Land Cover Mapping through Optimizing Remote Sensing Data for SVM Classification

Anthony Gidudu

Thesis presented for the degree of Doctor of Philosophy in the School of Architecture, Planning and Geomatics University of Cape Town

February 2006

Abstract

Support Vector Machines (SVMs) are a new supervised classification technique that has its roots in statistical learning theory. It has gained popularity in fields such as machine vision, artificial intelligence, digital image processing and more recently remote sensing. The three commonly used SVMs include: linear, polynomial and radial basis function (i.e. Gaussian) classifiers. Polynomial and radial basis function SVMs have been observed to be superior to traditional classifiers e.g. minimum-distance-to-means and maximum likelihood classifiers and new generation classifiers e.g. decision trees and fuzzy classifiers. The successful application of the listed SVM classifiers is dependant on the identification of their respective optimum parameters. The voluminous nature of remote sensing data renders the determination of these parameters slow and tedious. This research explores optimization of remote sensing data through data reduction as a possible solution. To this effect data reduction techniques have been categorized into feature and instance reduction methods. Feature reduction refers to the means of reducing the number of satellite bands needed to determine the optimum SVM parameters as well as to carry out the eventual classification. Instance reduction refers to the means of reducing the quantity of data needed to determine the optimum SVM parameters. Traditional data reduction techniques were investigated and novel ones proposed. The traditional reduction technique evaluated was the Principal Component Analysis (PCA) and the novel methods proposed were: exhaustive search, Population Based Incremental Learning (PBIL) and an instance reduction technique called center splitting. Key to the successful utilization of the exhaustive search and PBIL is the choice of criterion function. In this research a criterion function new to remote sensing was proposed called Thornton's separability index. For each SVM classier and data reduction technique a corresponding land cover map was derived and compared with ground truth data. The measures of comparison between the derived maps and the ground truth data were statistical measures such as the KHAT statistic, overall, user and producer accuracies. Binomial tests of significance were also carried out to ascertain the consequence of optimizing remote sensing data on classification accuracy. This thesis presents and analyses the interplay between the different SVM classifiers, optimization techniques and classification accuracy against the backdrop of the three selected study areas.

Acknowledgements

"Now to Him who is able to do exceedingly abundantly more than we can ask or imagine ... be all glory ... Amen!" Ephesians 3:20

The list of people to whom I am indebted for this work is honestly endless, for many people in numerous ways have contributed to my progress in ways they may never know. Some folks however, deserve special mention. I would like to express my gratitude to my supervisor Prof. Heinz Ruther who gave me the privilege and worked tirelessly to see that I get the opportunity of studying at the Department of Geomatics under his tutelage. Your approach, dedication and insight will remain with me for ever. My introduction to the world of soft computing from where I learnt about Support Vector Machines is attributed to Prof John Greene, to whom I am eternally indebted for his patience. Time and again I looked to him for help and not once did he turn me away. I owe this work to him. I would also like to extend my gratitude to Linda Mthembu who was Prof. Greene's research student and in this work I would like to describe as a 'trench-mate'. Linda was a discussion mate, helped give programming insight and helped demystify the ravels of soft computing. The staff in the Department of Geomatics deserve special mention for the difference they made in this my academic journey. To this effect I would like to make special tribute to Prof Merry, Dr. Bhunu, Jenny, Terry, Sandra, Chifamba, Dirk and of course Sydney. My officemates with whom we shared the 'cooler' that Room 531 in Menzies building was, are a bunch of friends am glad I met. The stories, jokes and laughter we shared surely made the odious journey bearable. Shempemba, Jacinta, Kahonde, Batungi, Ambogo, Eric, Katherine, Muzondo, Otukei, Keith, Gichobi, Ewelina, Kahilu, Wanyona, Kolade, Solomon, Themba, you all gave me reason to trek the long hill to Upper campus. The opportunity to study at UCT was made possible by the financial support of the German Academic Services (DAAD) and Makerere University Staff Development committee to whom I am sincerely grateful. I commit myself to making good the investment they have made.

Thankfully life at the UCT was not only about books and there was a life outside of research. I am grateful to my bible study group who provided immense spiritual and social support. You stood by me and my family in very trying times and I can only pray God's blessings upon you. I would like to thank Pastor Paul for showing genuine interest in my welfare and going out of his way for us. Thank you so much for the opportunity and privilege of being part of the bible study groups and participating at The Message. Lesotho Ntlama, meeting you was a God send and those Saturday prayer meetings surely brought us far, indeed: '...Ebenezeer...'!!!

A special thanks goes to Grace my wife who literally gave up every thing and held back nothing to see me this far. I am lost for words at how much you have willingly sacrificed for me. And Thabo, thanks for the joy you have brought to our lives. You are one reason that all this study has meaning. I would also like to express a special thanks to my late dad who would have loved to see me this far. His death during the course of this research was certainly a low moment of my life. I would also like to thank my mum for believing in me and for being the inspiration in my life from years ago. To the rest of my family, in-laws and friends who have thought, prayed and stood by us, ***mwanyala nabi***!

Declaration of Free License

"I hereby

a. grant the University free license to reproduce this thesis in whole or in part, for the purpose of research;

b. declare that:

(i) the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that stated below.

(ii) except as stated below, neither the substance or any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree in the University or any other University.

(iii) I am now presenting the thesis for examination the thesis for examination for the Degree of PhD"

Signed: Signed by candidate Signature removed

Anthony Gidudu

February 2006

To Him in whom we live, move and have our being

To Grace, Thabo, mum, dad and family

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background

Land cover refers to the composition and characteristics of land surface elements (Cihlar 2000). These land surface elements may include, but are not limited to: vegetative cover such as forests, shrubs, grasslands and wetlands; water bodies such as lakes, rivers and snow cover; built up areas such as roads and buildings. The ability to collect accurate and up-to-date land cover information is of benefit to numerous disciplines for example: regional and urban planning; environmental applications such as global change studies, environmental modeling and policy formulation; natural resource management such as forestry management, wetlands monitoring and precision farming. Over the last three decades remote sensing has become a prime source of land cover information (Kramer, 2002; Foody and Mathur, 2004a). This has been occasioned by improvements in satellite sensor technology thus enabling the acquisition of land cover information over large areas at various spatial, temporal spectral and radiometric resolutions.

Satellite imagery is a major product of remote sensing which facilitates the mapping of land cover. The process of relating pixels in a satellite image to land cover is called image classification. Image classification may be categorized in several ways such as: supervised or unsupervised, parametric or nonparametric, contextual or noncontextual classification (Keuchela et al, 2003). This research considers the application of nonparametric supervised classification algorithms called Support Vector Machines (SVMs). Any supervised classification process involves defining different classes of land cover in advance, from which their properties are learned using training samples. All the data points are then classified according to the models derived from these training samples (Richards and Jia, 1999).

One of the earliest supervised classification algorithms to be used in remote sensing was the maximum likelihood classifier (Landgrebe, 1997). Since then various other algorithms have been developed such as the minimum-distance-to-means and parallelepiped classifiers. Some of the newer classification algorithms include neural networks, decision trees, fuzzy logic and SVMs, the latter being the focus of this research. An overview of the above mentioned methods is presented in Chapter Two.

SVMs represent a group of theoretically superior machine learning algorithms making them one of the state-of-the-art classifiers in remote sensing (Huang et al, 2002). Unlike traditional classifiers such as maximum likelihood, minimum-distance-to-means and the parallelepiped classifiers which are parametric, SVMs are nonparametric. They therefore do not attempt to model the distribution of the data, but try to separate the different classes by directly searching for adequate boundaries between them (Keuchel, 2003). A theoretical overview of SVMs is presented in Chapter two.

SVMs have their roots in statistical learning theory (Vapnik, 1995) and have gained prominence in machine vision fields such as character, handwriting digit and text recognition (Vapnik, 1995; Joachims, 1998). These applications found SVMs to be competitive with the best available classification methods (Huang et al, 2002). Buoyed by the success of SVMs in machine vision, the remote sensing community has over the last approximately five years picked interest in applying them to land cover classification tasks.

Huang et al (2002), Guobin and Blumberg (2002) and Mahesh and Mather (2003) represent some of the earliest works on the application of SVMs to land cover mapping. These initial studies have primarily focused on the comparison of SVMs with traditional classifiers such as the maximum likelihood classifier and other new generation classifiers such as decision trees and neural network classifiers. The evaluation of these classifiers has focused on the determination of their relative accuracy, speed, stability and comprehensibility (Huang et al, 2002). In the studies, there is consensus regarding the superiority of SVM classifiers over traditional classifiers such as maximum likelihood and newer classifiers such as decision trees. Of contention however is the comparison of SVMs with neural networks. Whereas

Mahesh and Mather (2003) claim superiority of SVMs over neural networks, Huang et al (2002) are more guarded in their assessment citing instances where neural networks have performed better than SVMs.

1.2 Problem Statement

One of the key factors that lends itself to the successful deployment of SVMs for land cover mapping is the determination of appropriate SVM parameters. The major setback regarding the application of SVMs to land cover mapping is that the voluminous nature of remote sensing data renders this crucial stage a tedious and slow process. A possible solution is to reduce the quantity of data needed to determine these parameters. The challenge is therefore to derive methods that can optimally reduce the quantity of data required to derive these SVM parameters while simultaneously maintaining classification accuracy

1.3 Aim and Objective of Study

As alluded to in the above problem statement, the aim of this research is to propose methods/techniques through which remote sensing data may be optimized through data reduction for SVM classification. In this research, two aspects of data reduction are considered namely: reducing the number of bands needed to carry out SVM classification and reducing the quantity of data needed to determine the SVM parameters. To this effect, the objective of this research is to propose methods that enable band and instance reduction for SVM classification.

1.4 Justification

Since the early stages of the establishment of remote sensing as a land cover classification approach, various attempts have been made at feature (band) reduction (Langrebe, 1997). Remote sensing researchers such as Bruzzone and Serpico (2000) and Tso and Mather (2001) contend that reducing the number of bands needed to classify a given study area results in significant savings in computational time. The necessity to reduce the number of features (bands) in land cover classification stems from the fact that some of the sensor bands are highly correlated (Schowengerdt, 1983; Bruzzone and Serpico, 2000; Tso and Mather, 2001; Kavzoglu and Mather, 2002). Schowengerdt (1983) attributes this correlation to natural spectral correlation e.g. in Landsat MSS, vegetation exhibits low reflectances in the fourth and fifth bands

and high reflectances in sixth and seventh bands. Another cause mentioned in Schowengerdt (1983) is the issue of overlapping spectral sensitivities between adjacent spectral bands. Schowengerdt (1983) further asserts that whereas this factor is reduced as much as possible in the Landsat MSS design and engineering, it can not be completely eliminated. In view of this, the analysis of the original spectral bands can be extremely inefficient in terms of the actual amount of non-redundant data present in the Landsat MSS image (Schowengerdt, 1983). To this effect, two main approaches to solving this problem have gained prominence in remote sensing studies. The first approach involves compressing the correlated bands through a linear transformation to a subset of uncorrelated components retaining most of the variance of the original set in fewer dimensions (Borak and Strahler, 1999; Bruzzone and Serpico, 2000). One such transformation is the Principal Component Analysis (PCA). Given the success of PCA in its application to traditional parametric classifiers, this research explores its applicability to SVM classification. To wit the first research question is: *"Can the PCA approach be used to reduce the number of bands needed to classify an image while maintaining the SVM classification accuracy?"*

The other approach to feature (band) reduction is feature selection. The aim of feature selection is to identify a subset of bands that minimize the overall error of the classifier (Bruzzone and Serpico, 2000). To productively effect feature selection, it is important to identify an appropriate search algorithm and criteria function (Bruzzone and Serpico, 2000). The purpose of the search algorithm is to generate and compare possible "solutions" of the subset of features by using the criterion function as a measure of the effectiveness of each considered feature subset (Bruzzone and Serpico, 2000). The subset yielding the best measure is considered the output of the feature selection algorithm (Bruzzone and Serpico, 2000). Given that the aim of feature selection is to identify the subset of features that minimize the overall classification error, the criterion function adopted should be related to the behavior of the error made by the classifier used (Bruzzone and Serpico, 2000). Separability indices provide one of the common criterion functions used in remote sensing feature selection studies. Bruzzone and Serpico (2000) list some of the separability indices that have been used as criterion functions for parametric classifiers and include; the divergence criterion (Richards and Jia, 1999; Swain, 1978), Bhattacharyya distance (Richards and Jia, 1999; Swain, 1978; Kailath, 1967), Jeffreys-Matusita (J-M)

distance (Thomas et al, 1987; Swain, 1978), Scatter Matrices (Fakunaga, 1990) and transformed divergence (Thomas et al, 1987; Swain, 1978). In this research, a new separability index tailored for nonparametric classifiers is proposed called Thornton's separability index (Thornton, 2000). New search algorithms are also proposed to effect feature selection namely exhaustive search and Population Based Incremental Learning (PBIL) (Baluja, 1994). The separability index and both search algorithms have their roots in machine learning. The second research question addressed in this study is: *"Can the exhaustive and PBIL search methods using Thornton's separability index as a criterion function be used to reduce the number of bands needed to classify an image while maintaining the SVM classification accuracy?"*

One of the positive attributes of SVMs is the fact that they require small sample quantities of training data (Huang et al, 2002). In this research, this advantage is capitalized upon to propose an instance reduction technique, to reduce the quantity of data needed to train a given SVM classifier. This instance reduction method also has its roots in machine learning and was proposed by Thornton (2000). The third research question investigated is: *"Can the proposed instance reduction method be used to reduce the quantity of data needed to train an SVM classifier while maintaining the SVM classification accuracy?"*

1.5 Thesis Structure

Chapter One begins by giving an introductory overview of this thesis highlighting the importance of land cover mapping and the role played by remote sensing in this regard. This is followed by a brief overview of some of the traditional classification algorithms. Support vector machines are introduced as a new supervised classification technique with mention being made of the current research interests regarding their comparison with traditional classifiers. The problem statement is then presented, from which the research aim and objectives are derived. Justification is also given regarding the research direction pursued and the research questions under investigation. The chapter concludes with an overview of the study areas and the satellite imagery used in the course of this research.

In Chapter Two a historical perspective to Earth observation focusing on the events that influenced interpretation techniques is given. A chronological sequence from aerial photographic interpretation to automatic image classification is also briefly presented. The chapter proceeds to give an overview of traditional and current classifiers. SVMs are then introduced as a new supervised classification technique and the application of SVMs to linear and nonlinear data structures is discussed. The various SVMs plus their respective parameters that ought to be optimized are discussed and current research in the application of SVMs to land cover classification is highlighted.

Chapter Three describes in detail the data reduction methods that have been applied in the course of this research. The methods are categorized into feature and instance reduction techniques. Under feature reduction, the two approaches discussed earlier namely Principal Component Analysis (PCA) and feature selection are explored. Two feature selection techniques are presented herewith namely exhaustive search method and Population Based Incremental Learning (PBIL), which is an adaptation of the genetic algorithm.

Chapter Four presents an overview of issues surrounding the assessment of land cover classification accuracy. Accuracy methods and measures are highlighted.

Chapter Five outlines the methodology developed to effect this research. It weaves together the theory presented in Chapters Two, Three and Four. Issues that are highlighted include data preparation, data reduction of satellite imagery, SVM classification and accuracy assessment.

Chapter Six shows the results of applying the reduction methods proposed in Chapter Three. This chapter indicates the bands selected following exhaustive and PBIL search for each study area's corresponding land cover classes. It continues to present the various derived land cover maps following the application of the different reduction techniques. A visual and quantitative analysis of the derived maps is presented. The discussion is essentially centered on how/if the various reduction methods have maintained, retarded or improved the accuracy of the SVM classification.

Chapter Seven discusses the results from the three datasets from which conclusions are drawn and recommendations given for further research.

1.6 Study Areas

Figure 1.1 depicts the geographical distribution of the study areas used in this research. As can be observed, three study areas have been identified namely: The East Rand in Johannesburg - Gauteng Province, South Africa; The Mara Basin in Northern Tanzania and Malmesbury in Western Cape, South Africa.

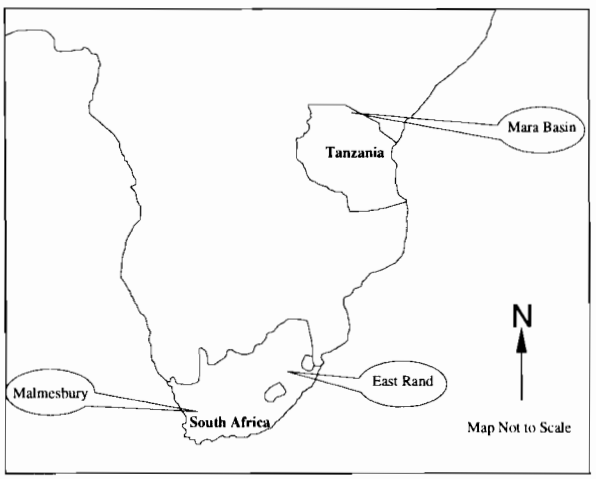


Figure 1.1: Geographical distribution of study areas

The East Rand study area is located South East of Johannesburg between the towns of Benoni and Springs. It covers an area of approximately one thousand six hundred square kilometers meters (1,600km^2). This area is better known as the Witwatersrand gold field, renown for being the biggest gold field in the world. The mines have been exploited for over a century which has resulted in an accumulation of thousands of hectares of mining residue thus posing environmental risk to the environs (Chevrel et al, 2003). The dominant land cover patterns in the area include agricultural fields, water, roads, soils, tailings, tailings dams, yellow tailings, trees, wetlands, 'soil and trees'.

Tailings dams are surface waste storage facilities and they represent the main source of contamination to the East Rand area. Some of the pollutants associated with these tailings dams include: sulphates, heavy metals, cyanides, arsenic and salts. The erosion of tailings dams causes the transportation of tailings (or contaminated suspended matter) over long distances and their redeposition in water sources in the form of large layers along wetlands in river embankments.

From an environmental and public health perspective there is therefore a need to be able to accurately identify the locations and the extents of these tailings and tailings dams. This would make it possible to plan mitigation measures for the surrounding land use activities. This research uses the classification of satellite imagery to determine the land cover types in the study area and compares the results of different classification and optimization techniques.

The second study area considered in this research is sectioned off from the Mara River Basin. The Mara River Basin stretches from Masai Mara in Southern Kenya to the Serengeti National Park in Northern Tanzania. The Mara River originates from the Mau escarpment in Kenya, traverses through the Serengeti plains and finally drains into Lake Victoria (Glows, 2005). Some of the land use activities that characterize this area include the Mau forests, subsistence and commercialized agriculture, pastoralism, Masai Mara and Serengeti Game reserves and wetlands (Mati et al, 2004). The last 50 years have seen changes in the Land Use Land Cover profile of the Mara River Basin such as massive deforestation for timber and cultivation, encroachment of riparian wetlands and unsustainable agricultural practices such as irrigation (Mati et al, 2004). As a consequence, increased pollution and high seasonal variabilities in the flow of the Mara and its tributaries are a glaring reality (Mati et al, 2004). Remote sensing provides a means through which historical and current land cover changes can be detected and monitored. This in turn facilitates the development of mitigation and remedial measures to curb the environmental degradation. As alluded to before, this research explores ways of enhancing the means through which land cover can be mapped. Based on prior field work, six land cover classes were sought in the course of this research namely; wetlands, water (lakes and rivers), Bush/shrub/trees, Grasslands, "bare ground" and Roads.

Malmesbury was the third study area considered in this research. Malmesbury is the largest town in what is referred to as Swartland. The Swart region is geographically located in the Western Cape and stretches from Darling near the Western Coast to Porterville in the East and Piketberg in the North. Malmesbury is essentially surrounded by agricultural fields. The provision of up-to-date land cover maps of urban and rural areas is a hallmark of remote sensing. Through this research the identification of image classification is used to identify the land surface elements of

the study area. The three land cover classes which were sought for in this study area were fields, trees and built up areas.

1.7 Satellite Imagery

The East Rand study area was classified using an ASTER image acquired on 16th October 2000. ASTER is an acronym for Advanced Space-borne Thermal Emission and Reflection Radiometer. ASTER has a temporal resolution of sixteen (16) days. Table 1.1 below summarizes the details of ASTER's spectral and spatial resolution.

Table 1.1: ASTER satellite specifications

Bands	Wavelength	Spectral Range	Resolution
1	0.52 – 0.60 μm	Green (Visible NIR)	15m
2	0.63 – 0.69μm	Red (VNIR)	15m
3	0.76 – 0.86μm	Near Infrared (VNIR)	15m
4	1.60 – 1.70μm	Short Wave Infrared (SWIR1)	30m
5	2.145 – 2.185μm	Short Wave Infrared (SWIR1)	30m
6	2.185 – 2.225μm	Short Wave Infrared (SWIR1)	30m
7	2.235 – 2.285μm	Short Wave Infrared (SWIR1)	30m
8	2.295 – 2.365μm	Short Wave Infrared (SWIR1)	30m
9	2.360 – 2.430μm	Short Wave Infrared (SWIR1)	30m
10	8.125 – 8475μm	Thermal Infrared (TIR)	90m
11	8.475 – 8.825μm	Thermal Infrared (TIR)	90m
12	8.925 – 9.275μm	Thermal Infrared (TIR)	90m
13	10.256 – 10.95μm	Thermal Infrared (TIR)	90m
14	10.950 – 11.650μm	Thermal Infrared (TIR)	90m

Both the Mara Basin and Malmesbury study areas were classified using Landsat ETM+ imagery. Like ASTER, Landsat ETM+ has a temporal resolution of 16 days. Table 1.2 below summarizes the details of Landsat ETM+'s spectral and spatial resolution that was available for this research.

Table 1.2: Landsat ETM+ Satellite Specifications

Bands	Wavelength	Spectral Range	Resolution
1	0.45 – 0.52 μm	Blue (Visible)	30m
2	0.52 – 0.60μm	Green (Visible)	30m
3	0.63 – 0.69μm	Red (Visible)	30m
4	0.76 – 0.90μm	Near Infra Red (NIR)	30m
5	1.55 – 1.75μm	Middle Infra Red (MIR)	30m
7	2.08 – 2.35μm	MIR	30m
6	10.40 – 12.50μm	Thermal Infra Red	60m

Chapter 2
Land Cover Classification

2.1 Background

Earth observation has become the prime source of data in the geosciences and many related disciplines, permitting research into the distant past, the present and into the future e.g. by assessing environmental impacts (Kramer, 2002). This has yielded new clarity and better awareness of the earth's dynamic nature. Earth observation is based on the premise that information is available from the electromagnetic energy field arising from the earth's surface (or atmosphere or both) and in particular from the spatial, spectral & temporal variations in that field (Kramer, 2002).

The earliest attempts at Earth observation began in 1858 in France by the French photographer Gasporal Felix Tournachon, who obtained the first aerial photographs over Paris from an altitude of 80 meters using a tethered balloon. He later applied this technique to map the countryside and in 1860 was commissioned to obtain reconnaissance photography in preparation for the battle of Solgerino in Northern Italy (Kramer, 2002). Aerial photographs taken from tethered balloons were also used during the US civil wars (1861 – 1865) to study enemy positions (Gibson, 2000). The 1880s and early 1900s saw attempts to use kites and pigeons to carry cameras to many hundred meters of altitude. Extraction of information from these aerial photographs was principally by photographic interpretation.

The 1890s through to the early 1900s witnessed groundbreaking progress in aviation research, which was to later provide a breakthrough in Earth observation. The advent of the airplane provided the capability of controlling the speed, altitude and direction required for the systematic use of the airborne camera (Campbell, 1996). In 1909, the first recorded airborne photographs were taken from a plane piloted by Wilber Wright over Centocelli in Italy (Elachi, 1987; Campbell, 1996). Aerial photography was to later on play a pivotal role in both World War I and II for military reconnaissance. During World War II, the electromagnetic spectrum was extended from almost

exclusive emphasis on the visible spectrum to other regions, most notably the infrared and microwave regions (Kramer, 2002). This development enabled research on the spectral reflectance patterns of natural terrain and photographic emulsions for aerial color and infrared photography. The main incentive was to develop techniques for camouflage detection (Campbell, 1996). In 1956, Colwell performed some of the early experiments on the use of special purpose aerial photography for the classification and recognition of vegetation types and the detection of diseased and damaged vegetation (Elachi, 1987).

Space age Earth observation, started with the launch of Russia's first Sputnik satellite in 1957. Measurements from Sputnik-1 permitted a first estimation of the density of the upper atmosphere. In 1958, the United States also launched its own satellite, the Explorer-1, which provided preliminary information on the environment and conditions in space outside Earth's atmosphere. The first space borne imaging sensors were fitted with film cameras, providing a "birds eye view" from space and into space. One of the famous US space missions was the Apollo program, which was purposely initiated to land humans on the moon and return them to earth safely. To facilitate this mission, maps of the lunar surface especially of the proposed landing sites were a major necessity. The mapping for this purpose was performed predominantly through photographic interpretation.

Plans for a deliberate and systematic approach to Earth observation did not start until the mid 1960s (Kramer, 2002). As a prelude to the launch of the Earth Resources Technology Satellite (ERTS), preliminary research was carried out by collaboration between the University of California Berkley, University of Michigan and Purdue University (Landgrebe, 1997). Their first development was an aerial scanner which was able to collect high quality data in twelve to eighteen spectral bands (Landgrebe, 2005). The availability of data in various bands necessitated the adoption of digital multivariate statistical methods for the extraction of land cover information (Landgrebe, 1997). The maximum likelihood and minimum distance to means classifiers were some of the early image classification methods experimented on (Wacker and Landgrebe, 1972; Landgrebe, 2005). Neural network analysis was popular at the time, however it had the disadvantage of requiring substantial computing time (Landgrebe, 1997). This collaboration served as a precursor to the

design of ERTS which was launched in 1972. ERTS was later renamed Landsat 1, thus beginning the Landsat series of Earth observation satellites. Other classification methods that have since gained prominence in the remote sensing field include parallelepiped, decision trees and fuzzy classifiers. In the next section, a brief overview of these methods is presented.

2.2 Overview of some common Image Classification Methods

Image classification provides a means through which an image is associated with a label representing a real world object (Mather, 1987). It involves analyzing spectral signatures and assigning image pixels to categories based on similar signatures (Sabins, 1978).

Supervised classification provides one of the common means through which image classification can be effected. In general, supervised classification methods are based upon prior knowledge of the number of classes in the study area. These methods require the user to collect representative samples of the known land cover classes, called training samples, and based on these samples determine the decision boundaries in feature space between the classes (Tso and Mather, 2001). Feature space essentially refers to the plotting of the radiance values defined by the axes of the image bands.

Sabins (1978) and Schowengerdt (1983) categorize supervised classification methods into parametric and nonparametric algorithms. Parametric algorithms basically assume a particular class statistical distribution, commonly the normal distribution, and then estimates the parameters of that distribution e.g. mean vectors and variance matrices to use in the classification algorithm (Schowengertdt, 1983; Paolo and Schowengerdt, 1995). Examples of parametric classifiers discussed in this thesis include maximum likelihood, minimum-distance-to-means and parallelpiped classifiers. Nonparametric algorithms are distribution free (Benediktsson et al, 1990) hence make no assumptions about the training samples' distribution. Examples of nonparametric classifiers include neural networks, decision trees and fuzzy classifiers.

Tso and Mather (2001) further categorize supervised classification methods into hard and soft classifiers (algorithms). In hard classification, a pixel can only belong to one class as is the case in maximum likelihood, minimum-distance-to-means and parallelepiped classifiers. Hard (conventional) classification in which pixels are assigned unambiguously to one class make a number of fundamental assumptions (Foody, 2002a). First and foremost, hard classification assumes that the classes are discrete and mutually exclusive (Congalton and Green, 1998) and secondly that the individual pixels or groups of pixels are homogeneous (Foody, 2002a). The shortcomings of these assumptions are that in reality many classes are continuous rather than discrete and mutually exclusive (Foody et al, 1992). Secondly, the coarse resolution associated with satellite sensors invariably includes mixed pixels. Since both continuous classes and mixed pixels can not be accurately presented by a conventional hard classification, much attention has been directed at the development of soft classifiers which permit multiple memberships across various classes (See fuzzy classifiers in section 2.2.5) (Foody, 2002a). Neural networks have the advantage of being adaptable for the production of both hard and soft classifications (Foody, 2002a).

In the next section, an overview of the commonly used supervised classification methods in remote sensing is presented with respect to the previous categorizations mentioned.

2.2.1 Parallelepiped Classifier

Given the training samples of each class, the user provides the minimum and maximum value of each respective class. These extreme values allow the estimation of the position of the boundaries of each parallelepiped (Mather, 1987) (see Figure 2.1). Alternatively, these boundaries may be defined by a number of standard deviations on either side of the mean vector of the training data for a given class (Tso and Mather, 2001). The decision rule is simply to check whether an unknown point representing a pixel in feature space lies inside any of the class parallelepipeds. The region outside the boundaries is referred to as "terra incognito" (Mather, 1987) and unknown pixels falling in this region are assigned the label 0. From the way

parallelepiped classifiers operate, they can also be likened to a high dimensional extension of grey level thresholds.

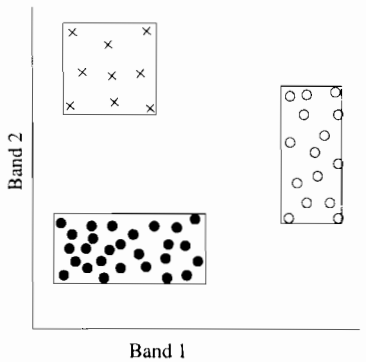


Figure 2.1: Parallelepiped decision boundaries

Parallelepiped classifiers have the advantage of being quick and easy to use. They also take into account the different variability of the classes. The rectangular shape accommodates the fact that variability may be different along different bands (Eastman, 2003). The disadvantage about these classifiers is that in feature space there is always the danger of overlapping parallelepipeds, hence leading to confusion as to which class the pixels in the region of overlap should belong. The other shortcoming is that by its nature, it doesn't make provision for pixels falling in the "terra incognito" region.

2.2.2 Minimum-Distance-to-Means-Classifier

The decision rule for this classifier is that given the training samples of known classes, an unknown pixel is labeled depending on its minimum distance to the class centroids measured by the Euclidean distance (Tso and Mather, 2001) as illustrated in Figure 2.2.

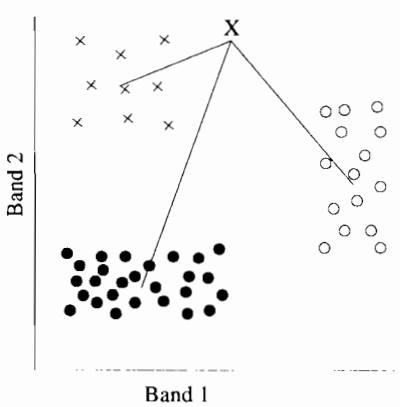


Figure 2.2: Minimum-distance-to-means classifier. The unknown pixel X is assigned the label of the class with the minimum distance

Minimum-distance-to-means classifiers are categorized as both parametric and hard classifiers. They have the advantage of being simple, possessing good classification accuracy and speed of performance. The disadvantage of these classifiers is that they do not consider signature variability. By clustering each class by its mean band reflectance only, it has no knowledge of the fact that some classes are inherently more variable than others, potentially leading to misclassifications (Eastman, 2003). To counter this impediment, Tso and Mather (2001) propose that the Mahalanobis distances be used rather than the Euclidean distance to determine the minimum distance between the unknown pixels and the class centers. The Mahalanobis distance takes into account the shape of the frequency distribution (assumed Gaussian) for a given cluster in feature space, resulting in ellipsoidal clusters. The Euclidean on the other hand, assumes equal variance and no correlation between the features (i.e. bands) giving circular clusters.

2.2.3 Maximum Likelihood Classifier

The Maximum Likelihood Classifier is well established in pattern recognition (Swain, 1978). In this classifier, the probability of a pixel belonging to each of the training data samples is first calculated. The decision rule is that the pixel will be assigned to the class for which the probability is the highest (Yool, 1998; Tso and Mather, 2001).

The Maximum Likelihood classifier can be expressed mathematically. The following is an adaptation from Yool (1998); Let $P(X|w_i)$ represent the probability density function for measurement vector X, where X derives from a pattern in class *i*. Further, let $P(w_i)$ be the prior probability of observing independently a pattern from class *i*. Formally, the maximum likelihood decision rule states: Decide $X \in w_i$ if and only if;

$$P(X|w_i)P(wi) \geq P(X|w_j)P(wj), \text{ for all } j = 1,2,\ldots,m \qquad 2.1$$

To classify a pattern X, a maximum likelihood classifier computes the product $[P(X|w_i)P(w_i)]$ for each class and assigns measurement vector X to the class giving the largest product.

Since these classifiers take into account the intercorrelation between bands, they yield high accuracies. In comparison to the previous classifiers discussed, their computation is more time consuming.

For more than three decades, Maximum likelihood and minimum-distance-to-means classifiers have been the main stay of pattern recognition. The two methods are collectively called the traditional classifiers and although they perform well, their general ability to resolve interclass confusion is limited (Tso and Mather, 2001). As a result, in recent years, and following advances in Information Technology, alternative strategies have been proposed, particularly the use of Decision Trees, Fuzzy classifiers and Artificial Neural Networks (Tso and Mather, 2001), which are the subject of discussion in the next sections.

2.2.4 Decision Trees

Decision trees use a stratified or layered approach in discriminating between classes (Schowengertdt, 1983; Mather, 1987), which Tso and Mather (2001) refer to as a hierarchical splitting mechanism. It is based on the spectral properties of each class and the relationships between them. Tso and Mather (2001) describe a decision tree as being composed of a root node, a set of interior nodes and terminal nodes called leaves. The root node and interior nodes, referred to as non-terminal nodes, are linked into decision stages. The terminal nodes represent final classifications, which provide the most detail in terms of the classification being performed. The classification process is implemented by a set of rules that determine the path to be followed. The path starts from the root node and ends at one terminal node, which represents the label for the object being classified. At each non-terminal node, a decision has to be taken about the path to the next node. Figure 2.3 illustrates a simple hierarchical decision tree using pixel reflectance as input. The aim of classification is to identify pixels as either vegetation or water.

The efficiency and performance of decision tree classifiers is strongly affected by the tree structure and choice of feature subsets (Tso and Mather 2001). The disadvantage about decision tree classifiers is that considerable analyst effort may be required to design an effective tree.

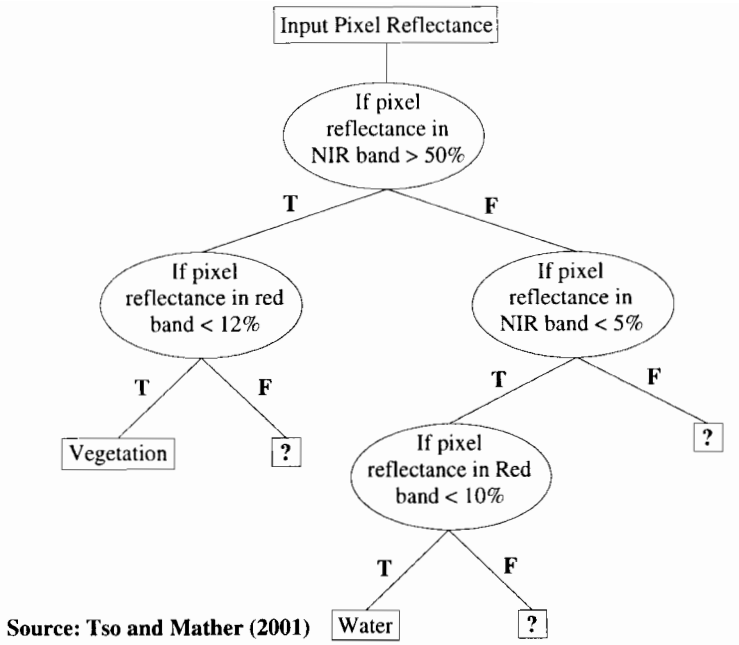


Figure 2.3: Decision tree classifier

2.2.5 Fuzzy Classifiers

Fuzzy classification allows for class membership of a pixel between several classes. Thus rather than being associated with a single class as is the case in conventional hard classifiers, a pixel can have multiple and partial class membership to all classes (Foody, 1998; Zhang and Foody, 2001). Construction of fuzzy rules from numerical data for image classification requires the fuzzy partition of input features (bands) based on user-defined fuzzy membership functions (Tso and Mather, 2001). Membership grades typically range from 0 for non membership to 1.0 for full membership, with intermediate values signifying partial membership in one or more other classes (Campbell, 1996). A graphical illustration using a trapezoidal membership function to divide the feature space into fuzzy partitions is depicted in figure 2.4.

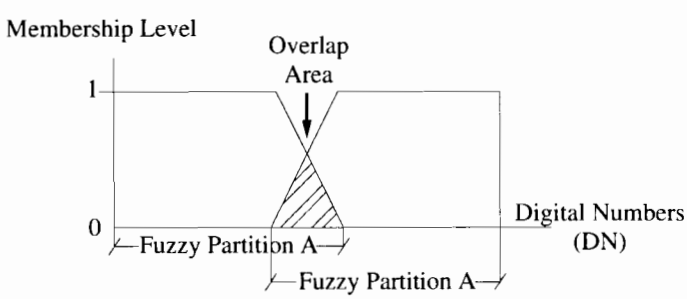


Figure 2.4: Fuzzy classification. In each partition, the class membership certainty varies with DN values (Tso and Mather, 2001)

The disadvantage of fuzzy classifiers is that whereas the output describes the class composition of every pixel, it provides no indication as to how this is distributed spatially within the area it represents (Foody, 1998)

2.2.6 Artificial Neural Networks

Artificial Neural Networks (ANNs) have been studied for years and are well established in image analysis, speech and handwriting recognition (Tso and Mather, 2001). ANNs can be conceived as a complex mathematical function that converts input data to a desired output (Zhang and Foody, 2001). In the case of land cover mapping, the input data would be image bands, while the desired output is a land cover map. Similarly to the other supervised classification techniques discussed before, ANNs need training data to determine the function parameters.

The basic building block of an ANN is the neuron, and a collection of neurons forms a layer. Every ANN has at least three layers with all neurons on a given layer being linked to all neurons on the previous and subsequent layers as shown in Figure 2.5. The idea is to simulate how the brain functions.

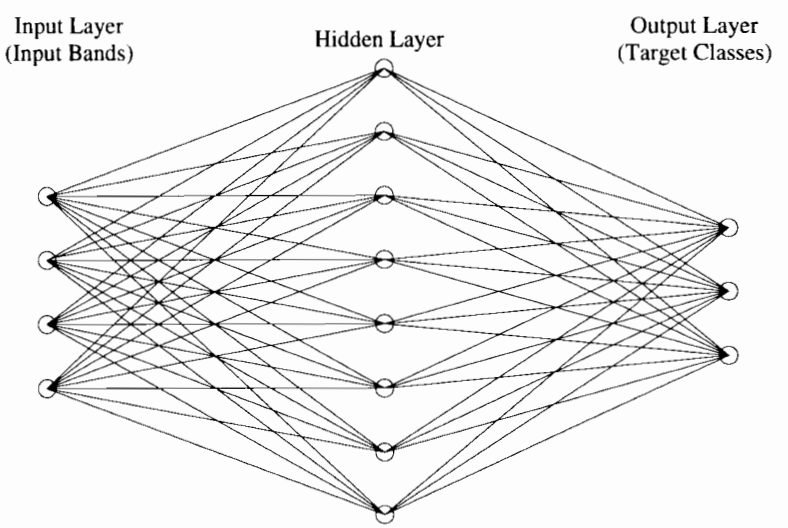


Figure 2.5: A simple ANN architecture

Each connection stands as a linear or nonlinear weighted function and each node operates as a weighted sum of its inputs (Zhang and Foody, 2001). Each processing node calculates a weighted sum of the outputs from the nodes in the preceding layer to which it is connected, passes this through a transfer function to derive its own output which is fed on to the nodes in the next layer (Zhang and Foody, 2001). In

some cases, a more complex mathematical operation may be chosen for a particular need in place of the summation operator (Tso and Mather, 2001). The number of input nodes is determined by the number of image bands at hand while the number of nodes in the output layer is determined by the number of known classes. The number of hidden nodes is determined subjectively on the basis of trial runs (Zhang and Foody, 2001). ANNs differ in terms of learning strategy and neuron characteristics. A learning strategy specifies the initial set of weights and indicates how those weights and neuron values should be updated in order to improve performance (Tso and Mather, 2001). The goals of simplifying the design procedure and optimizing learning strategies have attracted the attention of many researchers (Tso and Mather, 2001)

The fact that ANNs are distribution free renders them robust in recognizing patterns from non-Gaussian probability densities, unlike Maximum Likelihood Classifiers which require patterns to be normally distributed (Paolo and Schowengerdt, 1995). ANNs are therefore more resistant to noise and incomplete or under sampled data (Yool, 1998). The numerous neurons are also thought to add to the ANNs robustness and fault tolerance since damage to few elements or links will generally not minimize overall performance (Tso and Mather, 2001). Most ANNs also use adaptive connection weights to improve performance based on current results. Adaptation also serves a degree of robustness by compensating for minor variability in the characteristics of the data (Tso and Mather, 2001). It is in view of these positive attributes of ANNs that several researchers e.g. Hepner et al (1990), Peddle et al (1994) and Zhang and Foody (2001) acclaim them to be superior to traditional classifiers.

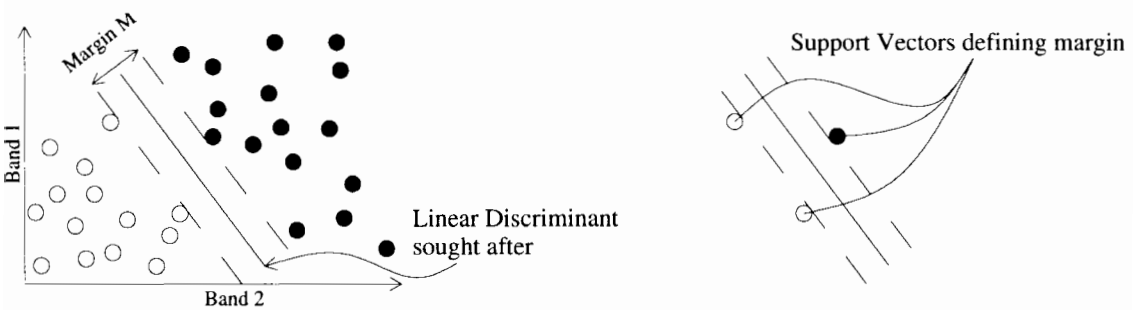
The main challenge regarding the use of ANNs is that they are perceived to be difficult to train and questions of the type of network, network architecture, initial values of parameters, learning rate and momentum, the number of iterations required to train the network and the choice of initial weights are all difficult to answer (Yool, 1998).

2.3 Support Vector Machines

SVMs are non-parametric hence they boast the robustness associated with ANNs (Foody and Mathur, 2004a; Foody and Mathur, 2004b) and other nonparametric classifiers. In the following sections, an introductory discussion is given on the theoretical background of SVMs. The general overview attempts to provide an intuitive non mathematical presentation of SVMs, which is followed by a more technical discussion of SVMs.

2.3.1 General Overview

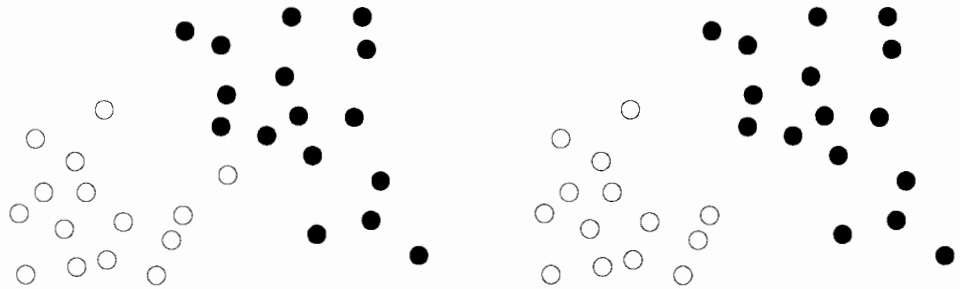
Linear SVMs define the decision boundary between two classes of interest by placing a linear discriminant midway between them. The selection of this decision boundary involves the identification of data points at the edge of each class that defines the maximum possible margin. A margin is essentially the perpendicular distance from a linear discriminant to the nearest data points in either class. The data points in either class that define the maximum margin are called support vectors and are identified through quadratic programming using Lagrange Multipliers. The operation of SVMs on a binary classification task of interest is illustrated in figure 2.1.



a) Illustration of linear discriminant and margin b) Illustration of Support Vectors

Figure 2.6: A binary classification task illustrating the operation of SVMs

One of the inevitable challenges for any classification task is the issue of outliers or rogue points. The utilization of such points in the determination of the maximum margin clearly will result in a less than optimum decision boundary which could potentially compromise classification accuracy. Figure 2.7a showcases this scenario. In Figure 2.7b it can be seen that the removal of the rogue points renders the dataset linearly separable.



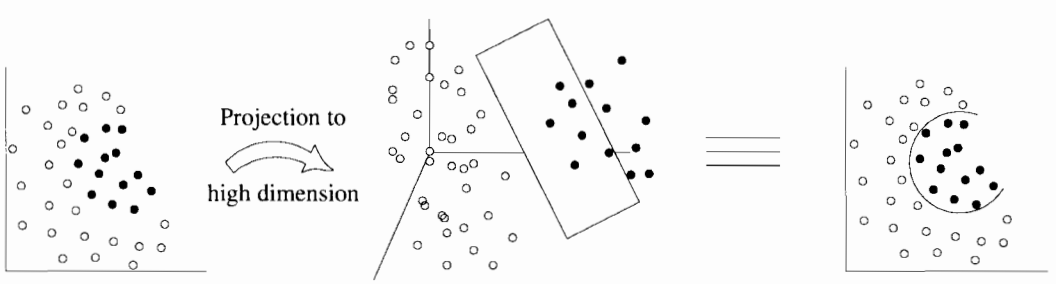
a) Outlier rendering dataset linearly inseparable

b) Removing outlier renders the dataset linearly separable

Figure 2.7: Influence of outliers

SVMs counter the influence of outliers by imposing an upper bound constant, nominally symbolized by C, which constrains the influence of the outliers. The bigger the value of C the bigger the influence of the outlier and conversely the smaller the value of C the smaller the influence accorded to the outlier points. The parameter C is usually determined through cross validation.

In many practical situations, classes of interest are not linearly separable and a nonlinear decision boundary, rather than a linear one, is required to separate the two classes. SVMs adopt the classical approach of nonlinearly projecting the input space to a high dimensional feature space sometimes referred to as an infinite space. In the infinite space, the originally nonlinearly separable dataset becomes separable and placing a linear discriminant in this space will have the same effect as placing a nonlinear discriminant in the input space as illustrated in Figure 2.8.



a) Scatter plot of linearly inseparable dataset in feature space

b) In high dimension, the dataset is linearly separable hence a linear discriminant can be fitted to define the decision boundary

c) Performing classification in this higher dimension feature space is equivalent to performing a nonlinear classification in the previous feature space

Figure 2.8: Classification of nonlinearly separable datasets

The functions used to project the data from input space to feature space are sometimes called kernels (or kernel machines), examples of which include polynomial, Gaussian (more commonly referred to as radial basis functions) and sigmoid functions. Each function has parameters which have to be determined prior to classification and they are usually determined through a cross validation process.

2.3.2 Technical Overview

2.3.2.1 Linearly Separable Classes

Consider a binary classification task in which a linear discriminant is sought for training data represented by: $\{x_i,y_i\}$, where i = 1,..., n and $y_i \in \{+1,-1\}$. (See Figure 2.9)

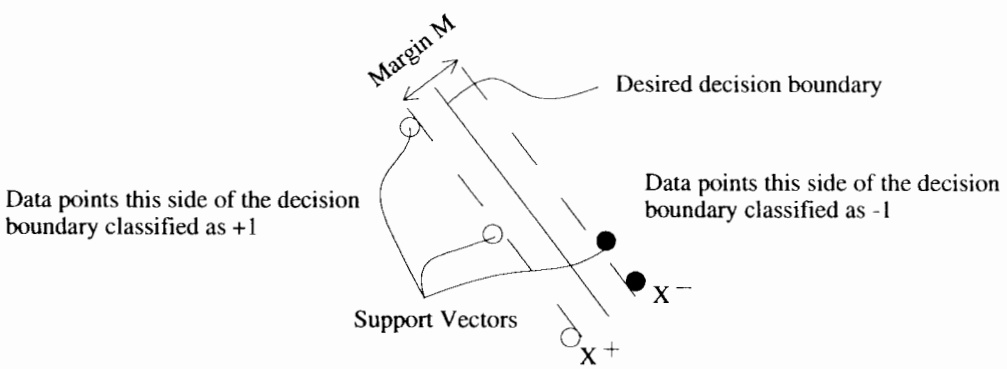


Figure 2.9: Binary classification task for which a linear decision boundary is sought

The linear discriminant in feature space is defined by the equation;

$$w^T.x_i + b = 0 \qquad (2.1)$$

where x_i is a point lying on the linear discriminant, w is the normal to the linear discriminant and b the bias.

The plane defining the positive space is defined by; $w^T.x_i + b = +1$ (2.2)

The plane defining the negative space is defined by; $w^T.x_i + b = -1$ (2.3)

A point is classified as +1 if; $w^T.x_i + b \geq +1$ (for $y_i = +1$) (2.4)

or it is classified as –1 if; $w^T.x_i + b \leq -1$ (for $y_i = +1$) (2.5)

Equations 2.4 and 2.5 can be combined to give: $y_i(w^T.x_i + b) - 1 \geq 0$ (2.6)

Of the possible linear discriminants, the most suitable is that which maximizes the margin M depicted in Figure 2.9. Suppose x^- is a point on the negative plane and x^+ is the closest on the positive plane nearest to x^- (see Figure 2.9) then M is determined by

$$M = |x^+ - x^-| \tag{2.7}$$

From equation 2.2, $w^T.x^+ + b = +1$

$$\Rightarrow x^+ = \frac{1-b}{\|w\|} \tag{2.8}$$

Similarly from equation 2.3, $w^T.x^- + b = -1$

$$\Rightarrow x^- = \frac{-(1+b)}{\|w\|} \tag{2.9}$$

Substituting for x^+ and x^- in equation 2.7 yields the following;

$$M = \left\| \frac{(1-b)}{w} - \frac{-(1+b)}{w} \right\| \tag{2.10}$$

$$\therefore M = \frac{2}{\|w\|} \text{ or } M = \frac{2}{\sqrt{w^T.w}} \tag{2.11}$$

In order to maximize the margin M, $\frac{\|w\|}{2}$ has to be minimized subject to $y_i(w^T.x_i + b) \geq 1$.

Through quadratic programming it is possible to search the spaces of w and b to determine the combination of values that yields the maximum margin M. Quadratic programming using Lagrange Multipliers is the most commonly used technique because it enables optimization of a function subject to given constraints. Equation 2.12 gives the Primal Lagrangian to be optimized;

$$L(w) = \frac{1}{2}\|w.w\| - \alpha_i \sum_{i=1}^{n} y_i (w^T.x_i + b) - 1 \tag{2.12}$$

Where α_i are the Lagrange Multipliers for each corresponding data point x_i

From Wolfe's theorem (Wolfe, 1961), the derivatives of equation 2.12 with respect to w and b are equated to zero and re-substituted into equation 2.12 to give the Wolfe dual Lagrange. Below is the derivative of the dual Lagrange from the Primal Lagrange;

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} y_i \alpha_i x_i = 0 \tag{2.13}$$

$$\Rightarrow w = \sum_{i=1}^{n} y_i \alpha_i x_i \tag{2.14}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{2.15}$$

$$=> \sum_{i=1}^{n} \alpha_i y_i = 0 \qquad (2.16)$$

Expanding equation 2.12;

$$L(w) = \frac{1}{2}\left\| w^T.w \right\| - \sum_{i=1}^{n} (\alpha_i y_i x_i) w - b\sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i \qquad (2.17)$$

From equations 2.12 and 2.14, equation 2.17 is modified to

$$L(w) = \frac{1}{2}\left\| w^T.w \right\| - (w^T.w) - (b.0) + \sum_{i=1}^{n} \alpha_i \qquad (2.18)$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\left\| w^T.w \right\| \qquad (2.19)$$

Substituting for w from equation 2.14 yields the following Wolfe dual Lagrange:

$$W(\alpha_i) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad (2.20)$$

With $\alpha_i \geq 0$ under constraints $\sum_{i=1}^{n} \alpha_i y_i = 0$

This is a quadratic programming problem which will yield a series of Lagrange Multipliers (α_i) corresponding to each training data point. The data points x_i for which $\alpha_i \neq 0$ are the support vectors, which will be the data points defining the maximum margin of the desired decision boundary. The Lagrange Multiplier distribution through out the dataset is illustrated in Figure 2.5.

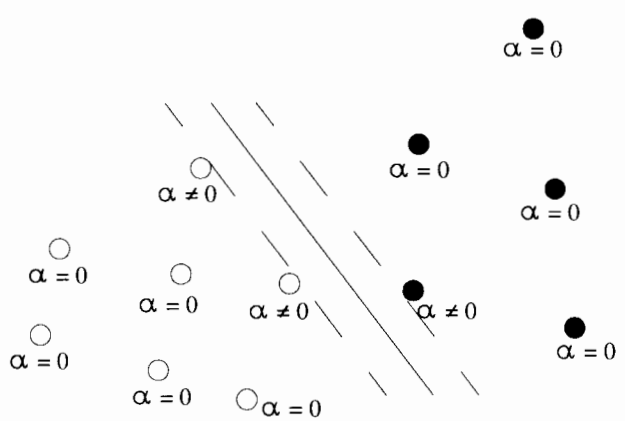


Figure 2.10: Lagrange Multiplier distribution in the training dataset

w can be derived from;

$$w = \sum_{i=1}^{s.v} \alpha_{s.v} y_{s.v} x_{s.v} \qquad (2.21)$$

where s.v are the support vectors

The bias b can be derived from;

$$b = \frac{1}{2} w[x_{+1} + x_{-1}] \qquad (2.22)$$

where x_{+1} and x_{-1} are the support vector values in either class

From equations 2.21 and 2.22, an unknown data point z can thereafter be classified from the function;

$$f(z) = \text{sign}(w^T.z + b) \qquad (2.23)$$

2.3.2.2 Imperfectly Separable case

The influence of outliers has been illustrated in Figure 2.7. Figure 2.11 modifies figure 2.2 by introducing "slack Variables" (ξ) as they are known in optimization theory.

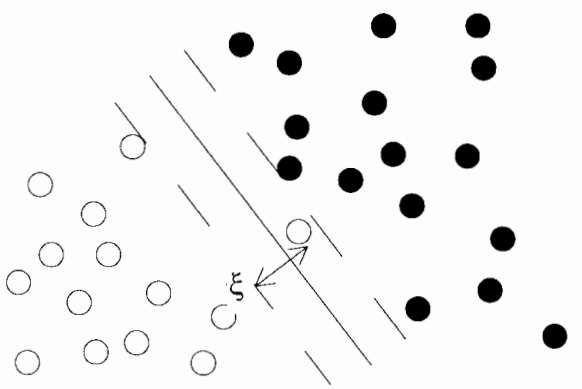


Figure 2.11: "Slack variables"

In the case that there is no classification error due to an outlier or rogue point, $\xi_i = 0$. Incorporating the "slack variables" into equations 2.4 and 2.5;

A point is classified as +1 if; $w^T.x_i + b >= +1 - \xi_i$ (2.24)

or it is classified as –1 if; $w^T.x_i + b <= -1 - \xi_i$ (2.25)

where $\xi_i \geq 0$ (2.26)

The optimization problem becomes;

$$\text{Minimize } \frac{\|w\|}{2} + C\sum_{i=1}^{n} \xi_i \qquad (2.27)$$

Subject to: $y_i(w^T.x_i + b) >= 1 - \xi_i$

Where $\xi_i \geq 0$

The constant C controls the magnitude of the penalty/error associated with training samples that lie on the wrong side of the linear discriminant.

The Wolfe dual Lagrange consequently becomes;

$$W(\alpha_i) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad (2.28)$$

Subject to: $C \geq \alpha_i \geq 0$ under constraints $\sum_{i=1}^{n} \alpha_i y_i = 0$

Similarly to equation 2.21, w can be derived from

$$w = \sum_{i=1}^{s.v} \alpha_{s.v} y_{s.v} x_{s.v} \qquad (2.29)$$

where s.v are the support vectors

The only difference with the linear case is that there is an upper bound C on the Lagrange Multipliers (α_i).

2.3.2.3 Nonlinear Decision Boundaries

The projection of the input space to a higher dimension space where the previously nonlinearly separable dataset becomes linearly separable is illustrated in Figure 2.8. In feature space, the Wolfe dual Lagrange would be

$$\text{Minimize } W(\alpha_i) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \varphi(x_i)^T . \varphi(x_j) \qquad (2.30)$$

Subject to: $C \geq \alpha_i \geq 0$ under constraints $\sum_{i=1}^{n} \alpha_i y_i = 0$

Working in this very high dimension poses two main challenges. The first one being the rhetorical "curse of dimensionality". Whereas a linear discriminant may be found at very high dimensions it may result in over-fitting within the input space and hence yield poor generalization. SVMs counter this optimization theory problem by seeking the linear discriminant with the largest margin in the high dimension. By so doing, over fitting in the input space is minimized. The second challenge of working in high dimension is the correspondingly high computational burden that ensues. SVMs counter this by way of the so called "kernel trick".

The kernel trick is based on the observation that the vectors in the linear case occur only in dot products. For instance, consider a point z that needs to be classified. In a linearly separable situation, the following function can be used to determine its class;

$$f(x) = \text{sign}(w^T.z + b) \qquad (2.31)$$

$$f(x) = \text{sign}(\sum_{i=1}^{n}(\alpha_i y_i x_i.z) + b) \quad (2.32)$$

In high dimension, the classification function would be;

$$f(x) = \text{sign}(\sum_{i=1}^{n}(\alpha_i y_i < \varphi(x_i).\varphi(z) >) + b) \quad (2.33)$$

Fortunately it can be shown that there exists a large class of functions K(x,z) such that;

$$K(x,z) = < \varphi(x_i).\varphi(z) > \quad (2.34)$$

For such a function to be applicable it ought to meet Mecer's conditions (Courant and Hilbert, 1953). These functions are called kernels machines. The advantage of using kernels is that they can be deployed without explicitly knowing the transformation space due to the transform φ. The classification function henceforth becomes;

$$f(x) = \text{sign}(\sum_{i=1}^{n}(\alpha_i y_i K(x_i, z)) + b) \quad (2.35)$$

2.3.2.4 Examples Of kernels

Listed below are three of the main kernels used to project data from the input space to feature space;

Polynomial Kernel: $K(x,y) = (xy + 1)^d$

- The parameters that need to be determined in this kernel are polynomial degree d and C

Gaussian/Radial Basis Functions Kernel: $K(x,y) = e^{\frac{-\|x-y\|^2}{2\sigma^2}}$

- The two parameters that need to be determined are the Gaussian width and C

Sigmoid Kernel: $K(x,y) = \tanh(k(xy) + \theta)$

- This kernel is least used probably because it doesn't satisfy Mercer's condition for all values of k and θ.

2.3.2.5 Multi-class Classification

SVM classification is essentially a binary (two-class) classification technique. Many practical applications, however, entail multiple class classification tasks. There exist two main strategies that can be employed to adopt SVMs to multiple classification tasks. The first approach is called the 'one against one' approach and involves constructing a machine for each pair of classes resulting in N(N-1)/2 machines. When

applied to a test point, each classification gives one vote to the winning class and the point is labeled with the class having most votes. The second approach involves the 'one against all' approach where by the N class dataset is divided into N two-class cases. Proponents of the 'one against one' approach contend that it is has the advantage of avoiding highly unbalanced training data (Gualtieri and Cromp, 1998). It is however acknowledged that this approach involves constructing more SVMs as compared to the 'one against all' approach, hence it is more computer intensive. One problem associated with the 'one against all' approach is that unbalanced training data may potentially compromise classification accuracy.

In conclusion, "Support Vector Machines are a rare example of a methodology where geometric intuition, elegant mathematics, theoretical guarantees and practical mathematics meet" (Bennet and Campbell, 2002).

2.3.2.6 Summary & Review of Land Cover Mapping Using SVMs

A review of literature reveals a growing interest in the application of SVMs to land cover mapping. In their ground-breaking work, Huang et al (2002) assessed the potential of SVMs for land cover mapping. Their study area had six (6) land cover classes and two datasets; one dataset with three (3) Landsat bands and another one with seven (7) Landsat bands. Some of the main issues they investigated were the factors influencing the efficiency of polynomial and Radial Basis Function (RBF) SVMs. From their assessment they observed that whereas changing the polynomial degree affected the classification accuracy, changes in the kernel width did not significantly impact on the classification accuracy. Their results continued to reveal a positive correlation between SVM classification accuracy and both the size of training samples and the number of bands.

Another aspect of their work involved comparing the performance of SVMs to other classification algorithms such as maximum likelihood, decision trees and neural networks. In comparison to decision trees and maximum likelihood classifiers, SVMs were seen to yield better classification accuracies. When the classification accuracies of SVMs were compared with neural networks the results were observed to be comparable. Where seven bands were used as input, SVMs performed significantly

better than neural networks in the majority of the test areas. Even though SVMs performed better than neural networks in the rest of the test areas, the differences were insignificant. When three bands were used, neural networks performed better than SVMs in terms of classification accuracy. Another key comparison made in the course of their research was the speed of training the classifier. From their results, SVMs were found to be the slowest to train followed by neural networks. Both maximum likelihood and decision trees were comparably faster to train than both neural networks and SVMs. Of the classifiers investigated SVMs were found to be the most stable i.e. even with small training sample data, they posted significantly better results than other classifiers.

Guobin and Blumberg (2002) represent another of the earliest publications on the application of SVMs to land cover mapping. In their work they compared the performance of polynomial and RBF SVMs in classifying the new ASTER imagery. They used two datasets: one with 15-meter resolution and the other 30-meter resolution. From their study it was observed that in regards to the 15-meter resolution dataset, while RBF SVMs were faster in terms of reaching convergence, polynomial based SVMs had better classification accuracy results. In the 30-meter resolution dataset however, this difference was not as apparent. They also observed that the results of RBF SVM classification had a comparatively better visual outlook as compared to those of the polynomial SVM. They attributed this to the fact that the projection by RBF SVMs to higher dimension resulted in more compact classes than those resulting from polynomial SVMs. They therefore concluded that there was an inherent cluster capability in RBF based SVM classification mechanisms.

Mahesh and Mather (2003) assessed the utility of SVMs for the land cover classification of multi and hyperspectral datasets in comparison with the most frequently used maximum likelihood and neural network classifiers. Their work consisted of two study areas: one in Littleport – England and another in the La Mancha region of Spain. In Littleport there were seven classes to be classified while in La Mancha there were eight classes. Landsat ETM+ data was used for Littleport while hyperspectral data was used for La Mancha region. For each classifier corresponding land cover maps were derived. The resulting error matrices were used to extract Kappa values and overall classification measures from which comparisons

were made to ascertain relative performance. The Z statistic was used to evaluate the significance of apparent difference between the three classification algorithms. The RBF SVM was used to classify the study areas and the kernel width and C determined by trial and error were found to be 2 and 5000 respectively. From the results of study, SVMs were seen to outperform maximum likelihood and neural network classifiers in terms of classification accuracy achieved, even when a small number of training areas was used. It was also seen that SVM parameters are easier to tune compared to neural networks.

In a more recent publication, Foody and Mathur (2004a) proposed a single multi-class SVM classification technique in place of the traditional multiple binary analysis for land cover classification. The SVM classification results were compared with those derived from the following classification algorithms; discriminant analysis, decision trees and neural networks. Their research was particularly geared at the correlation between training data size and classification accuracy. From their findings it was generally observed that for all training data sizes, the SVM classification results were significantly better than those derived from the decision tree classifier. The only time the SVM classifier yielded significantly better classification results than the discriminant analysis was when the largest training set was used. In the other comparisons between SVMs and the discriminant analysis, the differences were statistically insignificant. Similarly the difference in the classification accuracies between the neural network and SVMs were all statistically insignificant.

Foody and Mathur (2004b) introduce the concept of intelligent training of supervised image classification by directing training data acquisition for SVM classification. Their proposal is based on the premise that the only training data necessary for an SVM classification are the ground points corresponding to the support vectors in feature space. They derive their motivation from Tambouratzis (2002) who postulates that it is possible to reduce the size of a training set by reducing the number of training samples, hence speeding up the time it takes to train a neural network, without compromising classification accuracy. In this their ground breaking work they successfully come up with a way through which they are able to identify the ground points corresponding to support vectors in feature space.

From the previous review, the majority of the research has focused on investigating the factors that influence the application of SVMs to land cover mapping e.g. how the number of bands and quantity of training data influence the classification accuracy of SVM classifiers. There has also been interest regarding how SVMs compare with traditional classifiers and it has generally been observed that their performance is superior. Huang et al (2002) however raise a pertinent issue regarding the slower pace of training of SVMs in comparison to the traditional classifiers. As a result, this research focuses on exploring and proposing ways through which SVM classification can be optimized through data reduction. The challenge at hand is: are there ways through which remote sensing data can be optimized (i.e. reduced) without compromising classification accuracy? The next chapter discusses the data reduction methods that were proposed to effect this research.

Chapter 3

Data Reduction

3.1 Introduction

This chapter explores the various data reduction techniques proposed and investigated in the course of this research. Figure 3.1 shows a schematic representation of the proposed methods.

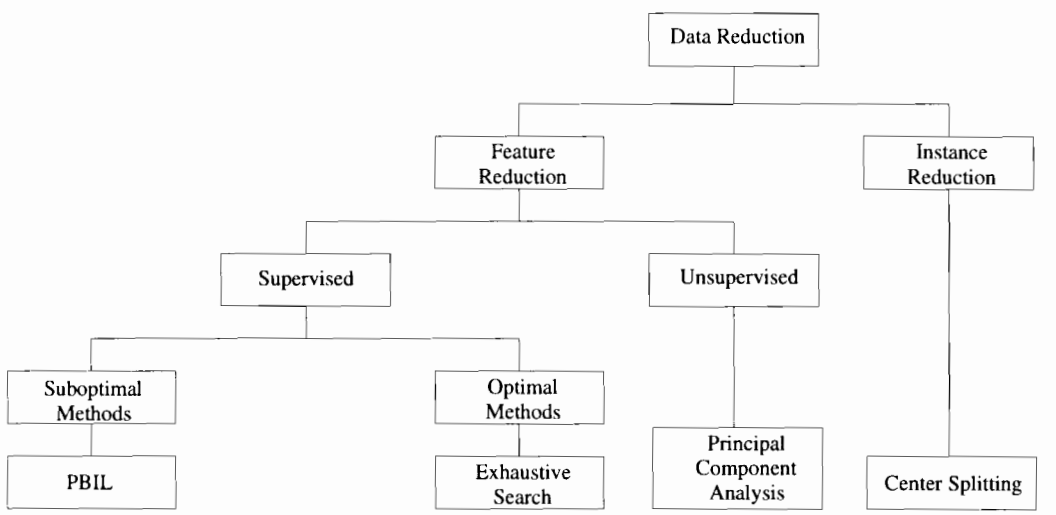


Figure 3.1: Schematic representation of data reduction methods used

3.2 Feature Reduction

In any remote sensing task, data is available for classification in many frequency bands. The number of bands varies from sensor to sensor e.g. Landsat ETM+ collects data in seven bands while ASTER collects data in fourteen bands. Whereas one could use all the sensor bands to classify a given area, these bands may have a high degree of correlation hence rendering some of the bands redundant or even irrelevant (Fangming and Guan, 2004). The factors that may potentially lead to correlation in remotely sensed satellite imagery have been presented in Chapter One. Using redundant bands inevitably leads to increased processing times and volumes of stored data. Another disadvantage of using redundant data is that it may diminish the ability to classify the study area accurately if all the bands are to be used. The focus of feature reduction therefore is to lower the dimensionality i.e. reduce the number of bands within which classification is to be carried out while preserving or improving

the discriminative ability of a given classifier (Weston et al, 2000). Feature reduction is further motivated by the fact that it is easier to visualize low dimensional data as compared to high dimensional data. It is also more computationally efficient to analyze low dimensional data (Fangming and Guan 2004; Gilad-Bachrach et al, 2004).

Feature reduction may be categorized into unsupervised and supervised feature reduction. With unsupervised feature reduction (which is also referred to as feature extraction), the original feature space (total number of bands) is projected to a subspace (reduced number of bands) using an orthogonal transformation such as the Karhunen-Loeve transformation which is better know as the Principal Component Analysis (PCA) (Pudil and Somol, 2005). The PCA transformation of the original feature space to a subspace with fewer bands enables the compression of the information available in the original feature set into a subset of uncorrelated features (Dean and Hoffer, 1983). Working with a reduced dataset has an advantage of reducing computational time, however, Bruzzone and Serpico (2000) contend that PCA results in losses to the significance of features, i.e. following a PCA transformation it is not possible to correlate the physical parameters used by a classifier with the information classes. Bruzzone and Serpico (2000) further contend that in some remote sensing applications this may represent a limitation on the understanding of the behavior of the implemented system and hence on the validation of its performance.

On the other hand, supervised feature reduction involves deriving a criterion by which a subset of the original features may be selected (Tso and Mather, 2001; Pudil and Somol, 2005). Supervised feature reduction is also commonly referred to as feature selection. Pudil and Somol (2005) postulate that the choice between supervised and unsupervised feature reduction is dependant on the application domain and nature of training data. Supervised feature reduction saves computational costs since some of the features (bands) are discarded and those identified retain their original physical meaning (Pudil and Somol, 2005). Of the two methods, Gilad-Bachrach et al (2004) favor supervised over unsupervised feature reduction techniques on account of their simplicity. This is particularly so because the supervised techniques do not require the evaluation of new complex functions of the irrelevant or redundant features. One

advantage Pudil and Somol (2005) attribute to unsupervised feature reduction is that it may provide better discriminative ability than the supervised feature reduction, due to the use of the resultant uncorrelated bands for classification. This section (3.2) examines both unsupervised and supervised feature reduction techniques.

3.2.1 Principal Component Analysis (PCA)

The technique provides a mapping of input data (bands in the case of remote sensing) to a feature space in which the new features (bands) are linear functions of the original attributes (original bands) and are sorted by the amount of variance that the data exhibited in each direction (Christianini and Shawe-Taylor, 2000). PCA can be effected in three basic steps;

1. It involves calculating the variance-covariance matrix of the multi-band image

$$C = \frac{\sum_{i=1}^{n} (x_i - m_k)(x_i - m_k)^T}{n - 1} \qquad (3.1)$$

where

x_i – pixel value in band k

m_k – mean of band k

n – number of pixels in a band

$(x_i - m_k)$ – deviation of pixel values from the band's mean pixel value

2. By nature, C is a real and symmetrical matrix of dimension k (where k refers to the original dimensionality/number of bands). This property renders it possible to extract eigenvalues and eigenvectors of derived matrix C through the following equation;

$$(C - \lambda_j I)A_j = 0 \qquad (3.2)$$

where

$A_j = (a_1, a_2, \ldots a_k)$ is the eigenvector corresponding to the eigenvalue λ_j

k – total number of feature space dimensions

I – is an identity matrix

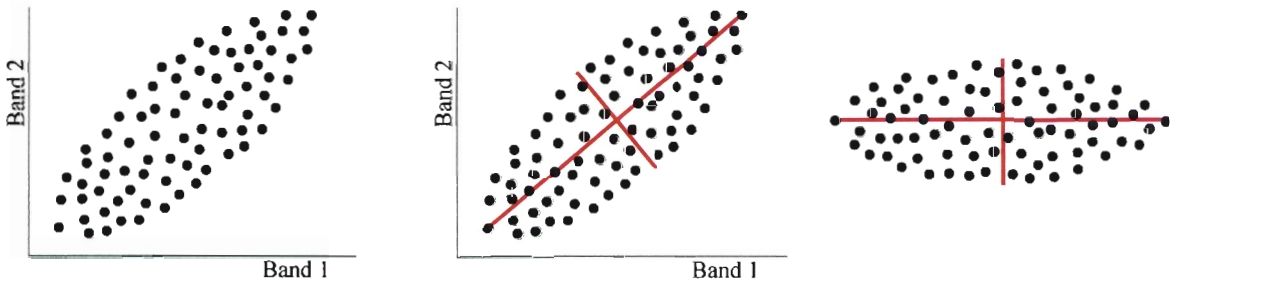
3. Having derived these eigenvectors, the third step involves projecting the original pixel values onto these vectors. The mapping of each pixel value X onto the eigenvector axes is given by

$$X_j A_j = x_1 a_1 + x_2 a_2 + \ldots + x_k a_k \qquad (3.3)$$

The corresponding eigenvalues yield a measure of importance of the eigenvectors in terms of the amount of variance or variability of the original data that they retain.

Typically PCA finds the directions in the data that express maximum variability and this is exhibited by the first few eigenvectors. The rest are mostly noise and may safely be rejected. The least number of bands containing most of the information is called the intrinsic dimensionality. For remote sensing images, the intrinsic dimensionality is usually 2 or 3. By reducing the whole array of bands to their intrinsic dimensionality, feature reduction is attained. However caution must be observed because some of the seemingly insignificant eigenvectors may represent components important to the task at hand. To illustrate this, Schott (1997) gives an example of the thermal band in Landsat TM imagery. Whereas the thermal band contains unique information not depicted in other bands, its low variance results in this vital information being mixed with noise from other bands and lost in the later principal components (Schott, 1997). Schott (1997) recommends normalizing the data by the variance in each band before performing the PCA transformation.

PCA can also be expressed graphically. Consider a two-band image with a scatter plot as depicted below:



a) Two-band image with an eliptical plot illustrating correlation between the two bands

b) The superimposed axes depict the maximum variability of the original data. These axes are determined using the eigen vectors

c) This illustration depicts the re-orientation of the original data set to the new axes devoid of correlation

Figure 3.2: Graphical representation of principal component analysis

The inclined elliptical distribution in figure 3.2a indicates correlation between bands 1 and 2. A linear transformation can be applied to these two bands in order to rotate this ellipse so that the axes are aligned according to the variability within the cluster of points. PCA is one such linear transformation. The resultant transformation consists of new uncorrelated bands referred to as principal components. The principal component containing the most variation is called the first principal component. The second principal component would have the second highest variation. There can only be as many components as the original number of bands though intrinsic dimensionality dictates that only a few would be relevant. Dean and Hoffer (1983) suggest that the sensitivity of a PCA is highly dependant upon the structure of the dataset. The effectiveness of the PCA is to a great extent dependant on the degree of correlations among the bands for a given dataset; thus the greater the interband correlation the more effective the transformation in dimensionality reduction (Dean and Hoffer, 1983).

Dean and Hoffer (1983) suggest that the PCA transformation can be used as a means of feature selection in which individual band loadings to each of the principal components are evaluated. The bands contributing the best loadings in all components are used for classification. In this research the traditional application of the PCA transformation is adopted.

3.2.2 Feature Selection

Swain and Davis (1978) define feature selection as "the search for a subset of the original measurement features that provide an optimal trade off between probability error and cost of classification". It involves selecting a subset of features (e.g. bands) from the original set of bands that captures the relevant properties of the data (Gilad-Bachrach et al, 2004) to enable adequate classification (Wu and Linders, 2000). Whereas feature selection in this research is motivated by the desire to reduce the quantity of data needed to perform image classification, there is evidence in remote sensing literature to the effect that increasing the number of bands does not automatically translate into improved classification accuracy. This is attributed to Hughes phenomenon (Hughes, 1968), which postulates that in the presence of a limited training sample size, contrary to intuition, the mean accuracy doesn't always increase with additional measurements. Rather it exhibits a peaking effect, i.e. the

accuracy will increase until a certain point beyond which increase in additional measurement will yield no significant improvement in classification accuracy (Muasher and Landgrebe, 1982). This further gives credence to the need for feature selection methods preferably those which are able to predict when/if this phenomenon occurs (Muasher and Landgrebe, 1982). In this research, Kohavi and John (1991)'s categorization of feature selection into wrapper and filter models is adopted.

The wrapper model selects features (bands) by directly optimizing the performance of the classifier. In other words, this model seeks the subset of features yielding the best classification accuracy result. Depending on the classifier adopted, different subsets will be selected. Since many subsets may have to be tried out, this approach is computationally inefficient (Kavzoglu and Mather, 2002) and to the best of the researcher's knowledge has not taken root in remote sensing research.

In the filter model on the other hand, the selection is done as a preprocessing application, without trying to optimize the performance of any specific classifier directly. This may be done through an evaluation function using a search method in order to select a set that maximizes this function. Kavzoglu and Mather (2002) define filter models as feature selection algorithms which use a performance metric (i.e. evaluation function) based entirely on the training data without reference to the classifier for which the features are to be selected. Some of the evaluation functions (also called the criterion functions) encountered in machine learning include info gain, Gamma test and naïve Bayes. Some of the criterion functions encountered in remote sensing literature include statistical separability measures such as Wilk's Λ, Hotelling's T^2 and more commonly separability indices. Separability indices give a measure of how separable classes are hence giving an indication of how easily the dataset may be correctly classified. In essence the filter model searches for subsets of bands that give the maximum separability index. Pudil and Somol (2005) further categorize these feature selection methods into optimal and suboptimal methods. Optimal methods search for the optimal subset out of all possible subsets e.g. exhaustive search methods. Suboptimal methods on the other hand make a trade off between the optimality of a selected subset and computational efficiency. Some of the search methods in this regard include greedy eliminative search, random search and guided random search methods which will be discussed in the following section.

3.2.2.1 Search Methods

In the greedy eliminative search method, the search is conducted by observing the effect of removing one band at a time on the separability index. If removing a given band reduces the separability index, that feature is restored. On the other hand, if removing the said band maintains or improves the separability index then the band is eliminated and the procedure recursively continued through the whole band set. One assumption this method makes is that each band makes an independent contribution to separability index, which may not necessarily be the case. Many times the significance of a band can only be appreciated in combination with other bands.

The exhaustive search method (also called enumerative search method) works in contrast to the greedy eliminative search method. In this method, all possible band combinations are enumerated and their separability indices calculated. Although this search method guarantees the optimality of a solution, it poses the problem of being computationally prohibitive (Pudil and Somol, 2005). For a dataset with d features (i.e. bands), $2^d - 1$ combinations are possible. This method is practicable if the number of bands is less than 10. The use of 10 or more bands would be costly in terms of computational speed. Dutra and Huber (1999) however are of the opinion that advancements in computer technology should eventually render exhaustive search an operational reality.

An alternative to the exhaustive search method is the random search method, which is especially applied when big feature sets are involved. In this method, a subset of bands is taken at random and their separability index calculated. If many bands are being considered the chances of hitting on the optimum subset in a limited random search will be small. However if good separability indices are possible from a variety of band combinations, there is a higher probability of encountering one of the optima quickly. The time it takes to converge to the optimum subset can be dramatically reduced by modifying the random search to a guided random search.

The guided random/stochastic search method is a randomized search in which attention is adaptively increased to focus on the band combinations that return promising results. In many cases this turns out to be an extremely effective strategy. It

has been known to yield results almost indistinguishable from the true optimum after only a few tries even where a large number of features are being investigated. Population Based Incremental Learning is a genetic algorithm that can be used to perform a guided random search.

3.2.2.2 Population Based Incremental Learning

The basic theory of genetic algorithms was pioneered by John Holland in the early 1970s and his full works were published in Holland (1992). Since their inception, genetic algorithms have found application in numerous engineering and scientific studies because of their ability to recognize trends and guide searches towards optimal solutions (Baluja, 1994). Some of the other remote sensing applications to which genetic algorithms have been applied to include: retrieval of land surface roughness and wetness parameters (Jin and Wang, 2001), feature mining in remotely sensed images (Luo et al 2003), development of genetic classifiers (Pal et al, 2001) and using genetic algorithms in sub-pixel mapping (Mertens et al, 2003).

Genetic algorithms are motivated by the evolutionary biology principle of natural selection and genetics. From a biological context, all living organisms are made up of cells characterized by the same set of chromosomes. Chromosomes consist of genes that encode peculiar traits. Possible settings for a trait are called alleles. During reproduction, genes from parents combine to form a new chromosome with traits from either parent. According to Darwin's theory, this breeding results in the fit traits being propagated at the expense of the weaker ones which end up being filtered away. Similarly genetic algorithms are tailored around this principle.

A genetic algorithm functions by randomly generating a population of 'chromosomes'. The 'chromosomes' are of equal length and may be represented by a special coding technique e.g. binary code (Tso and Mather, 2001). The fitness level of each 'chromosome' is then calculated based upon which random pairs of 'chromosomes' are bred to generate a new pool of 'chromosomes'. Breeding is effected through two mechanisms called crossover and mutation. Crossover involves the exchange of genes between two parent 'chromosomes' whereas mutation is carried out by randomly changing binary values that are representative of genes/traits (See Table 3.1 below).

Table 3.1: Crossover and mutation

Initial Pool		Crossover		Mutation	
Chromosome	Genes	Chromosome	Genes	Chromosome	Genes
A	1 1 0 1 1 0 1	A'	1 1 0 0 1 1 1	A''	1 1 0 1 1 0
B	1 0 0 0 1 1 1	B'	1 0 0 1 1 0 1	B''	1 0 1 1 1

Crossover and mutation facilitate genetic diversity without which the genetic algorithm would settle into a sub-optimal state. The process of selecting and breeding define a generation. The progression of genes from one generation to another is dependent on how well the 'chromosomes' pass a fitness test. The 'chromosomes' with high fitness levels may be programmed to have a higher probability of selection to ensure that strong traits are passed on to the next generation. The number of generations may be fixed or the breeding process allowed to continue until a satisfactory level of fitness is attained.

Population Based Incremental Learning (PBIL) is an adaptation of genetic algorithms whereby the population is replaced by a probability vector instead of storing each possible solution explicitly. It is this probability vector that is updated when one progresses from one generation to another rather than the fixed population (Gosling and Cromp, 2004).

Linking feature selection to PBIL, the elements of the probability vector define the probability of selection of each feature. The idea is to use PBIL to determine the subset of bands which when classified will give as good classification results (if not better) than when all bands are utilized. In the absence of *a priori* knowledge of the importance of the bands, the probability vector is initialized to a value for example 0.5. This would mean that in a randomized selection operation, each band has an equal chance of being chosen.

In each generation a population of trials is created by sampling the probability vector with a random vector. This ensures that the inclusion of a given feature follows the probabilities in the probability vector. The fitness of each trial is determined by calculating the separability index and the trial yielding the highest index is identified as the best (chromosome in genetic algorithm technology) in that generation. Based on these results the probability vector is adjusted to reflect the best trial. If for example bands 1, 3 and 4 ended up being the best trials out of 9 bands, the probability

vector corresponding to these bands would be increased slightly (by 10% from 0.5 to 0.55) while the other values would be reduced in the same proportion (from 0.5 to 0.45). This enables the subsequent generation to contain a greater proportion of trials that resemble to some degree the best trial of the previous generation.

Before proceeding to the next generation, mutation is applied to the probability vector to increase the search space in an attempt to avoid convergence towards a local optimum. Mutation may be implemented by a secondary adjustment to the probability vector in which the vector values are relaxed towards the neutral value (0.5 in this case) (Gosling et al, 2004).

Ultimately, after a series of generations the separability index of the best trial in each generation improves until the global optimum emerges. The final probability vector will also have converged towards 0 or 1 indicating the bands that had a higher or lower probability of being selected. The best trial at the end of the whole process would then represent the subset of bands with potential to yield classification accuracies comparable to those derived from using all the bands.

The interest in PBIL over the traditional genetic algorithm stems from the fact that PBIL results have been found to be more accurate, and are attained faster than the traditional genetic algorithms, both in terms of the number of evaluations performed and the clock speed. This gain in clock speed is attributed to the simplicity of the algorithm (Baluja, 1994). In view of this documented simplicity and associated accuracy, PBIL was selected to facilitate feature selection n this research.

3.2.2.3 Separability Indices

Separability indices give a measure of how well classes in a dataset are distinguishably or geometrically separate, and therefore are an indication of how well these classes can be identified. As mentioned in Chapter One, the choice of adopted separability index, i.e. evaluation function, should be related to the behavior of the error made by the classifier used if the optimal subset is to be selected (Bruzzone and Serpico, 2000). This separability analysis is performed on the training data to give an indication of the expected classification error for various band combinations (Swain and Davis, 1978). However, finding an appropriate definition of interclass separability

is not trivial (Schowengerdt, 1997). Distance measure between class means is one of the basic measures of separation encountered in remote sensing literature. Some of the distance measures include city block and Euclidean distance. Table 3.2 gives a summary of the separability indices discussed and their mathematical formulation. Schowengerdt (1997) expresses reservation about the city block and Euclidean distance measures as separability indices because they do not account for overlap of class distributions due to their variances. It is on this account that Schowengerdt (1997) further concludes that they are not particularly good measures of separability. Schowengerdt (1997) proposes the normalized city block as an improvement to the original city block since it is directly proportional to the separation of the class means and inversely proportional to their standard deviations. Schowengerdt (1997) also presents the Mahalanobis separability measure as a multivariate generalization of the Euclidean measure for normal distributions. Some of the more widely used separability indices encountered in the remote sensing research community include the divergence, transformed divergence, Bhattacharyya and Jeffreys and Matusita (JM) distance measures. Of these measures, the JM distance measure has been deemed to yield better results (Dutra and Huber, 1999). Some of the sources of a more detailed discussion of these distance measures include Swain and King (1973) and Schowengerdt (1997).

Table 3.2 displays the distance measures between two distributions a and b in feature space. μ_x refers to the mean of class a and b. As shown on the table, city block and Euclidean distance ignore the covariances of the distributions. The normalized city block and Mahalanobis measures are extensions of the city block and Euclidean distance respectively that include the covariance (C) information. The last four measures assume normal class distributions for classes a and b. All of the distance measures are scalars.

Table 3.2: Common distance measures

Distance Measures	Formulae
City Block	$L_1 = \lvert \mu_a - \mu_b \rvert = \sum_{k=1}^{k} \lvert m_{ak} - m_{bk} \rvert$
Euclidean distance	$L_2 = \lVert \mu_a - \mu_b \rVert = \left[(\mu_a - \mu_b)^T (\mu_a - \mu_b) \right]^{\frac{1}{2}}$
Normalized City block	$NL_1 = \sum_{k=1}^{k} \frac{m_{ak} - m_{bk}}{\frac{\sqrt{C_{ak}} + \sqrt{C_{bk}}}{2}}$
Mahalanobis	$MH = \left[(\mu_a - \mu_b)^T \left(\frac{C_a + C_b}{2} \right)^{-1} (\mu_a - \mu_b) \right]^{\frac{1}{2}}$
Divergence	$D = \frac{1}{2} tr\left[(C_a - C_b)(C_a^{-1} - C_b^{-1}) \right] + \frac{1}{2} tr\left[(C_a^{-1} + C_b^{-1})(\mu_a - \mu_b)(\mu_a - \mu_b)^T \right]$ Where tr – trace: sum of diagonal elements of the matrces
Transformed divergence	$D^t = 2\left[1 - e^{\frac{-D}{8}} \right]$
Bhattacharyya	$B = \frac{1}{8} MH + \frac{1}{2} \ln \left[\frac{\left\lvert \frac{C_a + C_b}{2} \right\rvert}{(\lvert C_a \rVert C_b \rvert)^{\frac{1}{2}}} \right]$
JM	$JM = \left[2\left(1 - e^{-B} \right) \right]^{\frac{1}{2}}$

(Source: Schowengerdt, 1997)

As mentioned before, it is important that separability measure is matched with the classifier whose classification error it matches best. To this effect, the city block and Euclidean distance are recommended for the minimum distance to means classifier while the other mentioned distance measures are better suited for the maximum likelihood classifier (Schowengerdt, 1997).

In light of the fact that nonparametric classifiers are gaining prominence in remote sensing studies, there is a corresponding need for separability measures tailored around the uniqueness of these nonparametric classifiers. The previously discussed distance measures work best for traditional classifiers because they can give an indication of the corresponding estimated classification error. Whereas Kavzoglu and Mather (2002) have used the Mahalanobis distance classifier to select features for artificial neural networks (which is a nonparametric classifier like the SVMs used in this research), this research proposes a separability index tailored for the uniqueness of nonparametric classifiers. This separability measure is called the Thornton's separability index and also has its origin in the machine learning field.

Thornton's separability index is defined as the fraction of data points whose classification labels are the same as those of their nearest neighbors. Thus it is a measure of the degree to which inputs associated with the same output tend to cluster together (Greene, 2001). Greene (2001) expresses this concept by the following formula;

$$\text{Thornton's separability index} = \frac{\sum_{i=1}^{n} (f(x_i) + f(x_i') + 1) \bmod 2}{n} \qquad (3.4)$$

Where x' is the nearest neighbor of x and n is the number of points.

With this separability measure, well separated classes will have high separability indices equal to or tending towards unity. As the clusters move closer to each other and points from opposing classes begin to overlap and the index begins to fall (Greene, 2001). A separability index of zero, therefore, would denote totally mixed up classes. The interest in this separability measure stems from its simplicity, speed and the fact that it is nonparametric which augers well with the SVM classifier which is also nonparametric.

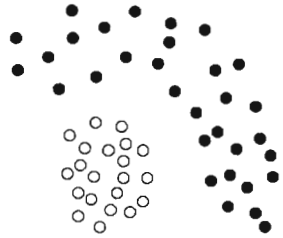
3.3 Instance Reduction

In feature reduction, the focus was on reducing the number of bands to be used in classification. In instance reduction, on the other hand, emphasis is put on deriving means through which the size of the training data may be reduced without compromising on the classification accuracy. This approach is sometimes referred to

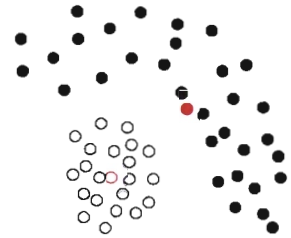
as data editing or pruning. The anticipated advantage of using this technique to train a classifier is that with the resultant sparse dataset, training can be accomplished much faster. Skalak (1997) categorizes these instance reduction techniques into; instance filtering, stochastic search and Instance averaging/abstraction methods.

In brief, instance filtering involves scanning the original dataset and singling out subsets of 'good' instances and filtering out the 'bad' instances. Stochastic search methods on the other hand involve randomly sampling the dataset to obtain representative instances. In instance averaging/abstraction algorithms, a collection of same class instances clustered together in feature space are represented by their mean attribute value called a prototype. This research was based on an adaptation of the instance averaging/abstraction technique by Thornton (2000), which he called the center splitting method.

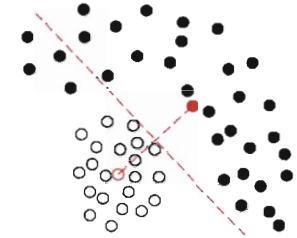
The center splitting algorithm begins by creating prototypes at the mean positions of each class dataset. The labels of the nearest instances are compared and any subsequent instances belonging to another class are assigned their own prototype point. This is iteratively continued until no prototype has instances nearest to it with an opposite class label. In an attempt to counter the effect of outliers, any prototype that has only one instance of the same class in its vicinity is deleted. The prototypes will then constitute the sparse dataset to be used to train the classifier. Figure 3.3 gives a graphical illustration of this process.



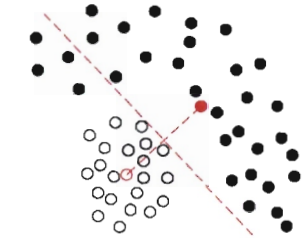
a) Consider a training dataset with two classes labelled +1 and -1



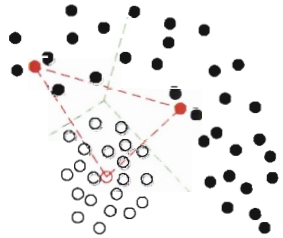
b) A prototype is located at the mean position of each class and assumes that class' label



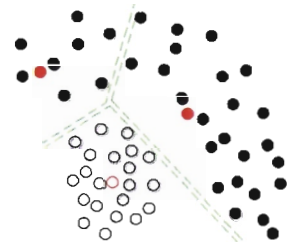
c) The prototypes then consider the instances closest to them. The perpendicular to the line joining them determines the search region for each prototype



d) The prototypes compare all the labels in their half until they encounter instances with different labels. The instances with different labels are assigned their own prototype center



e) New search regions are defined by the perpendiculars (green lines) to the (red) lines joining the prototypes



f) The search process is repeated until all surrounding instances share the same label as the prototype. Any prototype surrounded by only one like instance is ignored

g) The resultant prototypes representative of the original dataset described in a). Evidently it is a much smaller dataset, thus reducing on the data used to train classifier

Figure 3.3: Center splitting

Chapter 4

Accuracy Assessment

4.1 Introduction

The classification of remotely sensed data provides a means by which land cover maps can be derived detailing patterns, areal extent and quantification of various cover types (Stehman, 1997). Derived land cover maps contain different types of errors which should be determined for the benefit of both the map users and producers (Congalton and Green, 1998). The evaluation of these errors in remotely sensed spatial data is referred to as accuracy assessment. Accuracy assessment is motivated by the need to assess the quality of a given map and/or the need to compare different classification techniques, algorithms or interpreters. Congalton (2001) proposes a number of ways of investigating the accuracy/error in spatial data. These steps may be viewed as a progression and some of the mentioned methods are as follows:

- Visual inspection
- Non-site specific analysis
- Difference image creation
- Use of error matrices

In this chapter, an overview of issues surrounding accuracy assessment of derived land cover maps is given.

4.2 Accuracy Assessment Methods

4.2.1 Visual Inspection

This method involves the visual appraisal of the derived maps (Foody, 2002b). Visual inspection serves as an important preliminary step in any classification project. The map is considered accurate if it looks 'right' or 'good'. This may amount to the derived map's land cover distribution bearing similarity with the viewer's perception of the actual cover distribution. Congalton (2001) and Foody (2002b) both concur that this method's subjective nature renders it inappropriate to be used in isolation. Congalton (2001) continues to mention, however, that if the visual appraisal of the

derived map is unsatisfactory, then the analysis/classification should be redone before further accuracy assessment is undertaken.

4.2.2 Non-site Specific Analysis

Non-site specific analysis involves the comparison of overall amounts or areal extents of land cover types in the derived map relative to ground data, without referring to any locational component (Congalton, 2001; Foody, 2002b). The comparison may be measured in terms of acreage or percentage cover of the region mapped (Foody, 2002b). The non-site specific nature of this approach is, however, a major limitation as a map could easily display the classes in correct proportions but in incorrect locations, greatly limiting the value of the map for some users (Foody, 2002b).

4.2.3 Difference Image Creation

In this method, two registered images are directly compared hence incorporating spatial content in the evaluation of map error. This method is effected by comparing the derived image/map with the reference image pixel by pixel and representing the areas of agreement in one color (usually black) and those of disagreement in another color (usually white). The resultant difference image represents the exact pattern of error in the map.

4.2.4 Use of Error Matrices

Error matrices are key to the derivation of quantitative accuracy measures (Congalton, 2001) and are sometimes referred to as contingency tables or confusion matrices. Error matrices enable the comparison of the reference data, assumed accurate, with the predicted land cover classes in the derived map. An error matrix is arranged such that the columns represent the reference data while the rows represent the predicted classes. Each diagonal value represents the number of correctly predicted pixels while off-diagonal values represent the number of wrongly classified pixels. The error matrix serves as a starting point from which a series of descriptive and analytical statistical measures can be derived (Campbell, 1996; Congalton, 2001) and are discussed in section 4.3. An example of an error matrix can be seen in Appendix 1.

4.3 Quantitative Accuracy Assessment Measures

Stehman (1997) categorizes quantitative accuracy assessment measures as either being single summary measures of the error matrix or class-level accuracy measures.

4.3.1 Single Summary Measures of the Error Matrix

The simplest single summary measure of the error matrix is the overall accuracy. This measure gives a quantitative assessment of how the derived classification map compares with ground truth. It is derived by ratioing the sum of the main diagonal matrix values (i.e. the correctly classified pixels) to the total number of pixels used for the accuracy assessment.

$$\text{Overall Accuracy} = \frac{\text{Sum of Correctly Classified Pixels}}{\text{Total Number of Pixels}} \quad (4.1)$$

Another single summary measure of the error matrix is based on Kappa analysis, which yields a KHAT statistic. This is a discrete multivariate technique developed by Cohen (1960) and has been used for evaluating accuracies of the land cover maps derived from remotely sensed images. Kappa analysis gives a measure of agreement between the reference data and the derived map. It ranges from –1 to +1, however since there should be a positive relationship between the derived map and the reference data, positive values are expected. According to Rossiter (2004), the rationale for using the Kappa analysis is based on the premise that overall accuracy may be due to chance especially in a situation where some land cover types are more dominant hence making them more likely to be evaluated during field sampling than the less dominant cover types. Rossiter (2004) continues to illustrate this by giving an example of a mapping exercise where two classes A and B are to be evaluated. If class A covers 90% of the study area and Class B the remaining 10%, it is evident that statistically, a randomly selected sampling point has a 90% chance of being in class A and a 10% chance of being in class B. This implies that a high accuracy assessment would be expected irrespective of the actual pattern shown on the map, implying that it would be hard to distinguish a good map from a lucky one (Rossiter, 2004). A map with a dominant land cover class will therefore appear more accurate than one with many classes because of a simpler legend. The use of Kappa analysis is part of the

attempt to compensate for this bias. Kappa analysis is carried out by performing the following computation;

$$\hat{k} = \frac{N\sum_{i=1}^{r} p_{ii} - \sum_{i=1}^{r} p_{i+}p_{+i}}{N^2 - \sum_{i=1}^{r} p_{i+}p_{+i}} \qquad (4.2)$$

where

$\hat{k}$ = Agreement coefficient

N = Total number of pixels

p_{ii} = Total number of correctly classified pixels for given category $_i$

p_{i+}= Row total

p_{+i} = Column total

r = Number of rows in the error matrix

Landis and Kotch (1977) characterize the possible ranges for KHAT into three groups as shown in Table 4.1:

Table 4.1: Characterization of the KHAT statistic

Range	Remark
$\hat{k} \geq 0.8$	Very strong agreement
$0.4 < \hat{k} < 0.8$	Moderate to strong agreement
$\hat{k} < 0.4$	Poor agreement

In general, a $\hat{k}$ value of 1 would imply that there is no disagreement between the reference data and the derived map, while a $\hat{k}$ value of 0 would imply that the disagreements that occurred were exactly those expected by chance. If the $\hat{k}$ value was less than 0 it would imply that the derived map has more disagreement with the reference data than would be expected from a pure random assignment of classified pixels to classes (Rossitor, 2004).

Given that Kappa analysis is computed from all the elements in the error matrix, as opposed to the overall accuracy, it is deemed to be a better measure for the comparison of accuracies resulting from different classification algorithms. The use of Kappa analysis instead of overall accuracy for accuracy assessment becomes less

important as the number of classes increases as well as when the balance between classes becomes more even. Other single summary measures worth mentioning which will not be discussed are the modified Kappa statistic (Foody, 1992) and Kendell's Tau statistic.

These single summary measures discussed provide a means through which different classification algorithms can be compared or evaluated. The standard approach in remote sensing is to perform a binomial test of significance between the algorithms being compared. In the simple case of determining if there is a difference between two classifications (2 sided test), the null hypothesis (H_o) that there is no significant difference, would be rejected if $|Z| > 1.96$ (Congalton et al, 1983; Rosenfield and Fitzpatrick-Lins, 1986; Congalton and Green, 1998). Both overall accuracy and KHAT statictic values have been used in literature to compare algorithms in this manner. To calculate the Z score for the difference between two classifications using the overall accuracy, the following formula may be used;

$$Z = \frac{|O.A_1 - O.A_2|}{\sqrt{\frac{O.A_1(1-O.A_1)}{\text{No. of Sample Points}} + \frac{O.A_2(1-O.A_2)}{\text{No. of Sample Points}}}} \qquad (4.3)$$

Where

$O.A_1$ – Overall accuracy of first error matrix

$O.A_2$- Overall accuracy of second error matrix

While using the KHAT statistic values as a measure of comparison, the following formula may be used;

$$Z = \frac{\hat{k}_1 - \hat{k}_2}{\sqrt{\hat{\sigma}_{k_1}^2 + \hat{\sigma}_{k_2}^2}} \qquad (4.4)$$

Where $\hat{k}_1$ and $\hat{k}_2$ are assumed to be the independent KHAT statistic values for the respective error matrices under investigation. $\hat{\sigma}_{k_1}^2$ and $\hat{\sigma}_{k_2}^2$ represent the estimated variances of the derived KHAT statistic values.

In remote sensing studies the same set of sites are used in the assessment of the accuracy of the thematic maps to be compared. Consequently samples are not independent and the McNemar test provides a statistically rigorous means of handling

such data (Foody, 2004). The McNemar test is nonparametric and is applied to error matrices that are 2 X 2 in dimension. The constraint on the size of matrices is often not a problem because larger matrices can be collapsed to this size since attention is, in effect, focused on the binary distinction between correct and incorrect classes (Foody, 2004). The McNemar test is based on the standardized normal statistic;

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \qquad (4.5)$$

In which f_{ij} indicates the frequency of sites lying in error matrix positions i and j as depicted in Table 4.2.

Table 4.2: McNemar confusion matrix

		Classification 2		
Classification 1	Allocation	Correct	Incorrect	$\sum$
	Correct	f_{11}	f_{12}	
	Incorrect	f_{21}	f_{22}	
	$\sum$			

4.3.2 Class-level Accuracy Measures

Class-level accuracy measures enable the assessment of each individual land cover type. The two main class-level accuracy measures are Producer's and User's accuracies.

Producer's accuracy refers to the ratio of the total number of correct samples in a land cover category to the total number of samples of that category derived from the reference data (column total). Producer's accuracy gives a measure of the proportion of pixels in a test dataset that are correctly labeled/categorized by the classifier. It is hence referred to as the error of omission. Producer accuracy is calculated as follows:

$$\text{Producer's Accuracy} = \frac{\text{Correctly Classified Pixels}}{\text{Column Total}} \qquad (4.6)$$

User's accuracy on the other hand refers to the ratio of the total number of correct samples in a category to the total number of samples of that category derived from the resultant classification (row total). User's accuracy refers to the reliability of the derived map and gives an indication of the probability that a sample classified on the

map actually represents that cover type on the ground. It is hence referred to as the error of commission. User accuracy is calculated as follows:

$$\text{User's Accuracy} = \frac{\text{Correctly Classified Pixels}}{\text{Row Total}} \qquad (4.7)$$

It is also possible to calculate a measure of agreement between the reference data and the derived map for each category by using the conditional Kappa coefficient of agreement. Whereas the other accuracies only consider the diagonal elements of the error matrix, the Kappa value considers the off diagonal values as well, hence giving a more holistic interpretation of error matrix as compared to the producer's and user's accuracies (Fung and LeDraw, 1988). The conditional Kappa coefficient of agreement gives an indication of which land cover classes are well mapped and is evaluated as follows:

$$\hat{k}_i = \frac{N(p_{ii}) - (p_{i+})(p_{+i})}{N(p_{i+}) - (p_{i+})(p_{+1})} \qquad (4.8)$$

where

$\hat{k}_i$ = Agreement coefficient for the i^{th} category/land cover type

N = Total number of pixels

p_{ii} = Total number of correctly classified pixels for given category $_i$

p_{i+}= Row total

p_{+i} = Column total

4.3.3 The Normalized Error Matrix

Congalton et al (1983) propose an adaptation to the traditional error matrix in which the error matrix values are normalized using a technique called MARGFIT. In this technique, each row and column total (also called marginals) undergo an iterative proportional fitting procedure which forces the marginals to sum to one, hence the name MARGFIT. This procedure enables inter-class comparisons regardless of previous sample sizes. Given that the marginals sum to one, the resulting normalized matrix is more indicative of the off-diagonal matrix values than the original error matrix. The normalized overall accuracy can similarly be computed by summing the normalized matrix main diagonal.

4.3.4 Sampling Scheme

The assumption made when evaluating error matrices is that the reference data used is an accurate representation of the study area of interest. Since using the study area as a whole as reference data is impracticable, sampling is the only means by which accuracy assessment of a land cover map can be derived (Maingi et al, 2002). In order to ensure that the sampled reference data is representative of the study area, certain statistical considerations are required such as; an appropriate sampling scheme, sample size, sampling unit, maintaining independence between training and reference data and considering the effects of spatial correlation (Congalton, 1991).

A sampling scheme describes the way in which sampled pixels can be selected to facilitate comparison between reference data and the derived map. Reference data may be sourced from field observations, topographic maps, aerial photographs and high-resolution images. Verbyla (2002) categorizes the sampling of data into random sampling and vector vegetation polygons otherwise known as cluster sampling.

4.3.4.1 Random Sampling

Random sampling as the name suggests involves randomly selecting individual pixels in the derived image and locating these pixels in the field in a bid to compare the derived land cover class with the corresponding actual class. Whereas Congalton (2001) contends that random sampling has very good statistical properties, Verbyla (2002) highlights the following shortcomings:

- It is impossible to find single pixels in the field because of positional error in rectification and errors in field navigation
- By selecting pixels randomly, most sampled points will be from dominant land cover classes yet the most important cover classes could be relatively rare in the study area and not likely to be randomly selected.
- Randomly sampled points falling in remote areas may be expensive to travel to hence one may be led to compromise on the number of samples used which may affect the assessment negatively especially for uncommon land cover classes

The various random sampling techniques include: simple random sampling, systematic random sampling and stratified random sampling. In a random sampling scheme, the sampled pixels are randomly chosen from the image. It is relatively easy to perform and provides adequate estimates of the population parameters provided the sample size is big enough. The downside to this technique is that rare land cover classes may be under sampled or even missed out. In many cases, it is even necessary to remain close to roads, which ultimately biases the sampling (Thomas and Allcock, 1984). Systematic random sampling involves the sampling of the area of interest mostly in a gridded pattern. Ground truthing a systematic random sampling scheme is a lot easier than the simple random sampling scheme, however, in similar fashion small important classes may be missed out. On the other hand, stratified random sampling involves identifying the land cover classes of interest and randomly distributing the sampling points within that area.

4.3.4.2 Cluster sampling

Cluster sampling involves the selection of groupings of pixels representative of a land cover class of interest. This sampling approach has the advantage of saving time, reducing travel costs and enables the generation of a large sample size for a fixed cost (Stehman, 1997). Tso and Mather (2001) however cautions against the use of large cluster samples. This is because pixels within a group are not mutually independent, for example, a sample of 40 contiguous pixels does not represent 40 independent samples due to the autocorrelation effect. Congalton (1988) suggests that no cluster should be greater than ten (10) pixels. Some of the shortcomings of the cluster-sampling scheme are listed by Verbyla (2002) and include:

- There could be a danger of biased accuracy assessment if the clusters have larger minimum mapping units relative to the classified image.
- The reference data may be much older than the derived classification image hence leading to conservatively biased accuracy assessment.
- There is positional error inherent in both the clusters and classified images. The positional error may lead to positional differences between correctly classified pixels and correctly classified clusters, which may also lead to a conservative estimate of classification accuracy.

4.3.5 Sample Size

Sample size is dictated by the need to express accuracy in an error matrix (Congalton, 2001). As a general rule, at least 50 sample points per land cover class are required to adequately populate an error matrix (Curran, 1985; Story and Congalton, 1986). Curran (1985) however states that this is possible only where the study area changes over time, the field season is long and the sample sites are easily accessible. As this is rarely the case, Curran (1985) suggests that the sample size could be calculated with a much smaller sample size provided the results are treated as only approximations. Story and Congalton (1986) propose 30 sampling points as the absolute minimum sample size.

4.3.6 Sampling Unit

According to Stehman and Czaplewski (1998), the sampling unit serves as the fundamental unit on which accuracy assessment is based because it serves as the link between a spatial location on a derived map and its corresponding location on the ground. Sampling units may either be point or areal units. As opposed to areal sampling units, point units have no areal extents. Congalton (2001) and Stehman and Czaplewski (1998) mention three types of areal sampling units namely pixels, polygons and fixed area plots.

Pixels are usually uniform in shape and size. They represent small areas and are related to point sampling units, however their areal extent distinguishes them from point sampling units. Congalton (2001) expresses reservations about using point sampling because of the inability to accurately locate pixel positions on the image and their corresponding locations on the ground. Polygon sampling units on the other hand are irregular in shape and could differ in size. They represent homogenous land cover types in both the derived map and their corresponding ground locations. Fixed area plots are regularly shaped in an array format (e.g. 3x3). From the research carried out by Stehman and Czaplewski (1998), no consensus exists regarding the best sampling unit. Every mapping project should be evaluated on a case-by-case basis given the possible variation in project objectives, landscape characteristics, features being mapped and the practical constraints thereof.

4.3.7 Independence between Training and Reference Data

In order to ensure sound statistical analysis, care should be taken to ensure that the data used in the derivation of the classified map is different from the data used in accuracy assessment. Using the training data for accuracy assessment may result in higher accuracy results than should be.

Chapter 5

Methodology

5.1 Introduction

This chapter attempts to detail the methodology adopted in the execution of this research. Issues addressed include data processing, derivation of SVM parameters and SVM classification, optimization techniques explored and accuracy assessment. Figure 5.1 shows a simplified flow chart of the classification process.

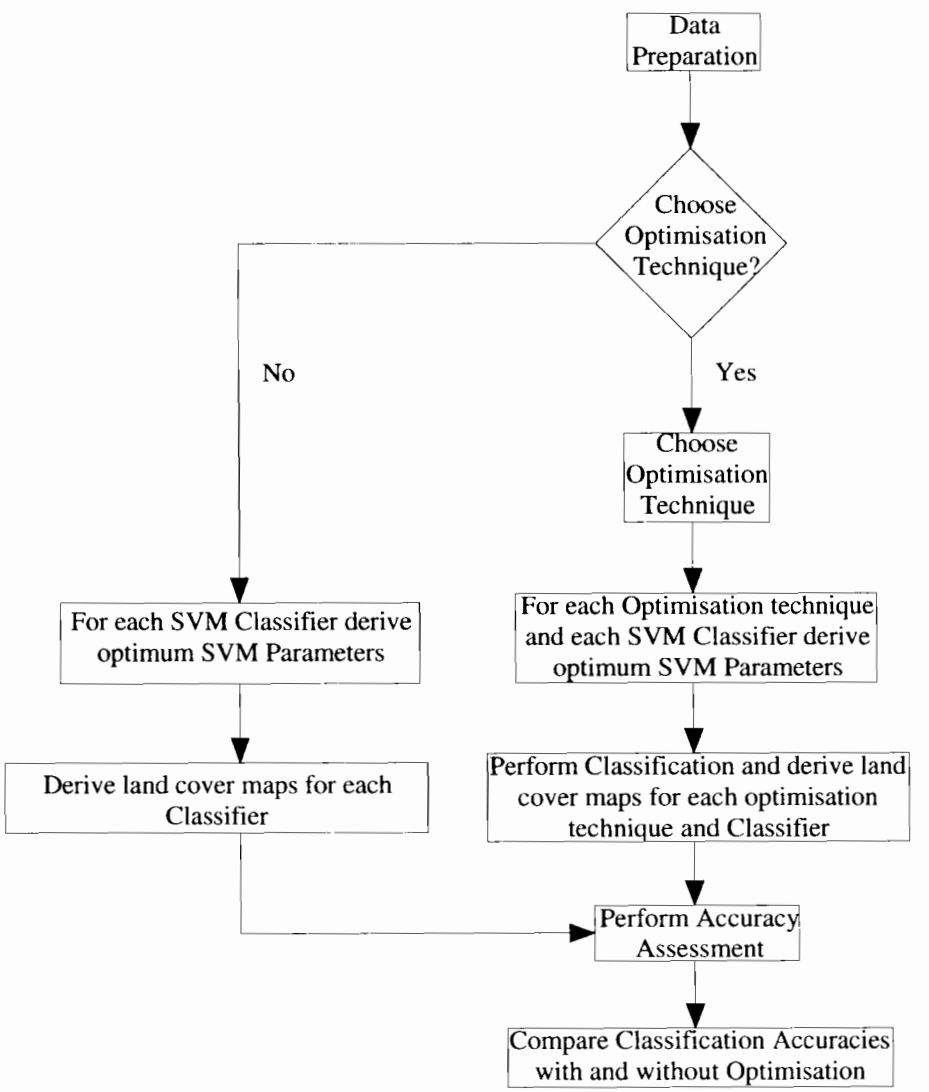


Figure 5.1: Simplified flow chart of adopted classification analysis process

5.2 Data Preparation

For each of the study areas, data preparation was effected in two stages. The first stage was carried out using IDRISI (Kilimanjaro edition). This stage basically involved importing the image scene into IDRISI, sectioning off the study area, georeferencing the resultant dataset and selecting representative training areas corresponding to known land cover types.

The East Rand study area was mapped using ASTER imagery. For the purposes of this research, only the visible and the Short Wave Infra Red (SWIR) bands were used. In total, nine bands out of the original fourteen bands were used in this classification task. Owing to differences in their spatial resolutions, the SWIR band images were resampled from 30 meters to 15 meters to match the visible band's spatial resolution. The selected classes included agricultural fields, water, roads, soils, tailings, tailings dams, yellow tailings, trees, wetlands, 'soil and trees'. The classes were selected based on the classification done by Chevrel et al (2003). Both the Mara Basin and Malmesbury datasets were mapped using Landsat Imagery. In both cases, only the visible and SWIR bands were used, leaving out the thermal band. In total, six out of the seven bands were used for classification. All the bands had a spatial resolution of 30 meters, hence no resampling was necessary. The classes identified in the Mara Basin dataset included wetlands, water, shrub/trees, grasslands, 'bare ground' and roads. While those in the Malmesbury dataset were built up areas, fields and trees. In both Mara Basin and Malmesbury, the land cover classes were identified through field observations.

In all three study areas, samples (training areas) were extracted from each band and for each land cover type and exported to a format that could be accessible by MATLAB software in which the SVM classification was to be carried out. The file format selected for this purpose was the bitmap file format. These training areas were used to train the SVM classifiers, following which eventual classification of the study area was carried out. This necessitated exporting each of the image bands of the study areas into a similar format accessible by MATLAB.

The second stage of data preparation involved processing the data using MATLAB. For each study area, the sampled areas are imported individually for each class corresponding to each band. In MATLAB this data was reconfigured and regrouped before being used to train the classifier i.e. derive optimum SVM parameters. Figure 5.2 gives a graphical illustration of the developed data preparation and SVM classification methodology (See No. 1 – 4 in Figure 5.2). To explain the illustration, consider the preparation of a water sample for classification from an ASTER image. For each water sample, there will be nine corresponding images relating to the ASTER bands under consideration (for Landsat there would be six bands). Each sample image is imported separately and reshaped from a two dimensional image to a column vector (see Figure 5.2). The column vectors are then regrouped to form a nine dimensional matrix corresponding to each band. In the case of Landsat imagery, the resultant matrix will be six dimensional. The water sample is then randomly divided into training and testing data so as to carry out cross validation to determine optimum SVM parameters. This process is repeated for each land cover class, following which the data is further processed to perform the SVM classification as discussed in the following section.

5.3 SVM Classification

SVMs are essentially binary classifiers i.e. they are primarily designed to separate two classes at a time. They however can be adapted to multi-class classification tasks which are common to remote sensing. Of the two approaches suggested in Chapter Two (see section 2.3.2.5), the one against all approach was used in this research. In this approach, for each study area, the classes of interest were classified separately and later on combined to form a land cover map as will be discussed later in this section. The OSU_SVM 3.0 toolbox (developed at the Oklahoma State University) was used to provide SVM functionality to MATLAB. This toolbox was selected for its simplicity and ease of use. The three classifiers experimented with included the linear, polynomial and radial basis function SVMs. For each classifier and optimization technique, a corresponding land cover classification map was derived. Table 5.1 lists the SVM parameters whose optimum values were searched for.

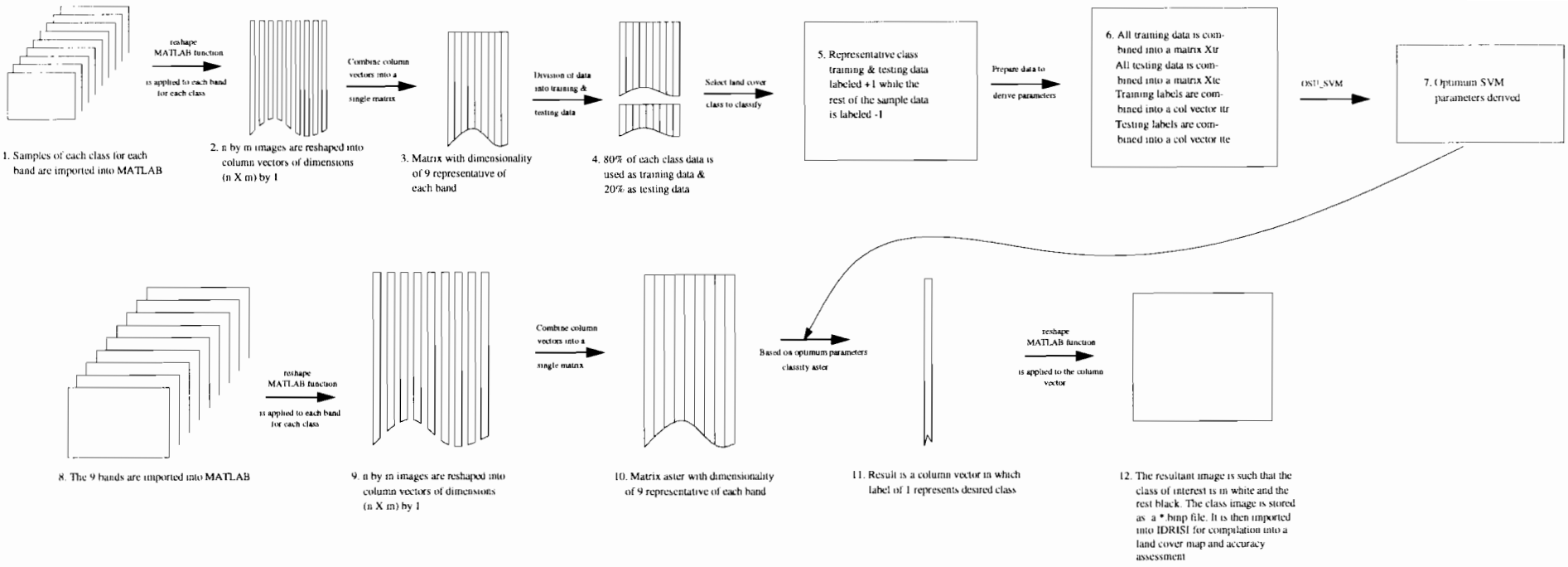


Figure 5.2: Data preparation and SVM classification methodology

Table 5.1: SVM parameters

SVM	Parameter Optimized
Linear	C
Polynomial	Polynomial Degree and C
Radial Basis Function	Kernel width and C

In order to determine the SVM parameters listed in Table 5.1, cross validation was used to develop a parameter search model. The concept of cross validation is that sample data of known land cover type is divided into training and testing data. The training data is used to train the classifier i.e. generate classification parameters. It is based on these derived classification parameters that predictions are made of the test data's land cover type. By comparing the predicted and actual testing data land cover types, the percentage of correctly classified data can be used as a measure of classification accuracy. The SVM parameters yielding the highest accuracy are then selected to effect classification of the study area. In Figure 5.2 this is illustrated by No. 5 – 7 and is discussed further in the next paragraph.

Figure 5.2 is an elaboration of Figure 5.1, and illustrates the classification procedure developed for this research. As mentioned previously, for any given study area, all the samples representative of the known land cover types were imported and processed as shown in Figure 5.2. The illustration is based on the East Rand dataset which had nine bands and ten land cover types (classes) to be distinguished. The developed methodology was generalized to suit the classification of the other study areas. Having imported all the class samples into MATLAB, they were divided into training and testing data. To effect the one against all approach, each class was classified against the non-class data, one at a time e.g. water class were classified against the non-water class. The non-water class in this case would be a combination of all the other classes. The water samples in both the training and testing data were then assigned a label +1 while the non-water samples (both training and testing) were assigned a label -1. The training data was then reconstituted into a nine dimensional matrix corresponding to the number of bands used (for Landsat imagery it would be a six dimensional matrix) consisting of both water and non-water areas and designated as Xtr. The corresponding labels were reconstituted into a column vector designated ttr. The testing data was also reconstituted into a matrix Xte consisting of water and non-water instances (samples). Its corresponding labels were reconstituted into a

column vector designated tte. Xtr and ttr were then used as input to explore various SVM parameters. For each set of derived classification parameters, Xte was used as input to derive new labels tte', which were then compared with the original labels (tte). The SVM parameters yielding the highest percentage of correct labels were used to classify the whole study area.

To finally classify the study area, the original bands had to be imported into MATLAB. IDRISI was used to export each band into a format accessible to MATLAB, which in this case was the bitmap file format. Each band was individually imported, reconfigured and reconstituted (as was done to the sample data) into a nine dimensional matrix called *aster* (see No. 8 – 10 in Figure 5.2). Based on the optimum SVM parameters so derived, the *aster* matrix was classified. The predicted labels of *aster* formed a column vector in which the sought after class had a label of 1 and the 'non sought after' class a label of –1 (see no. 11 in Figure 5.2). This column vector was then reshaped into the dimensions of the band images at the time of importation (i.e. before they were reshaped) and exported as a bitmap class image. Figure 5.3 depicts an example of one such class image that would correspond to No. 12 in Figure 5.2. Notice the white areas represent the class areas of interest (previously labeled +1) and the black areas the 'non sought after' areas (previously labeled -1).

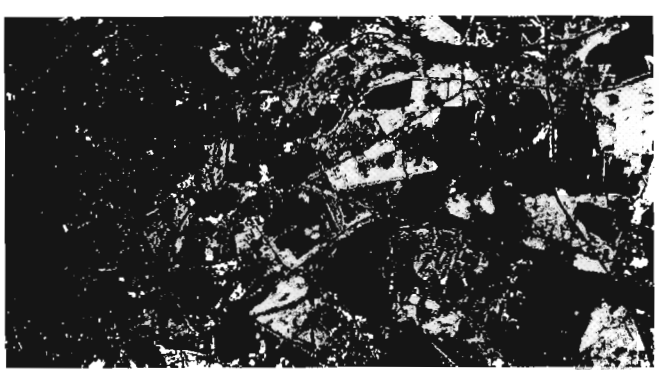


Figure 5.3: Class image

Each bitmap class image was imported into IDRISI for further processing into a land cover map corresponding to a given classifier and optimization technique. This processing stage involved georeferencing the class images, assigning new labels to the individual classes and then integrating them into a single land cover classification map. The resultant land cover map was then subjected to accuracy assessment discussed in section 5.5.

5.4 Optimization Techniques

5.4.1 Principal Component Analysis

For each of the study areas, Principal Component Analysis (PCA) was applied to the image bands using IDRISI's PCA functionality. The first three components were then selected for classification purposes. Training areas corresponding to known land cover were selected from each of the three components and exported as bitmaps for importation into MATLAB. In MATLAB, these data were then processed as described in the previous section (5.3). With reference to Figure 5.2, instead of working with nine dimensional matrices, during the processing of the PCA data, only three dimensional matrices are used corresponding to each of the PCA components. This had the effect of reducing the amount of data processed hence the interest in this method as an optimization technique. PCA enabled the reduction of ASTER imagery by a third and the Landsat imagery by a half. Of all the optimization techniques investigated, this was the only one done by IDRISI, the rest were done by MATLAB and are discussed hereafter.

5.4.2 Feature Selection

5.4.2.1 Exhaustive Search

As discussed in section 3.1.2, the idea behind feature selection was to identify the subset of bands from the parent dataset which had potential to yield equally good classification accuracy. The exhaustive search method was used to compare all possible band combinations with the original band set. For the ASTER dataset, the number of subsets was 511 while for both Landsat datasets there were 63 subsets. The measure of comparison was the Separability Index (S.I) discussed in section 3.1.2.3. For each study area and each land cover type, the S.I of each subset was compared with the S.I of the original bands. The subset with the least number of bands yielding the closest S.I to the original bands was selected for the classification process. This subset of bands was then used to derive optimum SVM parameters as described in section 5.3, which were subsequently used to classify the study area. From Figure 5.3, the training and testing matrix dimensions depended on the number of bands identified. Each class image was exported as a bitmap for importation into IDRISI for further processing as discussed in section 5.3.

5.4.2.2 Population Based Incremental Learning (PBIL)

PBIL was used to identify the best subset of bands with potential to yield classification accuracies similar to when all the band sets were used. As opposed to the exhaustive search method which investigated all possible band combinations, PBIL performed a guided search for the best band combinations. The simplified flow chart (see Figure 5.4) presents the basis upon which the PBIL algorithm was developed.

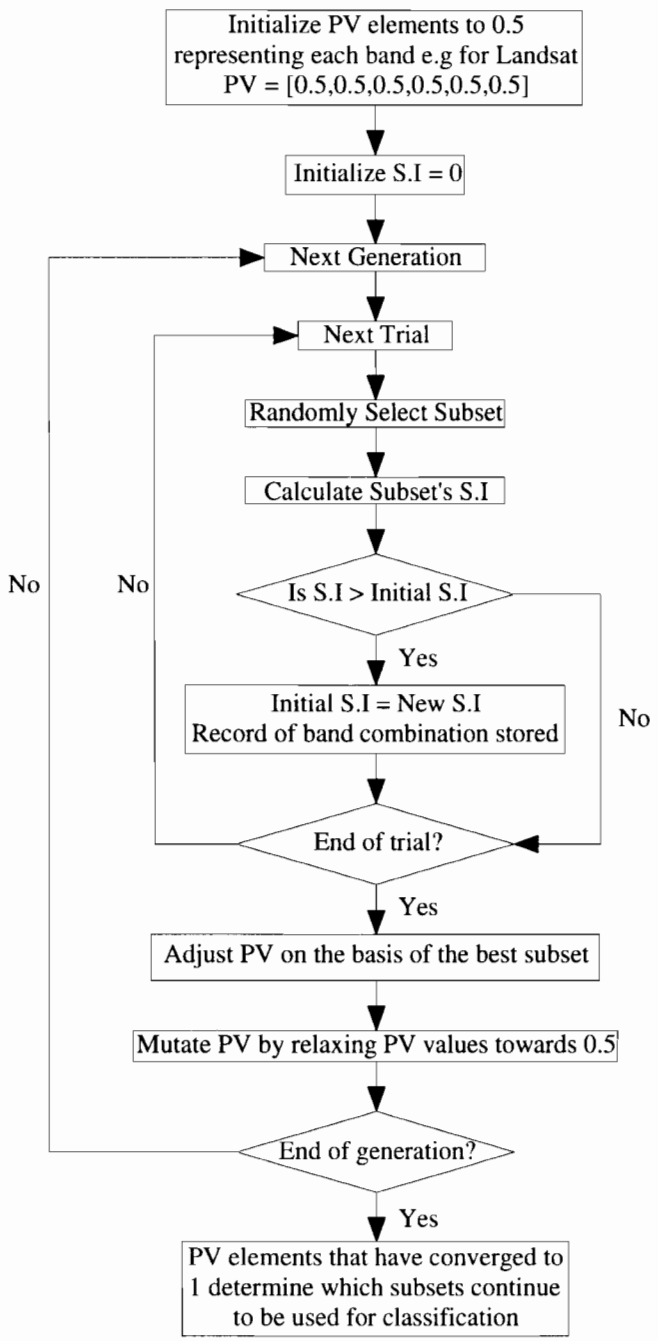


Figure 5.4: PBIL methodology
PV – *Probability Vector* SI – *Separability Index*

The resultant subset of bands identified through PBIL was then used to derive optimum SVM parameters as described in section 5.3, which were subsequently used to classify the study area. From Figure 5.2, the training and testing matrix dimensions depended on the number of bands identified. For each study area and each land cover type, the resultant class image was exported as a bitmap for importation into IDRISI (See section 5.3).

5.4.3 Instance Reduction

This method was used to reduce the number of instances i.e. the number of data points needed to derive optimum SVM parameters. The flow chart in Figure 5.5 conceptualizes the algorithm developed to effect data reduction as illustrated in Figure 3.3.

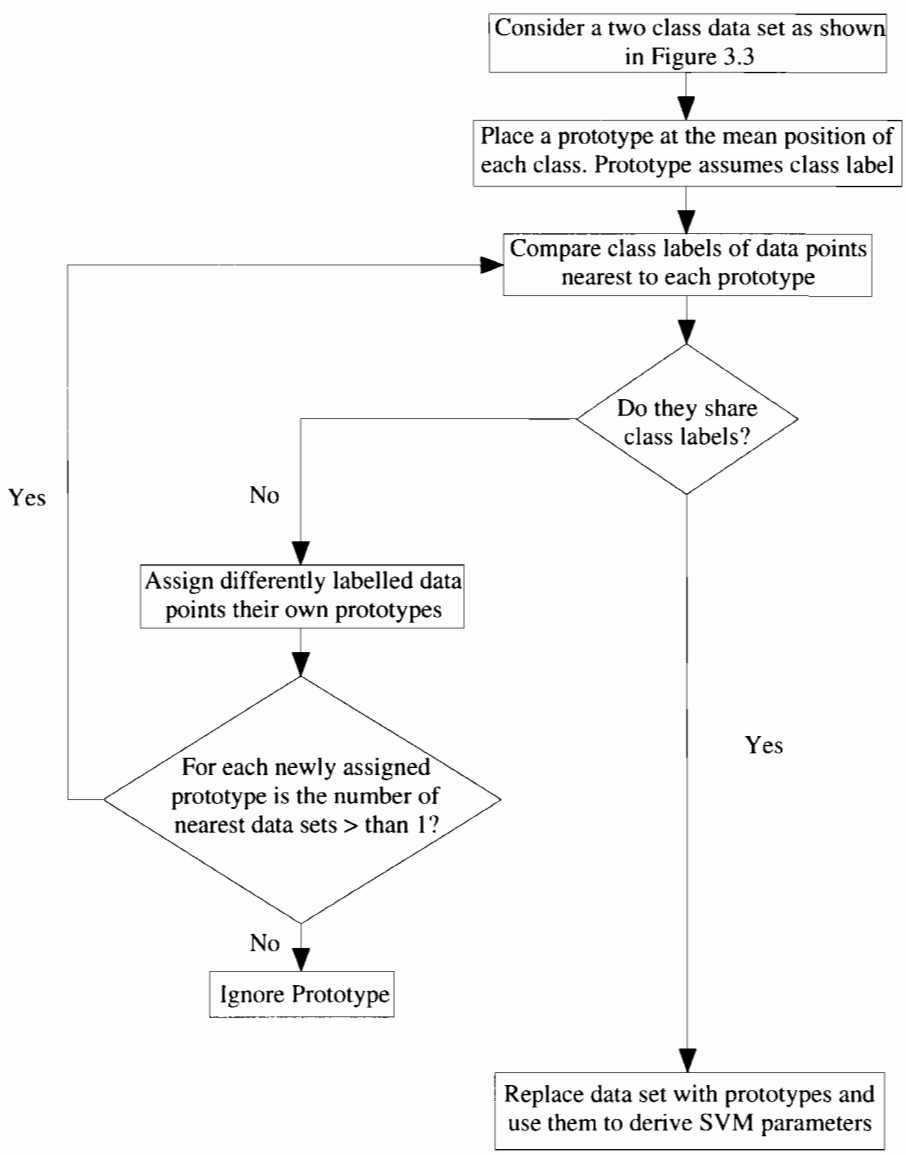


Figure 5.5: Instance reduction method

For each study area and for each land cover type, instance reduction was employed and the resultant prototypes were used to derive SVM parameters corresponding to each land cover type. From Figure 5.3, the training and testing matrix dimensions was nine dimensional (for Landsat it was six dimensional). Attempts were also made to apply instance reduction to the results of PCA transformation, in which case the training and testing matrix dimensions was three dimensional. For each study area and each land cover type, the resultant class image was exported as a bitmap for importation into IDRISI (See section 5.3).

5.5 Accuracy Assessment

As alluded to in section 5.3, accuracy assessment was carried out for each derived classification map. The aim of accuracy assessment was to obtain a measure of comparison between the derived classification maps before and after SVM optimization. It is based on these comparisons that conclusions could be drawn about the effectiveness and efficiency of the investigated optimization techniques. Accuracy assessment was carried out using IDRISI software and involved both visual inspection and quantitative assessment. For all study areas, visual inspection was done to get an overall impression of the derived land cover maps with particular emphasis on the number of classes successfully classified, wrongly classified classes and areas of overlap.

Quantitative assessment involved constructing an error matrix from the comparison between the derived land cover map and reference data. For the East Rand study area, reference data was extracted from 1:50 000 topographic maps, 1:10 000 orthophotos and previous classification done by Chevrel et al (2003) using IKONOS, Landsat and ASTER imagery of the study area. For the Mara basin and Malmesbury study area, reference data was obtained from field observations. Cluster sampling instead of random sampling of reference data was selected because of its inherent advantages discussed in section 4.3.2.2. During the sampling, care was taken to ensure that the samples used for accuracy assessment were different from those used for classification. Caution was also taken to ensure that instead of big contiguous clusters, the samples were small in size (of about ten (10) pixels).

For each classified image, an error matrix was generated from which single summary measures and class level measures were extracted. The single summary measures extracted included the overall accuracy and the KHAT statistic. Class level accuracy measures extracted included producer's accuracy, user's accuracy and conditional Kappa coefficient. For each study area, ranking of classification accuracies was used to determine the relative performance of all classification algorithms and optimization techniques. The ranking in this case was based on the KHAT statistic. For each classifier (i.e. linear, polynomial and RBF SVM classier), comparisons were made before and after the application of the aforementioned optimization techniques based on binomial tests of significance.

Chapter 6

Results & Analysis

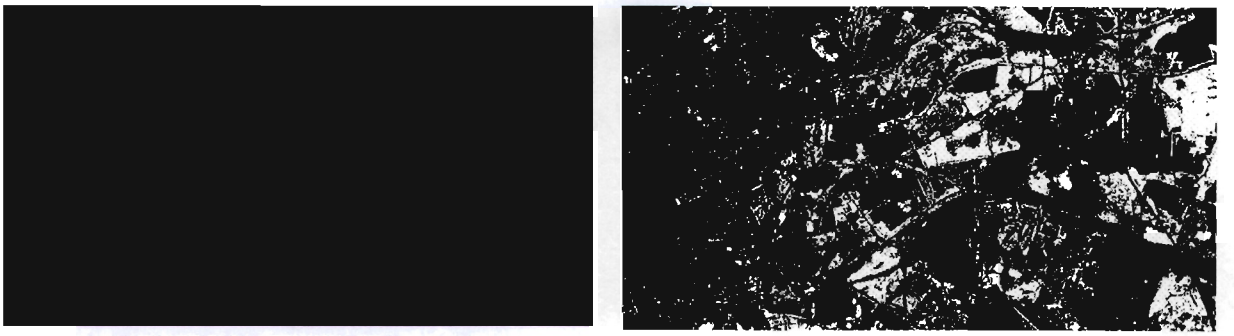
6.1 Introduction

In this chapter the results of the effect of various optimization techniques on SVM classification are presented for each of the study areas namely; East Rand, Mara Basin and Malmesbury. The chapter gives an overview of issues surrounding the derivation of optimum SVM parameters, data reduction through Principal Component analysis (PCA), feature selection and instance reduction. The results are presented in form of land cover maps from which statistical measures were extracted and comparisons made to give an indication of how optimization impacts on SVM classification.

6.2 Derivation of Optimum SVM Parameters

One of the key observations of this research regarding the application of SVMs to land cover mapping was the importance of deriving optimum parameters for each classifier. Using parameters less than or greater than optimum could potentially affect the classification results adversely. Figures 6.1 and 6.2 are extracts from the East Rand study area and depict both these mentioned scenarios. The class images were both derived from the classification of the PCA bands using a polynomial SVM classifier.

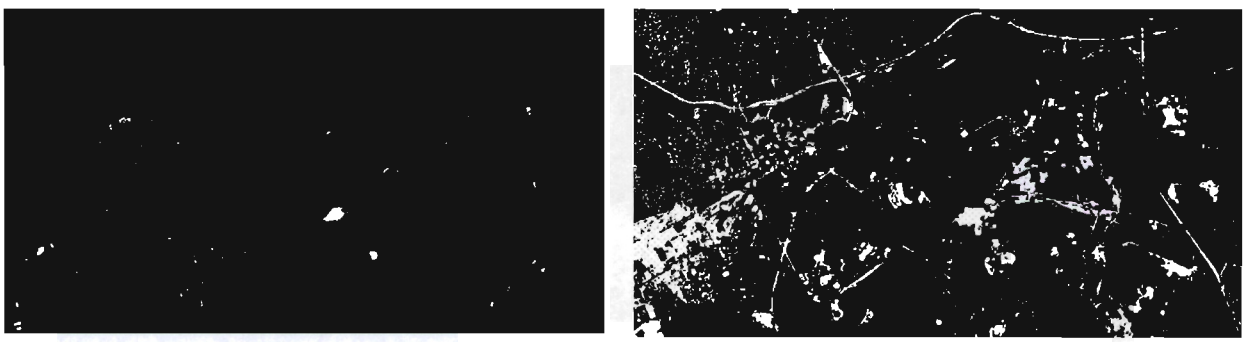
Figure 6.1 depicts the derivation of a soil class image. From cross validation, the optimum polynomial degree for this class was found to be 4 and the corresponding class image is shown in Figure 6.1b. Figure 6.1a shows the contrast resulting from using less than the optimum polynomial degree, in this case 3. Clearly using less than optimal SVM parameters may result in "under" classification of the class in interest. "Under" classification in this context refers to assigning fewer pixels than expected to a known land cover class. By contrast, "over" classification would refer to assigning more pixels than expected to a known land cover class.



a) Polynomial Degree 3, C = 100 b) Polynomial Degree 4, C = 100

Figure 6.1: Under-classification resulting from the use of less than optimum SVM parameters

Figure 6.2 on the other hand depicts the derivation of a yellow tailings class image. From cross validation, the optimum polynomial degree for this class was found to be 2 and the corresponding class image is shown in Figure 6.2a. Figure 6.2b shows the contrast resulting from using a polynomial degree greater than the derived optimum polynomial degree. Figure 6.2b illustrates a classic example of "over" classification stemming from the use of a polynomial degree over and above the optimum SVM parameter.



a) Polynomial Degree 2, C = 100 b) Polynomial Degree 10, C = 100

Figure 6.2: Over-classification resulting from the use of more than optimum SVM parameters

In view of the previous results, it must be mentioned, however, that different classes had different sensitivities to the different SVM parameters. Since this sensitivity wasn't detectable *a priori*, it was always prudent to work with optimum parameters at all times. As illustrated in these two examples, the accuracy with which an SVM may classify a dataset is dependent on the identification of the corresponding optimum SVM parameters. In any supervised classification task, the aim is to derive classifier parameters from training data which will enable the classifier to generalize well on unseen data. Classifier parameters which lead to over-fitting of the training data invariably generalize unseen data poorly, hence should be avoided. To this effect, when selecting SVM parameters for the various classifiers, in the case that cross

validation yielded numerous parameters with the same accuracy, preference was given to larger kernel widths and lower polynomial degrees (for the RBF and polynomial SVMs respectively) which in effect reduced the occurrence of over fitting.

Of the three SVM classifiers experimented on in this research, linear SVMs were the easiest and fastest to optimize and use owing to the fact that only one parameter C had to be sought. The search space for the polynomial and RBF SVMs consisted of two variables, hence considerably more time was needed to derive their optimum parameters.

6.3 Results of Data Reduction

In this section a summary of the results of feature and instance reduction are presented for each of the study areas. The feature reduction methods explored, as explained in Chapters Three and Five, were Principal Component Analysis (PCA), exhaustive search method and Population Based Incremental Learning (PBIL). The center-splitting method was used to effect instance reduction.

To each of the band sets of each study area, PCA transformation was applied. Figure 6.3 below gives an illustration of a color composite of the first three components of the PCA transformation of the East Rand study area. In addition to reducing the number of classification bands from nine to three, the PCA transformation enabled the visualization of the transformation results. In the case of the Mara Basin and Malmesbury dataset, PCA transformation resulted in reducing the original six bands to three which were subsequently used for classification.

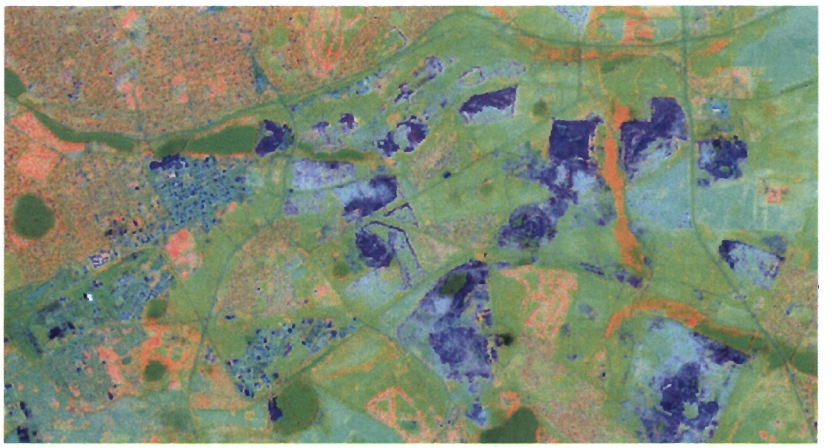


Figure 6.3: RGB color composite of the PCA transformation of the nine bands

For each study area, feature selection was carried out by employing the exhaustive search method and PBIL and the results of the bands selected for each class are hereby presented. Tables 6.1 – 6.3 list the bands that were selected using both the exhaustive search method and PBIL. As mentioned in Chapter Five, the one against all approach was adopted hence for each class, different bands had to be identified. The first column of each table lists the classes of interest at each study area. The second column presents the bands identified following the exhaustive search while the third column highlights the bands selected after applying PBIL to the original band sets.

Table 6.1: Exhaustive search of the East Rand dataset

Class	Bands Identified using Exhaustive Search	Bands Identified Using PBIL
Agriculture	2nd and 3rd	2nd, 3rd, 6th, 7th, 8th and 9th
Soil and Trees	3rd and 8th	3rd, 6th, 7th and 8th
Tailings	2nd	1st and 9th
Tailings Dam	1st	2nd, 4th, 7th and 8th
Water	4th	3rd, 4th, 5th, 6th, 7th, 8th and 9th
Wetlands	6th	2nd, 4th, 7th and 8th
Yellow Tailings	4th	1st,5th, 6th,7th and 9th
Trees	4th, 7th and 9th	1st, 2nd, 3rd, 4th, 5th and 6th
Roads	1st, 2nd, 3rd and 7th	2nd, 3rd, 6th, 7th, 8th and 9th
Soils	4th, 7th and 9th	1st, 2nd, 3rd and 7th

Table 6.2: Exhaustive search of the Mara Basin dataset

Class	Bands Identified using Exhaustive Search	Bands Identified Using PBIL
Wetlands	1st, 2nd and 6th	3rd, 4th,5th and 7th
Water	5th	1st and 7th
Bush/Shrub/Trees	3rd, 4th and 5th	1st,3rd, 4th, 5th and 7th
Grasslands	3rd and 4th	3rd, 4th and 5th
Bare ground	3rd and 5th	3rd, 4th and 5th
Roads	2nd and 5th	3rd, 4th and 5th

Table 6.3: Exhaustive search of the Malmesbury dataset

Class	Bands Identified using Exhaustive Search	Bands Identified Using PBIL
Built up areas	2nd, 4th, 5th and 6th	1st, 3rd, 4th, 5th and 6th
Trees	3rd, 4th, 5th and 6th	3rd, 4th, 5th and 6th
Fields	1st, 3rd, 4th and 5th	1st, 3rd, 4th and 5th

Looking through Tables 6.1 – 6.3, it is evident that the exhaustive search method was more efficient at feature selection since it identified the least number of features i.e. bands, as compared to PBIL. The only exception to this observation is in the two classes (i.e. trees and fields) listed in Table 6.3. The main shortcoming of the

exhaustive search method was that it was highly computer intensive hence consuming a lot of processing time, thus losing out on the anticipated advantages of feature selection. It was in this respect that PBIL had an edge over the exhaustive search method. Of particular benefit to PBIL was the fact that the search process converged within less than five generations. Hence the process of selecting the appropriate band combinations was much faster. As shown in Table 6.3, PBIL was even able to match the exhaustive search method in terms of number of identified bands. The one disadvantage of PBIL was that the each search process resulted in different band combinations being selected. By implication therefore, chances of replicating the selection of a given band combination were minimal.

Figure 6.4 below illustrates the effect of instance reduction to the results of PCA transformation for the Malmesbury study area. The other study areas were similarly treated and the figure below is used as an illustration of how the method worked. The green instances represent fields, the red ones trees and the blue ones the built up areas. The left and right images respectively portray the training data scatter plots before and after application of the PCA transformation. Both scatter plots have been reoriented to enable the best possible visualization of all the classes.

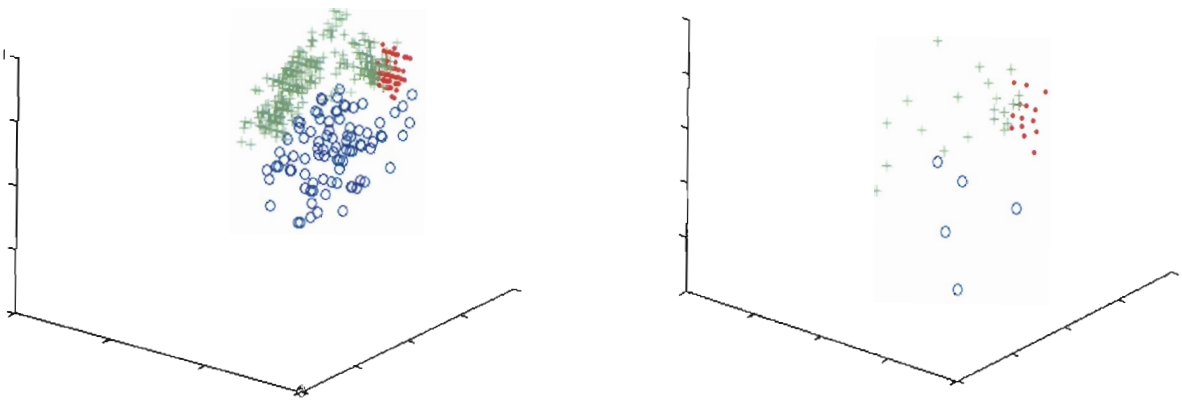


Figure 6.4: PCA scatter distribution before and after instance reduction

As explained in Chapter Three, the aim of instance reduction is to reduce the number of instances used to derive SVM parameters. From Figure 6.4, this method was successfully applied. Results from this research showed that instance reduction enabled much faster optimization of the SVM parameters. This enabled the exploration of a larger search space and more reruns of the experiment. The SVM

parameters derived using the results of instance reduction were then used to classify the study area and the derived land cover maps are presented in section 6.4.

6.4 Derivation and analysis of Land Cover Maps

6.4.1 Derived Land Cover Maps for East Rand Study area

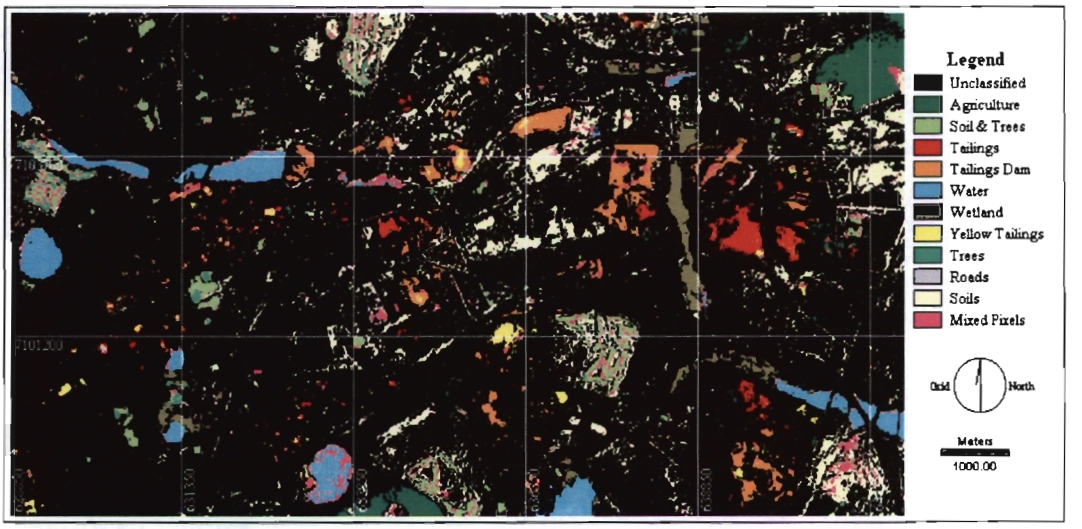
Figures 6.5 - 6.10 illustrate the classification results of various reduction techniques. Table 6.4 below will act as a legend to the derived maps. In each figure, three land cover maps represent classification due to linear, polynomial and RBF SVM classifiers respectively. Each Figure, apart from Figure 6.5, depicts the results of data reduction on the mentioned classifiers.

Figure 6.5 shows the SVM classification results before the application of any optimization technique. These derived maps in effect acted as the 'control' or 'reference' land cover maps to which all the other corresponding maps were compared. Table 6.4 shows the order in which the derived maps are presented, based on the reduction technique.

Table 6.4: Legend of derived maps

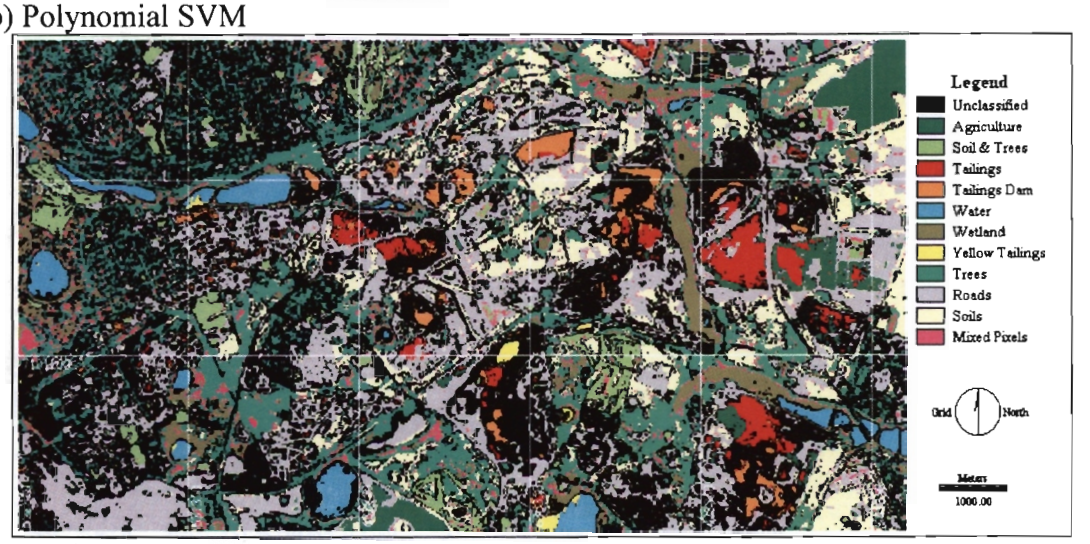
Figure	Reduction Method
Figure 6.5	None – classification of all nine bands
Figure 6.6	PCA
Figure 6.7	Exhaustive Search
Figure 6.8	PBIL
Figure 6.9	Instance reduction of all nine bands
Figure 6.10	Instance reduction of PCA

A detailed evaluation of these derived images is presented in Sections 6.4.2 and 6.4.3. The two sections respectively give a visual and quantitative assessment of these derived images.



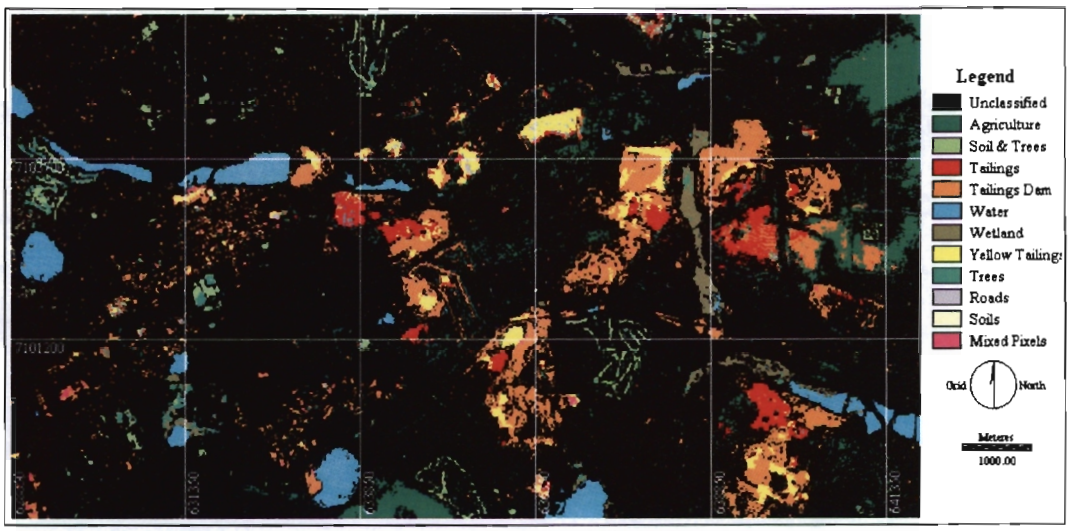
a) Linear SVM

b) Polynomial SVM

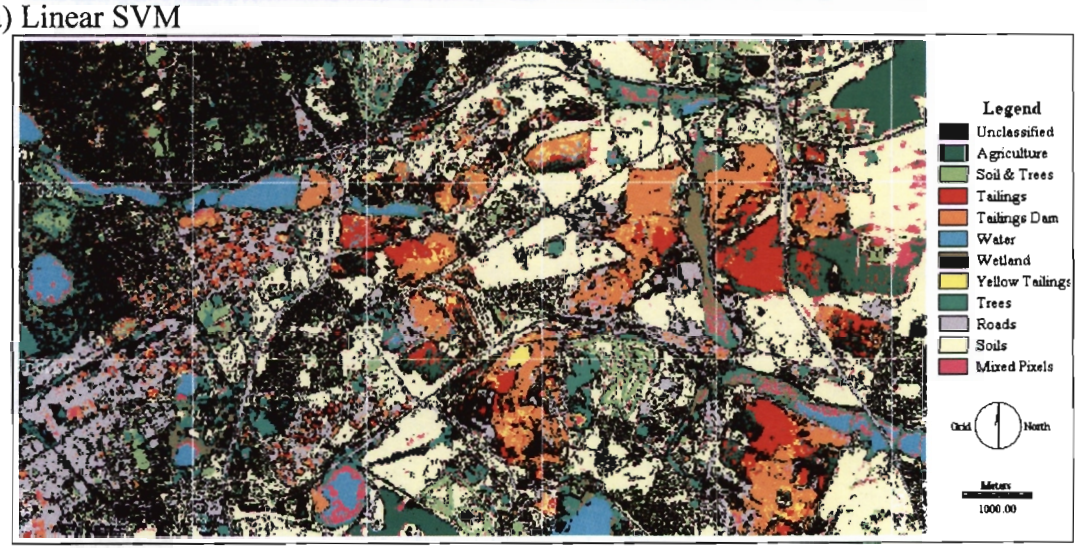


c) Radial Basis Function SVM

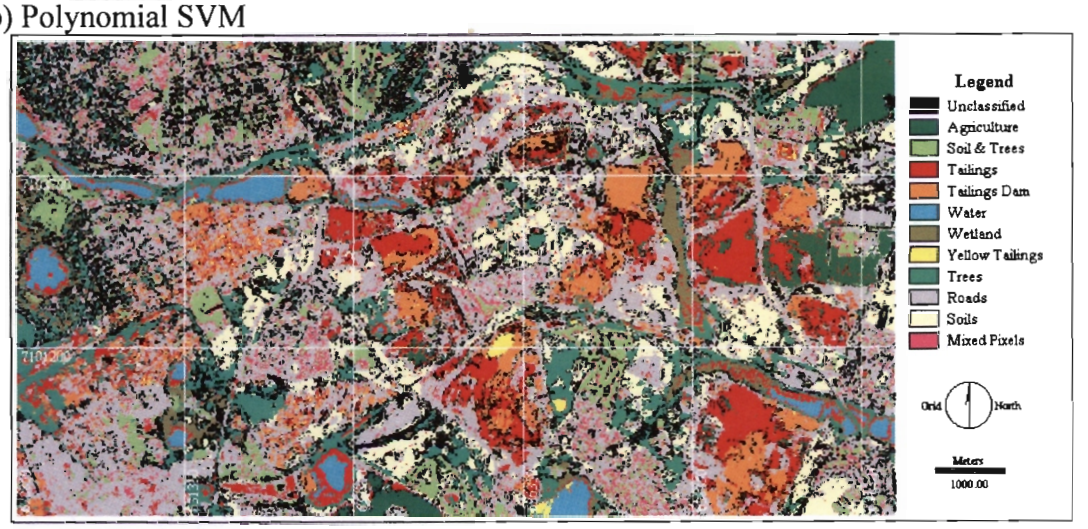
Figure 6.5: East Rand - Classification of all nine bands



a) Linear SVM

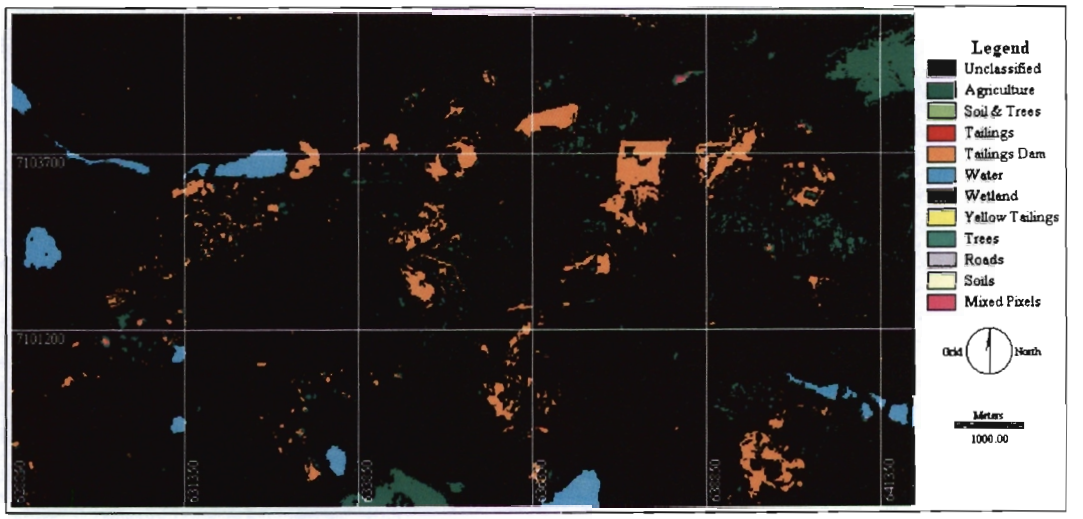


b) Polynomial SVM

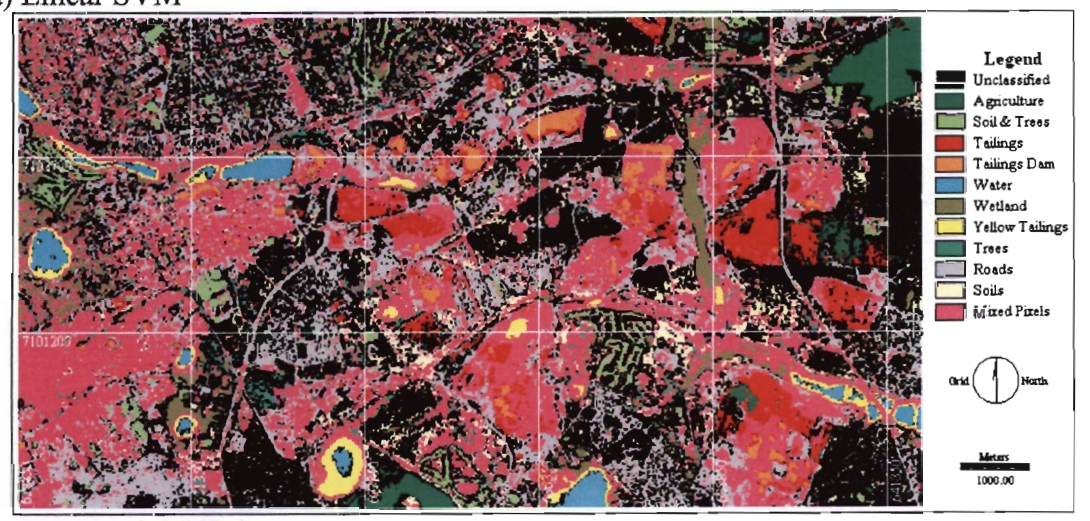


c) Radial Basis Function SVM

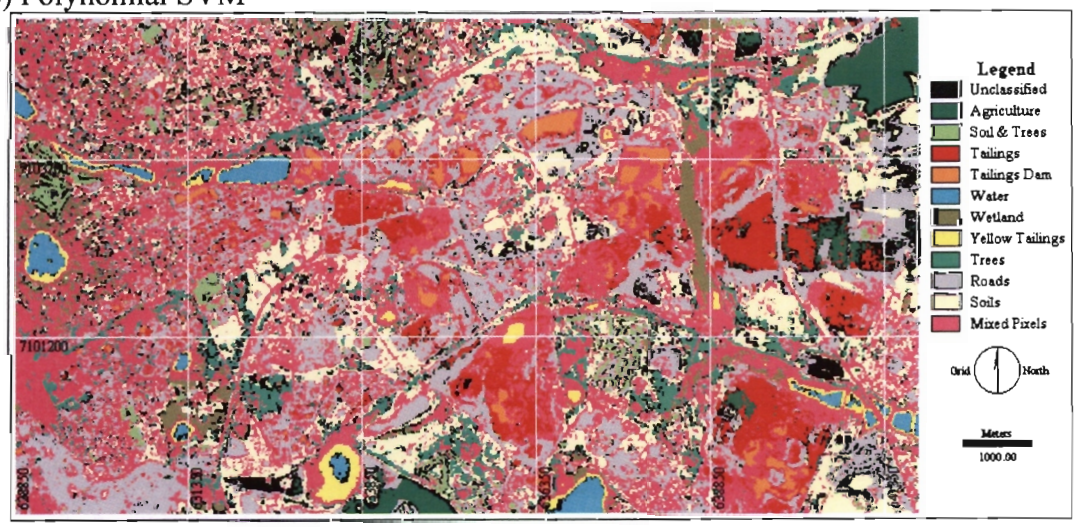
Figure 6.6: East Rand - Classification of PCA bands



a) Linear SVM

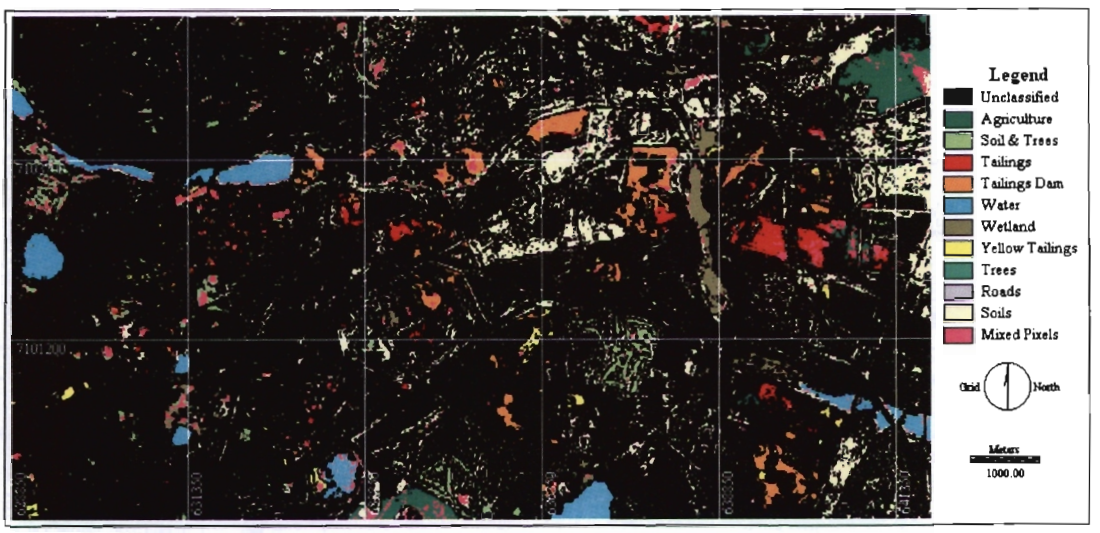


b) Polynomial SVM

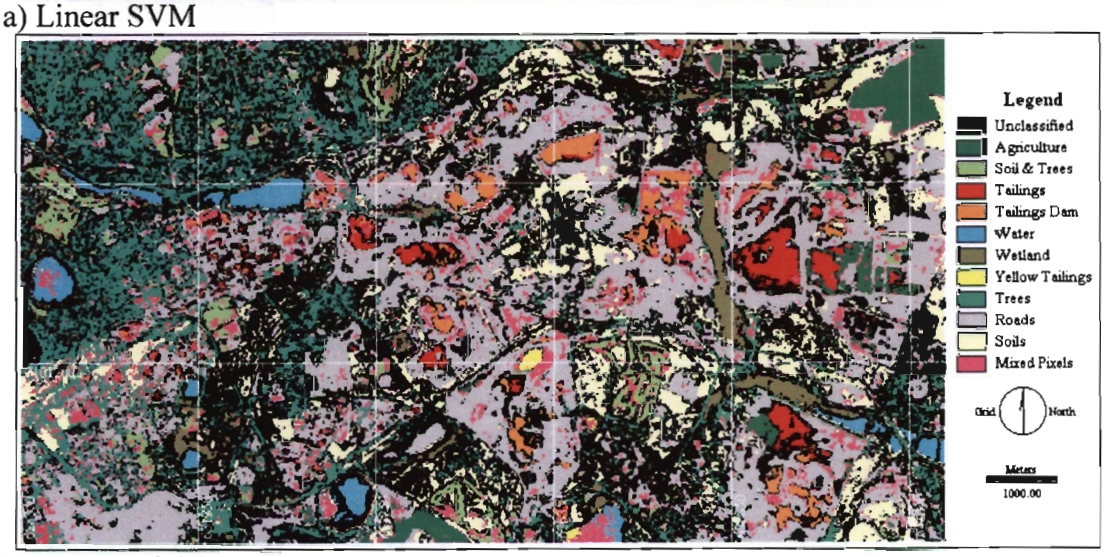


c) Radial Basis Function SVM

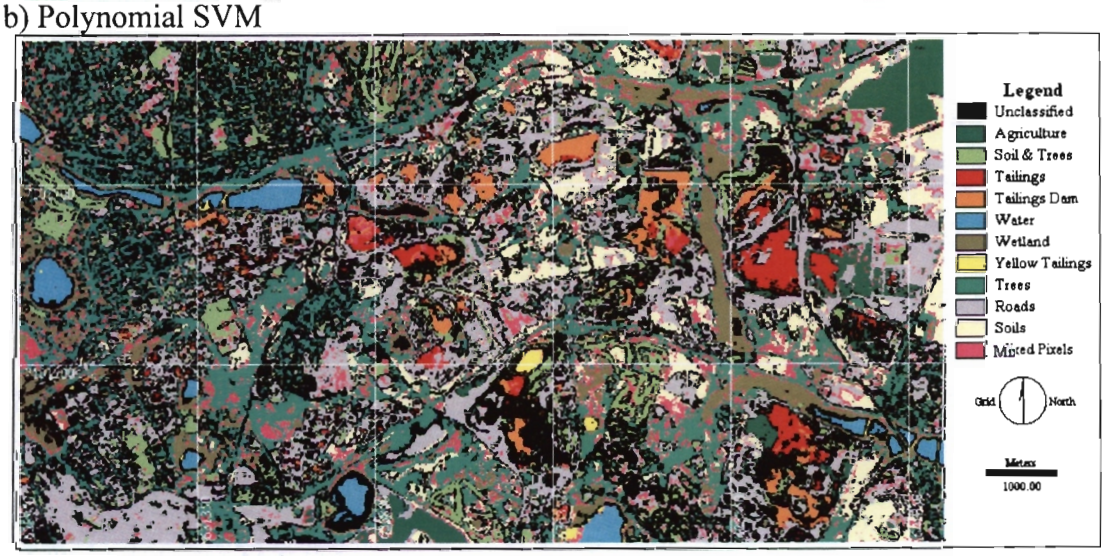
Figure 6.7: East Rand - Classification following exhaustive search



a) Linear SVM

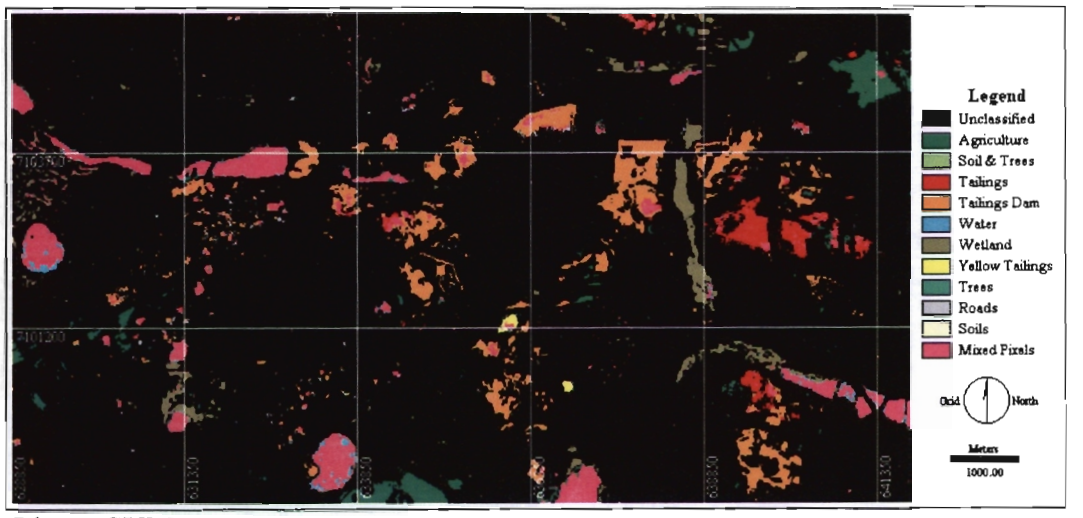


b) Polynomial SVM

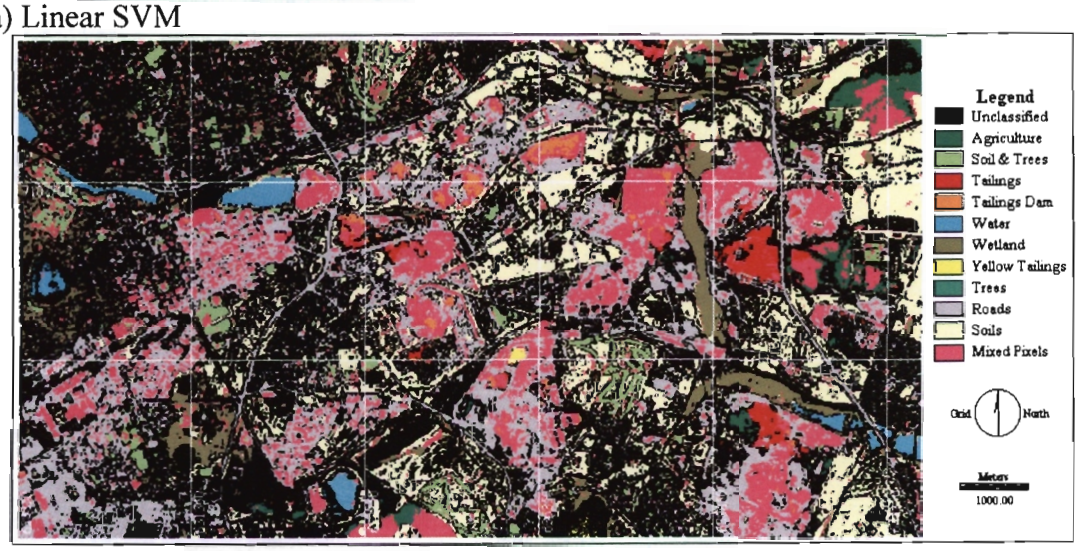


c) Radial Basis Function SVM

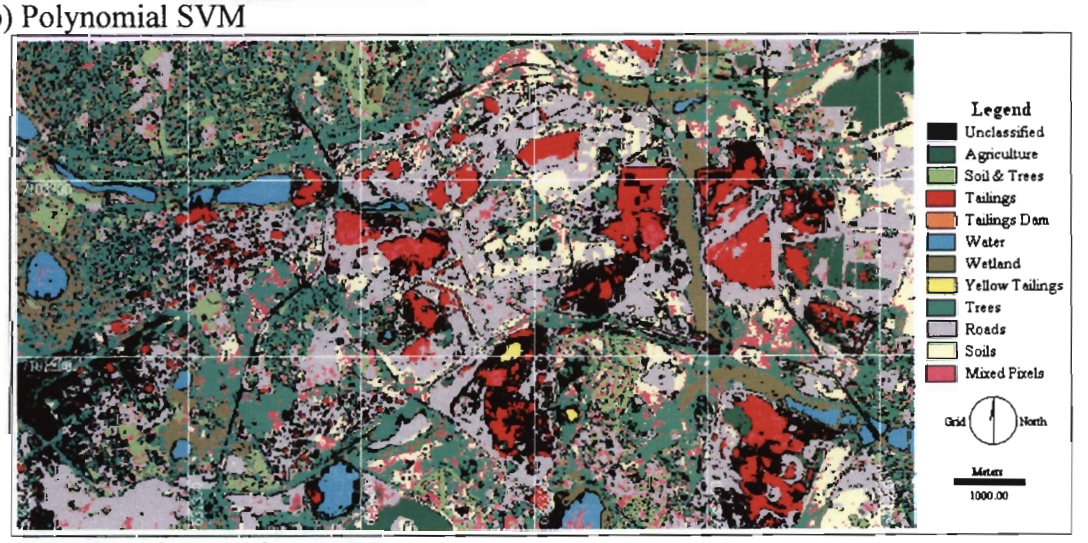
Figure 6.8: East Rand - Classification following PBIL



a) Linear SVM

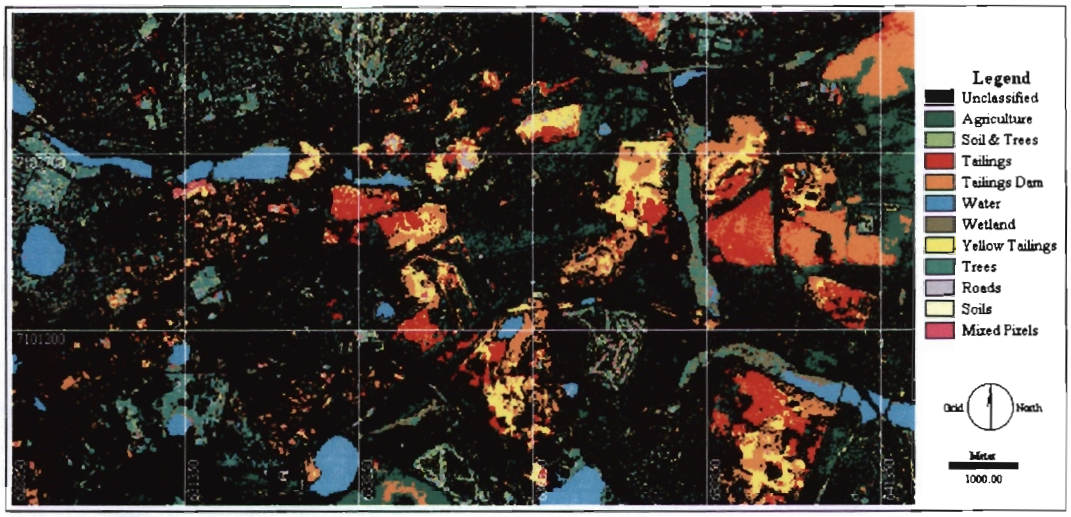


b) Polynomial SVM



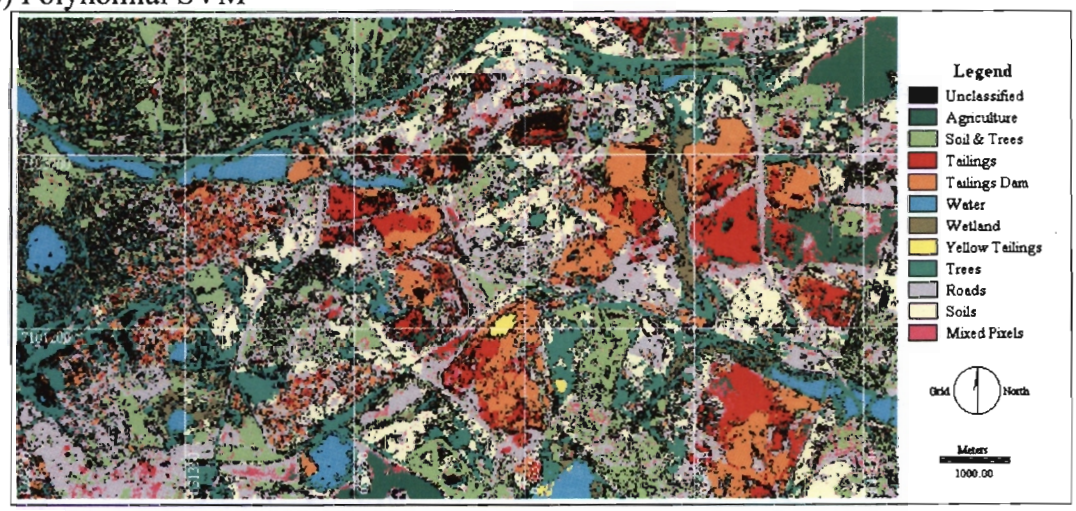
c) Radial Basis Function SVM

Figure 6.9: East Rand - Classification following instance reduction of all bands



a) Linear SVM

b) Polynomial SVM



c) Radial Basis Function SVM

Figure 6.10: East Rand - Classification following instance reduction of the PCA bands

6.4.2 Visual Analysis

From visual inspection of the derived maps in Figures 6.5 – 6.10, it is consistently apparent that the linear SVM has under performed. The land cover types were either not successfully classified e.g. roads and trees, poorly classified e.g. soils or misclassified. In most cases, water and wetlands were successfully classified. Evidently none of the data reduction techniques helped improve or maintain the visual appreciation of the linear SVM classification results.

In all cases, the polynomial and radial basis function SVMs were more effective than their linear counterparts, though they exhibited varying degrees of success. In Figure 6.5, all land cover types are where they are expected to be in comparison with the ground truth data. Figure 6.5b has as many mixed pixels as Figure 6.5c has unclassified pixels, however the main classes don't look too dissimilar.

Of all the derived maps, Figure 6.6b gives the best visual outlook. Most classes are clearly defined and it has few mixed pixels. The unclassified pixels in the upper right quarter may represent a class not defined as a training sample. Figure 6.6c is also a seemingly well-classified map. Figures 6.6b and 6.6c are evidence of the fact that PCA as a data reduction technique can improve the visual outlook of an SVM classification result.

The outcome of the SVM classification of exhaustive search results is exhibited in Figure 6.7. Clearly, reducing the number of bands to the bare minimum has not yielded visually appreciable results. This may have been due to "over" classification of the individual land cover classes. As a result, the integration of the corresponding class images into a single derived map yielded overlap hence the extensive number of mixed pixels.

The results of the PBIL search are exhibited in Figure 6.8 and can be seen to be much better than the results following Exhaustive Search in Figure 6.7. In the first case Figures 6.8b and 6.8c have fewer mixed pixels. Secondly they are very similar to the SVM classification of the original bands portrayed in Figures 6.5b and 6.5c.

Figure 6.9 shows the derived maps ensuing from the classification of instance reduction results of all the bands. Of the three Figures 6.9a, 6.9b and 6.9c, the latter shares closest resemblance to the corresponding original derived map in Figure 6.5c. Though Figure 6.9b is riddled with several mixed pixels, many of the land cover types are visually distinguishable.

Figure 6.10c looks a lot similar to Figure 6.6c, which is clearly an improvement on Figure 6.5c. Whereas Figure 6.9 resulted from the classification of instance reduction of all bands, Figure 6.10 demonstrates the classification of instance reduction of the PCA bands.

In assessing the studied data reduction techniques, from a visual perspective the exhaustive search method has resulted in dismal classification performance.

6.4.3 Quantitative Assessment

Further assessment was performed on the derived maps, the results of which are listed in Tables 6.5, 6.6 and 6.7. In these tables are statistical extracts following the comparison of the derived maps with ground truth data.

Appendix 1 details all the error matrices for each derived map from which the statistical results in Tables 6.5, 6.6 and 6.7 were derived. Table 6.5 shows a summary of the overall and class-by-class accuracy measures. Table 6.6 shows the producer's accuracy assessment, which measures the proportion of pixels in a test dataset that are correctly labeled/categorized by the classifier. In Table 6.7 the user's accuracy is shown and it refers to the reliability of the derived map to map a given land cover type. The three tables formed the basis of the following quantitative assessment.

Table 6.5: Summary of overall and class-by-class KHAT statistic

Kappa Index of Agreement		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
Cl	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Agriculture	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8896	1.0000	1.0000	1.0000	0.1509	1.0000	0.8889	1.0000	1.0000
2	Soil & Trees	0.5219	0.7990	0.9839	0.4579	0.5745	0.9050	0.0000	0.8590	0.8138	0.4082	0.7990	0.9839	0.0099	0.6148	0.8438	0.1124	0.1877	0.7120
3	Tailings	0.7397	0.8359	1.0000	0.8036	0.9668	0.9668	0.0000	1.0000	1.0000	0.5218	0.9668	0.9668	0.8036	0.8683	1.0000	0.8036	0.2254	0.9668
4	Tailings Dam	0.6871	0.4859	0.5360	0.4830	1.0000	1.0000	0.8427	0.7908	0.7908	0.7908	0.7393	0.7908	0.6370	0.3865	0.0000	0.1727	0.5828	0.5864
5	Water	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9036	0.9036	1.0000	1.0000	1.0000	0.0000	0.8718	1.0000	1.0000	0.7466	1.0000
6	Wetland	0.9205	0.7994	0.9590	0.7233	0.8807	1.0000	0.0000	0.8400	0.4513	0.3618	1.0000	0.9581	0.9141	1.0000	1.0000	0.0676	0.3145	0.7621
7	Yellow Tailings	1.0000	0.7437	0.6592	0.5629	1.0000	1.0000	0.0000	0.4916	1.0000	0.4916	0.6592	1.0000	0.8286	0.4916	1.0000	-0.0430	0.0000	1.0000
8	Trees	-0.0866	0.1824	0.1600	-0.1269	0.7083	0.8556	0.0000	0.0760	0.0805	0.0000	-0.0551	-0.0468	0.0000	0.0000	0.0034	-0.2381	0.8250	0.8892
9	Roads	0.0000	0.2000	0.1285	0.0000	0.9813	0.4799	0.0000	0.4434	0.0614	0.0000	0.2190	0.0482	0.0000	0.7839	0.5115	-0.0282	0.1632	0.4024
10	Soils	0.0821	0.1015	0.1847	0.0000	0.9736	0.6221	0.0000	0.1251	0.4324	0.1367	0.4118	0.1631	0.0000	0.6561	0.1607	0.0000	0.5505	0.5269
Over Kappa		**0.4618**	**0.5599**	**0.6136**	**0.3974**	**0.8663**	**0.8263**	**0.1805**	**0.6233**	**0.5730**	**0.3699**	**0.6193**	**0.5989**	**0.2435**	**0.6206**	**0.6249**	**0.2160**	**0.4002**	**0.7216**
Overall Accuracy		**0.5000**	**0.5989**	**0.6511**	**0.4511**	**0.8819**	**0.8462**	**0.1923**	**0.6566**	**0.6099**	**0.4011**	**0.6566**	**0.6374**	**0.2610**	**0.6566**	**0.6648**	**0.2637**	**0.4313**	**0.7500**
O.A (%)		**50.00**	**59.89**	**65.11**	**45.11**	**88.19**	**84.62**	**19.23**	**65.66**	**60.99**	**40.11**	**65.66**	**63.74**	**26.10**	**65.66**	**66.48**	**26.37**	**43.13**	**75.00**

Table 6.6: Producer's accuracy summary

Producer Accuracy		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
Cl	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Agriculture	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	89.47	100.00	100.00	100.00	15.99	100.00	89.47	100.00	100.00
2	Soil & Trees	58.23	83.54	98.73	51.90	63.29	92.41	0.00	88.61	84.81	46.84	83.54	98.73	1.27	67.09	87.34	13.92	22.78	75.95
3	Tailings	75.76	84.85	100.00	81.82	96.37	96.97	0.00	100.00	100.00	54.55	96.97	96.97	81.82	87.88	100.00	81.82	24.24	96.97
4	Tailings Dam	70.00	50.00	55.00	50.00	100.00	100.00	85.00	80.00	80.00	80.00	75.00	80.00	65.00	40.00	0.00	20.00	60.00	60.00
5	Water	100.00	100.00	100.00	100.00	100.00	100.00	100.00	91.18	91.18	100.00	100.00	100.00	0.00	88.24	100.00	100.00	76.47	100.00
6	Wetland	92.59	81.48	96.30	74.07	88.89	100.00	0.00	85.19	48.15	40.74	100.00	96.30	92.59	100.00	100.00	11.11	33.33	77.78
7	Yellow Tailings	100.00	75.00	66.67	58.33	100.00	100.00	0.00	50.00	100.00	50.00	66.67	100.00	83.33	50.00	100.00	0.00	0.00	100.00
8	Trees	0.00	28.13	34.78	0.00	75.00	87.50	0.00	9.38	15.63	0.00	0.00	6.25	0.00	0.00	9.38	0.00	84.38	90.63
9	Roads	0.00	24.62	16.92	0.00	98.46	53.85	0.00	49.23	9.23	0.00	27.69	7.69	0.00	81.54	56.92	0.00	20.00	46.15
10	Soils	16.28	11.63	20.93	0.00	97.67	65.12	0.00	13.95	46.51	16.28	46.51	18.60	0.00	69.77	18.60	0.00	58.14	55.81

Table 6.7: User's accuracy summary

User Accuracy		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
Cl	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Agriculture	100.00	100.00	100.00	90.48	95.00	100.00	86.36	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	89.47	100.00	100.00
2	Soil & Trees	100.00	100.00	100.00	100.00	100.00	100.00	-	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
3	Tailings	100.00	100.00	100.00	100.00	100.00	100.00	-	100.00	100.00	100.00	100.00	100.00	100.00	100.00	66.00	100.00	100.00	100.00
4	Tailings Dam	93.33	100.00	100.00	83.33	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	-	33.33	80.00	100.00
5	Water	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	-	100.00	91.89	70.83	100.00	100.00
6	Wetland	100.00	78.57	100.00	86.96	96.00	87.10	-	85.19	65.00	42.31	72.97	61.90	50.00	93.10	77.14	17.65	90.00	87.50
7	Yellow Tailings	92.31	100.00	100.00	41.18	100.00	100.00	-	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	0.00	-	100.00
8	Trees	0.00	20.45	20.51	0.00	46.15	57.14	-	42.86	16.67	-	0.00	5.26	-	-	9.09	0.00	69.23	51.79
9	Roads	-	76.19	64.71	-	100.00	85.37	-	100.00	50.00	-	66.67	45.45	-	100.00	86.05	0.00	81.25	83.33
10	Soils	21.88	83.33	81.82	-	97.67	100.00	-	100.00	95.24	63.64	60.61	80.00	-	68.18	72.73	-	100.00	100.00

To appreciate the relative performance of the classifiers and the data reduction techniques better, Table 6.8 was drawn to show the ranking of the results based on the classifier used and the derived maps' KHAT statistical measures.

Table 6.8: Ranking of classification results

Rank	Classifier (SVM)	Reduction Method	Overall Kappa
1	Polynomial	PCA	0.8663
2	Radial Basis Function	PCA	0.8263
3	Radial Basis Function	Instance reduction - PCA	0.7216
4	Radial Basis Function	Instance reduction – All bands	0.6249
5	Polynomial	Exhaustive search	0.6233
6	Polynomial	Instance reduction – All bands	0.6206
7	Polynomial	PBIL	0.6193
8	Radial Basis Function	None – All bands	0.6136
9	Radial Basis Function	PBIL	0.5989
10	Radial Basis Function	Exhaustive search	0.5730
11	Polynomial	None – All bands	0.5599
12	Linear	None – All bands	0.4618
13	Polynomial	Instance reduction - PCA	0.4002
14	Linear	PCA	0.3974
15	Linear	PBIL	0.3699
16	Linear	Instance reduction – All bands	0.2435
17	Linear	Instance reduction - PCA	0.2160
18	Linear	Exhaustive search	0.1805

One of the obvious observations of Tables 6.5 – 6.8 that tally with the visual inspection that preceded the quantitative analysis is that the linear SVM performed dismally. None of the reduction methods had accuracies close to or better than the results of classifying the whole dataset. Of the derived maps that were classified using the linear SVM, the map resulting from exhaustive search were the poorest – both visually and from quantitative assessment. Across the board, both polynomial and radial basis function SVMs did comparatively better than the linear SVMs. The best-ranked data reduction technique was the PCA, which also corresponded to the results of visual inspection. The only result that showed disparity between visual and statistical analysis was the polynomial SVM classification of the instance reduction of PCA. From the visual inspection, this derived map didn't look as bad as it is portrayed in the rankings.

Considering the class-by-class performance, the best land cover type classified across reduction techniques and SVM methods were agriculture and water. The only exception to this observation was the polynomial and linear SVM (respectively)

classification of instance reduction results of all the bands. These accuracy measures were less than acceptable.

The "soil and trees" class was consistently poorly classified where the linear SVM was applied, however there was a marked improvement when the polynomial and radial basis function SVMs were employed. In general, none of the reduction methods could better the results derived from classifying all the bands.

Tailings were generally well classified across classifiers and reduction methods. Again the only exception to this observation was with the outcome of classifying the results of instance reduction of the PCA bands. This derived map yielded poor agreement when compared with ground truth data.

Tailings dam were markedly better classified when PCA, PBIL and exhaustive search reduction techniques were employed in conjunction with polynomial and radial basis function SVMs. The derived map ensuing from the classification of the results of instance reduction were comparatively poorer.

In the original nine bands, wetlands were well delineated regardless of the SVM classifier method. The data reduction methods that posted similarly appreciable results regardless of the SVM used PCA and instance reduction of the all bands. PBIL's performance in classifying wetlands was also commendable when the polynomial and radial basis function SVMs were used. Exhaustive search was only able to classify wetlands well when the polynomial SVM was applied.

The linear SVM failed to post any good results during the classification of trees. Some of the reduction methods posted better accuracy measures as compared to the original bands. These reduction techniques included the PCA and instance reduction of PCA bands (when employing both polynomial and radial basis function SVM). The exhaustive search reduction method was only able to classify trees well when the polynomial SVM was used.

Roads were best classified when PCA and instance reduction of all bands were classified by polynomial SVMs. The rest of the classifiers and reduction methods performed poorly.

The land cover type – soils was poorly classified by all linear SVMs. The best classification results for soils were obtained when the polynomial SVM classified the PCA bands. The general trend was that data reduction helped improve on the classification of soils.

In an attempt to determine the significance of the results of data reduction in comparison to the original datasets, a binomial test of significance was carried out at the 95% confidence level (critical Z value = 1.96). Table 6.9 below summarizes the results of the test of significance. The test of significance was calculated using the overall accuracies as extracted from the error matrices in Appendix 1 and are summarized in Table 6.9 below. The comparisons were made for each classifier i.e. linear, polynomial and radial basis function classifiers, with a view of determining the effectiveness of data reduction on land cover classification.

Table 6.9: Calculated Z values for comparison between results of classifying all bands and the data reduction methods

	PCA	**EXHAUSTIVE**	**PBIL**	**CS_ALL**	**CS_PCA**
Linear	1.79	-9.22	-2.70	-6.85	-6.77
Comments	*Insignificant*	*Significantly Worse*	*Significantly Worse*	*Significantly Worse*	*Significantly Worse*
Polynomial	9.20	1.61	1.61	1.61	-4.59
Comments	*Significantly Better*	*Insignificant*	*Insignificant*	*Insignificant*	*Significantly Worse*
RBF	6.23	1.15	0.39	0.39	2.93
Comments	*Significantly Better*	*Insignificant*	*Insignificant*	*Insignificant*	*Significantly Better*

PCA – Principal component analysis
Exhaustive – Exhaustive search method
PBIL – Population Based Incremental search method
CS_ALL – Instance reduction of all the nine bands
CS_PCA – Instance reduction of the PCA bands

A cursory evaluation of Table 6.9 further confirms the limitation of the linear classification of the reduced datasets. Apart from the linear classification of the PCA dataset, the rest of the reduction methods have deposited significantly worse

classification accuracy results. The polynomial classification results on the other hand were significantly worse only when classifying the CS_PCA dataset. Exhaustive search, PBIL and CS_ALL posted insignificantly different classification accuracy results, while the polynomial classification of the PCA dataset yielded significantly better results. In general, the radial basis function classier fared best of the classifiers since none of the classification accuracies were significantly worse. PCA and CS_PCA datasets both resulted in significantly better results while the rest of the reduction methods were insignificantly different. In general the PCA dataset had the best performance. Exhaustive search method, PBIL and CS_ALL had generally similar significance evaluations and the CS_PCA only did significantly better when classified using a radial basis function classifier.

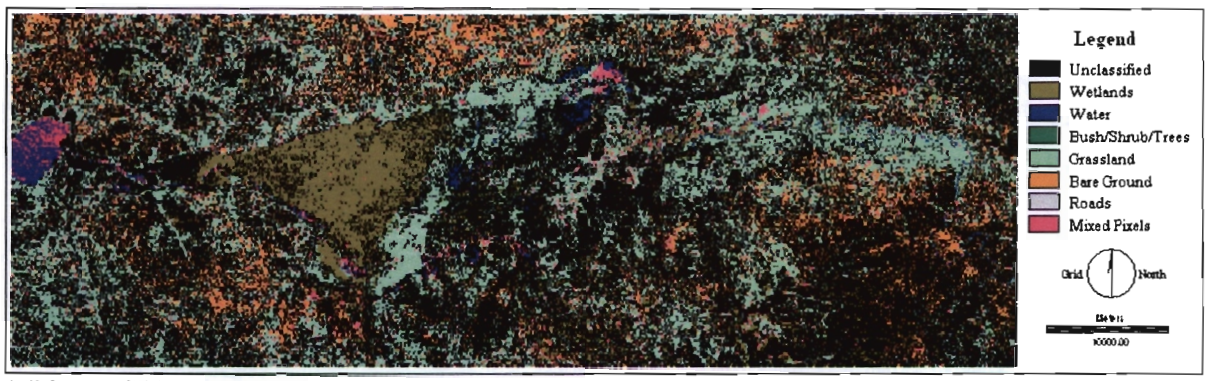
6.4.4 Derived Land Cover Maps for Mara Basin Study area

Figures 6.11 - 6.16 illustrate the classification results of various reduction techniques. Table 6.10 below will act as a legend to the derived maps. In each figure, three land cover maps are presented representing classification due to linear, polynomial and RBF SVM classifiers respectively. Each figure, apart from Figure 6.11, depicts the results of data reduction on the mentioned classifiers. Figure 6.11 shows the SVM classification results before the application of any reduction method. These derived maps in effect acted as the 'control' or 'reference' land cover maps to which all the other corresponding maps were compared. Table 6.10 shows the order in which the derived maps are presented, depending on the reduction technique.

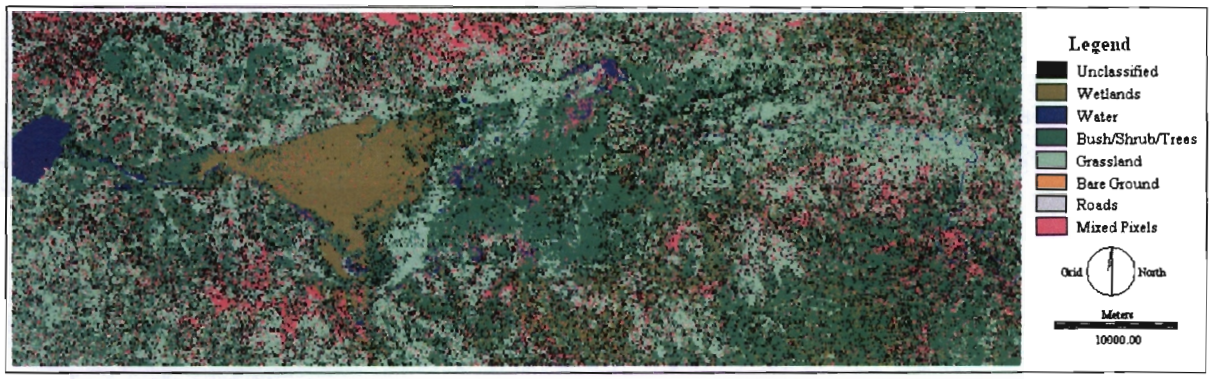
Table 6.10: Legend of derived maps

Figure	Reduction Method
Figure 6.11	None – classification of all nine bands
Figure 6.12	PCA
Figure 6.13	Exhaustive Search
Figure 6.14	PBIL
Figure 6.15	Instance reduction of all nine bands
Figure 6.16	Instance reduction of PCA

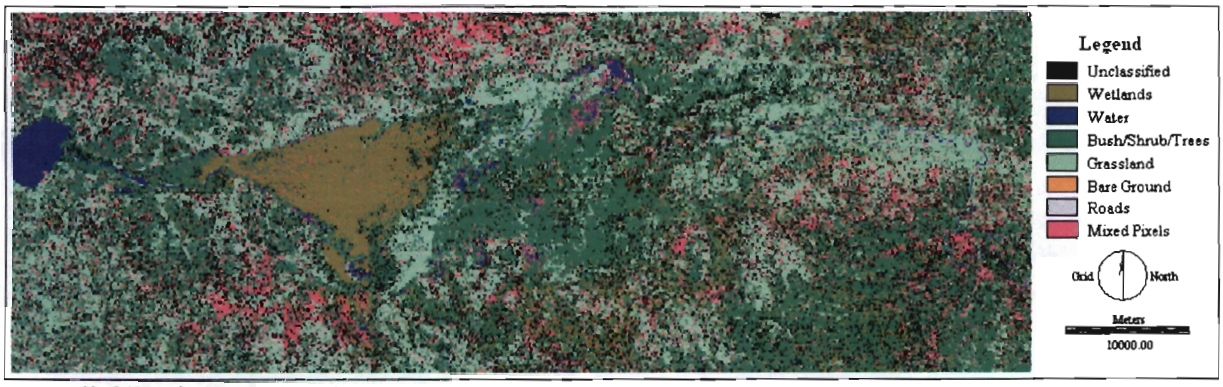
A detailed evaluation of these derived images is presented in Sections 6.4.5 and 6.4.6. The two sections respectively give a visual and quantitative assessment of these derived images.



a) Linear SVM

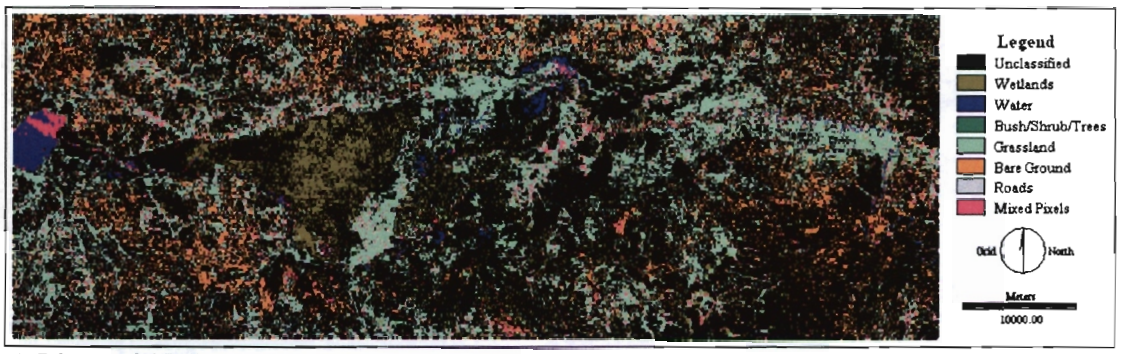


b) Polynomial SVM



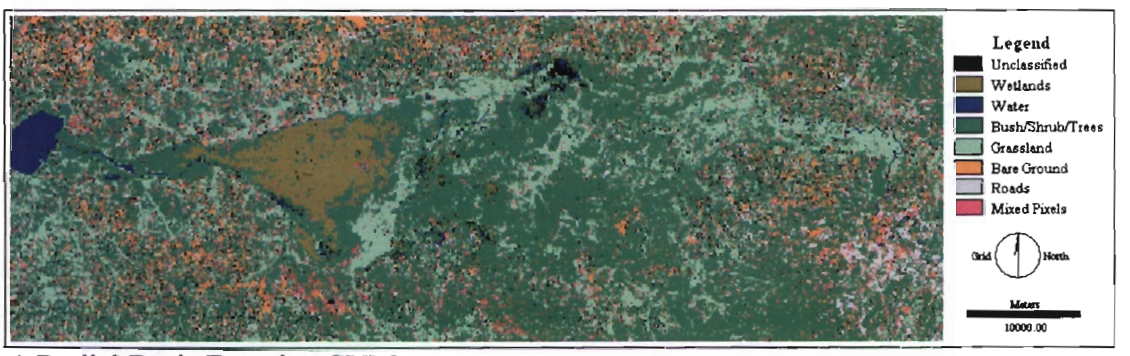
c) Radial Basis Function SVM

Figure 6.11: Mara Basin - Classification of all six bands



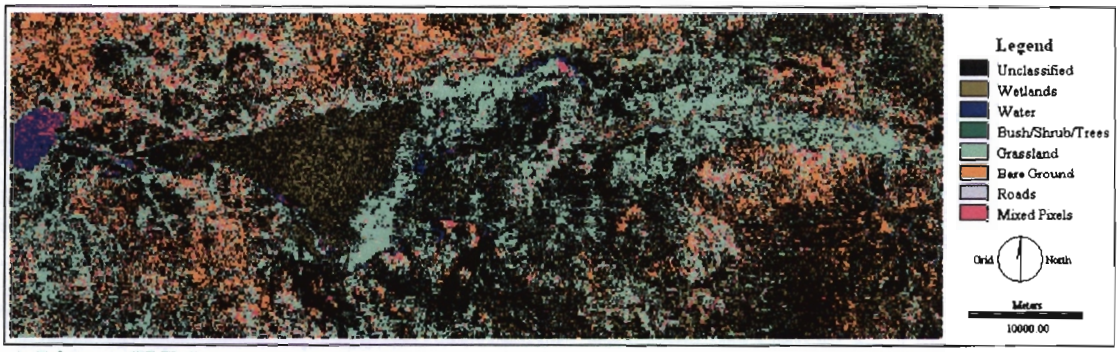
a) Linear SVM

b) Polynomial SVM

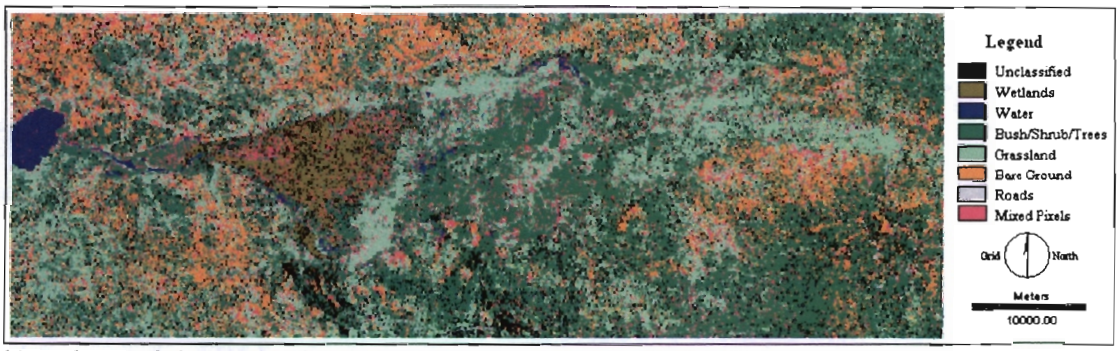


c) Radial Basis Function SVM

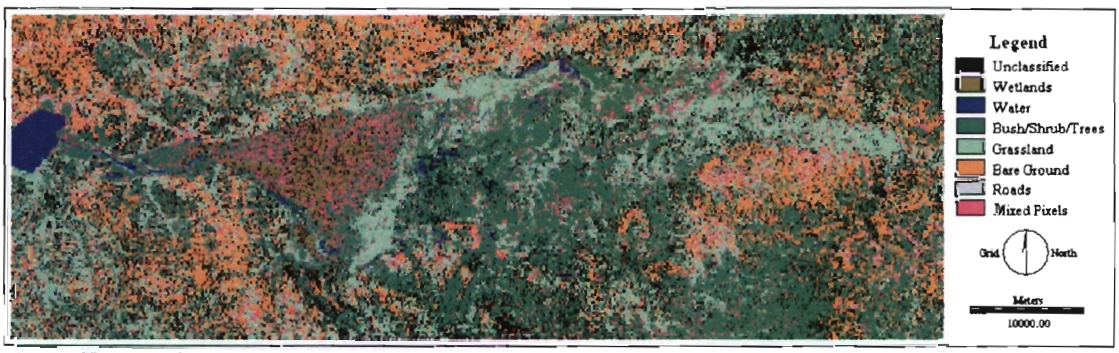
Figure 6.12: Mara Basin - Classification of PCA Bands



a) Linear SVM

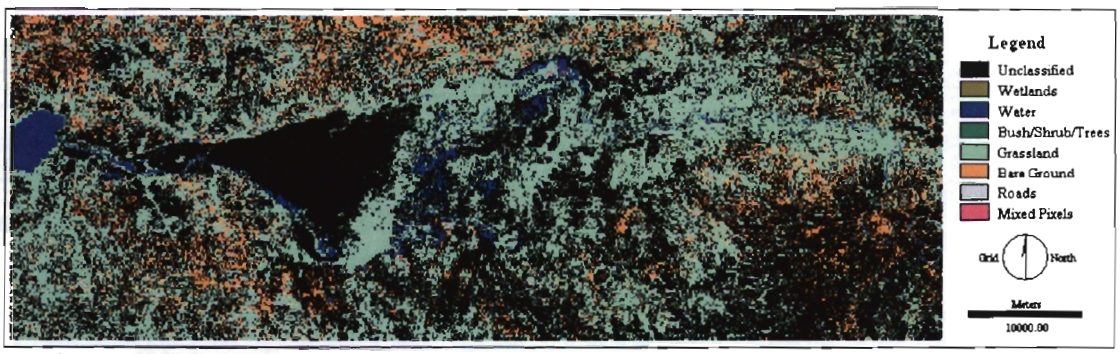


b) Polynomial SVM

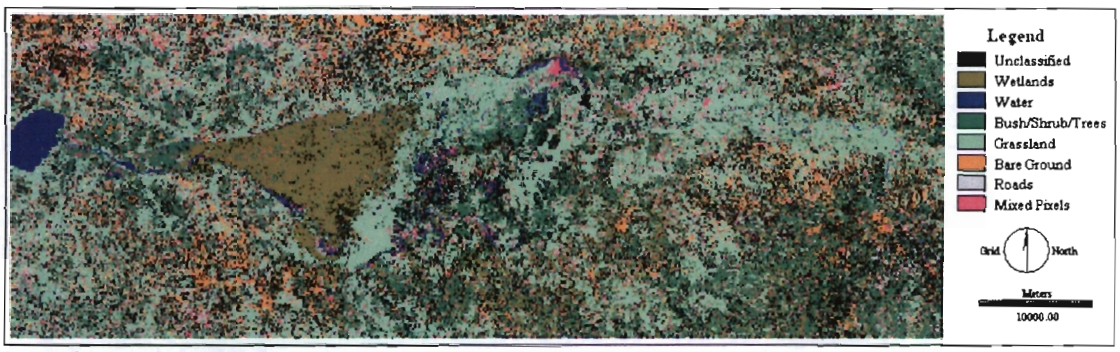


c) Radial Basis Function SVM

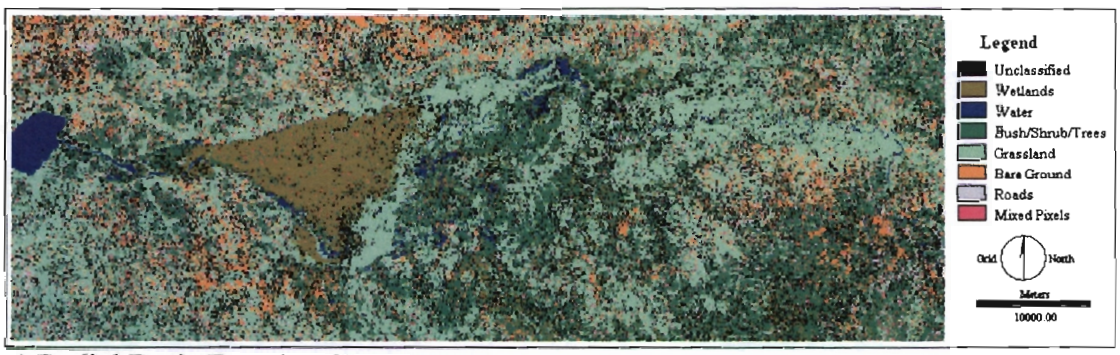
Figure 6.13: Mara Basin - Classification following exhaustive search



a) Linear SVM

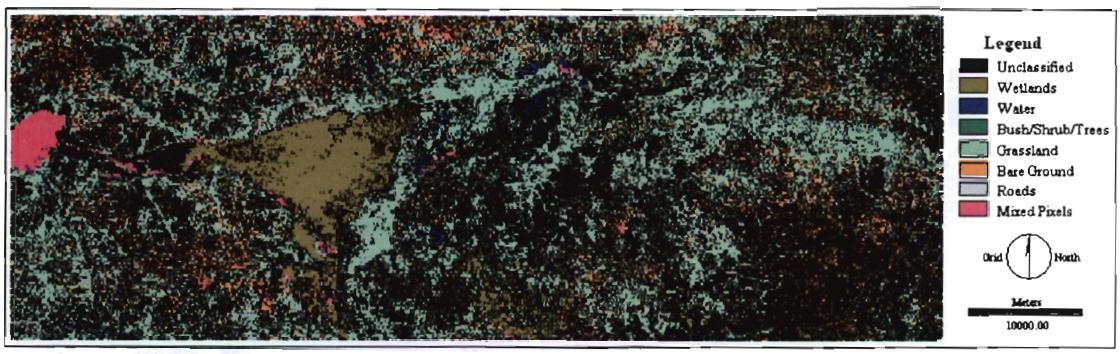


b) Polynomial SVM



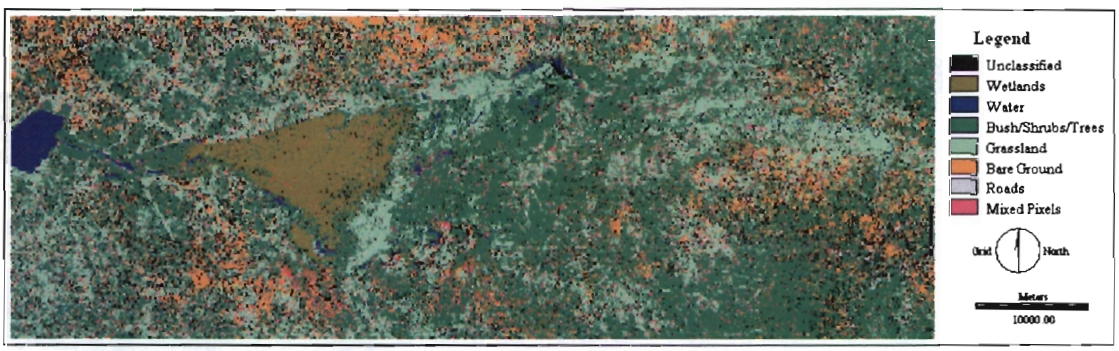
c) Radial Basis Function SVM

Figure 6.14: Mara Basin - Classification following PBIL



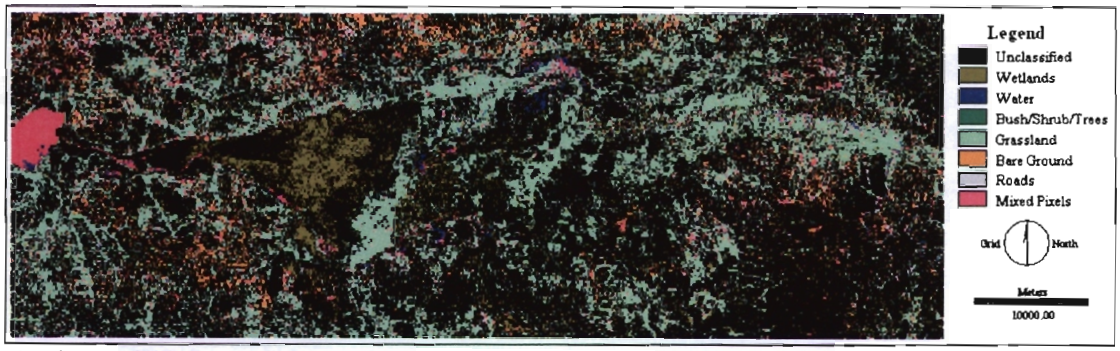
a) Linear SVM

b) Polynomial SVM

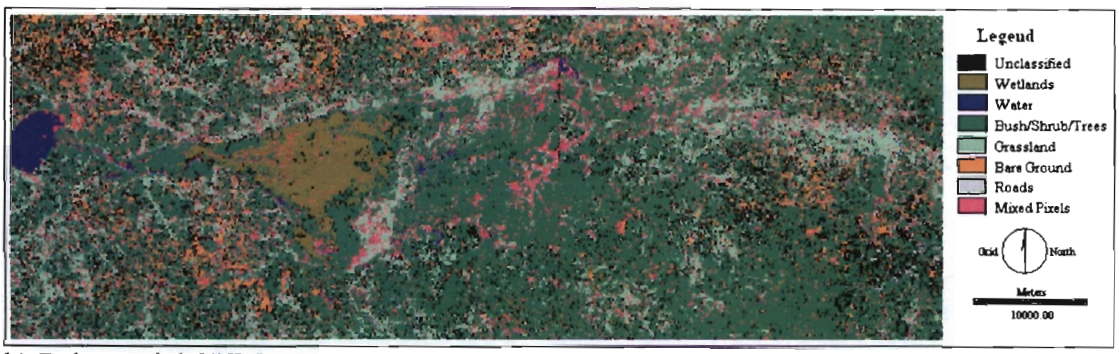


c) Radial Basis Function SVM

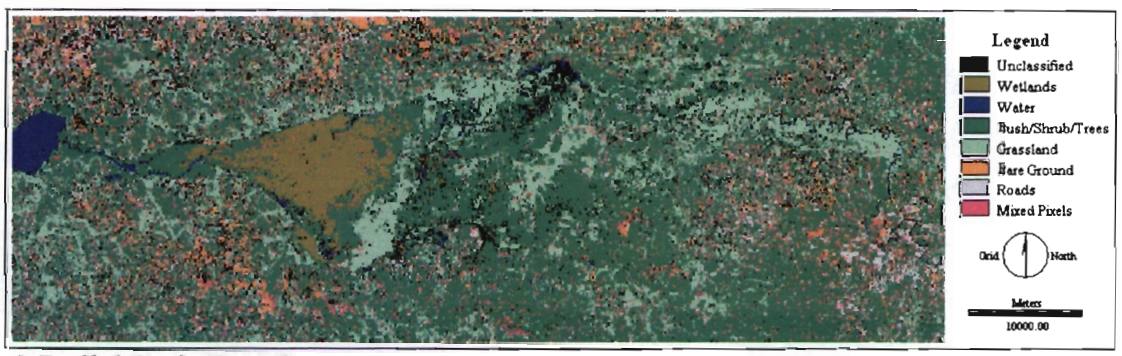
Figure 6.15: Mara Basin - Classification following instance reduction of all bands



a) Linear SVM



b) Polynomial SVM



c) Radial Basis Function SVM

Figure 6.16: Mara Basin - Classification following instance reduction of the PCA bands

6.4.5 Visual Analysis

Similarly to the visual analysis of the previous dataset, visual inspection of Figures 6.11 – 6.16 reveals the inadequacy of the Linear SVM. The land cover classes either remained unclassified or had overlaps. Evidently none of the data reduction methods improved the visual outlook of the derived maps due to the Linear SVM classifier.

In comparison, the Polynomial and RBF SVM classifiers were more effective in identifying all classes as compared to their Linear SVM counterparts. In Figure 6.11b and Figure 6.11c, all classes are as expected in comparison to ground truth data. In both derived maps, there is a spattering of overlapping classes. In comparison to the derived map due to the linear SVM, there are much fewer unclassified patches in the derived maps due to the polynomial and RBF SVM classifiers.

Figure 6.12 depicts the SVM classification of the PCA bands. Whereas Figure 6.12b has fewer unclassified patches as compared to Figure 6.12c, it does have more pixels characterized by overlap. The general land cover type distribution is pretty much satisfactory. Figure 6.12c does have fewer overlapping pixels as compared to the corresponding derived map due to the classification of all the bands. It can hence be deduced that the RBF SVM classification of the PCA bands has improved the visual appreciation of the derived map.

Figure 6.13 essentially depicts the classification of the dataset with the least number of possible band combinations derived through an exhaustive search process. Comparing Figures 6.13b and c with the original derived images in Figures 6.11b and c, one can argue that there isn't any visible improvement in what one sees. In fact the wetland, which is one of the dominant classes, has not been clearly delineated in both Figure 6.13b and c. Of all the derived maps, the exhaustive search method seemed to "over" classify the bare-ground class which in comparison to ground truth data was not as dominant as depicted.

The classification of the results of feature selection due to the PBIL search method is displayed in Figure 6.14. On account of there being fewer overlapping pixels in Figures 6.14b and c as compared to the corresponding Figures 6.11b and c, it can be

argued that PBIL has improved the visual outlook of the classification result. In similar measure, Figure 6.14 derived maps look comparatively better than the ones in Figure 6.13. In comparison to ground truth data, all the classes are pretty much as expected.

Figure 6.15 presents the derived maps due to instance reduction of all the bands. Whereas in both Figure 6.15b and c all the land cover types are as expected, Figure 6.15b has a spattering of mixed pixels in the water class. In comparison to the original classifications in Figure 6.11b and c, Figures 6.15b and c generally have fewer overlapping pixels and can hence be assumed to be better looking derived land cover maps.

Whereas Figure 6.15 displays the results of instance reduction on all the bands, Figure 6.16 illustrates the results of classifying instance reduction on the PCA bands. All classes are as expected, although Figure 6.16b has more unclassified pixels as compared to Figure 6.16c. Figure 6.16b however bears much resemblance with the corresponding Figure 6.11b. Figure 6.16c on the other hand has fewer overlapping pixels as compared to the original derived map in Figure 6.11c. Figure 6.16c could therefore be deemed an improvement on Figure 6.11c.

6.4.6 Quantitative Assessment

Quantitative assessment provided further analysis of the derived maps. Tables 6.11 – 6.13 give a summary of the statistical extracts following the comparison of the derived land cover maps with ground truth data. Appendix 2 details all the error matrices for each derived map from which the statistical results in Tables 6.11, 6.12 and 6.13 were calculated. Table 6.11 gives a summary of the overall and class-by-class accuracy measures. Table 6.12 shows the producer's accuracy assessment, which measures the proportion of pixels in a test dataset that are correctly labeled/categorized by the classifier. In Table 6.13 the user's accuracy is shown and it refers to the reliability of the derived map to map a given land cover type. The three tables formed the basis of the quantitative assessment discussed herewith.

Table 6.11: Summay of overall and class-by-class KHAT statistic

Kappa Index of Agreement		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
CI	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Wetlands	0.7171	0.7627	0.7805	0.3238	0.6667	0.5982	0.2941	0.3293	0.1732	0.0000	0.7604	0.7514	0.7721	0.7520	0.7523	0.2209	0.5401	0.6729
2	Water	0.4571	0.9125	0.9066	0.6071	0.8178	0.8861	0.5616	0.8554	0.9007	0.9790	0.9003	0.9492	0.0862	0.4962	0.8993	0.0652	0.8207	0.8788
3	Bush/Shrub	0.0000	0.7188	0.7366	0.0000	0.7750	0.8233	0.0000	0.7133	0.5141	0.0000	0.5351	0.5629	0.0000	0.6450	0.7756	0.0000	0.8341	0.8390
4	Grasslands	0.9569	0.9118	0.9004	0.8949	0.9494	0.9523	0.9722	0.9419	0.9326	0.9831	0.9756	0.9778	0.9531	0.8807	0.9381	0.9682	0.8116	0.9727
5	Bare Ground	0.4301	0.0000	0.6439	0.5429	0.4519	0.5209	0.7251	0.5264	0.5731	0.8594	0.5356	0.6039	0.5984	0.3632	0.5665	0.0141	0.5462	0.4849
6	Roads	0.0000	0.0000	0.7055	0.0343	0.8288	0.8674	0.0022	0.5577	0.5714	0.0000	0.5024	0.6500	0.0092	0.0076	0.7627	0.5074	0.4712	0.8484
Overall Kappa		**0.4029**	**0.6077**	**0.7880**	**0.3457**	**0.7769**	**0.7910**	**0.3498**	**0.6491**	**0.5734**	**0.3289**	**0.7055**	**0.7458**	**0.3853**	**0.5711**	**0.8012**	**0.3144**	**0.6830**	**0.8096**
Overall Accuracy		**0.4619**	**0.6709**	**0.8270**	**0.3997**	**0.8177**	**0.8312**	**0.4087**	**0.7068**	**0.6392**	**0.3800**	**0.7542**	**0.7902**	**0.4474**	**0.6384**	**0.8382**	**0.3712**	**0.7374**	**0.8454**
O.A(%)		**46.19**	**67.09**	**82.70**	**39.97**	**81.77**	**83.12**	**40.87**	**70.68**	**63.92**	**38.00**	**75.42**	**79.02**	**44.74**	**63.84**	**83.82**	**37.12**	**73.74**	**84.54**

Table 6.12: Producer's accuracy assesment

Producer Accuracy		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
CI	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Wetlands	76.56	80.51	82.11	38.68	71.93	65.64	35.66	38.43	21.22	00.00	80.32	79.58	81.62	79.64	79.52	27.39	60.09	72.42
2	Water	48.24	92.03	91.49	63.11	83.24	89.59	58.65	86.76	90.95	98.11	90.95	95.41	09.46	52.16	90.81	07.16	83.51	88.92
3	Bush/Shrub	00.00	00.79	00.80	00.00	00.83	00.88	00.00	00.79	00.63	00.00	00.62	00.65	00.00	00.73	00.83	00.00	00.88	00.88
4	Grasslands	96.71	93.19	92.28	91.77	96.08	96.31	97.90	95.57	94.84	98.75	98.18	98.35	96.42	90.69	95.23	97.56	84.90	97.90
5	Bare Ground	46.17	00.00	66.40	57.17	47.35	54.03	76.62	55.99	60.71	87.23	55.80	62.48	61.69	38.90	59.14	01.96	56.97	50.49
6	Roads	00.00	00.00	74.01	05.31	85.43	88.76	00.27	60.61	61.60	00.00	54.50	69.15	01.80	01.80	79.32	55.49	52.16	87.05

Table 6.13: User's accuracy assesment

User Accuracy		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
CI	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Wetlands	94.52	95.46	93.93	87.94	96.52	95.94	85.25	99.20	95.29	-	95.18	94.37	89.33	94.10	97.14	85.22	96.25	97.59
2	Water	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.38	100.00	94.29	96.01	97.38	100.00	100.00	100.00	100.00	100.00	100.00
3	Bush/Shrub	-	79.10	78.99	-	79.72	71.88	-	77.18	64.32	-	87.55	85.59	-	76.70	79.61	-	75.71	78.99
4	Grasslands	94.40	93.94	94.21	97.18	97.47	98.04	91.80	92.22	93.20	86.61	88.00	87.35	93.10	95.06	95.23	96.52	98.23	97.96
5	Bare Ground	55.29	-	78.06	60.37	80.07	88.71	34.06	52.58	50.66	53.34	76.96	78.71	88.95	63.87	68.56	23.26	73.05	86.24
6	Roads	-	-	91.34	39.60	83.19	84.58	75.00	80.33	85.95	-	92.52	84.60	29.41	25.00	89.73	83.60	79.56	86.82

From Table 6.11, the following Table 6.14 was derived to cast more light on the relative performance of all classifiers. Table 6.14 ranks the results of the various classification accuracies based on the derived maps' KHAT statistical measures.

Table 6.14: Ranking of classification results

Rank	Classifier (SVM)	Reduction Method	Overall Kappa
1	Radial Basis Function	Instance Reduction - PCA	0.8096
2	Radial Basis Function	Instance Reduction – All	0.8012
3	Radial Basis Function	PCA	0.7910
4	Radial Basis Function	None - All	0.7880
5	Polynomial	PCA	0.7769
6	Radial Basis Function	PBIL	0.7458
7	Polynomial	PBIL	0.7055
8	Polynomial	Instance Reduction – PCA	0.6830
9	Polynomial	Exhaustive Search	0.6491
10	Linear	None – All	0.6077
11	Radial Basis Function	Exhaustive Search	0.5734
12	Polynomial	Instance Reduction – All	0.5711
13	Polynomial	None – All	0.4029
14	Linear	Instance Reduction – All	0.3853
15	Linear	Exhaustive Search	0.3498
16	Linear	PCA	0.3457
17	Linear	PBIL	0.3289
18	Linear	Instance Reduction - PCA	0.3144

In spite of the subjective nature of the visual analysis, much of what was observed tallied with the quantitative analysis. For instance, in general polynomial and RBF SVM classifiers performed better than corresponding linear SVM classifiers. It can also be observed from both the visual analysis and ranking of the classification results that the Linear SVM classification retarded with application of optimization technique. On the other hand, polynomial and RBF SVM classification in general posted improved classification results following the application of different optimization techniques experimented with. The classification of the results of exhaustive search method (rank 11) were the only exception to this observation.

Tables 6.11 – 6.13 shed more light on the class level accuracies. From the tables, wetlands were best classified when all the bands were used as well as following the instance reduction. In both cases, all the three classifiers were deemed to be effective in delineating the wetlands. The worst classification of wetlands was posted by the exhaustive search method, whereby all the classifiers performed poorly. The PBIL method only posted good results when the polynomial and RBF SVM classifiers were

used. Both PCA reduction methods yielded averagely satisfactory results only when the polynomial and RBF SVM classifiers were in use. In general, none of the optimization techniques improved on the classification of the wetlands.

Across all the optimization techniques, the PBIL method posted the best accuracy measures for the water class for all classifiers. All the RBF SVMs delineated the water class well, which included the lake and rivers. Similarly, all optimization techniques of the polynomial SVMs, apart from instance reduction of all bands, classified the water class to a high accuracy level. The linear classification of the PBIL optimization method for this class also yielded impressive results, however for the PCA and exhaustive search method, the accuracy results were only averagely satisfactory. The optimization of the linear classifier through instance reduction of all bands and the PCA bands clearly posted dismal water class accuracies. In general, the PBIL optimization succeeded in improving the classification of water for the linear and RBF classifier. The Polynomial classifier's result for PBIL was appreciably close to the corresponding accuracy of all the bands.

All optimization of linear SVMs returned very poor accuracies for the bush/shrub/trees class. All optimization of the polynomial and RBF SVM classifiers on the other hand yielded good results for this class. Improvements in identifying this class using the two mentioned classifiers were observed when both PCA optimization techniques were used. The instance reduction of all bands also improved the classification accuracy when the RBF classifier was used.

Of all the classes, the grassland class posted the best classification accuracies across classifiers and optimization techniques. Whereas all the class accuracies were appreciably high, both feature selection techniques yielded accuracy improvements across all classifiers. The PCA optimization technique also posted improved accuracies for the polynomial and RBF SVM classifiers. Both instance reduction techniques returned improved accuracies when the RBF classifier was used, however only the linear classifier of the instance reduction of the PCA bands posted improved accuracies.

The bare ground class posted mixed results across optimization techniques and classifiers. On average the PBIL optimization of all the classifiers gave the best results in comparison to other techniques. All the other optimization techniques and classifiers yielded mixed results consequently no generalities can be made. All the optimization techniques improved the classification accuracies of this class following the utilization of the linear and polynomial SVMs. The exception to this observation was the result of optimizing the linear SVM classifier through instance reduction of all bands. None of the RBF classifiers improved the classification accuracies.

The linear SVM classifier generally faired poorly in delineating the roads with or without optimization, apart from instance reduction of the PCA bands where the accuracy returns were about satisfactory. The tables show that the polynomial classifier's performance for this class was enhanced following the application of the optimization techniques. The only deviation to this observation was the instance reduction of all bands. Improvements in delineating roads is equally scattered across optimization techniques and classifiers. Both feature selection optimization techniques improved the delineation of roads using the Polynomial SVM classifiers. Similarly, both the instance reduction techniques improved classification accuracy using the RBF SVM classifier, however only the Polynomial SVM of the PCA bands improved accuracy. In the same measure, the PCA optimization technique improved classification accuracy when using the Polynomial and RBF SVMs.

In an attempt to determine the significance of the results of data reduction in comparison to the original datasets, a binomial test of significance was carried out at the 95% confidence level (critical Z value = 1.96). The test of significance was calculated using the overall accuracies as extracted from the error matrices in Appendix 2 and summarized in Table 6.15 below. The comparisons were made for each classifier i.e. linear, polynomial and Radial Basis Function classifiers, with a view of determining the effectiveness of data reduction on land cover classification.

Table 6.15: Calculated Z values for comparison between results of classifying all bands and the data reduction methods

	PCA	EXHAUSTIVE	PBIL	CS_ALL	CS_PCA
Linear	-7.79	-6.65	-10.30	1.80	-11.43
Comments	*Significantly Worse*	*Significantly Worse*	*Significantly Worse*	*Insignificant*	*Significantly Worse*
Polynomial	20.37	4.65	11.07	-4.11	8.75
Comments	*Significantly Better*	*Significantly Better*	*Significantly Better*	*Significantly Worse*	*Significantly Better*
RBF	0.59	-23.74	-5.00	1.58	2.61
Comments	*Insignificant*	*Significantly Worse*	*Significantly Worse*	*Insignificant*	*Significantly Better*

PCA – Principal component analysis
Exhaustive – Exhaustive search method
PBIL – Population Based Incremental search method
CS_ALL – Instance reduction of all the nine bands
CS_PCA – Instance reduction of the PCA bands

Table 6.15 goes on to confirm the relative inadequacy of the linear classifier. Apart from the instance reduction of all bands, all other optimization techniques posted poorer results. Indeed, none of the optimization techniques helped improve the overall Linear SVM classification results. The Polynomial SVM classifier results on the other hand yielded significantly better results across optimization techniques, apart from the instance reduction of all bands. Both feature selection techniques performed significantly worse when the RBF classifier was applied. The PCA and instance reduction of all bands optimization techniques posted insignificantly different results however the instance reduction of the PCA bands resulted in improved classification accuracies.

6.4.7 Derived Land Cover Maps for Malmesbury Study area

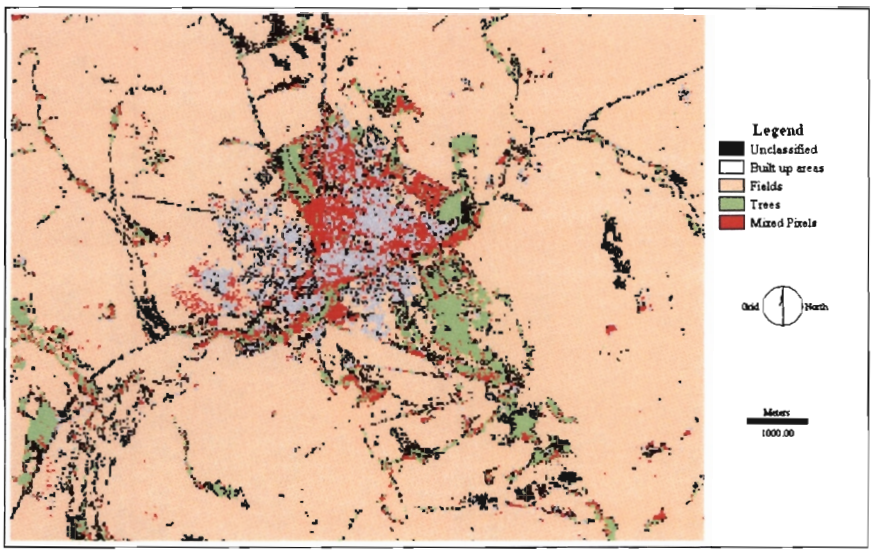
Figures 6.17 - 6.22 illustrate the classification results of various reduction techniques. Table 6.17 below will act as a legend to the derived maps. In each figure, three land cover maps are presented representing classification due to linear, polynomial and RBF SVM classifiers respectively. Each figure, apart from Figure 6.17, depicts the results of data reduction on the mentioned classifiers. Figure 6.17 shows the SVM classification results before the application of any reduction method. These derived maps in effect acted as the 'control' or 'reference' land cover maps to which all the

other corresponding maps were compared. Table 6.16 shows the order in which the derived maps are presented, depending on the reduction technique.

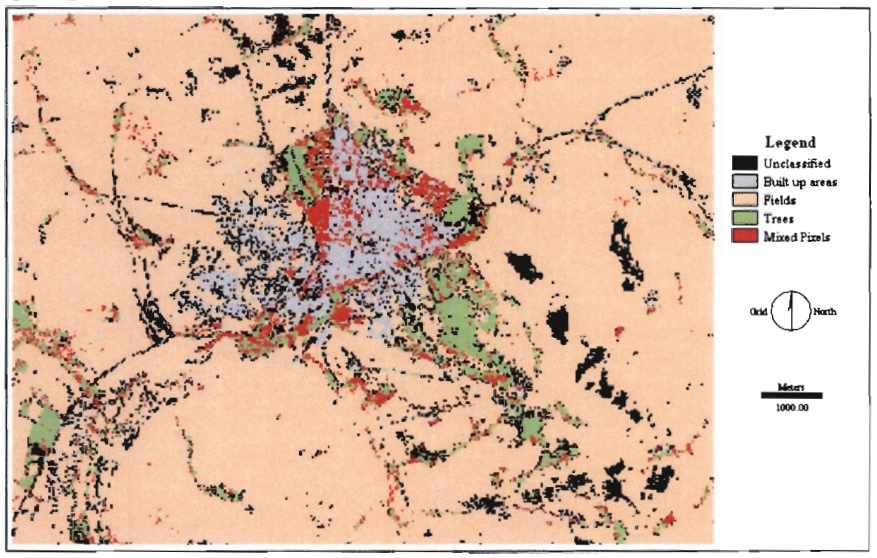
Table 6.16: Legend of derived maps

Figure	Reduction Method
Figure 6.17	None – classification of all nine bands
Figure 6.18	PCA
Figure 6.19	Exhaustive Search
Figure 6.20	PBIL
Figure 6.21	Instance reduction of all nine bands
Figure 6.22	Instance reduction of PCA

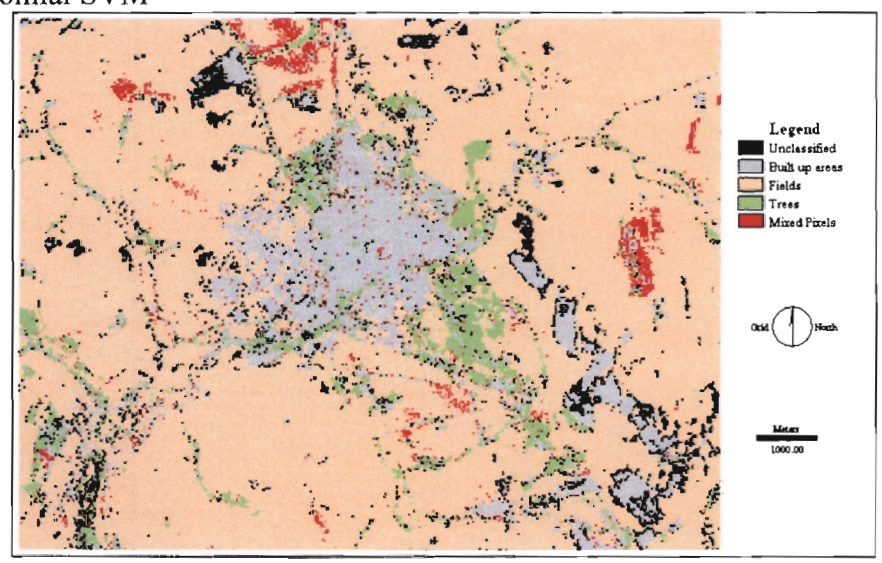
Sections 6.4.8 and 6.4.9 give a visual and quantitative assessment of these derived images.



a) Linear SVM

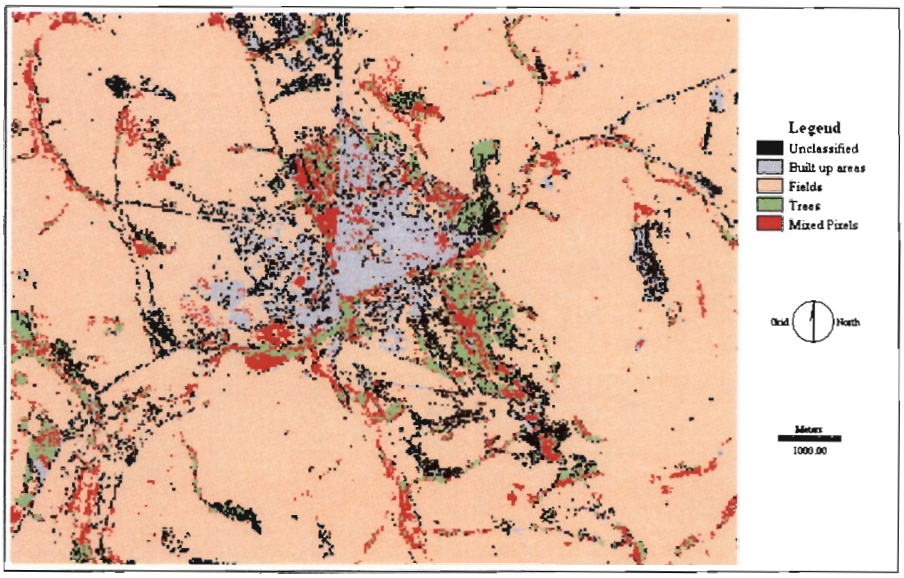


b) Polynomial SVM

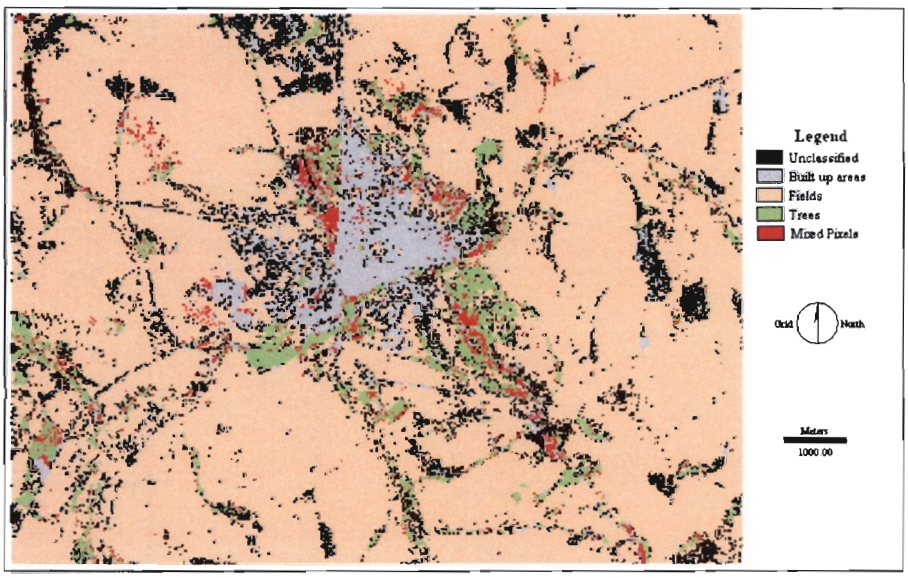


c) Radial Basis Function SVM

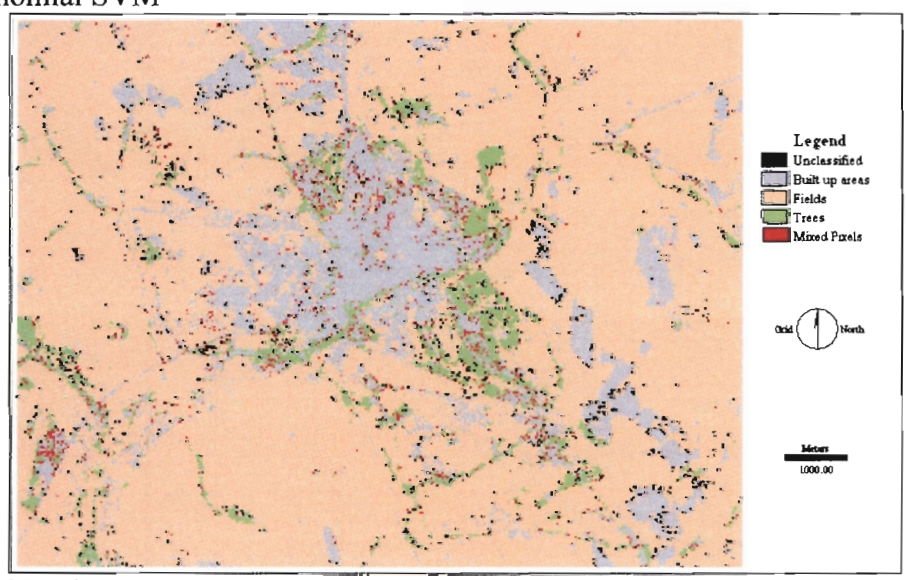
Figure 6.17: Malmesbury – Classification of all six bands



a) Linear SVM

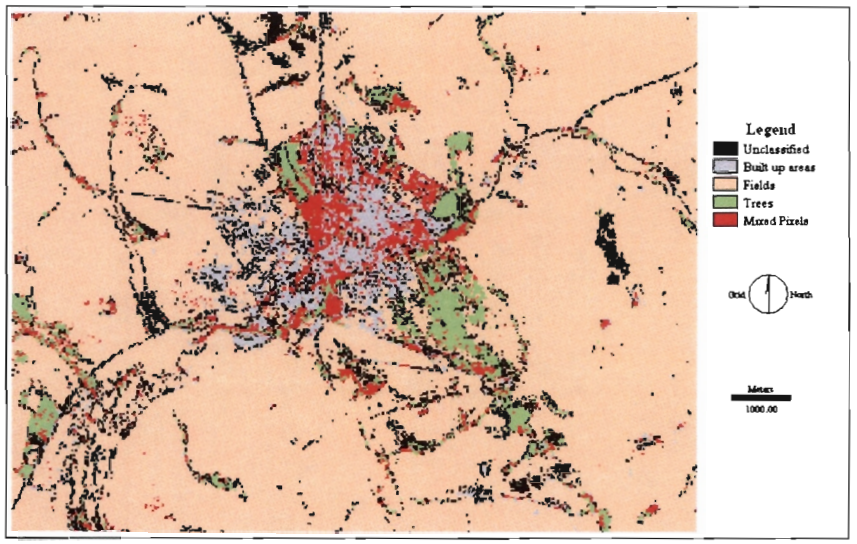


b) Polynomial SVM

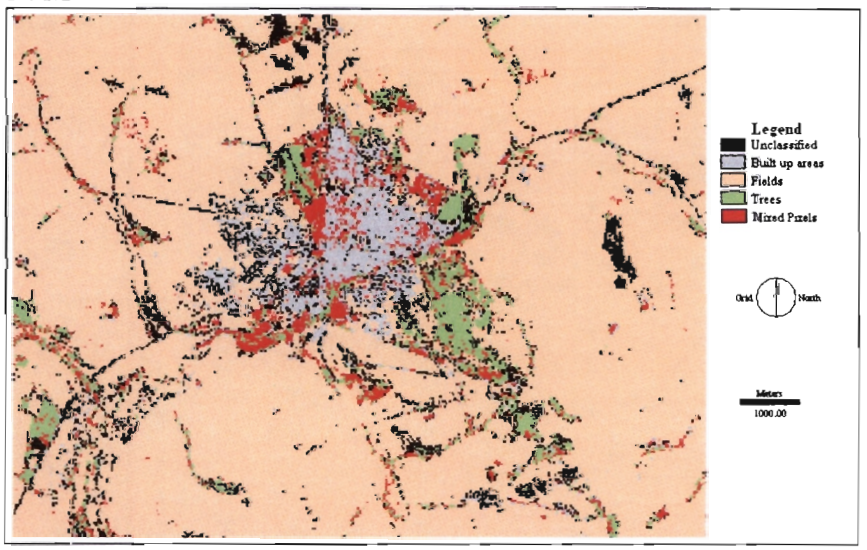


c) Radial Basis Function SVM

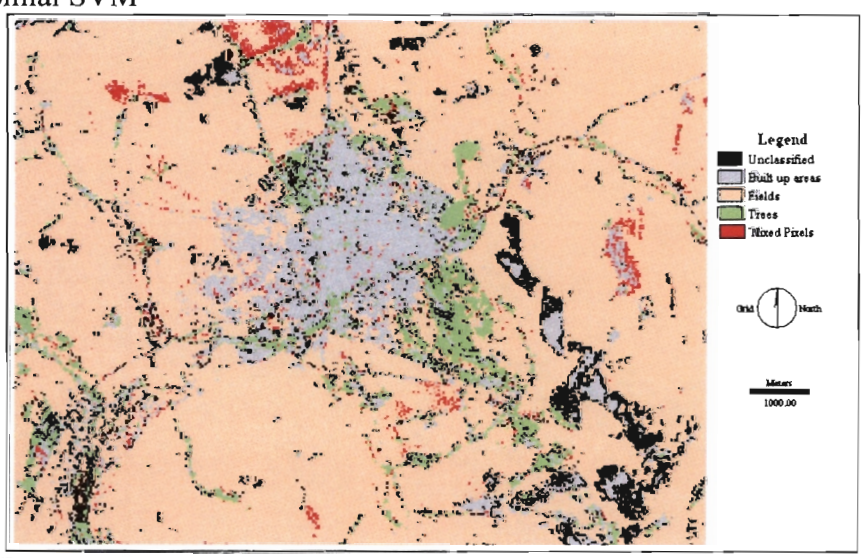
Figure 6.18: Malmesbury – Classification of PCA Bands



a) Linear SVM

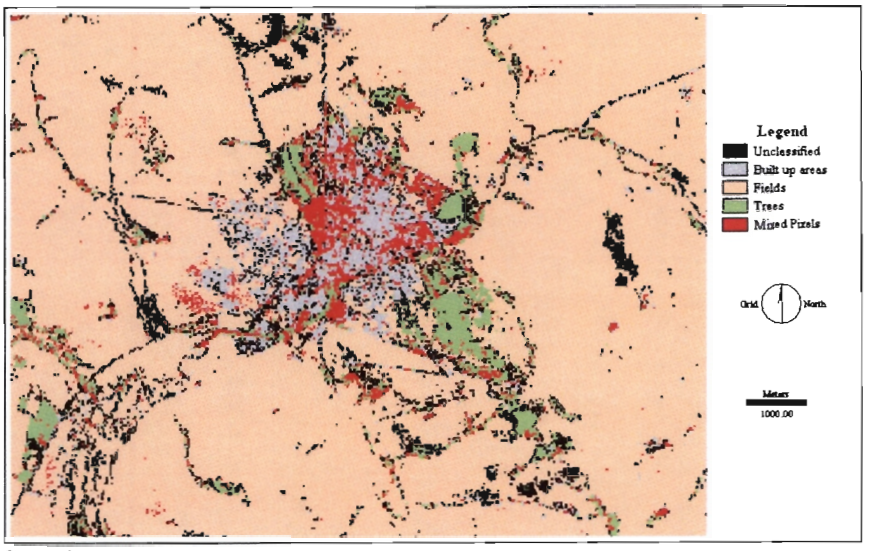


b) Polynomial SVM

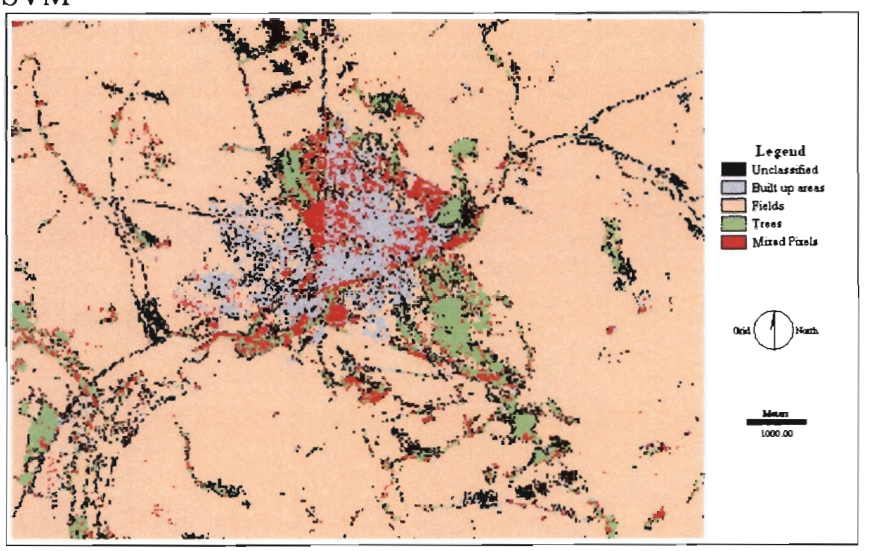


c) Radial Basis Function SVM

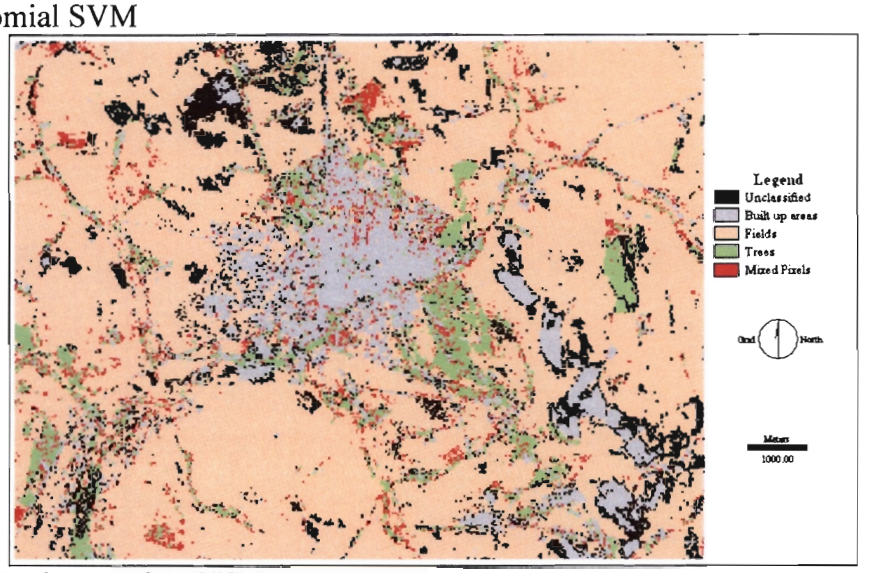
Figure 6.19: Malmesbusy – Classification following exhaustive search



a) Linear SVM

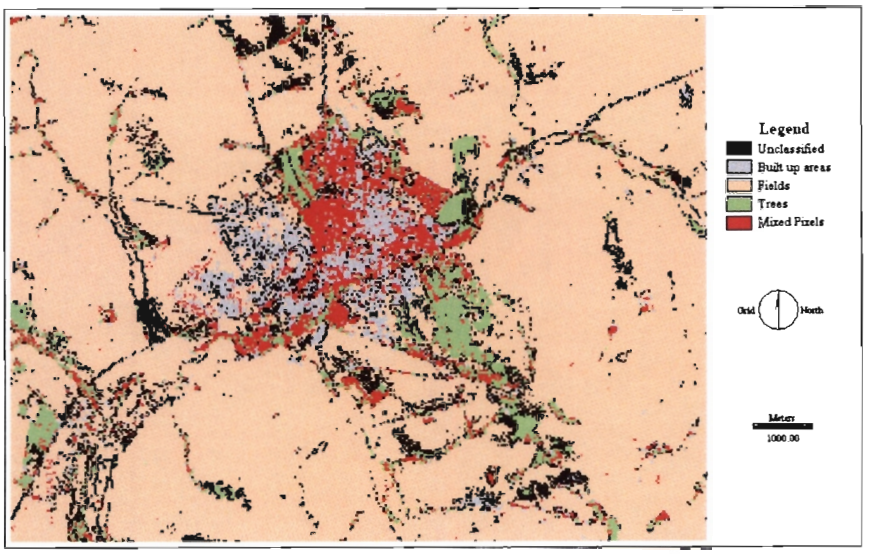


b) Polynomial SVM

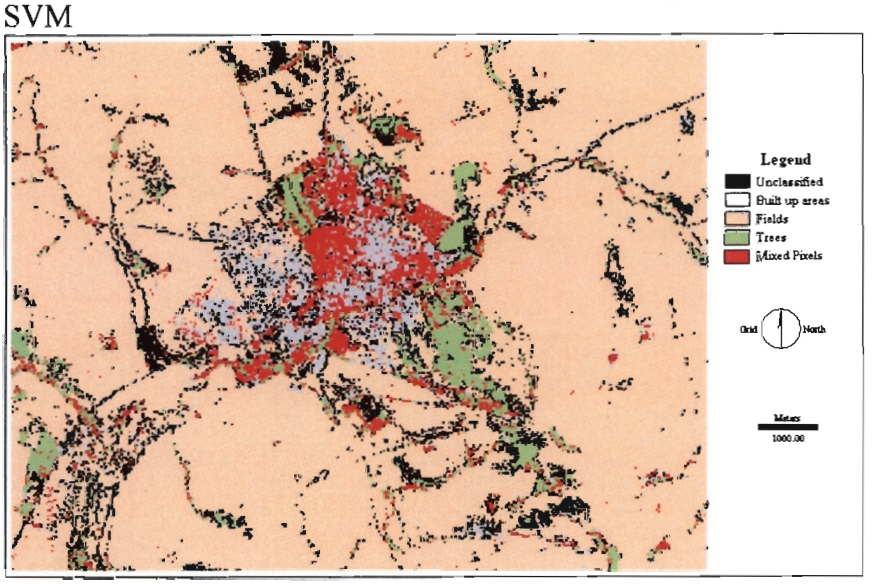


c)Radial Basis Function SVM

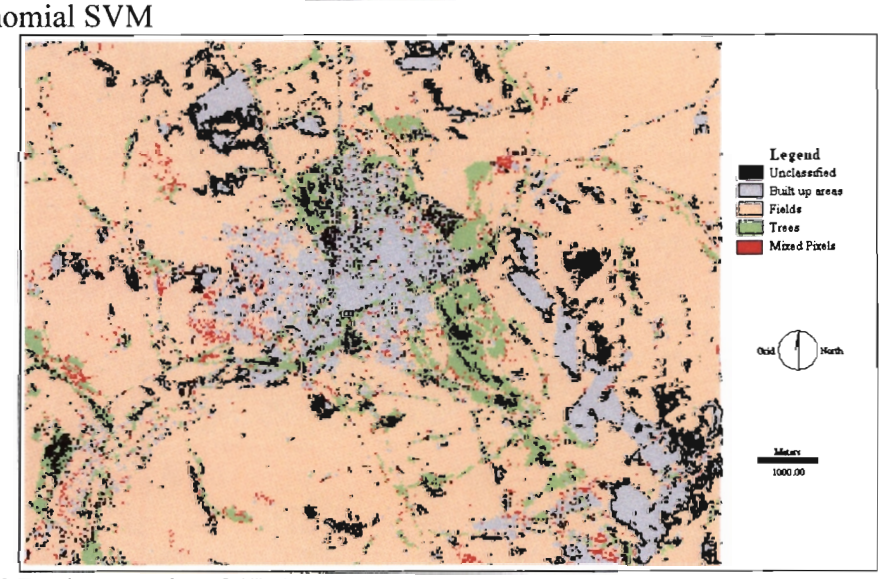
Figure 6.20: Malmesbury – Classification following PBIL



a) Linear SVM

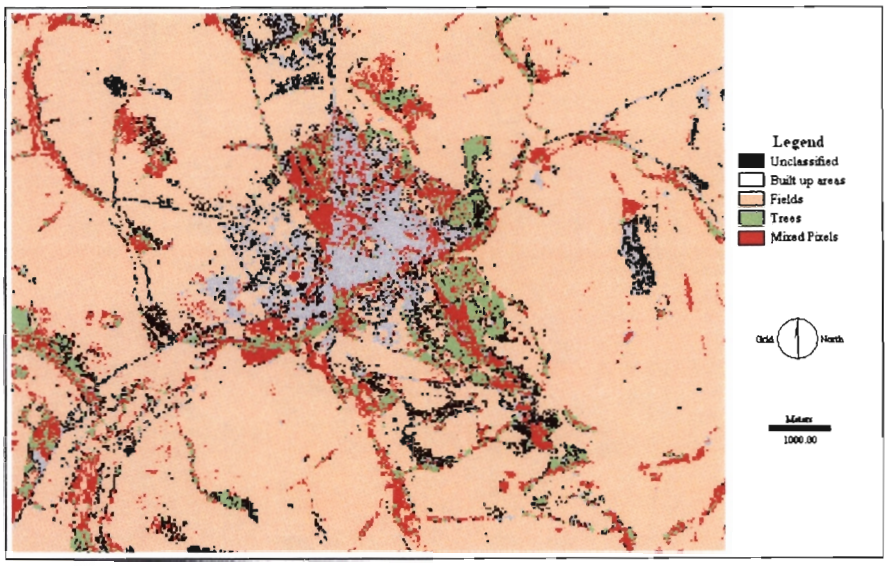


b) Polynomial SVM

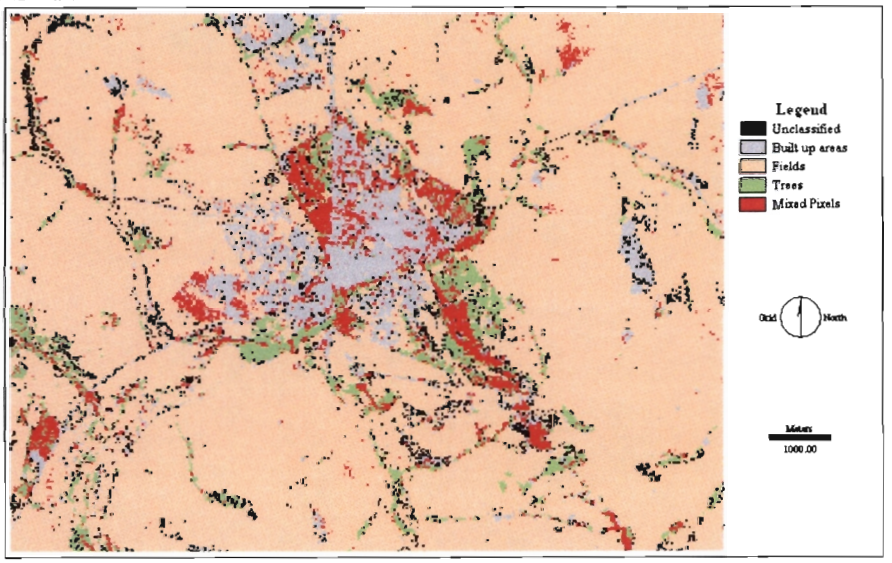


c) Radial Basis Function SVM

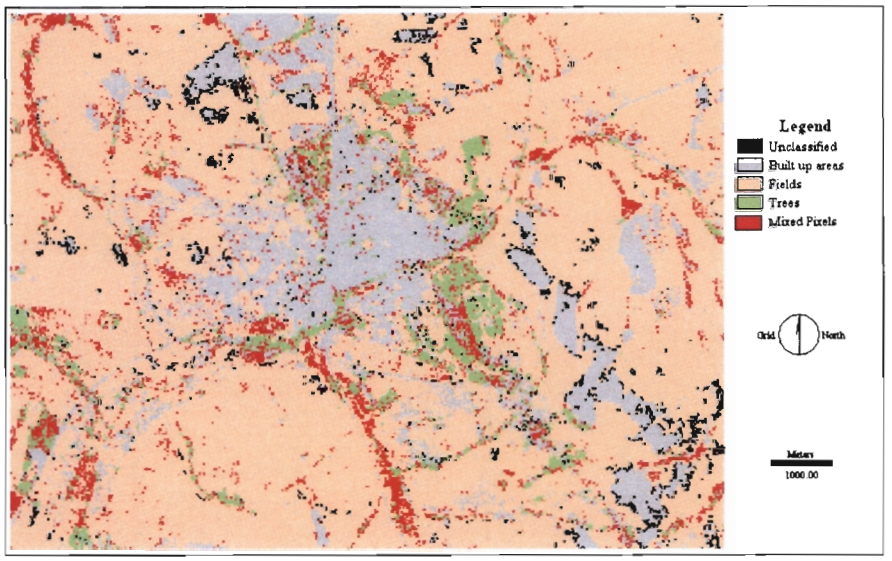
Figure 6.21: Malmesbury – Classification following instance reduction of all bands



a) Linear SVM



b) Polynomial SVM



c) Radial Basis Functions SVM

Figure 6.22: Malmesbury – Classification following instance reduction of the PCA Bands

6.4.8 Visual Analysis

In the two previous datasets, the linear SVM classification results were appreciably different from the results of the polynomial and RBF SVM classifications. From Figures 6.17 – 6.22, it is evident that little separates the visual appreciation of the various SVM classifier results.

Figure 6.17 depicts the classification of the original bands. Figures 6.17a - c all look pretty much the same apart from Figure 6.17c which has comparatively fewer unclassified and mixed pixels. Given the low resolution of the satellite imagery, it is understandable that the built up area is not clearly delineated.

Figure 6.18 displays the classification of the PCA bands. In all the derived maps, there is the same general outlay of the identified land cover types. The map derived using the RBF SVM classifier has the least number of mixed and unclassified pixels and could be considered the best map, even in comparison with all the other derived maps. From a visual perspective, Figure 6.18a has more mixed pixels as compared to the corresponding Figure 6.17a. In comparison, Figure 6.17a has more unclassified pixels. Comparing Figure 6.18b to Figure 6.17b, one can observe that there are more unclassified pixels hence no noticeable improvement can be recorded. Figure 6.18c is definitely an improvement of Figure 6.17c, and shows that through the PCA optimization technique, classification accuracy improvements are possible.

Figure 6.19 shows the classification results following feature selection using the exhaustive search method. Of the three maps, the RBF SVM classifier has yielded the most number of unclassified pixels, while the other classifiers could be argued to have more mixed pixels. In comparison, Figures 6.17a and 6.19a, are similar hence one could conclude that even with this optimization method classification accuracies can be maintained. Figure 6.19b definitely has fewer unclassified pixels in comparison to the original classification while Figure 6.19c has more unclassified pixels in comparison to the corresponding Figure 6.19c.

The SVM classification results of feature selection using PBIL are shown in Figure 6.20. Of the three maps, Figure 6.20b seems to have the least number of unclassified pixels, while Figure 6.20c seems to have the least number of mixed pixels. The general outlay of the classes is the same. In comparison to the original classification, Figures 6.20a and c are a lot similar to the corresponding classifications in Figures 6.17a and c. Figure 6.20b seems to be an improvement on Figure 6.17b since it has fewer unclassified pixels.

Figure 6.21 illustrates the SVM classification results of instance reduction of all bands. At a glance, Figure 6.21c has fewer mixed pixels hence could be deemed the best derived map. The other two maps look alike. In comparison to the corresponding original maps in Figure 6.17, one could say Figure 6.21 is pretty much similar.

Figure 6.22 shows the SVM classification results of instance reduction of the PCA bands. Of all the classification results, this category of derived maps seems to have the largest amount of mixed pixels. That said however, the layout of the expected land cover classes are as expected. In comparison to the original classifications, Figure 6.22 generally has more mixed pixels

6.4.9 Quantitative Assessment

Quantitative assessment provided further analysis of the derived maps. Tables 6.17 – 6.19 give a summary of the statistical extracts following the comparison of the derived land cover maps with ground truth data. Appendix 3 details all the error matrices for each derived map from which the statistical results in Tables 6.17, 6.18 and 6.19 were calculated. Table 6.17 gives a summary of the overall and class-by-class accuracy measures. Table 6.18 shows the producer's accuracy assessment, which measures the proportion of pixels in a test dataset that are correctly labeled/categorized by the classifier. In Table 6.19 the user's accuracy is shown and it refers to the reliability of the derived map to map a given land cover type. The three tables formed the basis of the quantitative assessment discussed herewith.

Table 6.17: Summay of overall and class-by-class KHAT statistic

Kappa Index of Agreement		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
CI	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Built up areas	0.2568	0.3414	0.3016	0.4921	0.5177	0.7601	0.2568	0.2984	0.3392	0.2984	0.3414	0.5278	0.2568	0.2568	0.3183	0.5018	0.6480	0.8955
2	Fields	0.8027	0.7984	0.5492	0.5758	0.5245	0.4276	0.6846	0.6780	0.6780	0.6846	0.6846	0.4545	0.6846	0.7424	0.8562	0.6364	0.6215	0.4380
3	Trees	0.6669	0.6669	0.6669	0.2293	0.2624	0.2966	0.6600	0.6669	0.5643	0.6669	0.6600	0.5769	0.6600	0.6600	0.5370	0.1974	0.1974	0.1974
Overall Kappa		**0.5582**	**0.5892**	**0.5131**	**0.4108**	**0.4154**	**0.4271**	**0.5223**	**0.5391**	**0.5310**	**0.5420**	**0.5553**	**0.5195**	**0.5223**	**0.5386**	**0.5491**	**0.4180**	**0.4475**	**0.4176**
Overall Accuracy		**0.6618**	**0.6912**	**0.6471**	**0.5294**	**0.5441**	**0.6029**	**0.6324**	**0.6471**	**0.6618**	**0.6471**	**0.6618**	**0.6618**	**0.6324**	**0.6471**	**0.6765**	**0.5294**	**0.5735**	**0.5735**
O.A(%)		**66.18**	**69.12**	**64.71**	**52.94**	**54.41**	**60.29**	**63.24**	**64.71**	**66.18**	**64.71**	**66.18**	**66.18**	**63.24**	**64.71**	**67.65**	**52.94**	**57.35**	**57.35**

Table 6.18: Producer's accuracy assesment

Producer Accuracy		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
CI	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Built up areas	33.33	42.86	47.62	61.90	66.67	90.48	33.37	38.10	52.38	38.10	42.86	66.67	33.33	33.33	42.89	61.90	76.19	55.24
2	Fields	86.36	86.36	68.18	68.18	63.64	54.55	77.27	77.27	77.27	77.27	77.27	59.09	77.27	81.82	90.91	72.73	72.73	54.55
3	Trees	76.00	76.00	76.00	32.00	36.00	40.00	76.00	76.00	68.00	76.00	76.00	72.00	76.00	76.00	68.00	28.00	28.00	28.00

Table 6.19: User's accuracy assesment

User Accuracy		Original Bands						Feature Selection						Center Splitting					
		All nine Bands			PCA			Exhaustive Search			PBIL			All nine Bands			PCA		
CI	Land Cover	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R	L	P	R
1	Built up areas	100.00	100.00	58.82	76.47	66.67	46.34	100.00	100.00	57.89	100.00	100.00	70.00	100.00	100.00	81.82	81.25	72.73	54.05
2	Fields	90.48	86.36	75.00	88.24	87.50	85.71	89.47	85.00	85.00	89.47	89.47	76.47	89.47	90.00	80.00	94.12	84.21	92.31
3	Trees	100.00	100.00	100.00	100.00	100.00	100.00	95.00	100.00	94.44	100.00	95.00	78.16	95.00	95.00	80.95	100.00	100.00	100.00

In the previous study areas, both visual and quantitative analysis showed that the linear SVM classifier across all optimization techniques performed dismally. In this study area, the quantitative analysis reveals that just as was portrayed in the visual analysis, the linear SVM classifier did as well as the polynomial and RBF classifiers.

Evaluating each optimization technique in turn, it is evident from Table 6.17 that improvements in classification accuracy were only observed when the RBF classifier was used to classify the results of both feature selection techniques (i.e. exhaustive search and PBIL) and instance reduction of the original bands. The other classifiers for the above mentioned techniques did not improve the classification accuracies and the same could be mentioned for both PCA optimization techniques (i.e PCA of the original bands and instance reduction of the PCA bands).

Table 6.18 depicts the producer accuracy which gives a measure of how well the classifiers predicted the individual classes. From this table it is evident that of the three classes the built up areas posted a higher number of improved results following optimization. Improved results were noted when the following optimization techniques and SVM classifiers used; PCA across all SVM classifiers, exhaustive search using linear and RBF SVM classifiers, PBIL when employing polynomial and RBF SVM classifiers, instance reduction of PCA bands utilizing linear and polynomial SVMs. Fields classification accuracy on the other hand only improved following the RBF SVM classification of instance reduction of all bands. The classification of trees was improved when linear and polynomial SVMs were applied to the both feature selection techniques and instance reduction of both bands.

Table 6.20 below was derived from Table 6.17 and casts more light on the relative performance of all classifiers. Table 6.20 ranks the results of the various classification accuracies based on the derived maps' KHAT statistical measures. As mentioned before, as opposed to the previous study areas the Linear SVM classifier in this study area did as well as the Polynomial and RBF SVM classifiers and this is depicted in the rankings in Table 6.20. Table 6.20 shows a broader spread of classifier performance as opposed to Tables 6.8 and 6.14 where the linear classifier dominated the poorest results. The one consistency about the table ratings is that the poorest accuracies are associated with the PCA optimization technique.

Table 6.20: Ranking of classification results

Rank	Classifier (SVM)	Reduction Method	Overall Kappa
1	Polynomial	None	0.6912
2	Radial Basis Function	Instance Reduction – All	0.6765
3	Linear	None	0.6618
4	Radial Basis Function	Exhaustive Search	0.6618
5	Polynomial	PBIL	0.6618
6	Radial Basis Function	PBIL	0.6618
7	Radial Basis Function	None	0.6471
8	Polynomial	Exhaustive Search	0.6471
9	Linear	PBIL	0.6471
10	Polynomial	Instance Reduction – All	0.6471
11	Linear	Exhaustive Search	0.6324
12	Linear	Instance Reduction – All	0.6324
13	Radial Basis Function	PCA	0.6029
14	Polynomial	Instance Reduction – PCA	0.5735
15	Radial Basis Function	Instance Reduction – PCA	0.5735
16	Polynomial	PCA	0.5441
17	Linear	PCA	0.5294
18	Linear	Instance Reduction - PCA	0.5294

Table 6.21 provides more insight into the relative performance of the optimization techniques in comparison to the original classification. Table 6.21 shows how significant the comparisons are by using a binomial test of significance carried out at the 95% confidence level (critical Z value = 1.96). The test of significance was calculated using the overall accuracies as extracted from the error matrices in Appendix 3. The comparisons were made for each classifier i.e. linear, polynomial and Radial Basis Function SVM classifiers, with a view of determining the effectiveness of data reduction on land cover classification.

Table 6.21 reveals that in spite of the relative rankings of the various classifiers and optimization techniques, there wasn't any significant difference between the optimization results and the original classification. As noted in the various analysis measures for this study area, even the linear SVM classifier performed as well as the polynomial and RBF SVM classifiers. Whereas the optimization techniques did not significantly improve the classification accuracies, at least the differences were insignificant.

Table 6.21: Calculated Z values for comparison between results of classifying all bands and the data reduction methods

	PCA	EXHAUSTIVE	PBIL	CS_ALL	CS_PCA
Linear	1.59	0.36	0.18	0.03	1.59
Comments	*Insignificant*	*Insignificant*	*Insignificant*	*Insignificant*	*Insignificant*
Polynomial	1.79	0.55	0.37	0.04	1.43
Comments	*Insignificant*	*Insignificant*	*Insignificant*	*Insignificant*	*Insignificant*
RBF	0.53	-0.18	-0.18	-0.03	0.88
Comments	*Insignificant*	*Insignificant*	*Insignificant*	*Insignificant*	*Insignificant*

PCA – Principal component analysis
Exhaustive – Exhaustive search method
PBIL – Population Based Incremental search method
CS_ALL – Instance reduction of all the nine bands
CS_PCA – Instance reduction of the PCA bands

6.5 Summary of Results

In an attempt to generate an overall view of the performance of the optimization techniques across SVM classifiers and datasets, a 'performance' matrix was proposed. The performance matrix was generated by appending scores to the KHAT and binomial tests of significance results for each SVM classifier and corresponding optimization technique (see Table 6.23). Table 6.22 gives the distribution of the proposed scores. The scores were selected with the intention of giving more weight to higher KHAT scores and binomial tests of significance.

Table 6.22: Proposed Scores for corresponding interval

KHAT	SIGNIFICANCE LEVEL	SCORE
$\hat{k} < 0.4$	Significantly worse	0
$0.4 < \hat{k} < 0.8$	Insignificant	1
$\hat{k} \geq 0.8$	Significantly better	2

Table 6.23 is a performance matrix in which the matrix elements are a summary of the scores from the KHAT values in Tables 6.8, 6.14 and 6.21. The performance matrix in Table 6.23 was then modified to give an equal score to the matrix elements with KHAT values greater than 0.4 since any value greater than this threshold would be considered to have moderate to very strong correlation (Table 4.1).Table 6.24 shows a performance matrix to this effect.

Table 6.23: Performance matrix of KHAT values

Classifier	PCA			Exhaustive			PBIL			CS_All				CS_PCA		Sum per Classifier
	DS1	DS2	DS3	DS1	DS2	DS3	DS1	DS2	DS3	DS1	DS2	DS3	DS1	DS2	DS3	
Linear	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	5
Polynomial	2	1	1	1	1	1	1	1	1	1	1	1	0	1	1	15
RBF	2	1	1	1	1	1	1	1	1	1	2	1	1	2	1	18
Sum per technique per dataset	4	2	3	2	2	3	2	2	3	2	3	3	1	3	3	
Sum per technique	9			7			7			8			7			

Table 6.24: Modification of Table 6.24

Classifier	PCA			Exhaustive			PBIL			CS_All				CS_PCA		Sum per Classifier
	DS1	DS2	DS3	DS1	DS2	DS3	DS1	DS2	DS3	DS1	DS2	DS3	DS1	DS2	DS3	
Linear	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	5
Polynomial	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	14
RBF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15
Sum per technique per dataset	2	2	3	2	2	3	2	2	3	2	2	3	1	2	3	
Sum per technique	7			7			7			7			6			

DS1 - East Rand dataset
DS2 - Mara Basin dataset
DS3 - Malmesbury dataset

Table 6.25: Performance matrix of binomial tests of significance

	PCA			Exhaustive			PBIL			CS_All				CS_PCA		Sum per
Classifier	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**Classifier**
Linear	1	0	1	0	0	1	0	0	1	0	1	1	0	0	1	7
Polynomial	2	2	1	1	2	1	1	2	1	1	0	1	0	2	1	18
RBF	2	1	1	1	0	1	1	0	1	1	1	1	2	2	1	16
Sum per technique per dataset	5	3	3	2	2	3	2	2	3	2	2	3	2	4	3	
Sum per technique	11			7			7			7			9			

Table 6.26: Modification of Table 6.26

	PCA			Exhaustive			PBIL			CS_All				CS_PCA		Sum per
Classifier	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**DS1**	**DS2**	**DS3**	**Classifier**
Linear	1	0	1	0	0	1	0	0	1	0	1	1	0	0	1	7
Polynomial	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	13
RBF	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	13
Sum per technique per dataset	3	2	3	2	1	3	2	1	3	2	2	3	1	2	3	
Sum per technique	8			6			6			7			6			

DS1 - East Rand dataset
DS2 - Mara Basin dataset
DS3 - Malmesbury dataset

The same approach was used to generate a performance matrix based on the binomial tests of significance extracted from Tables 6.9, 6.15 and 6.22. Table 6.25 depicts the result with Table 6.26 showcasing the modification. The modification was carried out to give any result with insignificant and significantly better levels equal weight since either significance level would be considered a satisfactory. The intention of developing this performance matrix was to find away of getting an overall assessment of the optimization techniques across the datasets. The performance matrix gives account of the relative successful of the classifiers and techniques. From the four performance matrices, it is evident that generally the linear SVM performed poorly in comparison to the polynomial and RBF SVM classifiers. The performance matrices also show that of the optimization techniques, the PCA had a slightly better score. This implies that across the datasets and in comparison to other techniques, the PCA was able to get more KHAT scores above the 0.4 threshold as well as levels of significance better than 'significantly worse'. In the next chapter, a discussion of these results is given detailing the importance of the findings.

Chapter 7

Discussion, Conclusions and Recommendations

7.1 Discussion and Conclusions

This research essentially focused on the potential of various data reduction techniques to optimize remote sensing data for SVM classification. From the results obtained, the success of the different optimization techniques was dependant on the classifiers used and the datasets under consideration. The general observation from these results show that linear SVMs generally performed poorly in comparison to the polynomial and RBF classifiers. This may be attributed to the fact that the scatter plot of remote sensing land cover classes is rarely linearly separable. This renders the application of linear SVMs inadequate for remote sensing tasks. The effectiveness of polynomial and RBF SVM classifiers for remote sensing tasks is based on the premise that they have the ability to project the initially nonlinearly separable feature space to a higher (or infinite) dimension space in which the classes are linearly separable. Classification in this very high dimension invariably results in high classification accuracies. As portrayed in the results, even with reducing the quantity of data needed to determine the SVM parameters, polynomial and RBF SVM classifiers were most of the time versatile enough to post satisfactorily classification accuracy results.

It was also observed that even with linear SVMs yielding poor classification results, accuracies further deteriorated with application of different optimization techniques. Reducing the quantity of data required to determine the SVM parameters clearly rendered the training data unstable for the linear SVM. The exception to this general observation was the Malmesbury dataset where the linear SVM performed comparably well with the polynomial and RBF SVM classifiers. This may be attributed to the fact that it had fewer classes as compared to the East Rand and Mara basin study areas which had ten and six classes respectively. From this it can be concluded that as long as the classes are potentially linearly separable, the linear SVM can perform as well as the polynomial and RBF SVM classifiers.

The two categorizations of optimization techniques investigated were feature and instance reduction as described in Chapter Three. Feature reduction was further categorized into unsupervised and supervised feature reduction. Principal Component Analysis (PCA) was used as an example of unsupervised feature reduction. The first research question was *"Can the PCA approach be used to reduce the number of bands needed to classify an image while maintaining the SVM classification accuracy?"* The results showed that the optimization of the original bands through PCA transformation generally resulted in satisfactory results, and it can conclusively be stated that it can be used to optimize remote sensing data for SVM classification.

Supervised feature reduction involved experimentation with two new feature selection techniques namely: exhaustive search and Population Based Incremental Learning (PBIL). These methods were the subject of the second research question: *"Can the exhaustive and PBIL search methods using Thornton's separability index as a criterion function be used to reduce the number of bands needed to classify an image while maintaining the SVM classification accuracy?"* In general, the visual appreciation of the exhaustive search classification was poorer than PBIL classification, as can be seen in the East Rand dataset (Compare Figure 6.7 with Figure 6.8). This may be attributed to the fact that the exhaustive search usually resulted in the bare minimum of bands for classification. This seems to have resulted in unstable band pairings hence the poor results. Conversely, the SVM classification results of PBIL were better than exhaustive search probably due to the fact that more bands were used thus ensuring training data stability. Of the two optimization techniques therefore, PBIL conclusively proved to be more effective. The Malmesbury dataset however again provided an exception to this general observation in the sense that exhaustive search and PBIL classifications yielded comparable results. The reason for this could be attributed to the fact that the number of bands selected following exhaustive search in this dataset were comparable to the bands selected using PBIL, hence ensuring an equally stable training dataset. Dean and Hoffer (1983) postulate that the number of features selected through any feature selection process should not be less than the intrinsic dimensionality. In the East Rand and Mara basin datasets, the number of features (bands) selected through the exhaustive search process were mostly less than the intrinsic dimensionality which in this research was taken as 3 (See Tables 6.1 and 6.2). The Malmesbury dataset on the

other hand had all features selected through exhaustive search being more than the intrinsic dimensionality. In light of Dean and Hoffer (1983)'s postulation, it is therefore understandable that the exhaustive search results for the Malmesbury study area were better than the corresponding results in the earlier datasets. Even as the results conclusively show that PBIL presents a viable means of optimizing data for SVM classification in terms of computational time, exhaustive search can be made competitive by restricting the search to band combinations equal to or greater than the intrinsic dimensionality.

On trial as well in this research question was Thornton's separability index which was used as an evaluation (criterion) function for the exhaustive and PBIL search methods. As mentioned in section 3.2.2.3, the evaluation function should be related to the behavior of the error made by the classifier. How well Thornton's separability index achieved this is evidenced by the results of the performance matrices in Tables 6.23, 6.24, 6.25 and 6.26. The fact that both exhaustive and PBIL search method had the same performance matrix scores, brings to mind how well this evaluation function related to the behavior of the error made by the SVM classifier. Questions would undoubtedly have been asked of this separability index if the results of the exhaustive and PBIL search methods yielded different significant levels. The point being made is that if with different band combinations resulting from different search methods, similar ranges of classification accuracy are obtained, then the criteria function with which the bands are selected models the classifier error well. From this research it can be concluded that the Thornton's separability index is an appropriate evaluation function for feature selection since it relates well to the behavior of the error made by the SVMs.

The third research question investigated was *"Can the proposed instance reduction method be used to reduce the quantity of data needed to train an SVM classifier while maintaining the SVM classification accuracy?"* Instance reduction was experimented on the original bands and the results of the PCA transformation. The visual appreciation of the instance reduction of the PCA transformation bands was generally poor, across classifiers, as compared to the instance reduction of the original bands. Actually in most cases, visual appreciation of the instance reduction of the PCA bands more often than not yielded the poorest results of all the optimization techniques. This

may be attributed to the instability of the training data due first to the reduction of dimensions through PCA transformation and secondly to the reduction of data through instance reduction. From a visual perspective therefore, it can be concluded that the instance reduction of the PCA bands has proved to be an inappropriate optimization technique.

Instance reduction of the original bands has in general yielded satisfactory results. The reduction of the quantity of training data resulted in a significantly faster means of determining the SVM parameters. It also had the advantage of increasing the search space for cross validation as opposed to the unreduced dataset for which the search space had to be kept small given the voluminous nature of remote sensing data. The ability for this data reduction technique to perform well in the face of small training data renders this technique an efficient optimization method. The other advantage associated with the instance reduction technique was that it reduced the effect of unbalanced training data which was listed as one of the weaknesses of the one-against-all SVM classification technique adopted in this research. This was made possible by using prototypes, which are representative of the training data, to derive the SVM parameters. The prototypes were significantly less unbalanced as compared to the original datasets. In conclusion therefore, instance reduction is an appropriate optimization technique for SVM classification only if the original bands are under consideration.

7.2 Recommendations

In view of the fact that the classification accuracy of linear SVMs improved as the number of bands increased, one could speculate that the linear SVM classifier would do well with hyperspectral data since training data ideally becomes more linearly separable as dimensions increase. This supposition however discounts the effect of inter-band correlation and Hughes phenomenon discussed earlier. Most of the studies regarding SVM classification of hyperspectral data have focused on the application of RBF SVM classifiers. From the observation of these results, it is recommended that the classification of hyperspectral data by linear SVMs be studied since from a theoretical perspective they should do well. If linear SVMs can be seen to perform well with hyperspectral data then appropriate optimization methods can be investigated.

Other recommended areas for further research originate from the fields of machine vision and soft computing and include combining SVM classifiers, feature selection in the projected infinite dimension and support vector clustering. One of the key factors that determine the effectiveness of the exhaustive and PBIL search methods is the evaluation (criterion) function. Whereas Thornton's separability index was successfully deployed in this research, further studies could explore the effectiveness of other known separability indices and the generation of new separability indices. The other recommendation is that with the advent of new classification techniques at the disposal of the remote sensing community, there is an increased need to have standardized databases of easily accessible data that can be used to evaluate these techniques as is the case in the machine vision field. Having these standardized databases would increase the objectivity with which new classification methods are evaluated.

The fact that less training data can be used when employing SVM classifiers also brings to mind the need for a new body of knowledge regarding training data for nonparametric classifiers or more specifically for SVMs. For parametric classifiers there is a whole body of knowledge regarding the quantity of training data needed in terms of the number and size of samples. This data is needed to accurately model the individual class distributions so as to determine the decision boundaries and rules. A reduced dataset as derived through instance reduction or inadequate sampling, would potentially result in poor classification results. From this research (and other related literature) it has been demonstrated that accurate results can be obtained from a small sample hence the proposition for a new outlook about how to collect training data for SVM classifications. As previously mentioned in related research (See section 2.3.2.6), Foody and Mathur (2004b) propose collecting training data corresponding to the support vectors. They postulate that this approach would be useful in case a given area needed to be reclassified. Instead of sampling the whole study area, the ground positions corresponding to the support vectors may be sampled and used to derive SVM parameters for classification of the study area. Foody and Mathur (2004b) and this research both point to a need to investigate how best to determine ground positions corresponding to the margin boundaries *a priori*.

Bibliography

Baluja, S. 1994. ***Population Based Incremental Learning – A Method for integrating Genetic Search Based Function Optimization and Competitive Learning.*** Technical Report No. CMU-CS-94-163. (Pittsburgh, Pennsylvania: Carnigie Mellon University).

Benediktsson, J. A., Swain, P. H., and Ersoy, O. K. 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. ***I.E.E.E.Transactions on Geoscience and Remote Sensing*, 28,** 540 - 552.

Bennet K., and Campbell, C. 2002. Support Vector Machines: Hype or Hallelujah? ***Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Explorations*, 2,** 1 - 13.

Borak, J. S., and Strahler, A. H. 1999. Feature selection and land cover classification of a MODIS-like dataset for a semiarid environment. ***International Journal of Remote Sensing*, 20,** 919 – 938.

Bruzzone, L., and Serpico, S. B. 2000. A technique for feature selection in multiclass problems. ***International Journal of Remote Sensing*, 21,** 549–563.

Campbell, C. 2000. ***An Introduction to kernel Methods, Radial Basis Function Networks: Design and Applications***. (Berlin: Springer Verlag).

Campbell, J. B. 1996. ***Introduction to Remote Sensing*** (2nd Edition). (London and New York: Taylor and Francis).

Chevrel, S., Courant, C., Cottard, F., Coetzee, H., Bourguignon A., and Ntsume G. 2003. ***Very High Resolution Remote Sensing coupled to GIS-based Environmental Assessment – East Rand Goldfield, South Africa.*** Report BRGM/RP-52724-FR.

Christianini, N., and Shawe-Taylor, J. 2000. ***An introduction to support vector machines: and other kernel-based learning methods***. (Cambridge and New York: Cambridge University Press).

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

Cohen, J. 1960. A coefficient of agreement for nominal scales. ***Educational and Psychological Measurement,* 20,** 37 – 46.

Congalton, R. G. 2001. Accuracy assessment and validation of remotely sensed and other spatial information. ***International Journal of Wildland Fire,* 10,** 321 - 328.

Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. ***Remote Sensing of Environment,* 37,** 35 – 46.

Congalton, R. G. 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. ***Photogrammetric Engineering and Remote Sensing*, 54,** 593 – 600.

Congalton, R. G., and Green, K. 1998. ***Assessing the Accuracy of Remotely Sensed Data: Principles and Practices.*** (Boca Raton, Florida: Lewis Publishers).

Congalton, R. G., Oderwald, R. G., and Mead, R. A. 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. ***Photogrammetric Engineering and Remote Sensing,*** **49,** 1671 – 1678.

Courant, R., and Hilbert, D. 1953. ***Methods of Mathematical Physics*** (volume 1). (New York: Interscience Publishers Inc.).

Curran, P. 1985. ***Principles of Remote Sensing***. (Harlow, Essex: Longman Scientific & Technical).

Dean, M. E., and Hoffer, R. M. 1983. Feature Selection Methodologies Using Simulated Thematic Mapper Data. In ***Proceedings of the 9th International Symposium of Machine Processing of Remotely Sensed Data***. Pp: 347 – 356, Purdue University, West Lafayette, Indiana. 21st – 23rd June 1983.

Dutra, L., and Huber, R. 1999. Feature Extraction and selection for ERS-1/2 InSAR Classification. ***International Journal of Remote Sensing,*** **20,** 993 – 1016.

Eastman, J. 2003. ***IDRISI Kilimanjaro: Guide to GIS and Image Processing***. IDRISI Kilimanjaro User Manual. (Worcester, Massachusetts: Clark labs)

Elachi, C. 1987. ***Introduction to the Physics and Techniques of Remote Sensing***. (New York: John Wiley and Sons).

Fangming, Z., and Guan, S. 2004. Feature selection for modular GA-based classification. ***Applied Soft Computing,*** **4,** 381 – 393.

Fitzgerald, R. W., and Lees, B. W. 1994. Assessing the classification accuracy of multiresource remote sensing data. ***Remote Sensing of Environment,*** **47,** 362 – 368.

Foody, M. G. 2004. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. ***Photogrammetric Engineering and Remote Sensing,*** **70,** 627 – 633.

Foody, M. G. 2002a. Hard and soft classifications by a neural network with a non-exhaustively defined set of classes. ***International Journal of Remote Sensing,*** **23,** 3853 – 3864.

Foody, M. G. 2002b. Status of land cover classification accuracy assessment. ***Remote Sensing of Environment,*** **80,** 185 – 201.

Foody, M. G. 1998. Sharpening fuzzy classification output to refine the representation of sub-pixel land cover distribution. ***International Journal of Remote Sensing,*** **19,** 2593 – 2599.

Foody, M. G. 1992. On the compensation for chance agreement in image classification accuracy assessment. ***Photogrammetric Engineering and Remote Sensing,*** **58,** 1459 – 1460.

Foody, M. G., and Mathur, A. 2004a. A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. ***IEEE Transactions on Geoscience and Remote Sensing,*** **42,** 1335 – 1343.

Foody, M. G., and Mathur, A. 2004b. Toward Intelligent Training of Supervised Image Classifications: Directing Training Data Acquisition for SVM Classification. ***Remote Sensing of Environment*, 93,** 107 - 117.

Foody, G. M., Campbell, N. A., Trodd, N. M., and Wood, T. F. 1992. Derivation and applications of probabilistic measures of class membership from the maximum likelihood classification. ***Photogrammetric Engineering and Remote Sensing*, 58,** 1335–1341.

Fukunaga, K. 1990. ***Introduction to Statistical Pattern Recognition*** (2nd Edition). (New York: Academic Press).

Fung, T., and Ledrew, E. 1988. The determination of optimal threshold levels for change detection using various accuracy indices. ***Photogrammetric Engineering and Remote Sensing*, 54,** 1449 – 1460.

Gibson, P. 2000. ***Introductory Remote Sensing – Principles and concepts***. (London and New York: Taylor and Francis).

Gidudu, A., and Ruther, H. 2005. Land Cover Mapping : Optimizing SVM Classification through data reduction. In ***Proceedings of the 9th International Symposium of Physical Measurements and Signatures in Remote Sensing (ISPMSRS05),*** Beijing, China, 17th – 19th October 2005.

Gidudu, A., and Ruther, H. 2005. Land Cover Mapping : Optimizing SVM Classification through feature selection. In ***Proceedings of the 1st International Conference on Advanced Remote Sensing for Earth Observation; Systems, Techniques, and Applications,*** Riyadh, Saudi Arabia. 7th – 11th May 2005.

Gidudu, A., and Ruther H. 2004. Land Cover Mapping: Exploring Support vector Machines. In ***Proceedings of the 15th Annual Symposium of the Pattern Recognition Association of South Africa***, Cape Town, South Africa. 27th – 29th November 2004.

Gilad-Bachrach, R., Navot, A., and Tishby, N. 2004. Margin based Feature Selection – Theory and Algorithms. In ***Proceedings of the 21st International Conference on Machine Learning,*** Banff, Alberta, Canada. 4th – 8th July 2004.

Global Water for Sustainability Program (Glows) 2005. ***Mara Basin: Overview and Issues***. http://glows.fiu.edu/Locations/Africa/OverviewIssues/tabid/331/Default.aspx (Accessed: 15th November 2005).

Gosling, T., Nanlin, J., and Tsang, E. 2004. ***Population Based Incremental Learning Versus Genetic Algorithms: Iterated Prisoners Dilemma***. Technical Report No. CSM-401. (Essex: University of Essex, England).

Greene, J. 2001. Feature subset selection using Thornton's separability index and its applicability to a number of sparse proximity-based classifiers. In ***Proceedings of the 12th Annual Symposium of the Pattern Recognition Association of South Africa.*** November 2001.

Gualtieri, J. A., and Cromp, R. F. 1998. Support vector machines for hyperspectral remote sensing classification. In ***Proceedings of the 27th AIPR Workshop: Advances in Computer Assisted Recognition,*** Washington, DC, Oct.14th -16th October, 1998 (Washington, DC: SPIE), pp. 221–232.

Guobin, Z., and Blumberg, D. G. 2002. Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. ***Remote Sensing of Environment,* 80,** 233– 240.

Gunn, S. 1997. ***Support Vector Machines for Classification and Regression.*** ISIS Technical Report (Southampton: University of Southampton).

Hepner, G. F., Logan, T., Ritter, N., and Bryant, N. 1990. Artificial neural network Classification using a minimal training set; comparison to conventional supervised classification. ***Photogrammetric Engineering and Remote Sensing*, 56,** 469 - 473.

Holland, J. H. 1992. Genetic algorithm. ***Scientific American*, 80,** 66 – 72.

Huang, C., Davis, L. S., and Townshed, J. R. G. 2002. An assessment of support vector machines for land cover classification. ***International Journal of Remote Sensing*, 23,** 725–749.

Hughes, G. F. 1968. On the mean accuracy of statistical pattern recognizers. ***IEEE Transactions on Information Theory*, 14,** 55–63.

Jin, Y. G., and Wang, Y. 2001. A genetic algorithm to simultaneously retrieve land surface roughness and soil wetness. ***International Journal of Remote Sensing*, 22,** 3093–3099.

Joachims, T. 1998. Text categorization with support vector machines—learning with many relevant features. In ***Proceedings of the 10th European Conference on Machine Learning, Chemnitz, Germany.*** (Berlin: Springer), pp. 137–142.

Kailath, T. 1967. The divergence and the Bhattacharyya distance measures in signal selection. ***IEEE Transactions on Communication Technology*, 15,** 52–60.

Kavzoglu, T. and Mather, P. M. 2002. The role of feature selection in artificial neural network applications. ***International Journal of Remote Sensing*, 15,** 2919–2937.

Keuchela, J., Naumanna, S., Heilera, M., and Siegmund, A., 2003. Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data. ***Remote Sensing of Environment,* 86,** 530–541.

Kohavi, R., and John, G. 1997. Wrapper for feature subset selection. ***Artificial Intelligence*, 97,** 273 – 324.

Kramer J. H., 2002. ***Observation of the earth and its environment: Survey of missions and sensors*** (4th Edition). (Berlin: Springer).

Landgrebe, D. 2005. Multispectral Land Sensing: Where From, Where to? ***IEEE Transactions on Geoscience and Remote Sensing,* 43,** 433 - 440.

Landgrebe, D. 1997. The evolution of Landsat Data Analysis. ***Photogrammetric Engineering and Remote Sensing*, 63,** 859 – 867.

Landis, J., and Kotch, G. 1977. The measurement of observer agreement for categorical data. ***Biometrics*, 33,** 159 – 174.

Lillesand, T., and Kiefer, R. 2000. ***Remote Sensing and Image Interpretation*** (4th Edition). (New York: John Wiley and Son).

Luo, C. J., Zheng, J., Leung Y., and Zhou H. 2003. A knowledge-integrated stepwise optimization model for feature mining in remotely sensed images. ***International Journal of Remote Sensing*, 24,** 4661 – 4680.

Ma, Z., and Redmond, R. L. 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. ***Photogrammetric Engineering and Remote Sensing*, 61,** 435 – 439.

Mahesh P., and Mather, P. M. 2003. Support Vector classifiers for Land Cover Classification. In ***Proceedings of the 6th Annual International Conference, Map India 2003, New*** Delhi, India. 28th – 31st January 2003.

Maingi, J. K, Marsh, S. E, Kepner, W. G., and Edmond, C. M. 2002. ***An Accuracy Assessment of 1992 Landsat-MSS Derived Land Cover for the Upper San Pedro Watershed (U.S./Mexico)***. U.S. Environmental Protection Agency.

Mather, P. 1987. ***Computer Processing of Remotely Sensed Images – An Introduction***. (New York: John Wiley and sons).

Mati, B., Mtalo, F., and Mtalo, G., E. 2004. Land Use Change and its impacts on Hydrological Determinants of Wetlands in the Mara Basin. In ***Proceedings of the Waternet – WARFASA Symposium***, Windhoek, Namibia. November 2004

Mertens, K. C., Verbeke, L. P. C., Ducheyne, E. I., and De Wulf, R. R. 2003. Using genetic algorithms in sub-pixel mapping. ***International Journal of Remote Sensing*, 24,** 4241 – 4247.

Muasher, M. J., and Landgrebe, D. A. 1982. A binary tree feature selection technique for limited training sample size. In ***Proceedings of the 8th International Symposium of Machine Processing of Remotely Sensed Data***. Purdue University, West Lafayette, Indiana. 7th – 9th July 1982. Pp. 130 – 137

Pal, S. K., Bandyopadhyay S., and Marthy, C. A. 2001. Genetic classifiers for remotely sensed images: Comparison with standard methods. ***International Journal of Remote Sensing*, 22,** 2545 – 2569.

Paola, J. D., and Schowengerdt, R. A., 1995. A detailed comparison of back propagation neural network and maximum likelihood classification for urban land use classification. ***I.E.E.E. Transactions on Geoscience and Remote Sensing*, 33,** 981 - 996.

Peddle, D. R., Foody, G. M., Zhang, A., Franklin, S. E., and LeDrew, E. F. 1994. Multisource image classification II: An empirical comparison of evidential reasoning and neural network approaches. ***Canadian Journal of Remote Sensing*, 20,** 397 - 408.

Pudil, P., and Somol, P. 2005. An Overview of Feature Selection Techniques in Statistical Pattern Recognition. In ***Proceedings of the 16th Annual Symposium of the Pattern Recognition Association of South Africa*** Nov 2005.

Richards, J. A. 1994. ***Remote Sensing Digital Image Analysis. An Introduction*** (New York: Springer).

Richards, J. A., and Jia, X. 1999. ***Remote sensing digital image analysis: An introduction*** (3rd Edition). (Berlin: Springer).

Rosenfield, G.H, and Fitzpatrick-Lins, K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. ***Photogrammetric Engineering and Remote Sensing***, **52,** 223 – 227.

Rossiter, D. G. 2004. ***Technical Note: Statistical methods for accuracy assessment of classified thematic maps.*** Technical Report. (Enschede, Netherlands: ITC). April 2004.

Sabins, F., Jr. 1978. ***Remote Sensing Principles and Interpretation.*** (New York: W. H. Freeman and Co).

Schott, J. 1997. ***Remote Sensing: The Image Chain Approach.*** (New York: Oxford University Press).

Schowengerdt, R. 1997. ***Remote Sensing: Models and Methods for Image Analysis*** (2nd Edition). (San Diego: Academic Press).

Schowengerdt, R. 1983. ***Techniques for Image Processing and Classification in Remote Sensing***. (Orlando, Florida: Academic Press).

Skalak, D. B. 1994. Prototype and feature Selection by Sampling and Random Mutation Hill Climb. In ***Proceedings of the 11th International Conference on Machine Learning.*** New Brunswick, New Jersey, Canada. Pp. 293 – 301

Smits, P. C, Dellepiane, S. G., and Schowegerdt, R. A. 1999. Quality assessment of image classification algorithms for land cover mapping: a review and a proposal for a cost-benefit approach. ***International Journal of Remote Sensing***, **8,** 1461 – 1486.

Stehman, S. V. 1997. Estimating Standard Errors of Accuracy Assessment Statistics under Cluster Sampling. ***Remote Sensing of Environment,*** **60,** 258 – 269.

Stehman, S. V., and Czaplewski, R. L. 1998. Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. ***Remote Sensing of Environment,*** **64,** 331 – 344.

Story, M., and Congalton, R. 1986. Accuracy assessment: A user's perspective. ***Photogrammetric Engineering and Remote Sensing***, **52,** 397 – 399.

Swain, P. H. 1978. Fundamentals of pattern recognition in remote sensing. In ***Remote Sensing: The Quantitative Approach*** edited by P. H. Swain and S. M. Davis (New York: McGraw-Hill), Chapter Three.

Swain, P.H., and Davis, S. M. 1978. ***Remote Sensing: The Quantitative Approach.*** (New York : McGraw-Hill).

Swain, P. H., and King, R. C. 1973. Two effective feature selection criteria for Multispectral Remote Sensing. In ***Proceedings of the 1st International Joint Conference on Pattern Recognition***. Washington D.C. November 1973.

Tambouratzis, T. 2002. Counter-clustering for training pattern selection. ***Computer Journal,*** **43,** 177 – 190.

Thomas, I. L., and Allcock, G. M. 1984. Determining the confidence level for a classification. ***Photogrammetric Engineering and Remote Sensing,*** **50,** 1491 – 1496.

Thomas, I. L., Ching, N. P., Benning, V. M., and D'Aguanno, J. A. 1987. A review of multi-channel indices of class separability. ***International Journal of Remote Sensing*, 8,** 331–350.

Thornton, C. 2000. ***Truth from Trash, How learning makes sense***. (Massachusetts: MIT Press).

Tso, B., and Mather, P. 2001. ***Classification Methods for Remotely Sensed Data***. (London and New York: Taylor and Francis).

Vapnik, V. N. 1995. ***The Nature of Statistical Learning Theory***. (New York: Springer-Verlag).

Verbyla, D. L. 2002. ***Practical GIS analysis***. (London and New York: Taylor and Francis).

Wacker, A. G., and Langrebe, D. 1972. Minimum Distance Classification in Remote Sensing. In ***Proceedings of the 1st Canadian Symposium for Remote Sensing February***. 7th – 9thFeb 1972.

Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T. and Vapnik, V. 2000. Feature selection for SVMs. In ***Proceedings of the 15th Conference on Neural Information Processing Systems (NIPS)***. 27th – 30th November 2000. pp. 668 - 674.

Wolfe, P. 1961. A duality theorem for nonlinear programming, ***Quarterly of Applied Mathematics,* 19,** 239 – 244.

Wu, D., and Linders, J. 2000. Comparison of three different methods to select features for discriminating forest cover types using SAR imagery. ***International Journal of Remote Sensing,* 21,** 2089 – 2099.

Yool, S. R. 1998. Land cover classification in rugged areas using simulated moderate resolution remote sensor data and an artificial neural network. ***International Journal of Remote Sensing,* 19,** 85 – 96.

Zhang J., and Foody, G. M. 2001. Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. ***International Journal of Remote Sensing,* 22,** 615 – 628.

Appendix 1
East Rand

A1.1 All Nine Bands

A1.1.1 linear

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	3	7	3	0	2	0	7	65	36	123		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**46**	0	0	0	0	0	0	0	0	46	100.00	0.0000
3	0	0	**25**	0	0	0	0	0	0	0	25	100.00	0.0000
4	0	0	1	**14**	0	0	0	0	0	0	15	93.33	0.0667
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	0	0	0	0	**25**	0	0	0	0	25	100.00	0.0000
7	0	0	0	1	0	0	**12**	0	0	0	13	92.31	0.0769
8	0	29	0	0	0	0	0	**0**	0	0	29	0.00	1.0000
9	0	0	0	0	0	0	0	0	**0**	0	0	-	-
10	0	0	0	0	0	0	0	25	0	**7**	32	21.88	0.7813
11	0	1	0	2	0	0	0	0	0	0	3		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	58.23	75.76	70.00	100.00	92.59	100.00	0.00	0.00	16.28		**50.00**	
E.O	0.0000	0.4177	0.2424	0.3000	0.0000	0.0741	0.0000	1.0000	1.0000	0.8372			

Legend

0. Unclassified pixels
1. Agriculture
2. Soil & Trees
3. Tailings
4. Tailings Dam
5. Water
6. Wetland
7. Yellow Tailings
8. Trees
9. Roads
10. Soils
11. Mixed Pixels

50.00 **Overall Accuracy**

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.1.2 Polynomial

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	3	3	0	0	0	0	13	0	8	27		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**66**	0	0	0	0	0	0	0	0	66	100.00	0.0000
3	0	0	**28**	0	0	0	0	0	0	0	28	100.00	0.0000
4	0	0	0	**10**	0	0	0	0	0	0	10	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	0	0	0	0	**22**	0	6	0	0	28	78.57	0.2143
7	0	0	0	0	0	0	**9**	0	0	0	9	100.00	0.0000
8	0	0	0	0	0	0	0	**9**	11	24	44	20.45	0.7955
9	0	0	0	3	0	0	0	0	**16**	2	21	76.19	1.0000
10	0	0	0	0	0	0	0	1	0	**5**	6	83.33	0.1667
11	0	10	2	7	0	5	3	3	38	4	72		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	83.54	84.85	50.00	100.00	81.48	75.00	28.13	24.62	11.63		**59.89**	
E.O	0.0000	0.1646	0.1515	0.5000	0.0000	0.1852	0.2500	0.7188	0.7538	0.8837			

A1.1.3 Radial Basis Function

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	1	0	9	0	0	4	7	45	4	70		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**78**	0	0	0	0	0	0	0	0	78	100.00	0.0000
3	0	0	**33**	0	0	0	0	0	0	0	33	100.00	0.0000
4	0	0	0	**11**	0	0	0	0	0	0	11	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	0	0	0	0	**26**	0	0	0	0	26	100.00	0.0000
7	0	0	0	0	0	0	**8**	0	0	0	8	100.00	0.0000
8	0	0	0	0	0	0	0	**8**	9	22	39	20.51	0.7949
9	0	0	0	0	0	0	0	0	**11**	6	17	64.71	1.0000
10	0	0	0	0	0	0	0	2	0	**9**	11	81.82	0.1818
11	0	0	0	0	0	1	0	6	0	2	9		
Total	19	79	33	20	34	27	12	23	65	43			
P.A%	100.00	98.73	100.00	55.00	100.00	96.30	66.67	34.78	16.92	20.93		**65.11**	
E.O	0.0000	0.0127	0.0000	0.4500	0.0000	0.0370	0.3333	0.6522	0.8308	0.7907			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.2 Principal Component Analysis

A1.2.1 Linear

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	2	2	0	0	1	0	30	65	43	143		1.0000
1	**19**	0	0	0	0	0	0	2	0	0	21	90.48	0.0952
2	0	**41**	0	0	0	0	0	0	0	0	41	100.00	0.0000
3	0	0	**27**	0	0	0	0	0	0	0	27	100.00	0.0000
4	0	0	2	**10**	0	0	0	0	0	0	12	83.33	0.1667
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	1	2	0	0	**20**	0	0	0	0	23	86.96	0.1304
7	0	0	0	10	0	0	**7**	0	0	0	17	41.18	0.5882
8	0	35	0	0	0	6	0	**0**	0	0	41	0.00	1.0000
9	0	0	0	0	0	0	0	0	**0**	0	0	-	-
10	0	0	0	0	0	0	0	0	0	**0**	0	-	-
11	0	0	0	0	0	0	5	0	0	0	5		1.0000
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	51.90	81.82	50.00	100.00	74.07	58.33	0.00	0.00	0.00		**43.41**	
E.O	0.0000	0.4810	0.1818	5.0000	0.0000	0.2593	0.4167	1.0000	1.0000	1.0000			

A1.2.2 Polynomial

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	0	0	0	0	1	0	2	1	0	4		
1	**19**	0	1	0	0	0	0	0	0	0	20	95.00	0.0500
2	0	**50**	0	0	0	0	0	0	0	0	50	100.00	0.0000
3	0	0	**32**	0	0	0	0	0	0	0	32	100.00	0.0000
4	0	0	0	**20**	0	0	0	0	0	0	20	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	1	0	0	0	**24**	0	0	0	0	25	96.00	0.0400
7	0	0	0	0	0	0	**12**	0	0	0	12	100.00	0.0000
8	0	28	0	0	0	0	0	**24**	0	0	52	46.15	0.5385
9	0	0	0	0	0	0	0	0	**64**	0	64	100.00	0.0000
10	0	0	0	0	0	0	0	1	0	**42**	43	97.67	0.0233
11	0	0	0	0	0	2	0	5	0	1	8		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	63.29	96.97	100.00	100.00	88.89	100.00	75.00	98.46	97.67		**88.19**	
E.O	0.0000	0.3671	0.0303	0.0000	0.0000	0.1111	0.0000	0.2500	0.0154	0.0233			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.2.3 Radial Basis Function

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	0	1	0	0	0	0	3	1	7	12		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**73**	0	0	0	0	0	0	0	0	73	100.00	0.0000
3	0	0	**32**	0	0	0	0	0	0	0	32	100.00	0.0000
4	0	0	0	**20**	0	0	0	0	0	0	20	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	4	0	0	0	**27**	0	0	0	0	31	87.10	0.1290
7	0	0	0	0	0	0	**12**	0	0	0	12	100.00	0.0000
8	0	2	0	0	0	0	0	**28**	17	2	49	57.14	0.4286
9	0	0	0	0	0	0	0	0	**35**	6	41	85.37	0.1463
10	0	0	0	0	0	0	0	0	0	**28**	28	100.00	0.0000
11	0	0	0	0	0	0	0	1	12	0	13		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	92.41	96.97	100.00	100.00	100.00	100.00	87.50	53.85	65.12		**84.62**	
E.O	0.0000	0.0759	0.0303	0.0000	0.0000	0.0000	0.0000	0.1250	0.4615	0.3488			

A1.3 Exhaustive Search

A1.3.1 Linear

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	79	30	3	0	27	12	32	65	43	291		
1	**19**	0	3	0	0	0	0	0	0	0	22	86.36	0.1364
2	0	**0**	0	0	0	0	0	0	0	0	0	-	-
3	0	0	**0**	0	0	0	0	0	0	0	0	-	-
4	0	0	0	**17**	0	0	0	0	0	0	17	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	0	0	0	0	**0**	0	0	0	0	0	-	-
7	0	0	0	0	0	0	**0**	0	0	0	0	-	-
8	0	0	0	0	0	0	0	**0**	0	0	0	-	-
9	0	0	0	0	0	0	0	0	**0**	0	0	-	-
10	0	0	0	0	0	0	0	0	0	**0**	0	-	-
11	0	0	0	0	0	0	0	0	0	0	0		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	0.00	0.00	85.00	100.00	0.00	0.00	0.00	0.00	0.00		**19.23**	
E.O	0.0000	1.0000	1.0000	0.1500	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.3.2 Polynomial

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	4	0	0	3	0	0	5	0	23	35		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**70**	0	0	0	0	0	0	0	0	70	100.00	0.0000
3	0	0	**33**	0	0	0	0	0	0	0	33	100.00	0.0000
4	0	0	0	**16**	0	0	0	0	0	0	16	100.00	0.0000
5	0	0	0	0	**31**	0	0	0	0	0	31	100.00	0.0000
6	0	3	0	0	0	**23**	0	1	0	0	27	85.19	0.1481
7	0	0	0	0	0	0	**6**	0	0	0	6	100.00	0.0000
8	0	0	0	0	0	0	0	**3**	0	4	7	42.86	0.5714
9	0	0	0	0	0	0	0	0	**32**	0	32	100.00	1.0000
10	0	0	0	0	0	0	0	0	0	**6**	6	100.00	0.0000
11	0	2	0	4	0	4	6	23	33	10	82		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	88.61	100.00	80.00	91.18	85.19	50.00	9.38	49.23	13.95		**65.66**	
E.O	0.0000	0.1139	0.0000	0.2000	0.0882	0.1481	0.5000	0.9063	0.5077	0.8605			

A1.3.3 Radial Basis Function

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	2	0	0	3	0	0	2	1	2	10		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**67**	0	0	0	0	0	0	0	0	67	100.00	0.0000
3	0	0	**33**	0	0	0	0	0	0	0	33	100.00	0.0000
4	0	0	0	**16**	0	0	0	0	0	0	16	100.00	0.0000
5	0	0	0	0	**31**	0	0	0	0	0	31	100.00	0.0000
6	0	3	0	0	0	**13**	0	4	0	0	20	65.00	0.3500
7	0	0	0	0	0	0	**12**	0	0	0	12	100.00	0.0000
8	0	0	0	0	0	0	0	**5**	11	14	30	16.67	0.8333
9	0	0	0	0	0	0	0	0	**6**	6	12	50.00	1.0000
10	0	0	0	0	0	0	0	1	0	**20**	21	95.24	0.0476
11	0	7	0	4	0	14	0	20	47	1	93		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	84.81	100.00	80.00	91.18	48.15	100.00	15.63	9.23	46.51		**60.99**	
E.O	0.0000	0.1519	0.0000	0.2000	0.0882	0.5185	0.0000	0.8438	0.9077	0.5349			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.4 PBIL

A1.4.1 Linear

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	19	14	4	0	16	6	15	65	36	175		
1	**17**	0	0	0	0	0	0	0	0	0	17	100.00	0.0000
2	0	**37**	0	0	0	0	0	0	0	0	37	100.00	0.0000
3	0	0	**18**	0	0	0	0	0	0	0	18	100.00	0.0000
4	0	0	0	**16**	0	0	0	0	0	0	16	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	8	0	0	0	**11**	0	7	0	0	26	42.31	0.5769
7	0	0	0	0	0	0	**6**	0	0	0	6	100.00	0.0000
8	0	0	0	0	0	0	0	**0**	0	0	0	-	-
9	0	0	0	0	0	0	0	0	**0**	0	0	-	-
10	0	0	0	0	0	0	0	4	0	**7**	11	63.64	0.3636
11	2	15	1	0	0	0	0	6	0	0	24		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	89.47	46.84	54.55	80.00	100.00	40.74	50.00	0.00	0.00	16.28		**40.11**	
E.O	0.1053	0.5316	0.4545	0.2000	0.0000	0.5926	0.5000	1.0000	1.0000	0.8372			

A1.4.2 Polynomial

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	1	1	0	0	0	0	4	25	17	48		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**66**	0	0	0	0	0	0	0	0	66	100.00	0.0000
3	0	0	**32**	0	0	0	0	0	0	0	32	100.00	0.0000
4	0	0	0	**15**	0	0	0	0	0	0	15	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	2	0	0	0	**27**	0	8	0	0	37	72.97	0.2703
7	0	0	0	0	0	0	**8**	0	0	0	8	100.00	0.0000
8	0	0	0	0	0	0	0	**0**	19	0	19	0.00	1.0000
9	0	0	0	3	0	0	0	0	**18**	6	27	66.67	1.0000
10	0	0	0	0	0	0	0	12	1	**20**	33	60.61	0.3939
11	0	10	0	2	0	0	4	8	2	0	26		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	83.54	96.97	75.00	100.00	100.00	66.67	0.00	27.69	46.51		**65.66**	
E.O	0.0000	0.1646	0.0303	0.2500	0.0000	0.0000	0.3333	1.0000	0.7231	0.5349			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.4.3 Radial Basis Function

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	1	0	4	0	0	0	2	45	3	55		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**78**	0	0	0	0	0	0	0	0	78	100.00	0.0000
3	0	0	**32**	0	0	0	0	0	0	0	32	100.00	0.0000
4	0	0	0	**16**	0	0	0	0	0	0	16	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	0	0	0	0	**26**	0	16	0	0	42	61.90	0.3810
7	0	0	0	0	0	0	**12**	0	0	0	12	100.00	0.0000
8	0	0	0	0	0	0	0	**2**	15	21	38	5.26	0.9474
9	0	0	0	0	0	0	0	0	**5**	6	11	45.45	1.0000
10	0	0	0	0	0	0	0	2	0	**8**	10	80.00	0.2000
11	0	0	1	0	0	1	0	10	0	5	17		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	98.73	96.97	80.00	100.00	96.30	100.00	6.25	7.69	18.60		**63.74**	
E.O	0.0000	0.0127	0.0303	0.2000	0.0000	0.0370	0.0000	0.9375	0.9231	0.8140			

A1.5 Instance reduction of all nine bands

A1.5.1 linear

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	34	5	3	0	2	0	32	65	43	184		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**1**	0	0	0	0	0	0	0	0	1	100.00	0.0000
3	0	0	**27**	0	0	0	0	0	0	0	27	100.00	0.0000
4	0	0	0	**13**	0	0	0	0	0	0	13	100.00	0.0000
5	0	0	0	0	**0**	0	0	0	0	0	0	-	-
6	0	25	0	0	0	**25**	0	0	0	0	50	50.00	0.5000
7	0	0	0	0	0	0	**10**	0	0	0	10	100.00	0.0000
8	0	0	0	0	0	0	0	**0**	0	0	0	-	-
9	0	0	0	0	0	0	0	0	**0**	0	0	-	-
10	0	0	0	0	0	0	0	0	0	**0**	0	-	-
11	0	19	1	4	34	0	2	0		0	60		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	1.27	81.82	65.00	0.00	92.59	83.33	0.00	0.00	0.00		**26.10**	
E.O	0.0000	0.9873	0.1818	0.3500	1.0000	0.0741	0.1667	1.0000	1.0000	1.0000			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.5.2 Polynomial

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	0	0	0	4	0	0	2	12	13	31		
1	**3**	0	0	0	0	0	0	0	0	0	3	100.00	0.0000
2	0	**53**	0	0	0	0	0	0	0	0	53	100.00	0.0000
3	0	0	**29**	0	0	0	0	0	0	0	29	100.00	0.0000
4	0	0	0	**8**	0	0	0	0	0	0	8	100.00	0.0000
5	0	0	0	0	**30**	0	0	0	0	0	30	100.00	0.0000
6	0	1	0	0	0	**27**	0	1	0	0	29	93.10	0.0690
7	0	0	0	0	0	0	**6**	0	0	0	6	100.00	0.0000
8	0	0	0	0	0	0	0	**0**	0	0	0	-	-
9	0	0	0	0	0	0	0	0	**53**	0	53	100.00	1.0000
10	0	0	0	0	0	0	0	14	0	**30**	44	68.18	0.3182
11	16	25	4	12	0	0	6	15	0	0	78		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	15.79	67.09	87.88	40.00	88.24	100.00	50.00	0.00	81.54	69.77		**65.66**	
E.O	0.8421	0.3291	0.1212	0.6000	0.1176	0.0000	0.5000	1.0000	0.1846	0.3023			

A1.5.3 Radial Basis Function

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	1	0	3	0	0	0	17	15	1	37		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**69**	0	0	0	0	0	0	0	0	69	100.00	0.0000
3	0	0	**33**	17	0	0	0	0	0	0	50	66.00	0.3400
4	0	0	0	**0**	0	0	0	0	0	0	0	-	-
5	0	3	0	0	**34**	0	0	0	0	0	37	91.89	0.0811
6	0	0	0	0	0	**27**	0	8	0	0	35	77.14	0.2286
7	0	0	0	0	0	0	**12**	0	0	0	12	100.00	0.0000
8	0	0	0	0	0	0	0	**3**	13	17	33	9.09	0.9091
9	0	0	0	0	0	0	0	0	**37**	6	43	86.05	1.0000
10	0	0	0	0	0	0	0	3	0	**8**	11	72.73	0.2727
11	0	6	0	0	0	0	0	1	0	11	18		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	87.34	100.00	0.00	100.00	100.00	100.00	9.38	56.92	18.60		**66.48**	
E.O	0.0000	0.1266	0.0000	1.0000	0.0000	0.0000	0.0000	0.9063	0.4308	0.8140			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.6 Instance reduction of PCA bands

A1.6.1 Linear

	1	2	3	4	5	6	7	8	9	10	Total	U.A %	E.C
0	0	0	0	0	0	0	0	16	65	41	122		
1	**17**	0	0	0	0	0	0	0	0	2	19	89.47	0.1053
2	0	**11**	0	0	0	0	0	0	0	0	11	100.00	0.0000
3	0	0	**27**	0	0	0	0	0	0	0	27	100.00	0.0000
4	2	0	6	**4**	0	0	0	0	0	0	12	33.33	0.6667
5	0	5	0	0	**34**	0	9	0	0	0	48	70.83	0.2917
6	0	0	0	0	0	**3**	0	14	0	0	17	17.65	0.8235
7	0	0	0	13	0	0	**0**	2	0	0	15	0.00	1.0000
8	0	46	0	0	0	24	0	**0**	0	0	70	0.00	1.0000
9	0	6	0	1	0	0	3	0	**0**	0	10	0.00	1.0000
10	0	0	0	0	0	0	0	0	0	**0**	0	-	-
11	0	11	0	2	0	0	0	0	0	0	13		
Total	19	79	33	20	34	27	12	32	65	43			
P.A %	89.47	13.92	81.82	20.00	100.00	11.11	0.00	0.00	0.00	0.00		**26.37**	
E.O	0.1053	0.8608	0.1818	0.8000	0.0000	0.8889	1.0000	1.0000	1.0000	1.0000			

A1.6.2 Polynomial

	1	2	3	4	5	6	7	8	9	10	Total	U.A %	E.C
0	0	0	0	0	1	0	6	0	2	8	17		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**18**	0	0	0	0	0	0	0	0	18	100.00	0.0000
3	0	0	**8**	0	0	0	0	0	0	0	8	100.00	0.0000
4	0	0	0	**12**	0	0	3	0	0	0	15	80.00	0.2000
5	0	0	0	0	**26**	0	0	0	0	0	26	100.00	0.0000
6	0	1	0	0	0	**9**	0	0	0	0	10	90.00	0.1000
7	0	0	0	0	0	0	**0**	0	0	0	0	-	-
8	0	0	0	0	2	0	0	**27**	8	2	39	69.23	0.3077
9	0	0	0	0	0	0	3	0	**13**	0	16	81.25	1.0000
10	0	0	0	0	0	0	0	0	0	**25**	25	100.00	0.0000
11	0	60	25	8	5	18	0	5	42	8	171		
Total	19	79	33	20	34	27	12	32	65	43			
P.A %	100.00	22.78	24.24	60.00	76.47	33.33	0.00	84.38	20.00	58.14		**43.13**	
E.O	0.0000	0.7722	0.7576	0.4000	0.2353	0.6667	1.0000	0.1563	0.8000	0.4186			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

A1.6.3 Radial Basis Function

	1	2	3	4	5	6	7	8	9	10	Total	U.A%	E.C
0	0	16	1	8	0	6	0	3	2	12	48		
1	**19**	0	0	0	0	0	0	0	0	0	19	100.00	0.0000
2	0	**60**	0	0	0	0	0	0	0	0	60	100.00	0.0000
3	0	0	**32**	0	0	0	0	0	0	0	32	100.00	0.0000
4	0	0	0	**12**	0	0	0	0	0	0	12	100.00	0.0000
5	0	0	0	0	**34**	0	0	0	0	0	34	100.00	0.0000
6	0	3	0	0	0	**21**	0	0	0	0	24	87.50	0.1250
7	0	0	0	0	0	0	**12**	0	0	0	12	100.00	0.0000
8	0	0	0	0	0	0	0	**29**	26	1	56	51.79	0.4821
9	0	0	0	0	0	0	0	0	**30**	6	36	83.33	1.0000
10	0	0	0	0	0	0	0	0	0	**24**	24	100.00	0.0000
11	0	0	0	0	0	0	0	0	7	0	7		
Total	19	79	33	20	34	27	12	32	65	43			
P.A%	100.00	75.95	96.97	60.00	100.00	77.78	100.00	90.63	46.15	55.81		**75.00**	
E.O	0.0000	0.2405	0.0303	0.4000	0.0000	0.2222	0.0000	0.0938	0.5385	0.4419			

U.A - User's Accuracy, P.A - Producer's Accuracy, E.C - Error of Commission, E.O - Error of Ommission

Appendix II
Mara Basin

A2.1 All Six Bands

A2.1.1 Linear

	1	2	3	4	5	6	Total	U.A%	E.C
0	380	31	1686	48	13	452	2610		
1	**1241**	3	69	0	0	0	1313	94.52	0.0548
2	0	**357**	0	0	0	0	357	100.00	0.0000
3	0	0	**0**	0	0	0	0	-	-
4	0	0	81	**1704**	1	19	1805	94.40	0.0560
5	0	0	75	0	**235**	115	425	55.29	0.4471
6	0	0	0	0	0	**0**	0	-	-
7	0	349	3	10	260	526	1148		
Total	1621	740	1914	1762	509	1112			
P.A%	76.56	48.24	0.00	96.71	46.17	0.00		**46.19**	
E.O	0.2344	0.5176	1.0000	0.0329	0.5383	1.0000			

A2.1.2 Polynomial

	1	2	3	4	5	6	Total	U.A%	E.C
0	18	1	211	8	121	126	485		
1	**1305**	1	61	0	0	0	1367	95.46	0.0454
2	0	**681**	0	0	0	0	681	100.00	0.0000
3	224	2	**1510**	103	0	70	1909	79.10	0.2090
4	0	0	77	**1642**	4	25	1748	93.94	0.0606
5	0	0	0	0	**0**	0	0	-	-
6	0	0	0	0	0	**0**	0	-	-
7	74	55	55	9	384	891	1468		
Total	1621	740	1914	1762	509	1112			
P.A%	80.51	92.03	0.79	93.19	0.00	0.00		**67.09**	
E.O	0.1949	0.0797	0.2111	0.0681	1.0000	1.0000			

A2.1.3 RBF

	1	2	3	4	5	6	Total	U.A%	E.C
0	20	27	106	15	92	91	351		
1	**1331**	6	80	0	0	0	1417	93.93	0.0607
2	0	**677**	0	0	0	0	677	100.00	0.0000
3	212	4	**1538**	121	0	72	1947	78.99	0.2101
4	0	0	66	**1626**	4	30	1726	94.21	0.0579
5	0	0	37	0	**338**	58	433	78.06	0.2194
6	0	0	4	0	74	**823**	901	91.34	0.0866
7	58	26	83	0	1	38	206		
Total	1621	740	1914	1762	509	1112			
P.A%	82.11	91.49	0.80	92.28	66.40	74.01		**82.70**	
E.O	0.1789	0.0851	0.1964	0.0772	0.3360	0.2599			

Legend

0 Unclassified Pixels
1 Wetlands
2 Water
3 Bush/Shrub/Trees
4 Grasslands
5 Bare Ground
6 Roads
7 Mixed Pixels

A2.2 PCA

A2.2.1 Linear

	1	2	3	4	5	6	Total	U.A%	E.C
0	994	47	1771	38	33	883	3766		
1	**627**	12	69	5	0	0	713	87.94	0.1206
2	0	**467**	0	0	0	0	467	100.00	0.0000
3	0	0	**0**	0	0	0	0	-	-
4	0	0	47	**1617**	0	0	1664	97.18	0.0282
5	0	0	21	0	**291**	170	482	60.37	0.3963
6	0	0	0	0	90	**59**	149	39.60	0.6040
7	0	214	6	102	95	0	417		
Total	1621	740	1914	1762	509	1112			
P.A%	38.68	63.11	0.00	91.77	57.17	5.31		**39.97**	
E.O	0.6132	0.3689	1.0000	0.0823	0.4283	0.9469			

A2.2.2 Polynomial

	1	2	3	4	5	6	Total	U.A%	E.C
0	116	7	174	12	68	53	430		
1	**1166**	2	40	0	0	0	1208	96.52	0.0348
2	0	**616**	0	0	0	0	616	100.00	0.0000
3	266	49	**1596**	57	0	34	2002	79.72	0.2028
4	0	0	44	**1693**	0	0	1737	97.47	0.0253
5	0	0	31	0	**241**	29	301	80.07	0.1993
6	0	0	8	0	184	**950**	1142	83.19	0.1681
7	73	66	21	0	16	46	222		
Total	1621	740	1914	1762	509	1112			
P.A%	71.93	83.24	0.83	96.08	47.35	85.43		**81.77**	
E.O	0.2807	0.1676	0.1661	0.0392	0.5265	0.1457			

A2.2.3 RBF

	1	2	3	4	5	6	Total	U.A%	E.C
0	0	15	40	0	114	18	187		
1	**1064**	10	35	0	0	0	1109	95.94	0.0406
2	0	**663**	0	0	0	0	663	100.00	0.0000
3	534	52	**1679**	50	1	20	2336	71.88	0.2813
4	0	0	34	**1697**	0	0	1731	98.04	0.0196
5	0	0	17	0	**275**	18	310	88.71	0.1129
6	0	0	63	0	117	**987**	1167	84.58	0.1542
7	23	0	46	15	2	69	155		
Total	1621	740	1914	1762	509	1112			
P.A%	65.64	89.59	0.88	96.31	54.03	88.76		**83.12**	
E.O	0.3436	0.1041	0.1228	0.0369	0.4597	0.1124			

Legend

0 Unclassified Pixels
1 Wetlands
2 Water
3 Bush/Shrub/Trees
4 Grasslands
5 Bare Ground
6 Roads
7 Mixed Pixels

A.2.3 Exhaustive Search

A2.3.1 Linear

	1	2	3	4	5	6	Total	U.A%	E.C
0	1043	48	1586	18	20	360	3075		
1	**578**	2	92	6	0	0	678	85.25	0.1475
2	0	**434**	0	0	0	0	434	100.00	0.0000
3	0	0	**0**	0	0	0	0	-	-
4	0	0	133	**1725**	2	19	1879	91.80	0.0820
5	0	0	92	0	**390**	663	1145	34.06	0.6594
6	0	0	0	0	1	**3**	4	75.00	0.2500
7	0	256	11	13	96	67	443		
Total	1621	740	1914	1762	509	1112			
P.A%	35.66	58.65	0.00	97.90	76.62	0.27		**40.87**	
E.O	0.6434	0.4135	1.0000	0.0210	0.2338	0.9973			

A2.3.2 Polynomial

	1	2	3	4	5	6	Total	U.A%	E.C
0	249	1	100	0	25	58	433		
1	**623**	0	5	0	0	0	628	99.20	0.0080
2	0	**642**	4	0	0	0	646	99.38	0.0062
3	344	45	**1505**	24	1	31	1950	77.18	0.2282
4	0	0	127	**1684**	1	14	1826	92.22	0.0778
5	0	0	107	0	**285**	150	542	52.58	0.4742
6	0	0	0	0	165	**674**	839	80.33	0.1967
7	405	52	66	54	32	185	794		
Total	1621	740	1914	1762	509	1112			
P.A%	38.43	86.76	0.79	95.57	55.99	60.61		**70.68**	
E.O	0.6157	0.1324	0.2137	0.0443	0.4401	0.3939			

A2.3.3 RBF

	1	2	3	4	5	6	Total	U.A%	E.C
0	127	3	313	3	76	43	565		
1	**344**	1	16	0	0	0	361	95.29	0.0471
2	0	**673**	0	0	0	0	673	100.00	0.0000
3	534	44	**1213**	50	0	45	1886	64.32	0.3568
4	0	0	115	**1671**	0	7	1793	93.20	0.0680
5	0	0	204	0	**309**	97	610	50.66	0.4934
6	0	0	0	0	112	**685**	797	85.95	0.1405
7	616	19	53	38	12	235	973		
Total	1621	740	1914	1762	509	1112			
P.A%	21.22	90.95	0.63	94.84	60.71	61.60		**63.92**	
E.O	0.7878	0.0905	0.3662	0.0516	0.3929	0.3840			

Legend

0 Unclassified Pixels
1 Wetlands
2 Water
3 Bush/Shrub/Trees
4 Grasslands
5 Bare Ground
6 Roads
7 Mixed Pixels

A2.4 PBIL

A2.4.1 Linear

	1	2	3	4	5	6	Total	U.A%	E.C
0	1617	14	1604	22	54	857	4168		
1	**0**	0	0	0	0	0	0	-	-
2	0	**726**	44	0	0	0	770	94.29	0.0571
3	0	0	**0**	0	0	0	0	-	-
4	4	0	211	**1740**	1	53	2009	86.61	0.1339
5	0	0	55	0	**444**	202	701	63.34	0.3666
6	0	0	0	0	0	**0**	0	-	-
7	0	0	0	0	10	0	10		
Total	1621	740	1914	1762	509	1112			
P.A%	0.00	98.11	0.00	98.75	87.23	0.00		**38.00**	
E.O	1.0000	0.0189	1.0000	0.0125	0.1277	1.0000			

A2.4.2 Polynomial

	1	2	3	4	5	6	Total	U.A%	E.C
0	138	2	343	0	6	227	716		
1	**1302**	0	66	0	0	0	1368	95.18	0.0482
2	0	**673**	28	0	0	0	701	96.01	0.0399
3	126	12	**1181**	22	0	8	1349	87.55	0.1245
4	4	0	187	**1730**	0	45	1966	88.00	0.1200
5	0	0	42	0	**284**	43	369	76.96	0.2304
6	0	0	2	0	47	**606**	655	92.52	0.0748
7	51	53	65	10	172	183	534		
Total	1621	740	1914	1762	509	1112			
P.A%	80.32	90.95	0.62	98.18	55.80	54.50		**75.42**	
E.O	0.1968	0.0905	0.3830	0.0182	0.4420	0.4550			

A2.4.3 RBF

	1	2	3	4	5	6	Total	U.A%	E.C
0	110	8	265	13	75	227	698		
1	**1290**	1	76	0	0	0	1367	94.37	0.0563
2	0	**706**	19	0	0	0	725	97.38	0.0262
3	172	15	**1235**	16	0	5	1443	85.59	0.1441
4	0	0	211	**1733**	3	37	1984	87.35	0.1265
5	0	0	32	0	**318**	54	404	78.71	0.2129
6	0	0	28	0	112	**769**	909	84.60	0.1540
7	49	10	48	0	1	20	128		
Total	1621	740	1914	1762	509	1112			
P.A%	79.58	95.41	0.65	98.35	62.48	69.15		**79.02**	
E.O	0.2042	0.0459	0.3548	0.0165	0.3752	0.3085			

Legend

0 Unclassified Pixels
1 Wetlands
2 Water
3 Bush/Shrub/Trees
4 Grasslands
5 Bare Ground
6 Roads
7 Mixed Pixels

A2.5 Center Splitting - All

A2.5.1 Linear

	1	2	3	4	5	6	Total	U.A%	E.C
0	298	61	1662	63	134	1019	3237		
1	**1323**	3	155	0	0	0	1481	89.33	0.1067
2	0	**70**	0	0	0	0	70	100.00	0.0000
3	0	0	**0**	0	0	0	0	-	-
4	0	0	81	**1699**	1	44	1825	93.10	0.0690
5	0	0	11	0	**314**	28	353	88.95	0.1105
6	0	0	0	0	48	**20**	68	29.41	0.7059
7	0	606	5	0	12	1	624		
Total	1621	740	1914	1762	509	1112			
P.A%	81.62	9.46	0.00	96.42	61.69	1.80		**44.74**	
E.O	0.1838	0.9054	1.0000	0.0358	0.3831	0.9820			

A2.5.2 Polynomial

	1	2	3	4	5	6	Total	U.A%	E.C
0	34	0	198	0	251	994	1477		
1	**1291**	1	80	0	0	0	1372	94.10	0.0590
2	0	**386**	0	0	0	0	386	100.00	0.0000
3	261	63	**1396**	73	0	27	1820	76.70	0.2330
4	0	0	66	**1598**	0	17	1681	95.06	0.0494
5	0	0	64	0	**198**	48	310	63.87	0.3613
6	0	0	0	0	60	**20**	80	25.00	0.7500
7	35	290	110	91	0	6	532		
Total	1621	740	1914	1762	509	1112			
P.A%	79.64	52.16	0.73	90.69	38.90	1.80		**63.84**	
E.O	0.2036	0.4784	0.2706	0.0931	0.6110	0.9820			

A2.5.3 RBF

	1	2	3	4	5	6	Total	U.A%	E.C
0	50	3	88	9	106	80	336		
1	**1289**	0	38	0	0	0	1327	97.14	0.0286
2	0	**672**	0	0	0	0	672	100.00	0.0000
3	254	61	**1597**	66	0	28	2006	79.61	0.2039
4	0	0	53	**1678**	1	30	1762	95.23	0.0477
5	0	0	77	0	**301**	61	439	68.56	0.3144
6	0	0	0	0	101	**882**	983	89.73	0.1027
7	28	4	61	9	0	31	133		
Total	1621	740	1914	1762	509	1112			
P.A%	79.52	90.81	0.83	95.23	59.14	79.32		**83.82**	
E.O	0.2048	0.0919	0.1656	0.0477	0.4086	0.2068			

Legend

0 Unclassified Pixels
1 Wetlands
2 Water
3 Bush/Shrub/Trees
4 Grasslands
5 Bare Ground
6 Roads
7 Mixed Pixels

A2.6 Center Splitting - PCA

A2.6.1 Linear

	1	2	3	4	5	6	Total	U.A%	E.C
0	1177	50	1769	26	0	398	3420		
1	**444**	8	64	5	0	0	521	85.22	0.1478
2	0	**53**	0	0	0	0	53	100.00	0.0000
3	0	0	**0**	0	0	0	0	-	-
4	0	0	61	**1719**	0	1	1781	96.52	0.0348
5	0	0	13	0	**10**	20	43	23.26	0.7674
6	0	1	0	0	120	**617**	738	83.60	0.1640
7	0	628	7	12	379	76	1102		
Total	1621	740	1914	1762	509	1112			
P.A%	27.39	7.16	0.00	97.56	1.96	55.49		**37.12**	
E.O	0.7261	0.9284	1.0000	0.0244	0.9804	0.4451			

A2.6.2 Polynomial

	1	2	3	4	5	6	Total	U.A%	E.C
0	180	2	118	0	0	370	670		
1	**974**	0	38	0	0	0	1012	96.25	0.0375
2	0	**618**	0	0	0	0	618	100.00	0.0000
3	378	54	**1689**	71	0	39	2231	75.71	0.2429
4	0	0	27	**1496**	0	0	1523	98.23	0.0177
5	0	0	21	0	**290**	86	397	73.05	0.2695
6	0	0	0	0	149	**580**	729	79.56	0.2044
7	89	66	21	195	70	37	478		
Total	1621	740	1914	1762	509	1112			
P.A%	60.09	83.51	0.88	84.90	56.97	52.16		**73.74**	
E.O	0.3991	0.1649	0.1176	0.1510	0.4303	0.4784			

A2.6.3 RBF

	1	2	3	4	5	6	Total	U.A%	E.C
0	25	28	51	22	145	44	315		
1	**1174**	6	23	0	0	0	1203	97.59	0.0241
2	0	**658**	0	0	0	0	658	100.00	0.0000
3	362	43	**1692**	5	0	40	2142	78.99	0.2101
4	0	0	36	**1725**	0	0	1761	97.96	0.0204
5	0	0	15	0	**257**	26	298	86.24	0.1376
6	0	0	40	0	107	**968**	1115	86.82	0.1318
7	60	5	57	10	0	34	166		
Total	1621	740	1914	1762	509	1112			
P.A%	72.42	88.92	0.88	97.90	50.49	87.05		**84.54**	
E.O	0.2758	0.1108	0.1160	0.0210	0.4951	0.1295			

Legend

0 Unclassified Pixels
1 Wetlands
2 Water
3 Bush/Shrub/Trees
4 Grasslands
5 Bare Ground
6 Roads
7 Mixed Pixels

Appendix III
Malmesbury

A3.1 All Six Bands

A3.1.1 Linear

	1	2	3	Total	U.A%	E.C
0	9	3	0	12		
1	**7**	0	0	7	100.00	0.0000
2	2	**19**	0	21	90.48	0.0952
3	0	0	**19**	19	100.00	0.0000
4	3	0	6	9		
Total	21	22	25			
P.A%	33.33	86.36	76.00		**66.18**	
E.O	0.6667	0.1364	0.2400			

A3.1.2 Polynomial

	1	2	3	Total	U.A%	E.C
0	8	3	0	11		
1	**9**	0	0	9	100.00	0.0000
2	3	**19**	0	22	86.36	0.1364
3	0	0	**19**	19	100.00	0.0000
4	1	0	6	7		
Total	21	22	25			
P.A%	42.86	86.36	76.00		**69.12**	
E.O	0.5714	0.1364	0.2400			

A3.1.3 RBF

	1	2	3	Total	U.A%	E.C
0	1	0	0	1		
1	**10**	1	6	17	58.82	0.4118
2	5	**15**	0	20	75.00	0.2500
3	0	0	**19**	19	100.00	0.0000
4	5	6	0	11		
Total	21	22	25			
P.A%	47.62	68.18	76.00		**64.71**	
E.O	0.5238	0.3182	0.2400			

Legend

0 Unclassified Pixels
1 Built up areas
2 Fields
3 Trees
4 Mixed Pixels

A3.2 PCA

A3.2.1 Linear

	1	2	3	Total	U.A%	E.C
0	5	5	1	11		
1	**13**	2	2	17	76.47	0.2353
2	2	**15**	0	17	88.24	0.1176
3	0	0	**8**	8	100.00	0.0000
4	1	0	14	15		
Total	21	22	25			
P.A%	61.90	68.18	32.00		**52.94**	
E.O	0.3810	0.3182	0.6800			

A3.2.2 Polynomial

	1	2	3	Total	U.A%	E.C
0	5	6	1	12		
1	**14**	2	5	21	66.67	0.3333
2	2	**14**	0	16	87.50	0.1250
3	0	0	**9**	9	100.00	0.0000
4	0	0	10	10		
Total	21	22	25			
P.A%	66.67	63.64	36.00		**54.41**	
E.O	0.3333	0.3636	0.6400			

A3.2.3 RBF

	1	2	3	Total	U.A%	E.C
0	0	0	0	0		
1	**19**	10	12	41	46.34	0.5366
2	2	**12**	0	14	85.71	0.1429
3	0	0	**10**	10	100.00	0.0000
4	0	0	3	3		
Total	21	22	25			
P.A%	90.48	54.55	40.00		**60.29**	
E.O	0.0952	0.4545	0.6000			

Legend

0 Unclassified Pixels
1 Built up areas
2 Fields
3 Trees
4 Mixed Pixels

A3.3 Exhaustive Search

A3.3.1 Linear

	1	2	3	Total	U.A%	E.C
0	9	5	0	14		
1	**7**	0	0	7	100.00	0.0000
2	2	**17**	0	19	89.47	0.1053
3	1	0	**19**	20	95.00	0.0500
4	2	0	6	8		
Total	21	22	25			
P.A%	33.33	77.27	76.00		**63.24**	
E.O	0.6667	0.2273	0.2400			

A3.3.2 Polynomial

	1	2	3	Total	U.A%	E.C
0	9	5	0	14		
1	**8**	0	0	8	100.00	0.0000
2	3	**17**	0	20	85.00	0.1500
3	0	0	**19**	19	100.00	0.0000
4	1	0	6	7		
Total	21	22	25			
P.A%	38.10	77.27	76.00		**64.71**	
E.O	0.6190	0.2273	0.2400			

A3.3.3 RBF

	1	2	3	Total	U.A%	E.C
0	4	0	2	6		
1	**11**	2	6	19	57.89	0.4211
2	3	**17**	0	20	85.00	0.1500
3	1	0	**17**	18	94.44	0.0556
4	2	3	0	5		
Total	21	22	25			
P.A%	52.38	77.27	68.00		**66.18**	
E.O	0.4762	0.2273	0.3200			

Legend

0 Unclassified Pixels
1 Built up areas
2 Fields
3 Trees
4 Mixed Pixels

A3.4 PBIL

A3.4.1 Linear

	1	2	3	Total	U.A%	E.C
0	8	5	0	13		
1	**8**	0	0	8	100.00	0.0000
2	2	17	0	19	89.47	0.1053
3	0	0	**19**	19	100.00	0.0000
4	3	0	6	9		
Total	21	22	25			
P.A%	38.10	77.27	76.00		**64.71**	
E.O	0.6190	0.2273	0.2400			

A3.4.2 Polynomial

	1	2	3	Total	U.A%	E.C
0	8	4	0	12		
1	**9**	0	0	9	100.00	0.0000
2	2	**17**	0	19	89.47	0.1053
3	0	1	**19**	20	95.00	0.0500
4	2	0	6	8		
Total	21	22	25			
P.A%	42.86	77.27	76.00		**66.18**	
E.O	0.5714	0.2273	0.2400			

A3.4.3 RBF

	1	2	3	Total	U.A%	E.C
0	1	6	1	8		
1	**14**	0	6	20	70.00	0.3000
2	4	**13**	0	17	76.47	0.2353
3	2	3	**18**	23	78.26	0.2174
4	0	0	0	0		
Total	21	22	25			
P.A%	66.67	59.09	72.00		**66.18**	
E.O	0.3333	0.4091	0.2800			

Legend

0 Unclassified Pixels
1 Built up areas
2 Fields
3 Trees
4 Mixed Pixels

A3.5 Center Splitting - All

A3.5.1 Linear

	1	2	3	Total	U.A%	E.C
0	8	5	0	13		
1	**7**	0	0	7	100.00	0.0000
2	2	**17**	0	19	89.47	0.1053
3	1	0	**19**	20	95.00	0.0500
4	3	0	6	9		
Total	21	22	25			
P.A%	33.33	77.27	76.00		**63.24**	
E.O	0.6667	0.2273	0.2400			

A3.5.2 Polynomial

	1	2	3	Total	U.A%	E.C
0	8	4	0	12		
1	**7**	0	0	7	100.00	0.0000
2	2	**18**	0	20	90.00	0.1000
3	1	0	**19**	20	95.00	0.0500
4	3	0	6	9		
Total	21	22	25			
P.A%	33.33	81.82	76.00		**64.71**	
E.O	0.6667	0.1818	0.2400			

A3.5.3 RBF

	1	2	3	Total	U.A%	E.C
0	3	2	6	11		
1	**9**	0	2	11	81.82	0.1818
2	5	**20**	0	25	80.00	0.2000
3	4	0	**17**	21	80.95	0.1905
4	0	0	0	0		
Total	21	22	25			
P.A%	42.86	90.91	68.00		**67.65**	
E.O	0.5714	0.0909	0.3200			

Legend

0 Unclassified Pixels
1 Built up areas
2 Fields
3 Trees
4 Mixed Pixels

A3.6 Center Splitting - PCA

A3.6.1 Linear

	1	2	3	Total	U.A%	E.C
0	6	4	0	10		
1	**13**	2	1	16	81.25	0.1875
2	1	**16**	0	17	94.12	0.0588
3	0	0	**7**	7	100.00	0.0000
4	1	0	17	18		
Total	21	22	25			
P.A%	61.90	72.73	28.00		**52.94**	
E.O	0.3810	0.2727	0.7200			

A3.6.2 Polynomial

	1	2	3	Total	U.A%	E.C
0	1	1	0	2		
1	**16**	5	1	22	72.73	0.2727
2	3	**16**	0	19	84.21	0.1579
3	0	0	**7**	7	100.00	0.0000
4	1	0	17	18		
Total	21	22	25			
P.A%	76.19	72.73	28.00		**57.35**	
E.O	0.2381	0.2727	0.7200			

A3.6.3 RBF

	1	2	3	Total	U.A%	E.C
0	0	0	0	0		
1	**20**	10	7	37	54.05	0.4595
2	1	12	0	13	92.31	0.0769
3	0	0	**7**	7	100.00	0.0000
4	0	0	11	11		
Total	21	22	25			
P.A%	95.24	54.55	28.00		**57.35**	
E.O	0.0476	0.4545	0.7200			

Legend

0 Unclassified Pixels
1 Built up areas
2 Fields
3 Trees
4 Mixed Pixels