2006) *The Use of Landsat ETM Imagery as a Suitable Data Capture Source for Alien Acacia Species for the WFW Program*. Unpublished Masters Thesis. Rhodes University.

Congalton, R.G. & Green, K. (1999) *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers. New York

Conservation of Agricultural Resources Act' (Act No. 43 of 1983) www.dwaf.gov.za/wfw/Docs/Articles/CARA.doc Date Accessed: 17/09/08

Clark Labs, (2000) IDRISI Andes Help System. *Clark Labs, IDRISI Andes*. Clark University, Worcester, USA

CSIR (2002) The SADC Regional Land Cover Database Project – Project description available at *http://www.csir.co.za/plsql/ptl0002/PTL0002_PGE100_LOOSE_CONTENT?LOOSE_PAGE_NO=7020305* Date Accessed: 04/02/2009

De Wit, M.P., Crookes, D.J., & van Wilgen, B.W. (2001) Conflicts of Interest in Environmental Management: Estimating the Costs and Benefits of Tree Invasion. *Biological Invasions*, 3, 197-178

Everitt, J., Hunt, R & Hamilton, R. (2005) *A Weed Manager's Guide to Remote Sensing and GIS - Mapping & Monitoring.* USDA Forest Service Remote Sensing Applications Center, Salt Lake City, UT

Froody, G.M. & Arora, M. K. (1997) An Evaluation of Some Factors Affecting the Accuracy of Classification by Artificial Neural Networks. *International Journal of Remote Sensing*, 18 pg. 799 – 810

GAEA Projects (2001) Phase One: Part 3 – Analysis Report. *uMgungundlovu IDP*.

Goodman, P.S. (2003) Assessing Management Effectiveness, and Setting Priorities in Protected Areas in KwaZulu Natal. *Bioscience*, 53 (9), 843-850

Henderson, L. (2001) *Alien Weeds, and Invasive Plants*. Agricultural Research Council. Cape Town.

Higgins, S.I., Richardson, D.M., Cowling, R.M. & Trinder-Smith, T.H. (1999) Predicting the Landscape-Scale Distribution of Alien Plants and Their Threat to Plant Diversity. *Conservation Biology,* 13 (2) pg. 303 - 313

Hu, X., & Weng, Q. (2009) Estimating Impervious Surfaces from Medium Spatial Resolution Imagery Using the Self-organizing Map and Multi-layer Perceptron Neural Networks. *Remote Sensing of Environment*, 133 (10), pg. 2089-2102

Ibrahim, A.A. (2008) *Using Remote Sensing Technique (NDVI) for Monitoring Vegetation Degradation in Semi-arid Lands and its Relationship to Precipitation: Case Study from Libya*. 3rd International Conference on Water Resources and Arid Environments.

Images Spot © Cnes (2005) Spot Satellite Technical Data. www.spotimage.com Date Accessed: 20/07/2008

IUCN/SSC Invasive Species Specialist Group. *Acacia mearnsii.* (last modified April 2006) http://www.issg.org/database/species/ecology.asp?fr=1&si=51 Date Accessed: 11/02/2008

Joshi, C., de Leeuw, J. & van Duren, I.C. (2004) Remote Sensing and GIS Applications for Mapping and Spatial Modelling of Invasive Species. (Ed. Altan MO). In: *International Society for Photogrammetry and Remote Sensing (ISPRS)* Istanbul. Volume XXXV part B7, pg 669-677.

Joshi, C., De Leeuw, J., Van Andel, J., Skidmore, A.K., Lekhak, H.D., Van Duren, I. & Morbu, N. (2006) Indirect Remote Sensing of a Cryptic Forest Understory Invasive Species. *Forest Ecology and Management*, 225 pg. 245 – 256

Ju, J., Gopal, S. & Kolaczyk, E.D. (2005) On the Choice of Spatial and Categorical Scale in Remote Sensing Land Cover Classification. *Remote sensing of Environment*, 96 pg. 62 – 77

Kavzoglu, T. & Mather, P. M. (2003) The use of Backpropogating Artificial Neural Networks in Landcover Classification. *International Journal of Remote Sensing*, 24, pg. 4907- 4938.

KwaZulu Natal Data. Cartography Department. University of KwaZulu Natal, Pietermaritzburg. Date Received: March 2008

Linderman, M., Lui,J., Qi, J., An, L., Ouyang, Z., Yang, J. & Tan, Y. (2004) Using Artificial Neural Networks to Map the Spatial Distribution of Understory Bamboo From Remote Sensing Data. *International Journal of Remote Sensing*, 25 (9), pg. 1685 - 1700

Lillesand, T.M., Kiefer, R.W. & Chipman, J.W. (2004) *Remote Sensing and Image Interpretation (5th Ed.)*. New York, John Wiley and Sons, Inc.

Lu, D., Moran, E. & Batistella, M (2003) Linear Mixture Model Applied to Amazonian Vegetation Classification. *Remote Sensing of Environment*, 87, pg. 456 – 469

Macdonald, I.A.W. (2004) Recent Research on Alien Plant Invasions and Their Management in South Africa: a Review of the Inaugural Research Symposium of the Working for Water Programme. *South African Journal of Science,* 100 (1-2), pg. 21 – 26

Mas, J.F., & Flores, J.J. (2008) The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data. *International Journal of Remote Sensing*, 29 (3), pg. 617 — 663

Markham, B. L. & Townshend, J. R. G. (1981). *Land cover accuracy as a function of sensor spatial resolution*. Proc. 15th Int. symp. Remote sensing Environment, Ann. Arbor.

Milton, S.J., Dean W.R.J. & Richardson, D.M. (2003) Economic Incentives for Restoring Natural Capital in Southern African Rangelands. *Frontiers in Ecology and Environment,* 1 (5), pg. 247- 254.

Mutanga, O & Skidmore, A.K. (2004) Integrating Imaging Spectroscopy and Neural Networks to Map Grass Quality in the Kruger National Park, South Africa. *Remote Sensing of Environment*, 90 pg. 104 – 115

Mutanga, O & Skidmore, A.K. (2007) Fuelwood Distribution Pattern in North Western Zimbabwe is Modelled Using a Neural Network. *International Journal of Geoinformatics*, 3 (2), pg. 21-28

Nagendra, H. & Rocchini, D. (2008) High Resolution Satellite Imagery for Tropical Biodiversity Studies: the Devil is in the Detail. *Biodiversity Conservation,* 17, pg. 3431–3442

Nyoka, B.I. (2003). *Biosecurity in forestry:* a *case study on the status of invasive forest trees species in Southern Africa.* Forest Biosecurity Working Paper FBS/1E. Forestry Department. FAO, Rome, Italy (*unpublished*).

Pasqualini, V., Pergent-Martini, C., Pergent, G., Agreil, M., Skoufas, G.,l Scourbes, L. & Tsirika, A. (2005) Use of SPOT 5 for mapping seagrasses: An application to Posidonia oceanica. *Remote Sensing of Environment*, 94, pg. 39-45

Qui, F. & Jensen, J. R. (2004) Opening the Black Box on Neural Networks for Remote Sensing Image Classification. *International Journal of Remote Sensing,* 25, pg. 1749 – 1768

Richardson, D.M., & van Wilgen, B.W. (2004) Invasive alien plants in South Africa: how well do we understand the ecological impacts? *South African Journal of Science*, 100, pg. 45-52

Rocchini, A. (2007) Effects of Spatial and Spectral Resolution in Estimating ecosystem α-diversity by Satellite Imagery. *Remote Sensing of Environment*, 111, pg. 423 – 434

Rowlinson, L.C., Summerton, M. & Ahmed, F. (1999) Comparison of remote sensing data sources and techniques for identifying and classifying alien invasive vegetation in riparian zones. *Water SA*, 25 (4), pg. 497-500

Serpico, S.B., Bruzzone, L. & Roli, F. (1996) An Experimental Comparison of Neural Network and Statistical Non-parametric Algorithms for Supervised Classification of Remote sensing Images. *Pattern Recognition Letters,* 17, pg. 1331 - 1341

Shackleton, C.M., McGarry, D., Fourie, S., Gambiza, J., Shackleton, S.E. & Fabricius, C. (2007) Assessing the Effects of Invasive Alien Species on Rural Livelihoods: Case Examples and a Framework from South Africa. *Human Ecology*, 35, pg. 113 – 127

Sherry, S.P. (1971) *The Black Wattle (Acacia mearnsii De Wild)*. University of Natal Press. Pietermaritzburg

Skidmore, A.K. (1999) Accuracy Assessment of Spatial Information. *Spatial Statistics for Remote Sensing*. Kluwer Academic Publishers. Netherlands pg. 197 - 209

Sunar Erbek, K., Ozkan, C. & Taberner, M. (2004) Comparison of Maximum Likelihood Classification Method with Supervised Artificial Neural Network Algorithms for Land use Activities. *International Journal of Remote Sensing,* 25 (9), pg. 1733 - 1748

The Working for Water Program Annual Report (1998/ 1999)

Theron, J.M., van Laar, A., Kunneke, A. & Bredenkamp, B.V. (2004) A Preliminary Assessment of Utilizable Biomass in Invading *Acacia* Stands on the Cape Coastal Plains. *South African Journal of Science*, 100, pg. 123 – 125

Turpie, J.K., Marais, C. & Blignaut, J.N. (2008) The Working for Water Programme: Evolution of Payments for Ecosystem Service Mechanism That Address Both Poverty and Ecosystem Service Delivery in South Africa. *Ecological Economics*, 65 (4), pg. 788 – 798

Tsai, F. & Chen, C. (2004) *Detecting Invasive Plants Using Hyperspectral and High Resolution Satellite Images.* In Proc. The XXth ISPRS Congress, Istanbul, Turkey.

Van Wilgen, B.W., Richardson, D.M., Le Maitre, D.C., Marais, C. & Magadlela, D. (2001) The Economic Consequences of Alien Plant Invasions: Examples of Impacts and Approaches to Sustainable Management in South Africa. *Environment, Development, and Sustainability*, 3, pg. 145 – 168

Van Wilgen, B.W., Reyers, B., Le Maitre, D.C., Richardson, D.M. & Schonegevel, L. (2008) A Biome-Scale Assessment of the Impact of Invasive Alien Plants on Ecosystem Services in South Africa. *Journal of Environmental Management*, 89 (4), pg. 336 – 349

Wynberg, R (2002) A Decade of Biodiversity Conservation and use in South Africa: Tracking Progress from the Rio Earth Summit to the Johannesburg World Summit on Sustainable Development. *South African Journal of Science*, 98, pg. 233 - 243

Chapter 8. APPENDIX

The following pages will display some of the error matrices which resulted from this study.

Tables

Table 1. Error matrix for the accuracy assessment of the classification of Study site 1 using the Minimum Distance algorithm

Classified Data	Mixed tree	Woodland	Urban/ bare	Water	Grassland	Wattle	Agric. Crops	Row Total	User Accuracy
Mixed tree	5	0	0	0	0	10	0	15	33.33%
Woodland	0	1	0	0	1	0	0	2	50.00%
Urban/ bare	0	0	7	0	1	0	0	8	87.50%
Water	0	0	0	8	0	0	0	8	100.00%
Grassland	0	2	1	0	11	0	0	14	78.57%
Wattle	4	0	0	0	0	9	1	14	64.29%
Agric. crops	0	0	0	0	0	0	10	10	100.00%
Column Total	9	3	8	8	13	19	11	71	
Producer Accuracy	55.56%	33.33%	87.50%	100.00%	84.62%	47.37%	90.91%		

Overall Accuracy = 71.83%

Overall Kappa Statistics = 0.6630

Table 2. Error matrix for the accuracy assessment of the classification of Study site 1 using the Parallel Piped algorithm

Classified Data	Mixed tree	Woodland	Urban/ bare	Water	Grassland	Wattle	Agric. Crops	Row Total	User Accuracy
Mixed tree sp	9	0	0	0	0	7	0	16	56.25%
Woodland	0	2	0	0	1	0	0	3	66.67%
Urban/ bare	0	0	8	0	3	0	0	11	72.73%
Water	0	0	0	8	0	0	0	8	100.00%
Grassland	0	1	0	0	9	0	0	10	90.00%
Wattle	0	0	0	0	0	12	0	12	100.00%
Agric. crops	0	0	0	0	0	0	11	11	100.00%
Column Total	9	3	8	8	13	19	11	71	
Producer Accuracy	100.00%	66.67%	100.00%	100.00%	69.23%	63.16%	100.00%		

Overall Accuracy = 83.10%

Overall Kappa Statistics = 0.7999

Table 3. Error matrix for the accuracy assessment of the classification of Study site 2 using the Minimum Distance algorithm

Classified Data	Urban/ bare	Wattle	Mixed tree	Grassland	Agric. crops	Water	Row Total	User Accuracy
Urban/ bare	5	0	0	0	0	0	5	100.00%
Wattle	0	3	1	0	0	0	4	75.00%
Mixed tree sp	0	2	13	0	0	0	15	86.67%
Grassland	0	0	0	8	2	0	10	80.00%
Agric. crops	0	0	0	0	5	0	5	100.00%
Water	0	0	0	0	0	4	4	100.00%
Column Total	5	5	14	8	7	4	43	
Producer Accuracy	100.00%	60.00%	92.86%	100.00%	71.43%	100.00%		

Overall Accuracy = 88.37%

Overall Kappa Statistics = 0.8530

Table 4. Error matrix for the accuracy assessment of the classification of Study site 2 using the Parallel Piped algorithm

Classified Data	Urban/ bare	Wattle	Mixed tree	Grassland	Agric. crops	Water	Row Total	User Accuracy
Urban/ bare	5	0	0	0	0	0	5	100.00%
Wattle	0	4	7	0	0	0	11	36.36%
Mixed tree	0	1	7	0	1	0	9	77.78%
Grassland	0	0	0	7	0	0	7	100.00%
Agric. crops	0	0	0	1	6	0	7	85.71%
Water	0	0	0	0	0	4	4	100.00%
Column Total	5	5	14	8	7	4	43	
Producer Accuracy	100.00%	80.00%	50.00%	87.50%	85.71%	100.00%		

Overall Accuracy = 76.74 %

Overall Kappa Statistics = 0.7175

Table 5. Error matrix for the accuracy assessment of the classification of the Study site 3 using the Minimum Distance algorithm

Classified Data	Mixed tree	Woodland	Grassland	Water	Urban/ bare	Agric crops	Wattle	Row Total	User Accuracy
Mixed tree	10	0	2	0	0	4	13	29	34.48%
Woodland	0	3	10	0	0	2	0	15	20.00%
Grassland	1	0	5	0	1	1	0	8	62.50%
Water	6	0	0	12	0	0	0	18	66.67%
Urban/ bare	0	0	1	0	12	0	0	13	92.31%
Agric. crops	1	0	1	0	0	11	0	13	84.62%
Wattle	5	0	2	0	0	0	11	18	61.11%
Column Total	23	3	21	12	13	18	24	114	
Producer Accuracy	43.48%	100.00%	23.81%	100.00%	92.31%	61.11%	45.83%		

Overall Classification Accuracy = 56.14%

Overall Kappa Statistics = 0.4849

Table 6. Error matrix for the accuracy assessment of the classification of the Study site 3 using the Parallel Piped algorithm

Classified Data	Mixed tree	Woodland	Grassland	Water	Urban/ bare	Agric crops	Wattle	Row Total	User Accuracy
Mixed tree sp.	8	0	2	0	0	2	9	21	38.10%
Woodland	0	3	2	0	0	0	0	5	60.00%
Grassland	1	0	13	0	0	1	1	16	81.25%
Water	0	0	0	12	0	0	0	12	100.00%
Urban/ bare	0	0	0	0	13	0	0	13	100.00%
Agric. crops	0	0	2	0	0	15	0	17	88.24%
Wattle	14	0	2	0	0	0	14	30	46.67%
Column Total	23	3	21	12	13	18	24	114	
Producer Accuracy	34.78%	100.00%	61.90%	100.00%	100.00%	83.33%	58.33%		

Overall Classification Accuracy = 68.42%

Overall Kappa Statistics = 0.6208

Table 7. Error matrix for the accuracy assessment of the ANN classification for Study site 2

Classified Data	Agric. Crops	Mixed Tree sp.	Grassland	Urban/ bare	Water	Wattle	Row Total	User Accuracy
Agric. Crops	15767	34	139	22	0	59	16021	98.41%
Mixed Tree sp.	157	2554	60	8	0	90	2869	89.02%
Grassland	972	37	14133	67	4	13	15226	92.82%
Urban/ bare	414	6	113	1807	20	3	2363	76.47%
Water	0	3	0	1	9466	0	9470	99.96%
Wattle	574	658	7	0	0	832	2071	40.17%
Column Total	17884	3292	14452	1905	9490	997	48020	
Producer Accuracy	88.16%	77.58%	97.79%	94.86%	99.75%	83.45%		

Overall accuracy = 92.79%

Overall Kappa = 0.9019

Table 8. Error matrix for the accuracy assessment of the ANN classification for Study site 3

Classified Data	Agric. Crops	Mixed Tree sp.	Grassland	Urban/ bare	Water	Wattle	Woodland	Row Total	User Accuracy
Agric. Crops	21050	391	804	23	1	473	13	22755	92.51%
Mixed Tree sp.	31	12461	0	0	0	1718	0	14210	87.69%
Grassland	2944	64	29330	222	2	19	452	33033	88.79%
Urban/ bare	1115	6	432	2712	0	3	129	4397	61.68%
Water	0	1	0	6	116335	0	0	116342	99.99%
Wattle	176	2760	115	3	18	4225	21	7318	57.73%
Woodland	84	65	1108	136	18	23	4809	6243	77.03%
Column Total	25400	15748	31789	3102	116374	6461	5424	204298	
Producer Accuracy	82.87%	79.13%	92.26%	87.43%	99.97%	65.39%	88.66%		

Overall Classification Accuracy = 93.45 %

Overall Kappa Statistics = 0.8959