RULE-BASED LAND COVER CLASSIFICATION MODEL: EXPERT SYSTEM INTEGRATION OF IMAGE AND NON-IMAGE SPATIAL DATA

by

DAWIT K. KIDANE

THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE AT THE UNIVERSITY OF STELLENBOSCH

Supervisor: MR A VAN NIEKERK

April 2005

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my original work and has not previously in its entirety or in part been submitted at any university for a degree.

Signature: Date: August 2004

Dawit K. Kidane

ABSTRACT

Remote sensing and image processing tools provide speedy and up-to-date information on land resources. Although remote sensing is the most effective means of land cover and land use mapping, it is not without limitations. The accuracy of image analysis depends on a number of factors, of which the image classifier used is probably the most significant. It is noted that there is no perfect classifier, but some robust classifiers achieve higher accuracy results than others. For certain land cover/uses, discrimination based only on spectral properties is extremely difficult and often produces poor results. The use of ancillary data can improve the classification process. Some classifiers incorporate ancillary data before or after the classification process, which limits the full utilization of the information contained in the ancillary data. Expert classification, on the other hand, makes better use of ancillary data by incorporating data directly into the classification process.

In this study an expert classification model was developed based on spatial operations designed to identify a specific land cover/use, by integrating both spectral and available ancillary data. Ancillary data were derived either from the spectral channels or from other spatial data sources such as DEM (Digital Elevation Model) and topographical maps. The model was developed in ERDAS Imagine image-processing software, using the expert engineer as a final integrator of the different constituent spatial operations. An attempt was made to identify the Level I land cover classes in the South African National Land Cover classification scheme hierarchy. Rules were determined on the basis of expert knowledge or statistical calculations of mean and variance on training samples. Although rules could be determined by using statistical applications, such as the classification analysis regression tree (CART), the absence of adequate and accurate training data for all land cover classes and the fact that all land cover classes do not require the same predictor variables makes this option less desirable. The result of the accuracy assessment showed that the overall classification accuracy was 84.3% and kappa statistics 0.829. Although this level of accuracy might be suitable for most applications, the model is flexible enough to be improved further.

OPSOMMING

Afstandswaarneming-en beeldverwerkingstegnieke kan akkurate informasie oorbodemhulpbronne weergee. Alhoewel afstandswaarneming die mees effektiewe manier van grondbedekking en grondgebruikkartering is, is dit nie sonder beperkinge nie. Die akkuraatheid van beeldverwerking is afhanklik van verskeie faktore, waarvan die beeld klassifiseerder wat gebruik word, waarskynlik die belangrikste faktor is. Dit is welbekend dat daar geen perfekte klassifiseerder is nie, alhoewel sekere kragtige klassifiseerders hoër akkuraatheid as ander behaal. Vir sekere grondbedekking en -gebruike is uitkenning gebaseer op spektrale eienskappe uiters moeilik en dikwels word swak resultate behaal. Die gebruik van aanvullende data, kan die klassifikasieproses verbeter. Sommige klassifiseerders inkorporeer aanvullende data voor of na die klassifikasieproses, wat die volle aanwending van die informasie in die aanvullende data beperk. Deskundige klassifikasie, aan die ander kant, maak beter gebruik van aanvullende data deurdat dit data direk in die klassifikasieproses inkorporeer.

Tydens hierdie studie is 'n deskundige klassifikasiemodel ontwikkel gebaseer op ruimtelike verwerkings, wat ontwerp is om spesifieke grondbedekking en –gebruike te identifiseer. Laasgenoemde is behaal deur beide spektrale en beskikbare aanvullende data te integreer. Aanvullende data is afgelei van, óf spektrale eienskappe, óf ander ruimtelike bronne soos 'n DEM (Digitale Elevasie Model) en topografiese kaarte. Die model is ontwikkel in ERDAS Imagine beeldverwerking sagteware, waar die 'expert engineer' as finale integreerder van die verskillende samestellende ruimtelike verwerkings gebruik is. 'n Poging is aangewend om die Klas I grondbedekkingklasse, in die Suid-Afrikaanse Nasionale Grondbedekking klassifikasiesisteem te identifiseer. Reëls is vasgestel aan die hand van deskundige begrippe of eenvoudige statistiese berekeninge van die gemiddelde en variansie van opleidingsdata. Alhoewel reëls met behulp van statistiese toepassings, soos die 'classification analysis regression tree (CART)' vasgestel kon word, maak die afwesigheid van genoegsame en akkurate opleidingsdata vir al die grondbedekkingsklasse hierdie opsie minder aantreklik. Bykomend tot laasgenoemde, vereis alle grondbedekkingsklasse nie dieselfde voorspellingsveranderlikes nie. Die resultaat van hierdie akkuraatheidsskatting toon dat die algehele klassifikasie-akkuraatheid 84.3% was en die kappa statistieke 0.829. Alhoewel hierdie vlak van akkuraatheid vir die meeste toepassings geskik is, is die model aanpasbaar genoeg om verder te verbeter.

DEDICATION

To The glory of God!!!

ACKNOWLEDGEMENTS

My deepest thanks and gratitude go to Mr. A van Niekerk, my supervisor, for his direction, help and support throughout the project, and for giving me the opportunity of an assistantship in the advanced GIS course practical sessions. I would like to thank Prof. Larry Zietsman for his insightful ideas and comments and for letting me use PCI Geomatic v-9 software. My heartfelt thanks are due to Wolfgang Luck for giving me access to his database, for the technical explanations that inspired me, and for many other things I couldn't find space to include in this little acknowledgment. My respect and gratitude go to Dr. PH de Necker, co-coordinator of postgraduate courses, for his tremendous help and most of all his kind concern. My thanks also go to Gerbrand Mans for giving me access and space on an FTP server to store my huge database backup. Special thanks are due to Mark Thompson for providing me with technical help and documentation on wetlands. Hennie van der Berg at the CSIR kindly supplied me with ground truth training samples. I thank Dr. EP Hees for editing my thesis. To all my classmates I extend my respect and love.

I would like to thank the Eritrean Human Resources Development Programme (HRD) for funding me, and Mrs. Lula Gebreyesus for her help in my transfer from Rhodes to Stellenbosch University. To my dear family I offer thanks with honour and love, and also to my fiancée, Samle, a wonderful woman, for her love, prayer and support, and most of all her patience. Above all, I thank God with all my heart for the grace and strength for each day.

CONTENTS

TABLES

FIGURES

CHAPTER ONE: LAND COVER RESOURCES INFORMATION FOR DECISION MAKING

The continual production of relevant and up-to-date spatial information on the distribution of land cover resources is a first step in dealing with important environmental issues. This information provides a better understanding of resource utilization problems and forms the basis for the identification of suitable strategies for sustainable development (Moller-Jensen 1998). However, the flexibility and speed with which such data are produced are crucial. In this context remote sensing and digital image processing are highly suitable tools for many problems that are associated with the spatial distribution of phenomena on the earth's surface (Mulders & Jordens 1993; Moller-Jensen 1998; Burrough 1993). The spatial and temporal distribution of land cover constitutes a fundamental dataset for a wide variety of studies in the physical and social sciences, as well as government agencies for land planning purposes (Stefanov, Ramsey & Christensen 2001). As a result, remote-sensing techniques have been the single most effective method for land cover and land use data acquisition (Thompson 1996). Remote sensing is required for continuous monitoring, change detection and map updating of land cover and land use data (Moller-Jensen 1998).

1.1 LAND COVER VERSUS LAND USE

In recent years the term 'land cover' has come to be commonly used in association with the term 'land use'. The two are not synonymous and bear different meanings, though some overlap is evident. Land use is an abstract concept, covering an amalgam of economic, social and cultural factors, and includes everything the land of a country or an area is used for by its residents. Thus it is defined in terms of *function* rather than *physical* property (Barr & Barnsley 2000; Tapiador & Casanova 2003). Land cover refers more to cover of the land surface, which includes mainly vegetation and artificial constructions (Lindgren 1985). Land use information must be inferred from land cover information and the associated patterns (Jansen & Gregorio 2003).

Land cover classification from high-resolution imagery, using existing pixel-based multi-spectral classification algorithms, is comparatively easier than land use classification. This is due to the abstract nature of land use, which implies that the relationship between land use and the multi-spectral signals detected by a satellite sensor is complex and indirect (Barnsley & Barr 1996 in Barr & Barnsley 2000).

1.2 LAND COVER DATABASES

Land cover classification from remotely sensed data is often used for purposes of mapping and inventorying of natural resources over relatively large areas, and such general land cover information is required for many environmental, land management and modeling applications (Langley, Cheshire & Humes 2001; Vogelmann *et al.* 2001). Data can be derived at a range of spatial scales and the scale at which information is extracted determines the appropriate usage of data. For example, data derived from the Advanced Very High Resolution Radiometer (AVHRR) may be well suited for global analysis, but of limited value for regional and local investigations (Vogelmann *et al.* 2001). The dynamic perspective over a range of spatial scales has been the real strength of remotely sensed data (Boyd, Foody & Ripple 2002).

Land cover classification processes have been focusing on a generation of specialized data products suitable only for specific needs of projects. The effect of such narrowly focused applications has been remote sensing datasets and methods with limited value for other uses and that are difficult to compare (spatially and temporally) with one another. This may have been partly due to technological limitations and funding (Collin, Huang, Yang & Wylie 2002). Historically, this has been the case in South Africa (Thompson 1996). The Council for Scientific and Industrial Research (CSIR) and the Agricultural Research Council (ARC) initiated the National Land Cover project (NLC) to provide land cover products that are suitable for GIS-based mapping and modeling applications at suitable scales. The NLC project is the first standardized land cover database that provides baseline information on national land cover, including Swaziland and Lesotho.

The 1994 NLC database was mapped from a series of precision-corrected satellite images at a scale of 1:250 000. The product indicates the dominant land cover within a 1-2 hectare unit and can be used for a wide variety of purposes. However, reliability and consistency have been the biggest obstacles in the 1994 NLC project as qualitative methods were used. Images were interpreted by different interpreters rather than being classified quantitatively (CGA 2003). The product is available to the public domain and can be purchased from the CSIR.

The Land Cover Classification Scheme for South African Remote Sensing Applications was standardized using known land-cover classes identifiable on high-resolution satellite imagery.

The 31 broad-level thematic land cover classes can be adapted to suite individual user requirements (CSIR 2003). The current NLC project is the NLC 2000, which is being contracted out to remote sensing contractors. The NLC 2000 is suitable for mapping and modeling at a scale of 1:50 000. The supervising institutes specify instructions and frameworks that the contractors should comply with. The decision as to what classification methodology to use lies with the contractors. Some contractors have made use of advanced object-orientated classifiers but, as far as the author is aware, no contractor is employing a per-pixel rule-based classification method.

1.3 EXPERT SYSTEMS

Several methods and algorithms have been developed to maximize the extraction of information from digital satellite imagery. These include statistical (e.g. maximum likelihood), contextual, textural, fuzzy sets and artificial neural networks (Stuckens, Coppin & Bauer 2000). Although it has been proven that no image classifier is perfect (Matsuyama in Liu, Skidmore & Oosten 2002), classification accuracy can be improved by using more rigorous classifiers. Pixel-based spectral classification approaches have been shown to be limited in nature and are only effective in cases where land cover classes are spectrally well differentiated (Lira & Maletti 2002; Oetter *et al.* 2000). Many approaches utilize ancillary non-image spatial data as a pre-classification procedure or as part of post-classification manipulations to increase classification accuracy. Failure to incorporate ancillary data during the actual classification process might, however, result in the under-exploitation of the full range of information available (Lawrence & Wright 2001).

In recently developed approaches, such as expert systems and neural networks, ancillary data can be incorporated directly into the classification algorithms that are usually not dependent on *a priori* weights (Bolstad & Lillesand 1992; Lawrence & Wright 2001). Expert models may use various decision criteria and operations to identify distinct land cover classes. A set of defined spatial data operators, which may be called image classification primitives, is used to build expert classifier models. Expert systems support both evidential and hierarchical inferences, and this combination is desirable both for increased classification accuracy and enhanced run-time efficiency (Bolstad & Lillesand 1992). There is no complete classification expert system that can start from the raw data and produce complete and correct classifications, although some ongoing research is developing prototypes that show the viability of such an approach. Expert systems have been applied to solve classification problems by guiding classification of features by a set of decision rules (Tsatsoulis 1993; Bolstad & Lillesand 1992).

1.4 RESEARCH AIMS AND OBJECTIVES

The aim of this study is to develop an expert classifier based on available ancillary data that will provide an increased classification accuracy of Landsat ETM+ images for the extraction of land cover classes of the study area. The developed classifier model is assessed in terms of its suitability for the National Land Cover (NLC) project and the wider remote sensing community. The model will be implemented in the widely used remote sensing software ERDAS Imagine (ERDAS 2001). To achieve these aims, the following objectives were set:

1. Acquire Landsat ETM+ images as well as training and ancillary data.
2. Correct images geometrically and radiometrically.
3. Prepare ancillary data.
4. Identify and calculate basic models that constitute the lowest level of the classifier.
5. Determine the rules and procedures to be implemented in compound models.
6. Develop the compound models that identify the land cover classes.
7. Implement the classifier model into the ERDAS Imagine knowledge engineer.
8. Assess accuracy of the model.
9. Report findings.

1.5 STUDY AREA

The study site is located in KwaZulu-Natal and partly in the Eastern Cape, South Africa. It was chosen for the diversity of the land cover that occurs in the area and because near cloud-free Landsat 7 ETM+ images were already available. The study site is on the eastern side of South Africa bordered by the warm Indian Ocean to the east and the high escarpment of the Drakensberg Mountains to the west. The climate is generally warm subtropical. Summers are usually hot and humid with temperature averaging 28 degrees, and majority of the annual rainfall rains in summer. Whereas winters are generally warm, dry and clear with average temperatures of 23 degrees. There is however occasional frost in the interior and snow often falls in the higher altitudes in winter (Tourism KwaZulu-Natal 2004). The average annual rainfall of the area ranges from 570 to 1150 mm.

In terms of vegetation it is home to some of the most diverse forests in the country. A range of agricultural activities takes place in the area, which includes commercial tree plantations; non-timber cultivation such as sisal, sugar cane and orchards; and other agricultural activities (Mucina *et al.* 2003). The study area includes two major South African cities, namely, Durban and Pietermaritzburg, and other small urban areas, formal and informal townships and rural villages. Numerous light and heavy industries and mining activities are also found in the area.

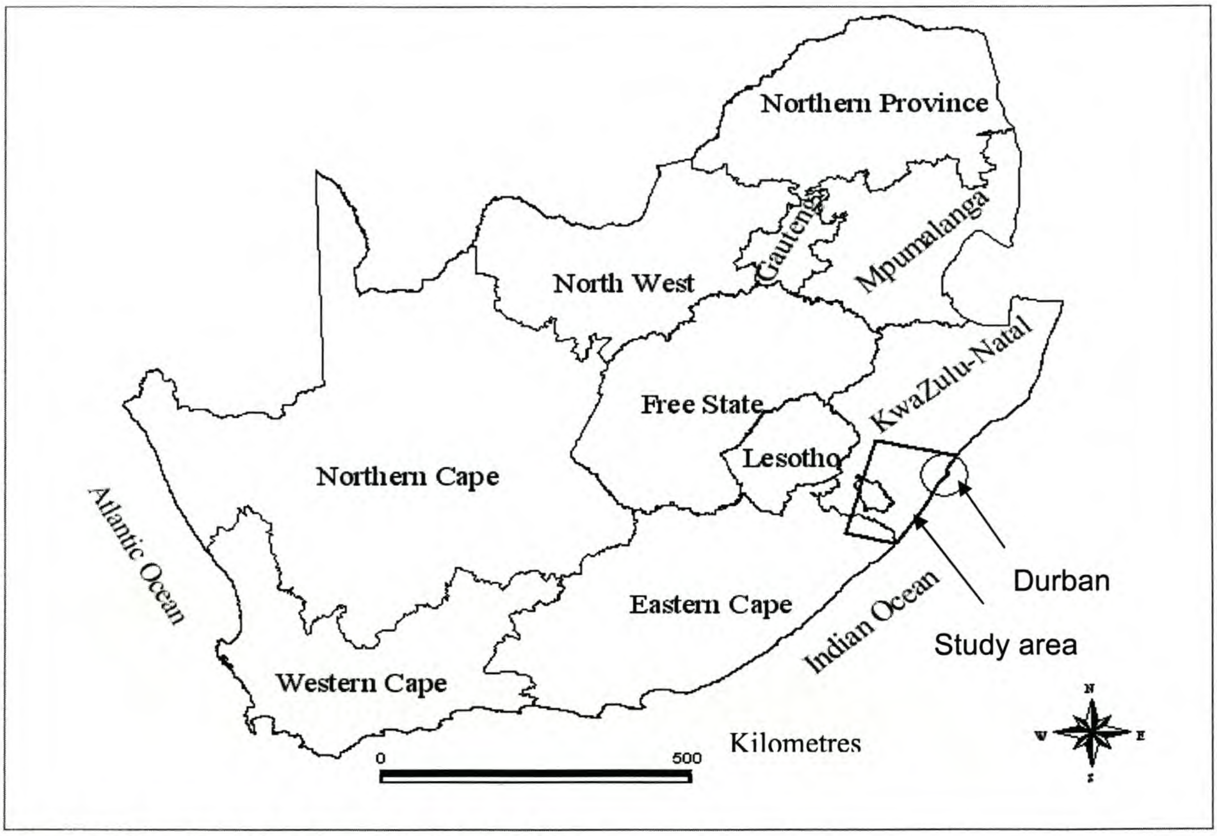


Figure 1.1 Location of study area in South Africa

1.6 REMOTELY SENSED IMAGERY AND PRE-PROCESSING

The Landsat ETM+ satellite images needed for this study were acquired from the Centre for Geographical Analysis, Department of Geography and Environmental Studies, University of Stellenbosch. The dates of the ETM+ images are 7 February and 17 July 2001. The images were orthorectified in ERDAS Imagine (ERDAS 2001) using the standard Landsat 7 ETM+ orthorectification module. A DEM of 30m, supplied by the CSIR (Environmentek), was used in the process. The DEM was generated from 20m (vertical) interval contour data. Input ground control points (GCPs) were collected from a fused image of +/- 15m resolution for better feature distinction. The fused image was generated from the multi-spectral and panchromatic Landsat ETM+ bands. A RMS error of 7.1537 meters was obtained, with no error exceeding 30m. Data were re-sampled with the nearest neighbourhood method. The UTM south zone 36 projection was used. The small amount of cloud cover and shadows present on the images was masked out by on-screen digitizing.

1.7 PREPARATION OF ANCILLARY AND TRAINING DATA

Ancillary data may be defined as data acquired by methods other than by remote sensing techniques and are used to assist in the classification or analysis of remotely sensed data (Campbell 2002). Topographic maps of 1:50 000 scale were supplied by the Chief Directorate, Surveys and Mapping. The toposheets for the study area were printed in 1984 and the maps indicated cultivated, built-up and forested areas along with other topographic information. The cultivated, forested and built-up areas were digitized on-screen and converted to a grid of 30m resolution. Other ancillary data that were obtained included a 30m Digital Elevation Model (DEM), as well as geological, rainfall and temperature data. Data in shapefile format were converted to a 30m resolution grid to match the resolution of the Landsat ETM+ images. Image-derived ancillary data were also used. These included: texture measures (local, first-order and second-order measures); NDVI; and synthetic bands, created either by means of arithmetic operations or by transformations. Training data of the study area were acquired from the CSIR in point shapefile format. These were converted to polygon themes based on the information supplied in the point theme's attribute table. To do this, a procedure for ArcView 3.2 was written in the Avenue programming language. Some manual adjustments of the polygon shapefiles based on the composite colour images (4-5-3) were necessary in some cases.

1.8 TEXTURE MEASURES

Textures are considered to be homogeneous patterns or spatial arrangements of pixels that regional intensity or colour alone do not adequately describe (Debeir *et al.* 2002). Texture analysis is often used to introduce spatial information of object classes into the classification of satellite images. Texture images are derived from the satellite images and they may be either classified directly or used as an additional band together with other multi-spectral bands in a classification (Berberoglu *et al.* 2000; He & Collect 1999; Wulder 2002). Texture analysis is done by calculating the grey value relationships between the current pixel and the pixel next to it using texture measures such as mean, variance, contrast and correlation (see figure 1.2). The output image's grey values represent the local measure of texture of an input image. A more effective and rigorous second-order texture measure is the grey level co-occurrence matrix method (Haralick 1986 in Zhang 1999). In this method the co-occurrence matrix of grey values of the input image is first calculated, which involves transforming the image's space into a co-occurrence matrix. Grey value relationships are then calculated on the co-occurrence matrix space (Zhang 1999). There are a range of texture measures, including homogeneity, contrast, entropy, dissimilarity, angular second moment and inverse difference (PCI Geomatica 2003). Texture measures are not only influenced by scale, but also by the size of the object features in the image (Ferro 1998). In this study the texture images were calculated in ERDAS Imagine (ERDAS 2001) based on local texture measure (variance) of the panchromatic image, while contrast and dissimilarity measures, based on a grey level co-occurrence matrix (GLCM) of the ETM+ band 4, were done in PCI Geomatica V9.0. Texture analysis is based on a single channel; a suitable channel that provides the best contrast among land cover features has to be selected. In this study ETM+ band 4 was selected as the basis for texture analysis, because this band gives better contrast among land cover classes than the other ETM+ bands. The panchromatic band could also have been used based on grey level co-occurrence, but would have to be degraded or re-sampled to 30m resolution as required by the PCI EASI/PACE module, which is not ideal. Another channel that is often used (Berberoglu *et al.* 2000) is the first component of the principal component analysis (PCA) of the six ETM+ bands (bands 1-5 and 7). It was however not used in this study.

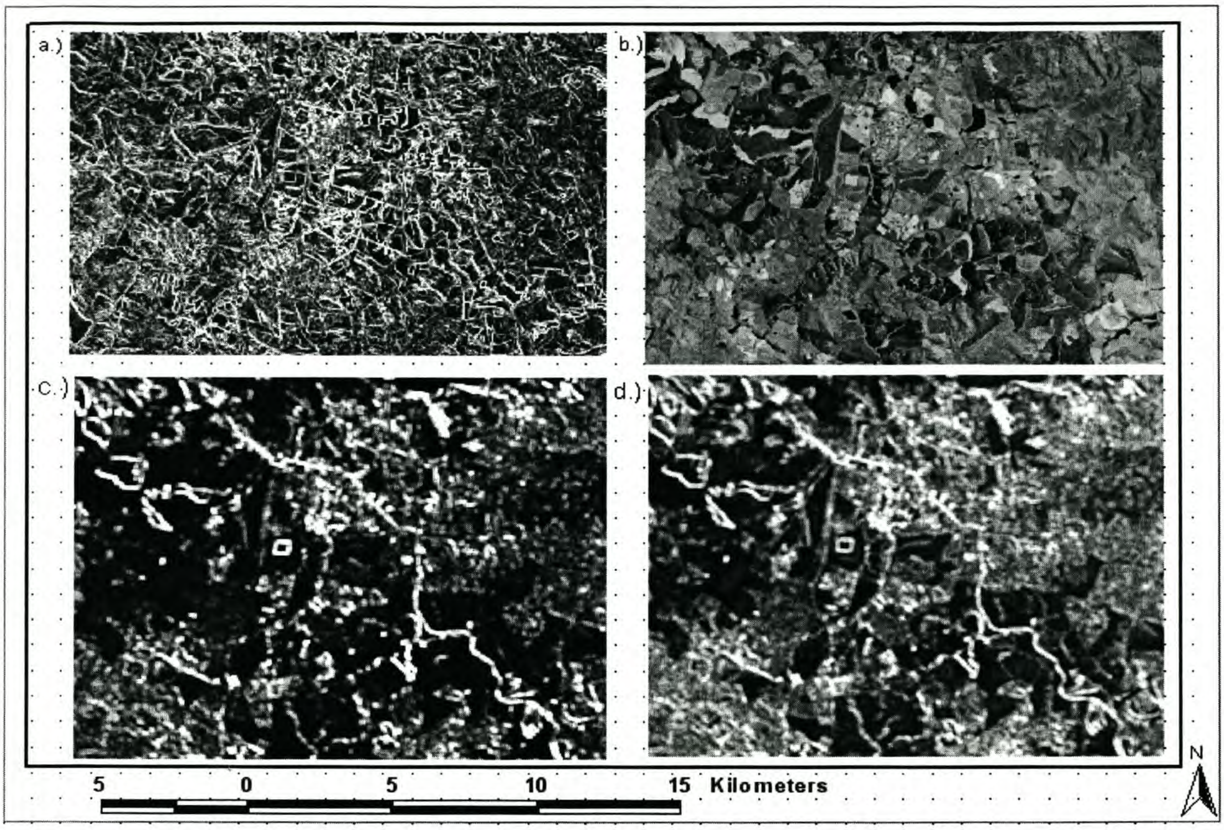


Figure 1.2: a) Variance, b) 4-5-3 composite image of the subset image, c) GLCM contrast and d) GLCM dissimilarity texture measure.

Textural classifiers are often used for urban land cover classification, which is characterised by a high heterogeneity within a pixel (Moller-Jensen 1998). The different texture measures are useful for contextual spatial properties of the various land covers. As a result most classification processes derive and use texture measures for improved classification accuracy. In the following chapter various classification techniques are discussed and reviewed.

CHAPTER TWO: REMOTE SENSING MAPPING OF LAND COVER AND A METHODOLOGICAL OVERVIEW

Remote sensing plays a fundamental role in land cover mapping and in long-term monitoring of changes in land cover and land use at multiple scales. Satellite remote sensing is used predominantly for large area land cover mapping. Diverse sensors are available with different technology and mapping capability (ACRES 2003).

The selection of an appropriate sensor for a specific land cover mapping and scale important implications for the accuracy of a classification process. In addition, the algorithm used to classify the particular satellite image is crucial.

This chapter discusses a range of topics that include remote sensors, image processing, pattern recognition, land cover classification schemes, image classifiers and land cover change monitoring. The main focus of this chapter is however the image classifiers and a variety of image classifiers are discussed and reviewed.

2.1 REMOTE SENSORS USED FOR LAND COVER MAPPING

The most commonly used satellite imagery for land use/cover mapping is acquired by the U.S. Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM) (including ETM+) sensors and the French Systeme Probatoire d`Observation de la Terra High Resolution Visible (HRV) Sensor (Smit 1993), which operate at different resolutions and area coverage (see Table 3.1).

It should be noted that the MSS system was designed less for land use/land cover mapping purposes than for geologic applications. The MSS has four spectral bands for which data are acquired and these four bands tend to be redundant. Bands 4 (0.5 to 0.6 micrometer) and 5 (0.6 to 0.7 micrometer) are the only two visible bands, and bands 6 (0.7 to 0.8 micrometer) and 7 (0.8 to 1.1 micrometer) are reflective infrared bands, providing information on vegetation and water resources, with band 7 being the superior one for this purpose.

The thematic mapper has a smaller pixel size than the MSS; it acquires data in eight bands, of which one was selected for geologic purposes (Lindgren 1985). The recently launched (by Space Imaging in

1999) IKONOS satellite has already started producing imagery that are black and white, colour and multi-spectral at higher resolution (~ 4m); with a scene size of 11x11km (Smith 2003).

Table 2.1: Principal specifications of common satellite media.

Specification	Landsat MSS	Landsat TM	SPOT XS	SPOT pan	IKONOS
Spatial resolution (m)	80	30	20	10	~ 4
Bands	4	7	3	1	multi
Area cover (km)	185 x 172	185 x 185	60 x 60	60 x 60	11x11

Source: Adapted from Edwards & Mumby 2000:66.

The ETM+ is the newest in the Landsat remote sensing satellites series. ETM+ has an additional panchromatic band with 15m resolution and a thermal band of an increased resolution (60m) compared to its predecessors. This sensor also has a five percent absolute radiometric calibration (ACRES 2003).

Table 2.2: Radiometric Characteristics of the ETM+ and TM Sensors

Band Number	Spectral Range (Microns)	EM Region	Generalized Application Details
1	0.45 – 0.52	Visible Blue	Coastal water mapping, differentiation of vegetation from soils
2	0.52 – 0.60	Visible Green	Assessment of vegetation vigor
3	0.63 – 0.69	Visible Red	Chlorophyll absorption for vegetation differentiation
4	0.76 – 0.90	Near Infrared	Biomass surveys and delineation of water bodies
5	1.55 – 1.75	Middle Infrared	Vegetation and soil moisture measurements; differentiation between snow and cloud
6	10.40 – 12.50	Thermal	Thermal mapping, soil

		Infrared	moisture studies and plant heat stress measurement
7	2.08 – 2.35	Middle Infrared	Hydrothermal mapping
8	0.52 – 0.90 (panchromatic)	Green, Visible Red, Near Infrared	Large area mapping, urban change studies

Source: (ACRES 2003)[1]

The new generation satellites are expected to provide unprecedented levels of spatial detail and, among other things, these data will be of particular use in urban studies, for which the relatively small size and complex spatial pattern of the component scene elements (e.g., building, roads and intra-urban open space) has had a limiting effect on the value of the previous generation's space-borne systems (Barnsley & Barr 1996 in Barr & Barnsley 2000).

2.2 IMAGE PROCESSING

Digital image processing consists of the computational processes applied to the image matrices with the aid of algorithms. The purposes of image processing are many and varied; however, they may be classified as follows:

- Image encoding: data compression or data reduction for efficient and reliable transmission or storage.
- Image enhancement: the processing of images to facilitate visual interpretations and further digital image processing.
- Image restoration: the removal or reduction of degradations (e.g., noise and distortions) that were incurred while the image was being obtained.
- Image analysis: extraction of information for measurements, pattern recognition, image interpretation (Gerbrands 1993; Mumby & Clark 2000).

[1] This table is available at the website: *http://www.ga.gov.au/acres/prod_ser/landdata.htm*

Encoding, improvement or restoration finally results in a new image, and this process can be described as image-to-image transformation. In contrast, computer-supported image analysis can be described as image-to-data transformation (Gerbrands 1993).

2.3 PATTERN RECOGNITION AND CLASSIFICATION

As part of pattern recognition, image interpretation is the recognition and identification of terrain objects using a digital approach. Terrain objects are characterized by their nature (thematics) and their position, shape and size (geometry). The nature of objects is determined by the spectral signature as represented in a remote sensing (RS) image (spectral pattern recognition); while their geometry is determined by the pixels making up the image of the object (spatial pattern recognition) (Molenaar 1993; Richards 1986). The basic functions needed to recognize objects in images include: pre-processing, feature selection and detection, segmentation, description, recognition and classification (Argialas & Harlow 1990). Pattern recognition is the categorization of data into identifiable classes and this is done by the extraction of significant features or attributes of the data (Tou & Gonzalez 1974 in Argialas & Harlow 1990).

Image information may be described at many levels of abstraction and the descriptions associated can range from one in terms of meaningful attributes of the scene captured in the image to one that describes only the spatial variation of intensity. These descriptions can be represented in a model that captures only the relevant features of the image at that level of abstraction, leaving the others unspecified. A model is helpful in converting the information in the image to usable forms, enabling the inference of objective properties of the objects under consideration (Argialas & Harlow 1990).

2.4 LAND COVER CLASSIFICATION SCHEME IN REMOTE SENSING APPLICATIONS

Spatial resolution of the satellite image affects the accuracy of land cover classification, and understanding the effect of scale on the spectral signatures of satellite data will help secure the correct interpretation of any classification results (Raptis, Vaughan & Wright In press). This effect is due to certain classes being spectrally heterogeneous at certain resolutions. A systematic framework for remote sensing-based classification is needed to avoid resolution-related errors of land cover representation. This hierarchical classification system is based on the assumption that there is a direct relation between the level of classification/categorization and the spatial resolution of the image data. For example, the U.S. Geological Survey classification scheme has four levels hierarchically organized, with each level having land cover classes that can be identified with at least 85% accuracy, from a specific group of sensors with similar spatial resolutions (Moller-Jensen 1998). Such a classification framework ensures that the final classification output structure and category definitions are appropriate for the objectives of the specific mapping. A significant problem associated with using a particular classification scheme is the lack of clear, precise and unambiguous class definitions. This lack may result in misinterpretation and erroneous data coding as well as difficulties of comparison of different thematic data sets based on the same classification scheme.

A similar systematic framework has been developed in South Africa, designed to suite the South African environment. This scheme has three levels (Level I, Level II and Level III) of which the first two levels are presented in Table 2.2. Level III categories are project specific. The standard classification scheme of South Africa is scale independent, while the expected operating range is between scales of 1:50 000 and 1:250 000 (Thompson 1996). The classes in the NLC 2000 field guide are a mix of land covers and land uses, and most land uses are poorly correlated with the imagery data, which makes it necessary to use ancillary data to isolate the land cover classes.

Table 2.3 Standard land cover classification for remote-sensing application in South Africa: class summary.

Level I	Level II
1. Forest and Woodland	Forest
	Woodland
	Wooded grassland
2. Thicket, bushland, scrub forest and high fynbos	Thicket
	Scrub forest
	Bushland
	Bush clumps
	High heathland (high fynbos)
3. Scrubland and low fynbos	Scrubland

	Low fynbos (heathland)
4. Herbland	
5. Grassland	Unimproved grassland
	Improved grassland
6. Forest plantations	Pine species
	Eucalypt species
	Wattle / other species
	Indigenous species
7. Water bodies	
8. Wetlands	
9. Barren lands	Bare rock / Soil
	Degraded land
10. Cultivated land	Permanent crops
	Temporary crops
11. Urban / built-up land	Residential
	Commercial
	Industrial / transport
12. Mines and quarries	

Source: (Thompson 1996:35).

2.5 SATELLITE IMAGE CLASSIFIERS

In land cover classification, land cover classes are the features to be extracted from the satellite images. This extraction is done with models that capture only the relevant features of the image at that level of abstraction. Models convert information in the image into usable forms so that inferences about the objects of interest can be made. Various models use different information abstraction and inference processes thus there are diverse approaches to pattern recognition such as mathematical or statistical, syntactic or structural, and heuristic or descriptive (Argialas & Harlow 1990).

The commonly employed pattern-recognition methodologies for land cover classification are statistical and contextual models (Molenaar 1993; Argialas & Harlow 1990). Of these, supervised, per-pixel, maximum-likelihood spectral classifiers are the most commonly used techniques in automated land cover classification (Bolstad & Lillesand 1991), while Foody (2000) used the fuzzy classification approach to improve accuracy. The per-pixel classification approach has however proved to be limited in nature and applicable only to spectrally well-differentiated cases (Lira & Maletti 2002; Buiten 1993).

Very recently, image interpretation techniques have been shifting from spectral classification to spatial, contextual classification, and to a recent and more powerful approach, namely knowledge-based interpretation (Argialas & Harlow 1990). The use of expert systems or knowledge-based systems within the field of remote sensing has been a topic of discussion in several studies and the trend is towards object-oriented methods, considering ancillary and multi-temporal data and spatial relations, with knowledge inferred via expert systems (Gumbricht, McCarthy & Mahlander 1996). Expert systems *for image processing* enable effective use of available image-processing techniques. This has to be distinguished from an *image understanding system,* which is designed for knowledge-based interpretation of visual scenes. Expert systems play an important role in simplifying the user's interaction with a complex search space, allowing the extraction of useful environmental information from complex sources of data, involving automatic data integration, and interpretation of results (Moller-Jensen 1998). As adjusted for a land cover classification, a knowledge-based system is composed of the following three main elements:

1.) A knowledge base: a set of simple facts composed of imagery and environmental information (database) and a set of rules describing relations between these facts.
2.) These rules are formalized through a learning phase, which aims at identifying distinguishable relationships between elements in the database and land cover type as represented in training areas. These relationships can then be generalized using expert knowledge and domain literature.
3.) A problem-solving mechanism (a recognition path): a set of production rules designated for accessing the knowledge base facts and rules, and which controls the way its rules are activated and utilized (Frost 1986 in Cohen & Shoshany 2002; Cordon, Del Jesus & Herrera 1999).

It has been shown that different classification algorithms can result in different classification products, even with the same training sets (Skidmore *et al.* 1997). This may be because algorithms implement procedures that recognize patterns based on certain properties of images, for example, contextual and spectral classifiers. The incorporation of ancillary data has proven not only to increase the accuracy but also the consistency of classifications (Marble & Peuquet 1983 in Bolstad & Lillesand 1992).

Incorporation can take place before, during or after the image classification process. Most classifiers do not support automated integration of non-image spatial data in image classification (Hutchinson 1982). Any such attempt results in inflexible hard-coded classifiers, which violate distributional conditions (Bolstad & Lillesand 1992; Mehldau & Schowengerdt 1990). However, classifiers such as expert systems and neural networks provide the means by which ancillary data can be incorporated into the classification process, allowing the utilization of the full information range of the ancillary data. Most non-expert approaches, such as maximum-likelihood classifiers, are self contained (Tsatsoulis 1993). In other words they can start from raw data and produce complete and correct classifications, therefore allowing only limited user interaction during the classification process.

The use of expert systems or knowledge-based systems for land cover classification has been investigated by several workers (Moller-Jensen 1998). Tsatsoulis (1993) gives an excellent review of expert systems in remote sensing applications. According to Tsatsoulis, expert systems applications in remote sensing have been classified into: (1) user assistance systems, (2) classifiers, (3) low-level processing systems, (4) data fusion systems, and (5) GIS applications. Expert systems classifiers may operate on the pixel level by applying rules to each pixel in the image or at regional level. Even though expert systems are not complete on their own, they have been found to improve land cover classification by allowing the integration of non-image spatial data (Argialas & Harlow 1990), and their flexibility, generality and intuitive appeal make them viable for remote sensing application (Bolstad & Lillesand 1992).

There are numerous reports in the literature on the usefulness of experts systems in remote sensing land cover classification. Sader, Ahl & Liou (1995) assessed the accuracy of GIS model/expert system classification against three other classifiers: unsupervised, supervised and Tasseled Cap classification. They found that GIS model classification improved accuracy over the unsupervised classification by 8%. Although this is not significant, the kappa value was the highest for the GIS model. The reasons for the low level of improvement could be attributed to many factors, of which the most crucial could be the use of inappropriate and/or inaccurate input explanatory variables.

Gumbricht, McCarthy and Mahlander (1996) employed expert systems for land cover classification in Cyprus. The expert system was pixel-based using a maximum-likelihood classifier based on Landsat TM data (bands 3, 4 and 7) and knowledge rules that considered Landsat-MSS data, elevation and geology. Manual and automatic (by extracting statistical data from training sets) representations of knowledge were tried and rules could be modified or added iteratively by domain experts for improved classification. Their findings showed that manually inferred knowledge rules performed better than automatic rules.

An expert system developed by Stefanov, Ramsey and Christensen (2001) had a primary objective to reclassify the initial maximum-likelihood classification and reduce errors of omission and commission. The model used TM bands together with ancillary data and an overall accuracy of 85% was attained. The model also achieved higher user accuracy for some land cover classes; validating the use of expert systems for land cover classification.

A study by Liu, Skidmore and Oosten (2002) integrated expert system classification with other sophisticated classifiers such as neural network classifiers (NNC). The study's objective was to investigate whether integrating individual classifiers improved classification accuracy. The integrated classifier produced better results than the maximum-likelihood, expert system and neural network classifiers when applied individually. Such findings reinforce the idea that incorporating complete, correct and relevant expert knowledge may lead to improved land cover classification.

Bolstad and Lillesand (1992) demonstrated that the integration of satellite imagery, thematic spatial data and artificial intelligence (AI) resulted in a significant increase of accuracy in land cover classification. In summary, rule-based expert systems had several advantages in comparison to standard approaches:

- the domain of discourse and control information is provided in an easily modified and understandable set of rules;
- specific feature type, thematic variable and image classification information can be persistent across different classifications of the same area, and can be modified for use in other regions or with different feature types;
- computationally expensive operations can be avoided using restriction operators, without resorting to manual image recoding, masking and image recombination;
- the modular rule-based approach allows the integration of evidential and deterministic discrimination techniques, and the incremental addition of new spatial data operators,

thematic data or knowledge, which aid land cover classification (Bolstad & Lillesand 1992; Gegg, Gunther & Riekert 1990).

Aware of the advantages of expert systems, Gegg, Gunther and Riekert (1990) developed a knowledge-based system for the extraction of environmental information from multi-spectral raster image data. Knowledge-based software architecture, which implements the integration of an image processing system, a geographic information system and an expert system, was chosen. The software, which they called RESEDA, contains a suit of image-processing operators, and the rule base contains the necessary strategic knowledge to apply the appropriate models for a required computation. The GIS serves as a representation framework and as a user interface to access the required information. The GIS software RESEDA was implemented in the SICAD-HYGRIS (Siemens Computer Aided Design-Hybrid Graphic Information System).

2.6 LAND COVER CHANGE DETECTION AND MONITORING

Human land use is an extremely dynamic process. Land use inventorying provides a basis against which to measure future changes in land use and to assess its temporal variability. In addition, land cover change monitoring is one of the objectives of long-term ecological research (Stefanov, Ramsey & Christensen 2001; Lindgren 1985). Land cover change monitoring requires a suitable methodology. Digital land use/cover change detection is a complex procedure and the accuracy may be as low as 50 percent or even less for some individual categories (Lindgren 1985). Digital change detection assumes that land cover change of a particular parcel of land will accordingly lead to a change in the spectral response of that parcel (Mongkolsawat & Thirangoon 1990; Lindgren 1985). A number of algorithms, including image differencing, image rationing, classification comparison and change vector analysis are used for change detection analysis. Algorithms operate differently, for example image differencing calculates the difference in reflectance values between different date images on a pixel-by-pixel basis, while image rationing calculates change by means of compensating for the difference in sun angle, sunlight intensity and shadows between data sets of different dates. Therefore the choice of using a particular algorithm should be based on factors such as familiarity with the region, precision of image registration and the algorithm's behaviour (Lindgren 1985). Change detection based on a thematic map should be treated with caution and intelligent approaches have been suggested which draw upon a broader knowledge of the directions, patterns and scale of the changes to be recorded in order to refine the assessments (Fuller, Smith & Devereux In press).

As the literature shows, expert systems in remote sensing application are being used increasingly for improved accuracy and consistency. The following chapter deals with the modular approach to expert systems for land cover classification developed in this study.

CHAPTER THREE: CLASSIFICATION MODEL, EXPERT SYSTEMS APPROACH

The modular expert systems approach is flexible as it allows for the modification and improvement of individual rule-based primitive models, while incorporating satellite and GIS data (Sader, Ahl & Liou 1995). An expert model has several advantages over other satellite image classifiers: a) it is non-parametric and therefore independent of the distribution of class signature; b) it can handle both continuous and nominal variables; c) it generates interpretable classification rules; and d) it is fast to train and is often as accurate and sometimes more accurate than many other classifiers (Hansen, Dubayah & DeFries 1996; Huang et al. 2002 in Huang *et al.* 2002a). The ability to handle data measured on different scales is another striking advantage (Pal & Mather 2002).

In this study the developed expert classification model consists of two hierarchical levels of sub-models: basic models and compound models (see Figure 3.1). Basic models constitute the lowest level of the expert classifier structure. Examples of these include NDVI, band ratios and texture. Basic models output results that can be used by compound models.

Basic models use algorithms and/or rules. Rules are simply recipes that can be followed to create a data product. They have a condition part that can contain one or more antecedent clauses and an action part (the consequent) that creates a data product (Argialas & Harlow 1990).

Rules were also used in the compound models. Each compound model consists of several basic models and may take as input output of other compound models and was designed to identify a single land cover class. Compound models may include intermediate results; data products derived in the process of identifying a certain land cover class. Intermediate geographic data are common in complex GIS models, and these by-products have great potential for use in other similar models (DeMers 2000). Some intermediate results were used as input to other compound models.

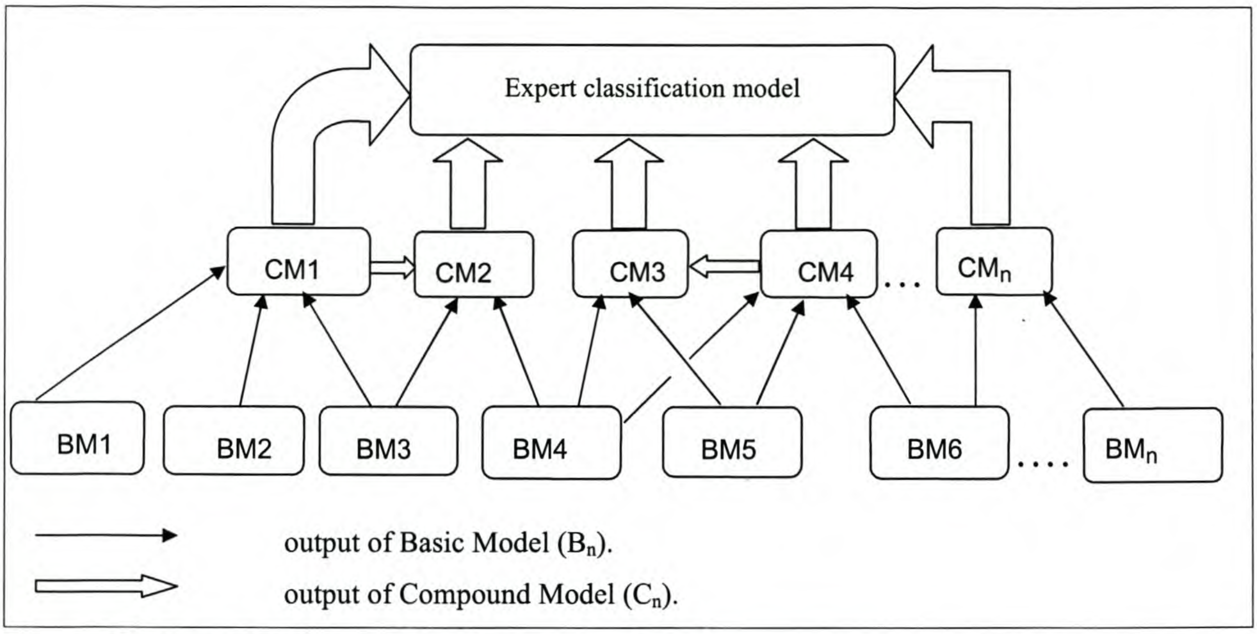


Figure 3.1: Expert classifier model structure.

3.1 SELECTION OF DATA LAYERS (VARIABLES) AND DETERMINATION OF THRESHOLDS FOR RULES

Since in modular expert systems each spatial operation model, also called a compound model, is designed to identify a single land cover category, the data layers used in the model need to be appropriate explanatory variables of the response variable (land cover). Explanatory variables could be spectral or ancillary data, continuous or categorical. Explanatory variables must be analyzed to find which thresholds best identify the response variable under investigation. One way of finding the thresholds is by applying expert knowledge. Another common and automated approach is statistical analysis. An example of such a statistical technique is the Classification And Regression Tree (CART) analysis, available in widely used statistical packages such as S-Plus. CART, using preset criteria, analyses explanatory variables by recursively splitting the data until terminal nodes (land cover categories) are obtained. The result is a dichotomous decision or classification tree. This classification tree can be viewed as a series of rules that may be used for predicting unknown response variables to likely class membership (Lawrence & Wright 2001). The use of a decision tree as a classifier has been explored within the context of global or continental-scale land cover classification. The majority of these studies have used data acquired by the Advanced Very High Resolution Radiometer (AVHRR) instrument at fairly coarse spatial scales ranging from 1 degree to 1 km (cell resolution).

A typical study would use AVHRR data acquired over a year and the classification would be based on either the temporal evolution of vegetation growth (phenology) using the NDVI index or individual spectral bands of the AVHRR as attributes for the classification (Brown de Colstoun *et al.* 2003).

A hybrid method was used in this study. This means that both expert knowledge and some statistical calculations were applied in determining suitable thresholds (to be used in the rules) for predicting response variables. Expert knowledge was used in cases where there is an expert understanding of the relationship between the explanatory and response variables. In cases where no accurate expert knowledge of the relationship between the explanatory and response variables is available, means and standard deviations were calculated from training samples of the particular response variable under investigation. Supervised classification was used for discriminating between classes that can be clearly differentiated spectrally, the result of which could be used as a data layer in the spatial operations.

3.2 KNOWLEDGE ENGINEER AND AGGREGATION OF LAYERS

The knowledge engineer in ERDAS Imagine 8.6 (see figure 3.2) is a powerful tool that can run several models to produce a single classification result. It implements a decision tree approach, called forward chaining in artificial intelligence (AI) terminology, to perform classification[2]. Decision trees and other expert systems such as neural networks are non-parametric in nature and can easily fit a variety of situations, whereas parametric classifiers have to meet certain statistical conditions (such as normality), which unfortunately most land cover objects do not assume (Brown de Colstoun *et al.* 2003; Pal & Mather 2002).

Although decision trees do not recover from classification errors like neural networks (NNC) and fuzzy classifiers do, they are computationally efficient and less expensive. Because of their hierarchical structure, decision trees give the analyst a simpler yet robust method to interpret, test and analyze the results (Brown de Colstoun *et al.* 2003).

---

[2] In rule-based systems two types of rules can be used: forward chaining and backward chaining. Forward chaining is used to establish new facts or hypotheses by matching rules against facts, whereas in backward chaining the system starts with what it wants to prove and proceeds to establish the facts it requires to prove it (Argialas & Harlow 1990).

Even though the knowledge engineer is a powerful tool, it has certain limitations. It firstly does not allow the use of several separate conditions in a hypothesis (a component model within the knowledge engineer that is equivalent to a compound model in this study) and secondly, the hypothesis must use input layers that are mutually inclusive (with the same geographical extent and pixel resolution and pixels must have values) to identify a single land cover. Since the compound models developed in this study included multiple conditions, input layers were not necessarily mutually inclusive. To solve this problem, each compound model was designed and executed separately.

The compound models were implemented in ERDAS modeller where the intermediate steps were simplified using separate spatial operations and the resulting layers were aggregated into suitable classes to match the NLC Level I categories. Results obtained by each compound model are mutually exclusive. In other words, there are no pixels that are members of two land cover classes. The knowledge engineer was used to combine all the results into one final land cover layer.

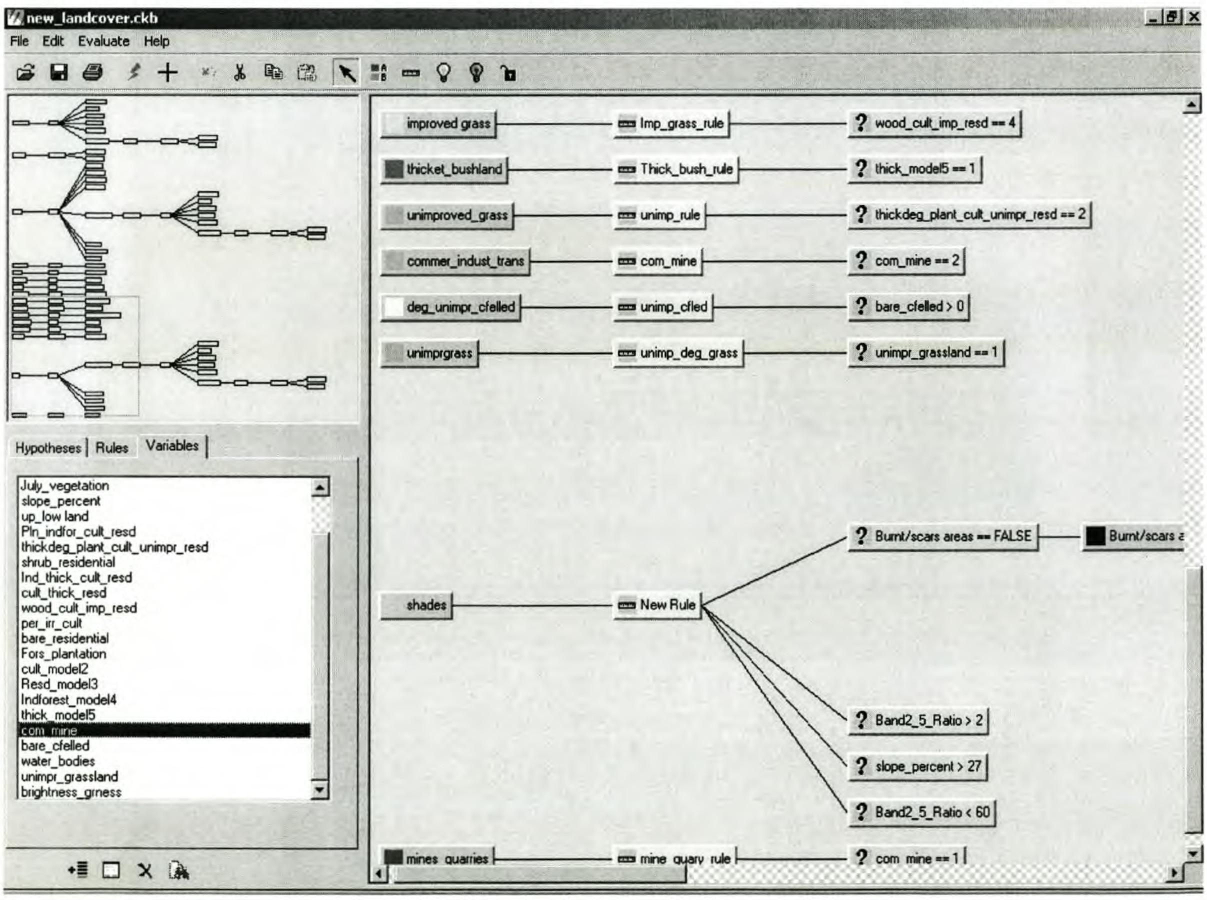


Figure 3.2: Example of the knowledge engineer window.

3.3 COMPOUND MODELS

A defined set of spatial operations is needed to constitute the image classification model. A classification model is defined by the sequence and combination of these compound models (Bolstad & Lillesand 1992). For this expert model, various compound models were used to build the classification model.

In expert models that apply progressive layered classification, the execution order of the compound models is important to ensure the mutual exclusiveness of the pixels that have to be assigned to a particular land cover class. It is logical and often practised to start a classification process with the less difficult and relatively easily identifiable classes. The logical order of classification operations is a key factor in the accuracy of the expert models. This is significantly important within compound models. Identified pixels are excluded from the next classification process by masking. This technique is applied in this study and the following compound models were developed and executed in the order they are presented:

1. Compound model 1: Water bodies;
2. Compound model 2: Burnt and fire scarred areas;
3. Compound model 3: Wetlands;
4. Compound model 4: Bare soils and degraded land;
5. Compound model 5: Forest plantations and commercial indigenous forests;
6. Compound model 6: Indigenous forests and woodlands;
7. Compound model 7: Coastal forests and woodlands;
8. Compound model 8: Thicket and bushland forests;
9. Compound model 9: Shrub land;
10. Compound model 10: Commercial permanently irrigated cultivated land;
11. Compound model 11: Temporary cultivated areas and unimproved grasslands;
12. Compound model 12: Residential areas;
13. Compound model 13: Mines and quarries;
14. Compound model 14: Commercial, industrial and transportation areas; and
15. Compound model 15: Shade.

These compound models incorporate one or more procedures based on basic models. The following sections discuss each compound model in terms of the techniques and basic models that it utilises.

3.3.1 Compound Model 1: Water bodies

Water bodies in the study area include farm dams, lakes, rivers and the ocean. In terms of its spectral characteristics water has low overall brightness and low reflectance in band 5 (near infrared). During summer days water bodies also have a lower temperature than the surrounding land.

Because of these qualities water is considered to be relatively easily identifiable on satellite images. This may not always be true due to suspended particulate matters on the water bodies that can cause considerable variation in the spectral and thermal signature (CGA 2003). In addition depth, dissolved particles and even substrate matter of the water bodies may increase variation. Foamy waters caused by wave action on seashores can cause high brightness and further complicate identification. Water can also be spectrally confused with opencast mines, shade and commercial and industrial land uses. Identifying water bodies based only on spectral bands may therefore not be accurate.

Topographical and hydrological models can be used to support the identification of water bodies. Such models are based on Digital Elevation Models (DEM) and may involve the calculation of topographic indices such as flow accumulation and slope. In addition to these explanatory variables, texture is another important discriminatory variable as water bodies are expected to have very little texture.

For compound model 1, water bodies were identified using a combination of spectral, topographical and spatial (distance) techniques. The following basic models were used:

Basic Model 1: Winter moisture

As can be seen in Figure 3.3 the spectral reflectance of water in ETM+ bands 2 (green) and 5 (near infrared) differs considerably. A simple ratio between these two bands should therefore highlight areas of high moisture. As non-water pixels are expected to show low moisture during winter, the ratio was calculated using the July image. The result was compared to known data and it was found that a ratio value of more than ten (band 2 has a reflectance of more than ten times that of band 5) represented areas of potential moisture.

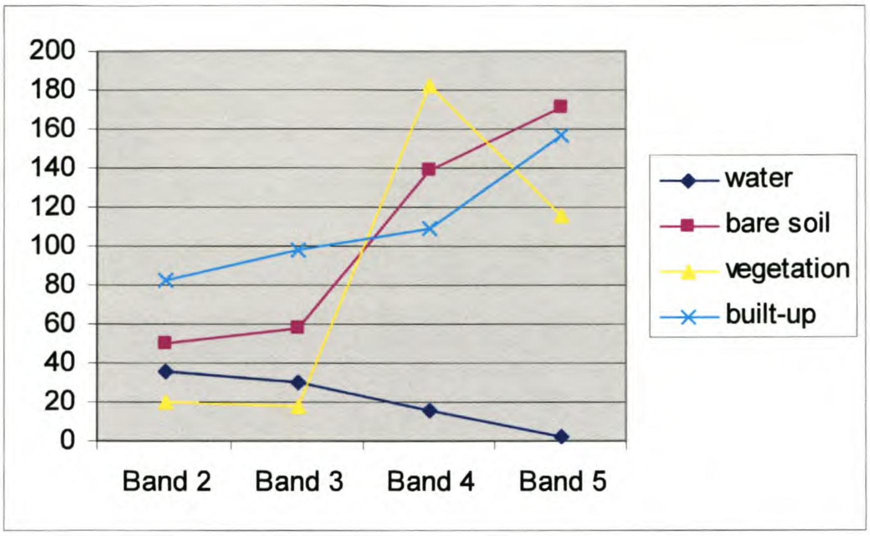


Figure 3.3: Spectral response of water, bare soil, vegetation and built-up areas

Basic Model 2: Summer NDVI

Since water bodies are not expected to be vegetated, one would expect areas of low vegetation during the summer to potentially represent water bodies. Vegetation's spectral reflectances for the different ETM+ bands are shown in Figure 3.3. These properties are used in the Normalized Difference Vegetation Index (NDVI) to indicate vegetation occurrence (Campbell 2002) and was subsequently calculated for the February image using formula 3.1.

$$NDVI = (TM4 - TM3) \div (TM4 + TM3) \qquad \ldots (3.1)$$

Based on known water bodies it was determined that NDVI values of below 0.1 potentially represented water bodies.

Basic Model 3: Slope percentage

Water bodies are usually found in depressions and areas with very low slope gradient and profiles. A DEM of the study area was used to calculate slope (in percent). Normally water bodies are expected to have 0% slope. But this is not the case with rivers, especially when the riverbeds are narrower than 2 pixel sizes, and slope for these pixels in the river course can be as high as 20% or more. The slope constraint was relaxed to include all the potential water pixels, while excluding shade (shadows of mountains) pixels. A slope of 27% was taken as the upper threshold.

Basic Model 4: Winter wetness

Water bodies are defined as perennial wet areas (Thompson 1996). Water bodies are expected to be wet even during the winter season when there is less precipitation. Seasonal wet areas should therefore be excluded. For basic model 4 and 5 a Tasseled Cap transformation was applied to the winter ETM+ image. This transformation condenses spectral information into meaningful thematic layers, the most important layers being 1, 2 and 3 representing scene brightness, greenness and wetness respectively. The coefficients used to calculate these transformations are those provided by Huang *et al.* (2002b) (see Appendix A). To differentiate water bodies from seasonal wet areas, the Tasseled Cap wetness layer (band 3) was used and a threshold of 130 was identified from known wet areas. Wetness values above this threshold were considered to be very wet and will potentially represent water bodies.

Basic Model 5: Winter brightness

Water bodies show low spectral reflectance. This characteristic is more pronounced in winter than in summer since the higher runoff during summer results in turbid and sediment laden water bodies. For discriminating, water bodies from other land cover classes, Tasseled Cap brightness was used. Brightness values of less than 70 were found to correlate well with known water bodies.

Basic Model 6: Shoreline distance

Foamy waters are mostly found within a few meters from the shoreline. This area constitutes a high-energy zone where a lot of wave action is present. To delineate this area the model calculates a buffer zone from the shore within which foamy waters are expected to be found. From visual inspection of satellite images it was decided that a 200m zone is sufficient. Anything within this distance was considered to potentially be a water body (the ocean).

Basic Model 7: NDVI change 1

NDVI and change in NDVI between seasons gives an important indication as to potentially what type of land cover a pixel can be. Most vegetation land cover types show significant NDVI variation. Water bodies on the other hand are expected to show no NDVI change and to maintain a negative NDVI value. Based on the ISODATA derived zones, NDVI change (February NDVI – July NDVI) between the two dates was calculated. ISODATA classification was performed to obtain 30 homogeneous zones in terms of texture, Tasseled Cap brightness, wetness and greenness. These layers were selected because of water's low reflectance, high wetness, low greenness and low texture properties. The texture used was a grey level, co-occurrence matrix based dissimilarity texture. Of the zones three showed negative mean NDVI change and were taken as potential water bodies.

Figure 3.4 illustrates how the basic models were combined to identify water bodies. The first procedure uses basic model 1 and 3 to identify the majority of the inland water bodies. The results showed that some water bodies were not detected, especially the ocean.

The second procedure identifies foamy waters. For this, basic model 1, 6 and 2 were utilized. Pixels within the buffer zone (200 metres) were classified, as foamy water if the winter moisture was greater than 1.8 and the NDVI was less than 0.1.

The remaining undetected water bodies, which include the ocean, estuaries and lagoons were identified by procedures three and four. Procedure three employs basic model 4, 5 and 7. Here pixels were identified as water if the winter wetness was greater than 130, winter brightness less than 70 and the zonal mean NDVI change less than 0.

The final procedure utilized basic model 3, 5 and 7. Pixels with winter brightness of less than 28, slope of less than 27% and a zonal mean NDVI change of less than 0 were considered to be water bodies. The outputs of the four procedures were combined to give the final layer representing water bodies. The result was compared to a supervised classification and was found to represent water bodies much more accurately. Pixels identified in this compound model were excluded in the further data processing. The next class identified was fire-scarred areas.

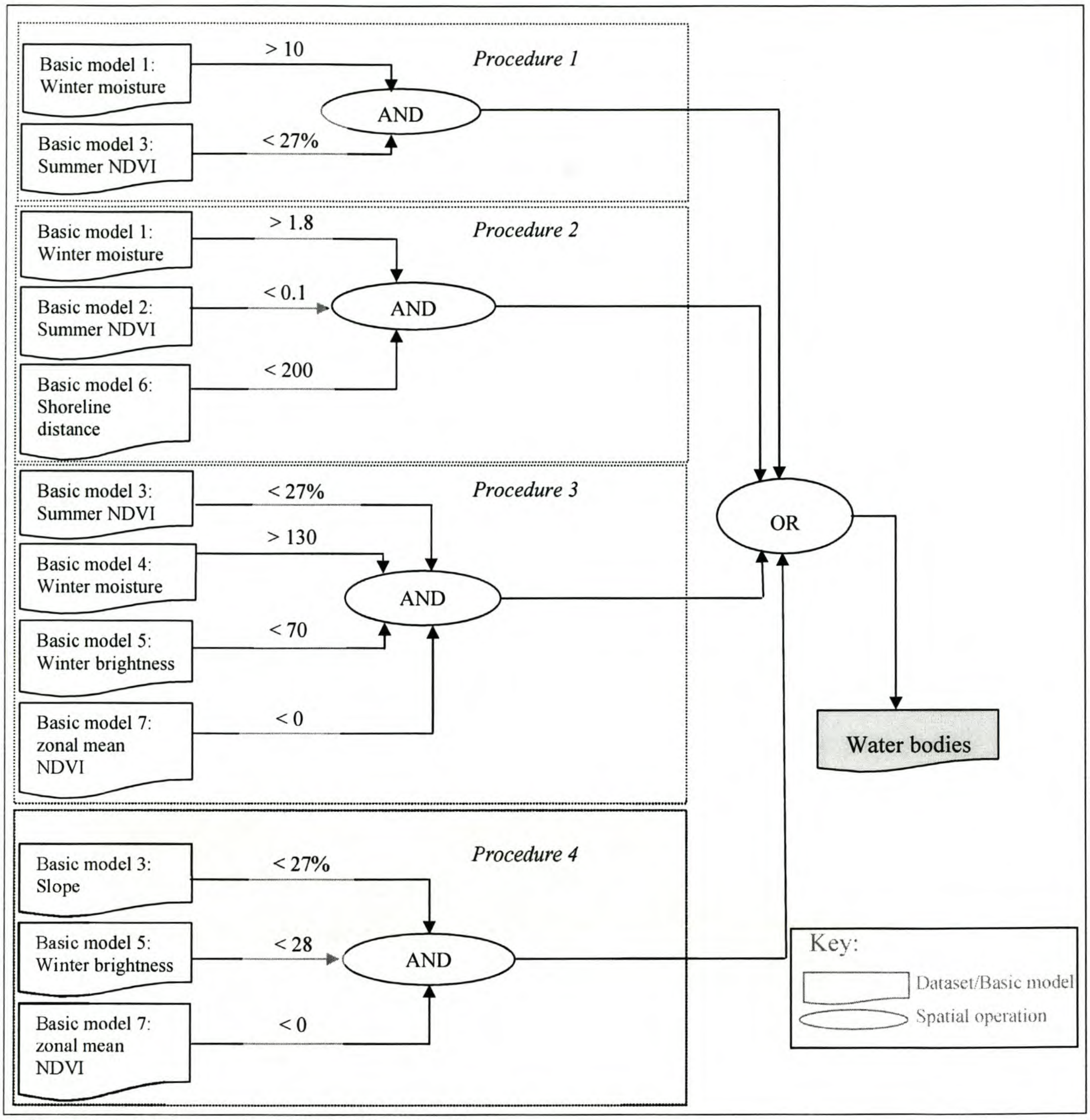


Figure 3.4: Compound model 1: Water bodies

3.3.2 Compound Model 2: Burnt and fire-scarred areas

Burnt areas were included as a class on its own, since the study area had a number of fires during 2001. Wild fires have a significant impact on the dynamics of vegetation and can disturb ecosystems. Fires also destroy timber resources and fire monitoring is therefore a critical aspect of sustainable forest management.

Smit (2001) showed that Landsat data can be used to identify and monitor fire-scarred areas. The low vegetation cover and dark burn residue in fire-scarred areas cause low reflection in the near-infrared (TM4) band (Lawrence & Wright 2001). As a result burnt areas usually have negative NDVI values (Smit 2001).

Burnt areas can easily be spectrally confused with water bodies, quarries and mines, since they have similar spectral properties. Identifying burnt areas based on satellite images alone can therefore lead to poor results. Single date imagery should also be avoided since fire occurrences are seasonal and fire-scarred areas can show different spectral responses at different stages of recovery. The use of multi-temporal images and ancillary data is therefore ideal.

Smit (2001) illustrated the value of uncorrelated pairs of bands in fire-scar mapping. To identify fire-burnt areas, an algorithm based on the NDVI equation was used. Good results were obtained by substituting the red and infrared bands with TM bands 4 (near infrared) and 7 (middle infrared). The formula used is as follows:

$$X = (TM4 - TM7) \div (TM4 + TM7) \qquad \text{...(3.2)}$$

Smit (2001) extended this formula to include the brightness and wetness bands derived from Tasseled Cap Transformation. The equation of the uncorrelated band pairs has a similar mathematical statement as in formula 3.2:

$$X = (Brighness - Moisture) \div (Brightness + Moisture) \qquad \text{...(3.3)}$$

When applied in the study area, these techniques did not produce acceptable results. This is probably due to the difference in vegetation, as Smit's (2001) study was specifically focused on fynbos vegetation. In addition, the underlying soils are also considerably different. The fire scars in Smit (2001) was on limestone soils, which produced fire scars that were in some cases lighter than the surrounding vegetated areas.

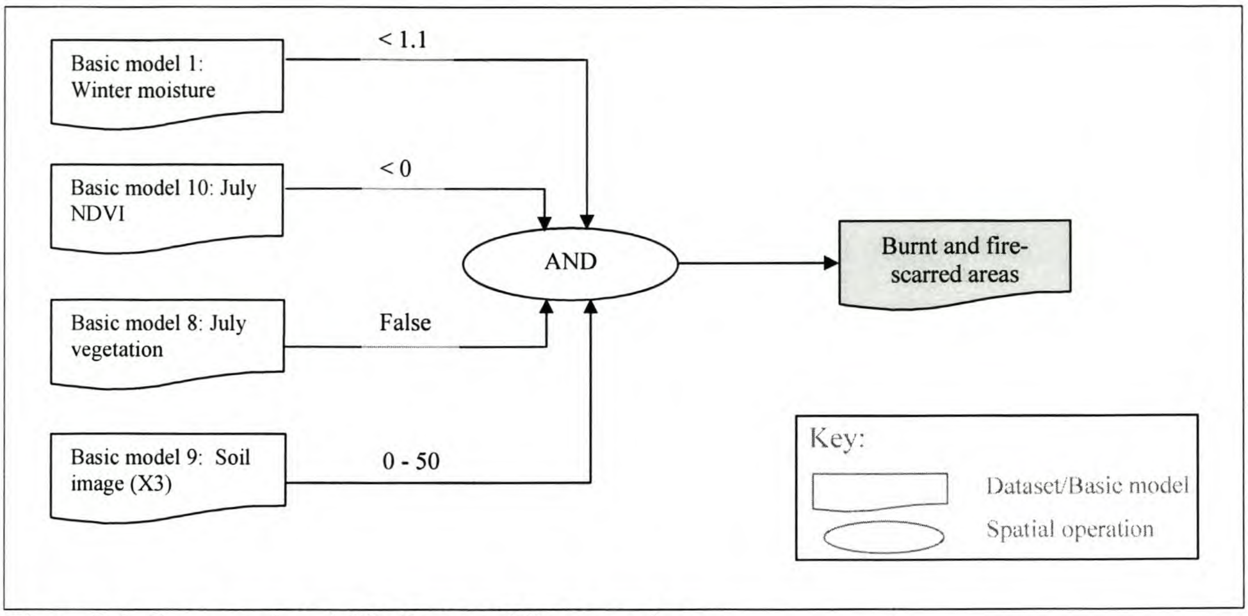


Figure 3.5: Compound model 2: Burnt and fire-scarred areas.

For this study a different approach was taken to identify fire-scarred areas. Compound model 2 (see figure 3.5) uses spectral and ancillary data. Rules were defined based on known burnt areas. In addition to basic model 1 (refer to section 3.3.1) the following basic models were employed:

Basic Model 8: July vegetation

Fire-scarred areas are usually dark and devoid of vegetation. Vegetated areas should therefore be excluded from further data processing. To do so, vegetated areas were identified on the winter image. Fire incidents predominantly occur in winter when berg winds blow from the interior plateau to the coast (Dilley *et al.* 2001). Figure 3.2 illustrates that vegetation pixels have a high reflectance in the ETM+ band 4 and a very low reflectance in band 3. Vigorous winter vegetation growth was obtained by the condition that if ETM+ band 4 (near infrared) is greater than ETM+ band 3 (red) and 5 (middle infrared), then the pixel is likely to represent vegetation. The resulting vegetation image was compared with known vegetated areas and it was found to be very representative. Areas with negative NDVI values were taken to potentially represent fire-scars.

Basic Model 9: Bare soils (X3)

As mentioned earlier, fire scars are usually dark in appearance. In this model the transformation developed by Shrestha (2000) was used to identify dark soils. According to Shrestha (2000), in a two-dimensional feature space defined by red and near-infrared bands, vegetation, soil and water occupy three distinct locations (see figure 3.6).

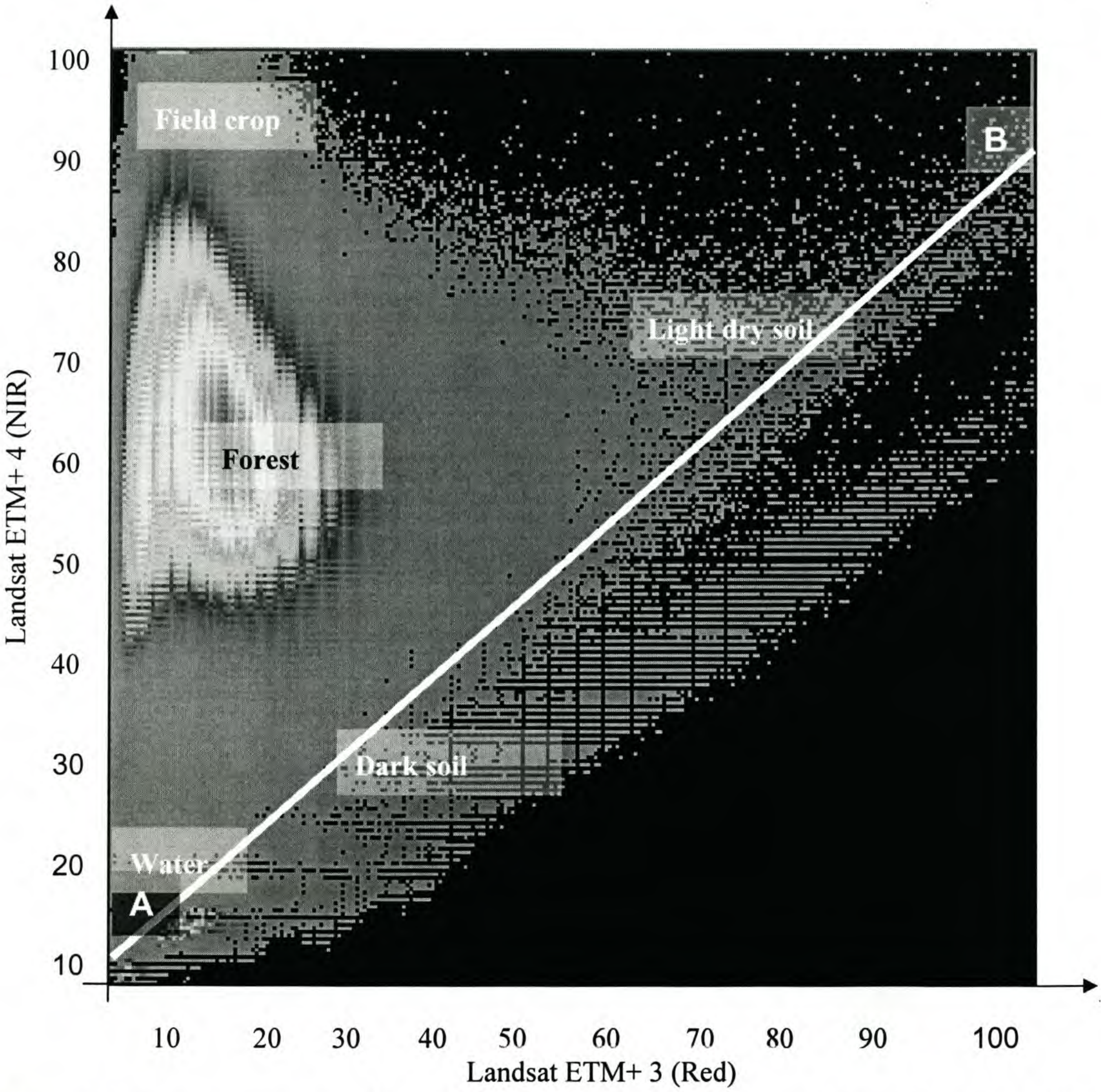


Figure 3.6: Feature space of TM4 and TM3 Source: Adapted from Shrestha (2000: 2).

The vegetation pixels occur in the upper left, implying high reflectance in the near infrared and high absorption in the red portion of the spectrum. Soil pixels occur in the upper right indicating high reflectance in the near infrared as well as red portion of the spectrum. Dark soil surfaces are represented in the lower-left of the feature space, while water bodies are found further down in the two-dimensional space. The diagonal line AB is considered to be the line showing variation in soil. Red and near infrared data can therefore be transformed into a soil index by means of band rotation.

The index is used to assist in soil feature mapping and it maximizes soil variation while suppressing spectral responses from vegetation. Formula 3.4 is used to transform the red and near infrared bands into a soil index. The formula achieves an anti-clockwise rotation of the bands through an angle to isolate soil pixels (Shrestha 2000).

$$\begin{pmatrix} X1 \\ X2 \end{pmatrix} = \begin{pmatrix} \cos\theta\sin\theta \\ -\sin\theta\cos\theta \end{pmatrix} \begin{pmatrix} ETM+ 3 \\ ETM+ 4 \end{pmatrix} \qquad \ldots (3.4)$$

This transformation creates two new bands, X1 and X2. Band X1 maximizes soil variation, while band X2 maximizes vegetation information. In order to find the suitable angle of rotation, a linear least square method must be applied on bare soil samples. To get the suitable angle of rotation for generating the soil line, 300 samples of bare soil surfaces were collected by visual observation from a true colour composite of the satellite images (bands 3-2-1).

Regression analysis of these samples showed a high positive correlation ($r^2 = 0.85$) between ETM+ band 3 and 4. Using the linear least square method, the slope and intercept points were calculated to find the best fitting line (formula 3.5).

$$y = ax + b \qquad \ldots (3.5)$$

Where b is the position where the line intercepts the Y-axis (ETM+ band 4) and a is the slope of the line. The values for a and b, based on the sample data, was 40 degrees and 14.5 respectively.

The algebraic form of equation (3.4) gives two distinct formulas (equation 3.6). The new bands, X1 and X2, were generated by substituting the cosine and sin values of 40 degrees in equation (3.4).

$$\begin{aligned} X1 &= 0.642 \ \text{Band } 3 + 0.642 \ \text{Band } 4 \qquad \ldots (3.6) \\ X2 &= -(0.642) \ \text{Band } 3 + 0.766 \ \text{Band } 4 \end{aligned}$$

To compensate for the small bias in the NIR band, which is the intercept value in the linear least square equation, a shift is applied in the NIR band by subtracting it from ETM+ band 4. The equations above were modified as follows:

$$X1 = 0.766 \text{ Band } 3 + 0.642 (\text{Band } 4 - b) \quad \ldots (3.7)$$
$$X2 = -(0.642) \text{ Band } 3 + 0.766 (\text{Band } 4 - b)$$

To get a soil image without water and vegetation, the following conditional expression was applied:

$$X3 = \text{if } X2 < 0 \text{ then } X1 \text{ ELSE } 0 \quad \ldots (3.8)$$

This expression generates a new band (X3) having only soil surface features with the rest of the area having pixels of value zero (see figure 3.7). By comparing the X3 band with known fire scars it was determined that values of between 0 and 50 corresponded very well with fire scars.

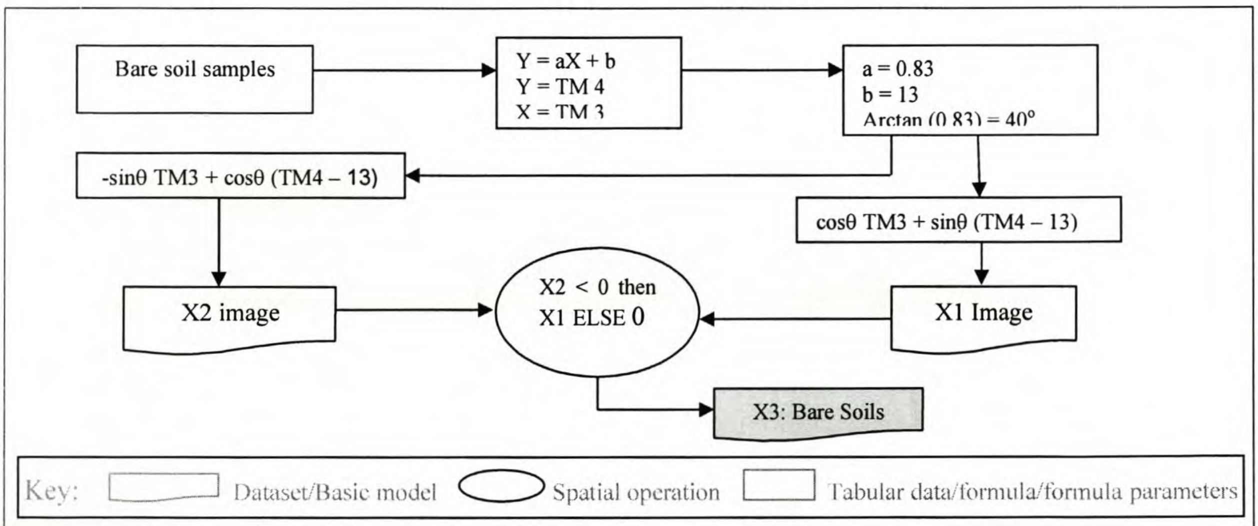


Figure 3.7: Basic model 9: Bare soils (X3).

Basic model 10: July NDVI
Burnt areas are expected to have low to negative NDVI values. NDVI was calculated for the July (winter) image, as burnt areas will recover more slowly during the winter season when precipitation is low.

A comparison between winter moisture (basic model 1) and fire scars showed that these areas have moisture values of less than 1.1. This observation together with the outputs of basic models 8 and 10 produced a result of acceptable accuracy when compared to the original images.

3.3.3 Compound Model 3: Wetlands

Various types of wetlands occur in the study area. These include seepage zones, saltmarsh, reedpan and highveld wetlands, which can be vegetated or not. Geographically they are scattered on the landscape and most of the individual wetlands are small in extent. Vegetated wetlands often have a similar physical appearance to the surrounding non-wetland vegetation, and in such circumstances can be difficult to identify based on single date spectral data (Thompson *et al.* 2002).

Wetlands usually occur on hydric soils, on lower slopes, in basins and near water bodies such as lakes, ponds and streams (Sader, Ahl & Liou 1995) and are often associated with either herbaceous or woody vegetation cover (Thompson 1996; Thompson *et al.* 2002). In terms of slope, Sader, Ahl & Liou (1995) showed that wetlands are likely to occur on slopes of less than 8% and that multi-temporal imagery is invaluable for the identification of seasonal wetlands such as dried pans.

An attempt was made to adapt the wetland model used by the national wetland inventory programme (Thompson *et al.* 2002), which uses the following procedures:

1) image classification using a combination of original and derived datasets (i.e. biomass and wetness indicators), in order to enhance seasonal differences in wetland and adjacent land covers' spectral characteristics, within each multi-temporal dataset;
2) terrain-based hydrological modeling to determine areas of 'potential wetness', where water, and thus wetlands, may be likely to accumulate, irrespective of land cover; and
3) spatial modeling to combine the terrain-based 'potential wetness' model with the image-derived wetland areas, in order to derive the final wetland distribution (Thompson *et al.* 2002).

For deriving a terrain-based hydrological model, a model called Landscape Wetness Potential (LWP) is recommended by Thompson *et al.* (2002). LWP makes use of a weighted overlay technique that combines several physical parameters influencing the formation of wetlands. The four parameters are:

occurrence of sinks or depressions, slope steepness, surface hydrological or flow accumulation, and relative slope position or topographic index (TPI) (Thompson *et al.* 2002).

An AML program for ARC/INFO GRID was given in Thompson *et al.* (2002) to derive TPI and it was implemented in ARC/INFO and rewritten in ERDAS modeller by the researcher. However, neither the AML in ARC/INFO nor its ERDAS equivalent produced a realistic result (the entire area was one value, i.e. 2) when executed on the DEM of the study area, which is approximately 184 x 143 km in size. When the programs were executed on a subset of the DEM, reasonable results were obtained. The algorithm is therefore not suitable for larger DEMs such as the one for this study area.

Thompson *et al.* (2002) recommended the TARDEM model for calculating flow accumulation. TARDEM is a suite of executable programs that can be downloaded from the Internet *(http://www.engineering.usu.edu/cee/faculty/dtarb/tardem.html)*. TARDEM was extremely slow (on a Pentium IV, 2.4 GHz computer) when applied to a DEM of geographical extent as large as the study area. For these reasons, as well as the fact that the wetland identification procedure requires some visual interpretation by the analyst during the unsupervised classification process, this approach was not pursued.

In this study a procedure that combines topographical as well as spectral properties was developed. For compound model 3 the following basic models were used:

Basic Model 11: Winter greenness

Thompson *et al.* (2002) noted that, for wetland classification, images taken during transitional wet-up or dry-down periods (when wetlands exhibit significantly different characteristics to the surrounding land-cover) are ideal, especially if the wetlands are primarily vegetated. Wetlands are less wet during winter, which can lead to lower densities in wetland vegetation.

To discriminate between vegetated wetlands and other spectrally similar vegetation the winter greenness band of the Tasseled Cap transformation (see section 3.2.1) was employed. Based on known wetland areas it was determined that winter greenness values of greater than 150 potentially represent wetlands.

Basic Model 12: Topographical profile

Wetlands occur in depressed and low profiled areas. A geomorphological model was adapted from Brabyn (1998) and executed on a 30m resolution DEM of the study area. As illustrated in the figure 3.8, to calculate the up-low land profile index the model derives slope (in percent), focal elevation range and focal maximum elevation based on a 5x5 moving window. Segmenting the landscape based on the combined index of these parameters has untapped potential for the prediction of many earth surface phenomena. To obtain the up-low land layer, the DEM elevation values are subtracted from the focal maximum elevation layer, which is then divided by focal elevation range. Pixels of the resulting layer are classified as up-land if the value is greater than 0.5 otherwise the pixels are low-land. This layer was combined with the slope layer, which was classified into classes of 0 – 15, 15 – 50, 50 – 60 and > 60 to produce the final topographical index. These class groups were adapted from Brabyn (1998) to match the landform components: plain/low hills, hills, high hills and mountains. Areas identified as low-land were considered to possibly represent wetlands.

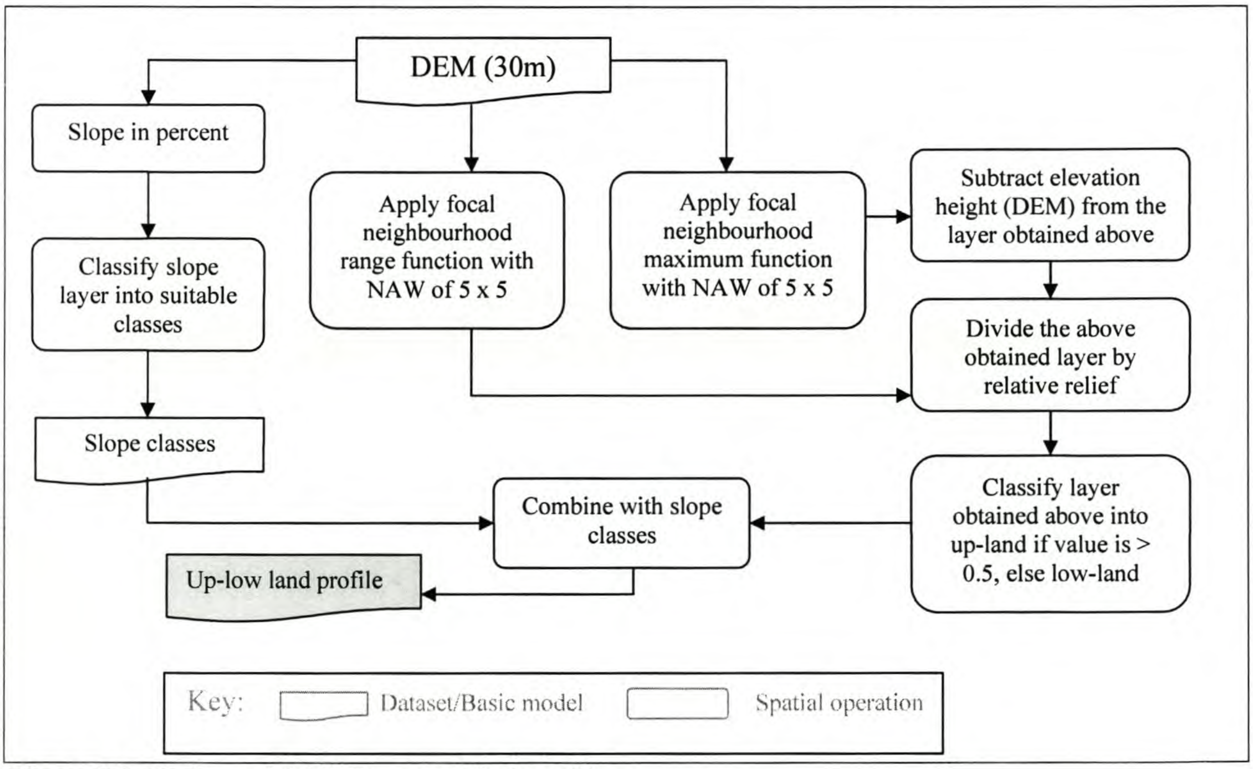


Figure 3.8: Basic model 12: Topographical profile.

In addition to these models, basic model 1, 3 and 8 were also used (refer to sections 3.2.1 and 3.2.2). Basic model 1 was used to determine the moisture content of a pixel, and pixels with moisture between

1.1 and 10 were considered to potentially represent wetlands. Since wetlands are predicted to occur on slopes of less than 8%, basic model 3 was employed to flag out potential wetland pixels that are below the threshold. Non-wetland vegetation were excluded by using basic model 8, as vigorous vegetation during the winter season are less likely to represent wetlands, when they are normally expected to be in their dry-down phase. Compound model 3 also utilized the output of compound model 1 and 2, in order, to exclude water bodies and fire scars from being considered as wetland areas (see figure 3.9).

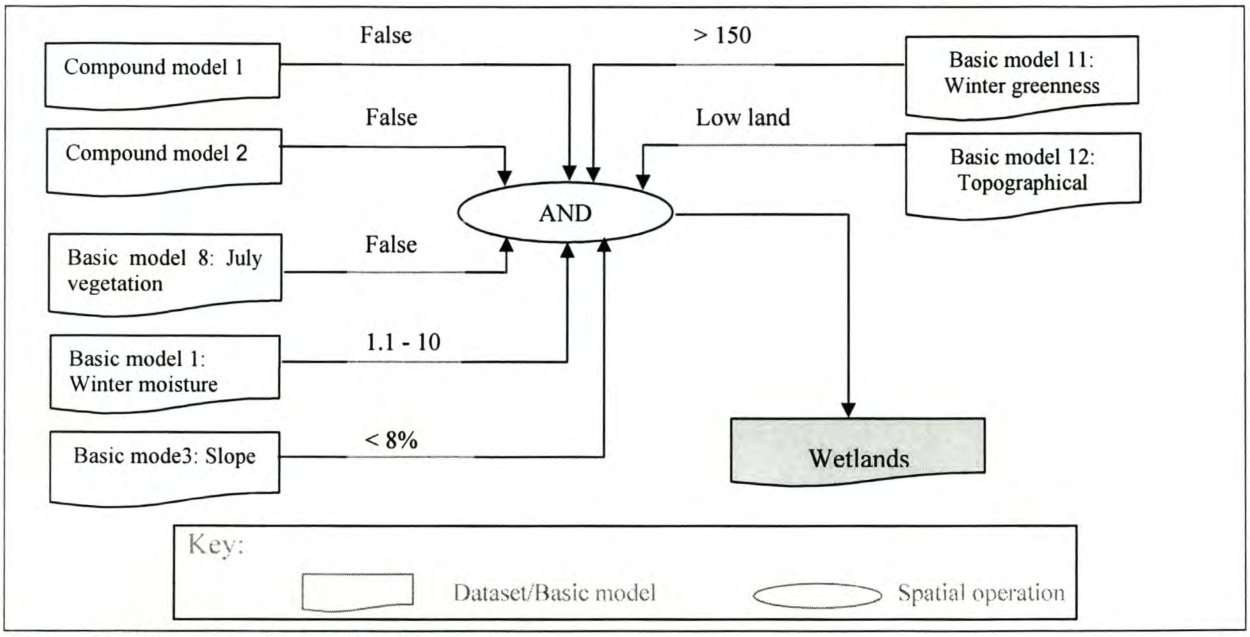


Figure 3.9: Compound model 3. Wetlands.

3.3.4 Compound Model 4: Bare soils and degraded land

Bare soils and rocks are naturally exposed sand, soil or rock with no or very little vegetation cover in any season. This category excludes agricultural areas without crop cover (CSIR 2000). The NLC 2000 field guide (CSIR 2000) describes degraded land as permanent or near-permanent areas of low vegetation cover induced by man; often caused by severe soil erosion. It is typically associated with subsistence agriculture. For this reason it is difficult to separate pixels of this category from agricultural fallow areas using spectral data only. A classifier that uses ancillary data is therefore ideal.

In this study bare soil and degraded land were identified using compound model 4 (see figure 3.10), which in turn employs the following basic models:

Basic model 13: February vegetation

As mentioned, degraded areas have very low vegetation cover throughout the year. Vigorously growing vegetation can therefore be excluded from further consideration. This model identifies areas of vigorous summer growth, by applying the following condition: if ETM+ band 4 (near infrared) is greater than ETM+ band 3 (red) and 5 (middle infrared), then the pixel represents vigorous vegetation. The summer image was selected as dry, seasonally un-vegetated areas on the winter image can easily be mistaken for bare soil.

Basic model 14: Built-up areas

Another land cover feature that can be mistaken for bare soil is built-up areas as these are hard surfaces that are devoid of vegetation. Fortunately, built up areas are usually well mapped and can therefore be easily excluded. Built-up areas were digitized from existing maps, at a scale of 1:50 000, obtained from Chief Directorate: Surveys and Mapping (CDSM) and converted to grid format.

To ensure that all built-up areas are excluded, buffered streets were used to supplement the built-up dataset. A buffer of 300m was identified as being an appropriate distance to include all built-up areas. The resulting layer was combined with the digitized built-up layer to form a layer that represents areas that should be excluded from further processing in this compound model.

Basic model 15: Cultivated land

Cultivated lands include fallow areas, which have similar spectral characteristics as degraded lands. Cultivated lands that may be found in the study area mostly constitute seasonal commercial or subsistence farming. This basic model attempts to avoid such confusion. Cultivated lands as shown on 1:50 000 topographical maps were digitized on-screen using ArcView 3.2 and the resulting shapefile was converted to a grid. These areas were not considered as degraded land.

Basic Model 16: Beach sand

Non-vegetated sand dunes found in the coastal areas are categorized as bare soils in the NLC 2000 field guide, even though they have different spectral properties. Beach sand is characterized by negative NDVI and high brightness. To identify beach sand a buffer was created around the coastal shoreline obtained from CDSM at 1:50 000. From visual inspection it was determined that beach sand does not occur more than 230m from the coastline. To differentiate between beach sand and other land covers in this zone, areas with a February NDVI (basic model 2) of less than 0.05 and winter moisture (basic model 1) of less than 1.8 were used to identify beach sand.

In addition to the basic models discussed here, basic model 2 (refer to section 3.3.1) and 8 (refer to section 3.3.2) as well as outputs from compound model 1, 2 and 3 were used to identify bare soil or degraded land. A mask and exclude approach was taken. Pixels that are least possible to represent degraded lands were excluded from further data processing. These pixels belonged to the output of basic model 8 (July vegetation) and 14 (built-up areas) as well as compound model 1 (water bodies), 2 (fire-scarred) and 3 (wetlands). From this exclusion one can expect the following land covers to remain: bare-soil, degraded land, mines, quarries and clear felled forest.

The next discriminatory step was to exclude vegetated areas using basic model 2. By close inspection of the summer NDVI image it was determined that pixels having a value greater than 0.05 represented cultivated areas, bare soil or degraded land and clear felled forest areas. By excluding these together with cultivated lands (basic model 15) and built-up areas (basic model 14) and by adding coastal dunes, a layer representing bare soils and degraded land was obtained.

In areas where there is little vegetation and high geological variation, the geologic index in formula (3.9) may be used to remove geological noise (CGA 2003) and to set rules. Since the study area is mostly covered by vegetation it was not used.

$$Baresoil\_index = \sqrt{(band7 - band2)/(band7 + band2)} \qquad \ldots (3.9)$$

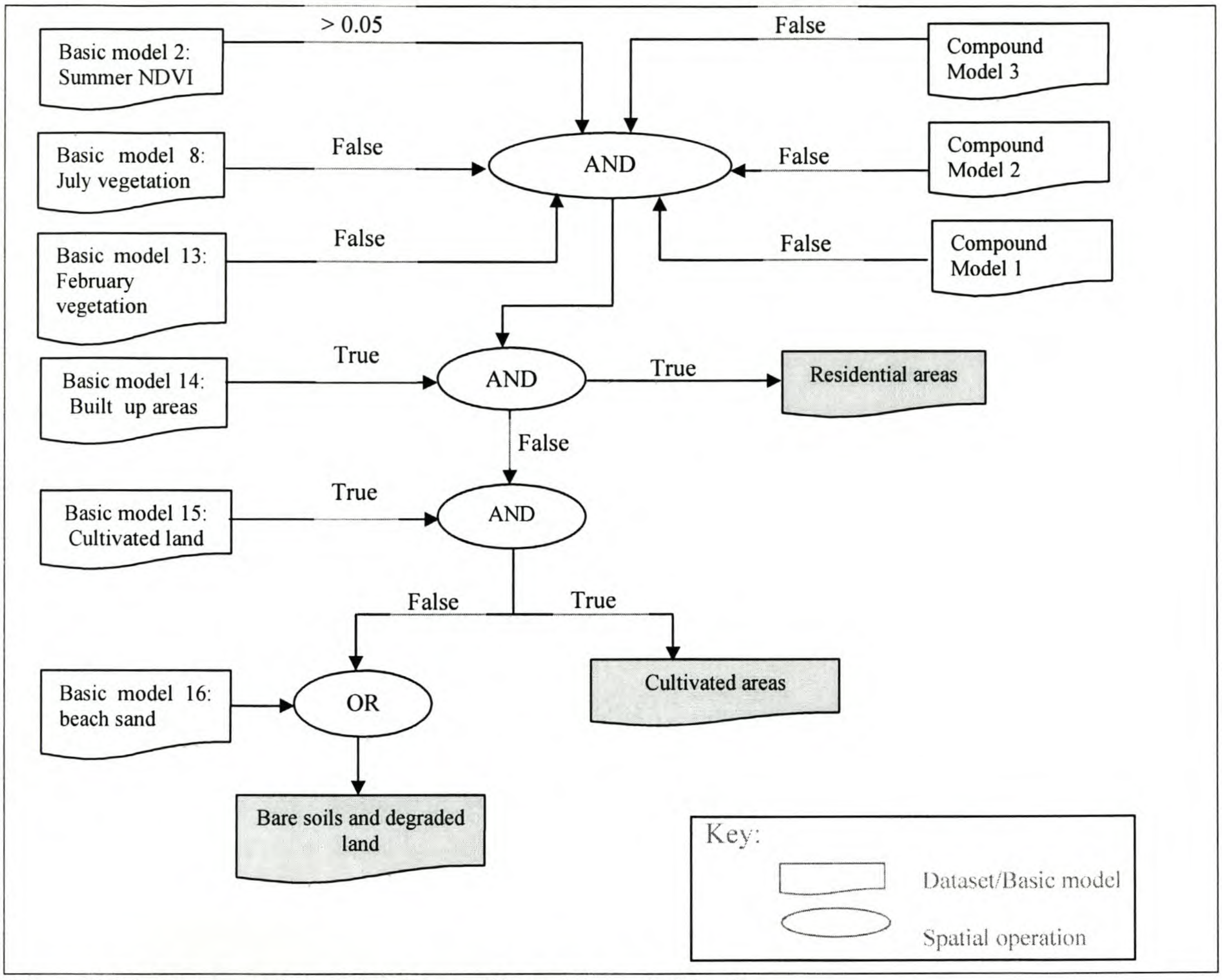


Figure 3.10: Compound model 4. Bare soils and degraded land.

3.3.5 Compound Model 5: Forest plantations and commercial indigenous forests

Forest plantations include all forests that are systematically planted, commercial and primarily composed of exotic trees. The most commonly planted trees in South Africa are: pine, eucalypts, wattle and indigenous species. This category excludes all non-timber-based plantations such as tea, sisal and orchards (Thompson 1996).

Forest plantations show low seasonal NDVI variation and are less textured than indigenous forests. Permanent dry land cultivation, which may include vineyards, nut and banana plantations, has similar reflectance and texture properties as indigenous and mixed timber plantations (CGA 2003).

The use of vegetation indices is one of the primary methods for discriminating forests from satellite images. Although their use has been successful, vegetation indices have limitations. Boyd, Foody and Ripple (2002) and Chen and Cihlar (1996) argue that vegetation indices could be environment-specific and may vary in applicability over space and time. Furthermore, since the indices are typically based on two or three spectral wavebands, their use could exclude useful information measured in other wavebands. This is a growing concern, since new-generation sensors are now operating at hyper-spectral resolutions.

A multi-step approach was taken in this compound model (see figure 11). Multi-temporal imagery was used to derive broad classes of annual and perennial vegetation covers. In addition to the basic models developed in this section basic model 8, 10 and 14 were also used (refer to sections 3.3.2 and 3.3.4).

The initial step in this compound model was to identify, mask and subsequently exclude pixels that would not contain vigorously growing vegetation from further data processing. Pixels that were not represented in a vegetation layer derived by the condition discussed in basic model 8 were excluded from subsequent data processing. A new layer-stacked dataset (refer to basic model 19) representing the sub-set layer obtained by the above condition was created to perform unsupervised classification to generate homogeneous zones in terms of the combination of the explanatory variables: texture variance (refer to basic model 17), spectral band difference between band 4 and 3 (refer to basic model 18), Tasseled Cap bands 1, 2 and 3 and NDVI layers of the two dates. The classification was performed (with 50 iterations and a 0.990 threshold) to get 20 classes. This was necessary to calculate zonal mean NDVI to extract zones that do not show significant NDVI change between seasons as potential forests.

Basic Model 17: Texture variance

Texture is one of the most important explanatory variables. Most land cover classes show characteristic texture patterns. Among forest vegetations, forest plantations show less texture due to the regular spacing of trees. Texture was calculated on the ETM+ panchromatic band using a 5x5 moving window. The 15m resolution of the panchromatic band enables the calculation of intra-pixel texture as it is twice the resolution of the multi-spectral bands. The resulting variance image was re-sampled to 30m to comply with the multi-spectral imagery.

Basic Model 18: Band difference

Different vegetation types require different wavelengths of electromagnetic energy for photosynthesis. Vigorously growing vegetations have higher reflectance in band 4 (near infrared) than in band 2 (green) (Dorren, Maier & Seijmonsbergen 2003). Band difference between ETM+ band 4 and 2, for both dates, was calculated to assist the discrimination between different vegetation types.

As explained in section 3.3.1, ISODATA derived zones are essential for calculating zonal statistics of explanatory variables. In this compound model a new layer-stacked dataset was prepared using the relevant explanatory variables to create spectrally homogeneous zones. The dataset included Tasseled Cap channels 1, 2 and 3, NDVI for the two dates, texture variance and band difference between ETM+ band 4 and 2. These layers were selected because they can individually and in combination increase the separability of land cover classes, especially among vegetated areas. This layer-stacked dataset was used as an input (see figure 3.11).

The mean NDVI change between February and July was calculated in percent. The result showed the range of change to be between -40 and 156.5 percent, with only two zones having negative percentages (-40 and -30.5% respectively). These negative values indicate that the NDVI of the zones were higher in winter than in summer. This implies that the areas represent irrigated lands, as one would expect higher vegetation growth during the rainy season.

Forests are permanently vegetated and are expected to show less seasonal NDVI variation. Based on this assumption 80% NDVI change was selected as a relaxed constraint to extract potential forests. Therefore only zones below this threshold and above 0% were considered for further discrimination. This criterion allows other non-forested vegetations to be included. NDVI was also used to exclude these.

Forest vegetations usually have higher NDVI values than most other vegetation types. Based on known forest areas a relaxed NDVI threshold of 0.5 was selected and pixel values greater than this were extracted as potential forests. On the output layer a maximum-likelihood supervised classification was performed to obtain classes representing forest plantations, permanent cultivated land and indigenous forests. The training data for this classification was obtained from the CSIR in point shapefile format, which was expanded to polygon format using the Avenue programming language.

Since residential areas, especially small-holdings, often contain vegetated areas it was necessary to isolate these for exclusion. Forests in residential areas that are larger than 7200m^2 (8 pixels) were not excluded, because some forest patches can be found within residential boundaries. To do so the ERDAS "clump" function was used. Therefore only clumps (connected pixels) of vegetation below the size of 8 pixels were excluded. Pixels with clump size below 8 pixels were used as input to compound model 12, which identifies residential areas (see section 3.3.12).

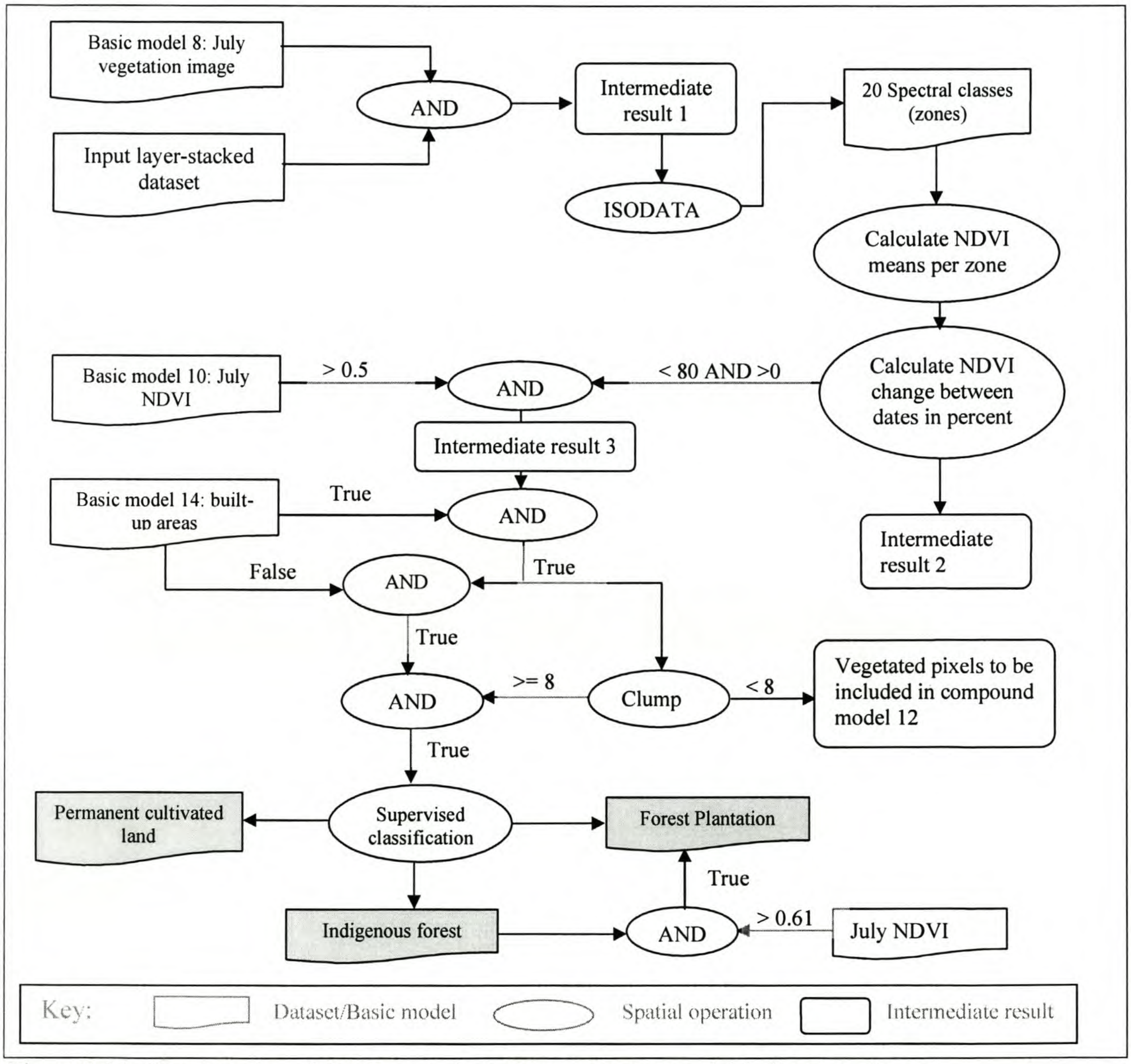


Figure 3.11: Compound model 5. Forest plantations and commercial indigenous forests.

The final step in this compound model was to reclassify indigenous forest pixels with July NDVI values greater than the threshold of 0.61 (determined from know areas in the study area) as forest plantations.

3.3.6 Compound Model 6: Indigenous forests and woodlands

In South Africa natural forests can be divided into two major categories namely inland temperate Afromontane and coastal subtropical Indian Ocean forests. At a more detailed level, natural forests may be grouped into: southern Afrotemperate, northern Afrotemperate, northern mist-belt, southern mist-belt, scarp, southern coastal, northern coastal, lowveld riverine, swamp, mangrove, and licuati sand forests. The greatest diversity of forest types is located in KwaZulu-Natal and the Eastern Cape (Mucina *et al.* 2003). The types of forest groups found in the study area are southern mist-belt, northern coastal, scrub and mangrove forests.

It has been reported that no definitive spatially explicit maps exist on the distribution of indigenous forests in South Africa (Geldenhuys 1994; Mucina *et al.* 2003). The most detailed available maps are those produced by Cooper in 1985 and the forest biome programme in 1987. These maps are however not available in GIS format and the scale at which they were mapped (approximately 1:1 000 000) is not appropriate for this study.

Another map of forests, derived from a combination of LANDSAT and field data, delineated plantation forestry (main focus) and indigenous forests on a portfolio of 31 map sheets at a scale of 1:250 000. The minimum mapping unit for plantations and indigenous forests was 25 and 50 ha respectively (Van der Zel 1988 in CSIR 2003). Although the spatial accuracy of this map is questionable, it is probably the best GIS map available to date (Mucina *et al.* 2003). The low accuracy can possibly be attributed to the wide range of spectral reflectances associated with forests, as it contains many shadows due to the terrain and the sun's azimuth and because the spectral properties can easily be confused with those from plantations. Or possibly it is due to the gradual transitions between forest and bushveld.

An alternative map is the 1994 National Land Cover (NLC) project map (Thompson 1999). Even though the map is spatially accurate, the accuracy with which it identifies forest patches raises some concern, as it did miss some known forests such as Island forest and other Eastern Cape dune forests. It is also possible that exotic wattle infestation, dense woodland and thicket may have been classified as forest in some areas (Mucina *et al.* 2003).

Geldenhuys (1994) hypothesized that environmental factors such as rainfall regime, geology and soils determine the potential limits of forest distribution in the east and southern parts of South Africa, but the actual location pattern of forest in the landscape is determined by the bergwind fire regime. The typical forest location pattern is described as follows:

i) The largest forests are found on the coastal platform in locations of immediately south of the southernmost mountain ridge and along the east-west running river valleys in the coastal platform. The platform forests occur on the west side of the north-south oriented river gorges cutting the platform. Forests are absent on the platform to the east of each gorge, but these sites are only vegetated with Fynbos or pine plantations replaced by Fynbos. The northern boundary of these forsts occurs on the steep foot slope of the southernmost ridge, resulting in the western north-south boundary being shorter than the eastern north-south boundary. These forests generally show a finger-pointed pattern on their southern boundary to the southwest, except when the forest borders the east-west river valleys through the coastal platform.

ii) On both the northern and southern sides of the valleys, forests occur close to the sharp edges formed by the coastal platform and the valley. However, forests occur at much lower level from the upper edge of the valley on ridges running from the coastal platform into the valley.

iii) Along the coastal scarp, forests, occupy positions that very similar to those in river valleys.

vi) The smallest forest patches are found in several localities in the mountains. Most of these forests occur west of the streams, near the bottom of the valleys. Such forests do not occur on the east side of these same streams. A few of these mountain forests, however, occur immediately below precipitous krantzes on concave slopes. Forests do not occur near the top of ridges, with the exception of those that occur in ridges that have gentler slope in the north than in the south, which can be straight or concave near the top of the ridge. Forests also occur near the lower end of some ridges, in the valley of a first-order stream within the forking end of the ridge (Geldenhuys 1994).

This pattern of indigenous forest distribution may be explained to some extent by using a geomorphological model (refer to basic model 12). The geomorphological model involves calculating local relief, local up and low land, slope classes and aspect.

In the *National Land Cover 2000: illustrated field guide*, indigenous forest areas are described as all wooded areas with a tree canopy (mainly composed of self-supporting, single stemmed, woody plants greater than 5m in height) cover of greater than 70%. A community of multi-layered with interlocking canopies composed of canopy, sub-canopy, shrub and herb layers.

For identifying natural indigenous forests, the compound model developed here (see figure 3.11) used basic model 2, 10, 12 and 14 and intermediate result 1 and 2 (refer compound model 5) from the previous sections and the basic model 20 developed here.

Basic Model 20: Relative relief

As pointed out by Geldenhuys (1994) indigenous forests are not likely to occur on crests or wind exposed altitudes. Relative relief[3] was calculated to identify potential forest areas in terms of positional occurrence on the terrain. Relative relief was derived from a DEM by calculating Focal Neighbourhood Range (FNR) using a 5x5 moving window. The resulting image (see Table 3.1) is classified into five classes adapted from Brabyn (1998). It was determined that indigenous forests are likely to occur in relative reliefs that range between 20 and100m.

Table 3.1: Relative relief classes

Relief	Class ranges
Flat/low relief	0 – 20m
Low hills	20 – 100m
Hills	100 – 600m
High hills	600 – 900m
Mountains	> 900 m

Source: adapted from Brabyn (1998:40).

Figure 3.12 illustrates the basic models were combined to identify indigenous forests and woodlands. The process starts with the masking of intermediate result 1 (stacked vegetation image) obtained in compound model 5 and intermediate result 3 (pixels in this image are already assigned to classes) from the same compound model to exclude pixels that have been already identified.

[3] Relative relief is an important topographic parameter used to explain the terrain of a landscape. Relative relief indicates a pixel's position in terms of its altitude relative to its neighbouring pixels.

Using basic model 12 (refer to section 3.3.3) potential indigenous forests was extracted from the resulting image. Basic model 12 provides the topographic orientation of a pixel.

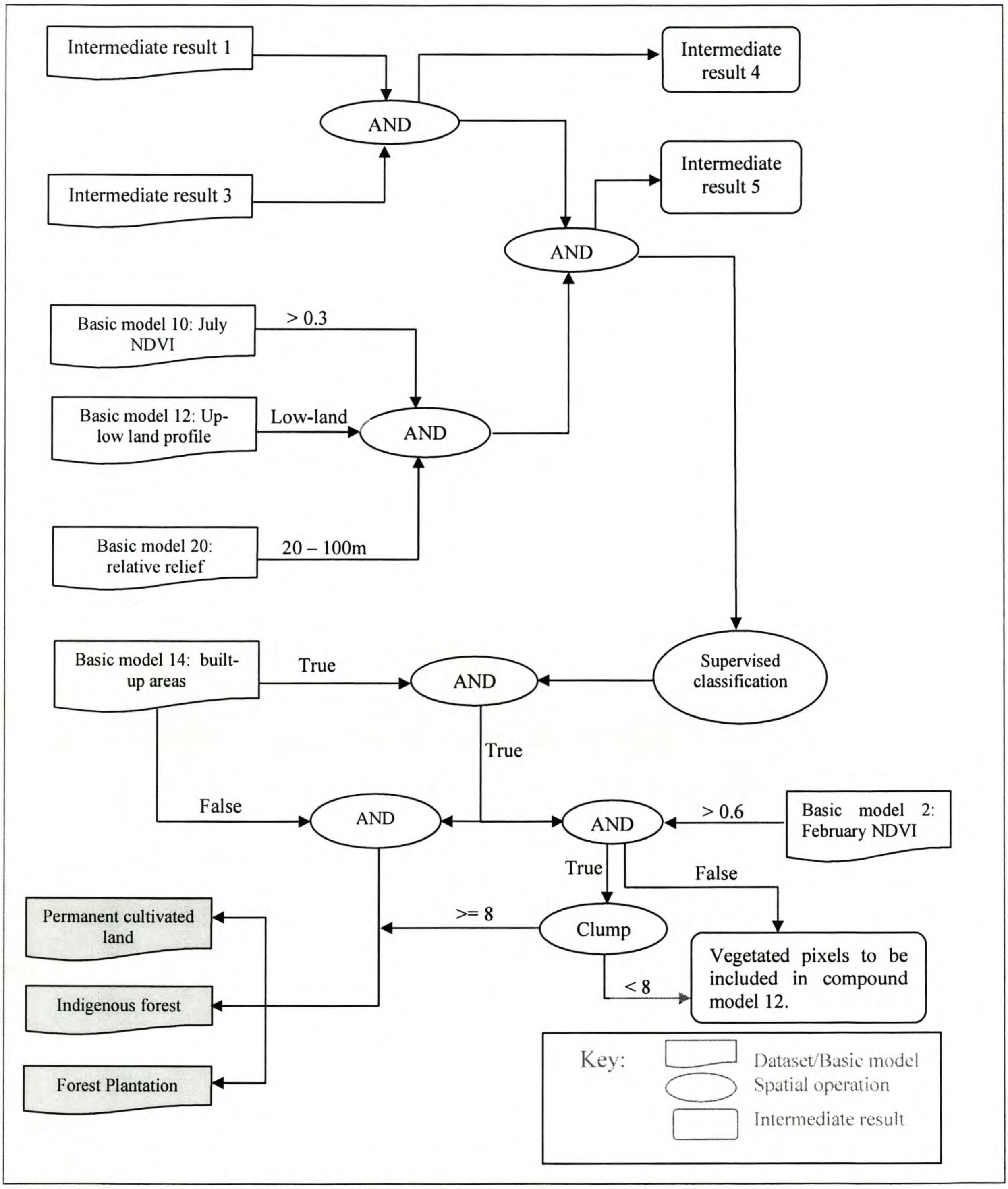


Figure 3.12: Compound model 6. Indigenous forests and woodlands.

As explained by Geldenhuys (1994) indigenous forests usually occur below precipitous krantzes on concave slopes and occupy positions that are less affected by berg winds. For example forests do not usually occur near the top of ridges. As a result pixels that were low-land (basic model 12) with relative relief (basic model 20) values of between 20 and 100m were selected as potential indigenous forests. In addition, a vegetation index, July NDVI, was used to retain perennial vegetation, excluding pixels that are less likely to be indigenous forests. To do this a suitable threshold (> 0.3) was selected based on known areas.

On the resulting image, a supervised maximum-likelihood classification was performed based on the training data. The following classes were obtained: indigenous forest, thicket and bushland, and commercial permanent cultivated. In some cases the forest class included woodlands and some dense thickets as they have similar spectral characteristics as indigenous forests (Mucina *et al.* 2003). Residential vegetation pixels were excluded using the same method described in section 3.3.5. Pixels that have summer NDVI values of greater than 0.6 and are residential (basic model 14) were excluded.

This model can be modified by including techniques to accurately predict the distribution of indigenous forests based on more explanatory variables and empirical data. Information on soil, ecology and other model variables may improve predictions.

3.3.7 Compound Model 7: Coastal forests and woodlands

Coastal forests differ from inland Afromontane forests in terms of species composition and distribution. They occur in a relatively low relief landscape and communities are usually not as dense as those of inland forests. In the study area, coastal forests are limited to a small number of localities. This is perhaps because most of the coastal forests' habitats are being used for agricultural purposes.

As mentioned earlier (see section 3.3.6) forests in general can be grouped into Afromontane, Scarp and Coastal forests. Coastal forests include coastal dune forests and they are positioned along the narrow, geologically young coastal strip of the Zululand coast (Mucina *et al.* 2003). The differentiation of these forests from other forests can be done based on floristic and biogeographical models. The models are ecological models that need biological samples, and are used by foresters.

The extraction of coastal forests and woodlands from remote sensing satellite images requires the use of vegetation indices and topographical information. The compound model developed in this study used NDVI and local relief to isolate potential coastal forests and woodlands. The model firstly considered all vegetations with relatively high July NDVI values of greater than 0.3 and low local relief of less than 20 meters as potentially coastal forests and woodlands. Vegetation pixels were identified by using basic model 8 (July vegetation). Pixels that were identified as forest plantations and indigenous forests in compound model 5 and 6 respectively were excluded. To do so, intermediate result 4 (from compound model 6) was masked with intermediate result 5 (also from compound 6) (see figure 3.13). From the resulting layer potential pixels were isolated by using the July NDVI, and pixels with NDVI value of greater than 0.3 were extracted. In the resulting layer other two vegetation types are contained namely commercial cultivated vegetation and improved grasslands. Maximum-likelihood supervised classification was employed based on the training data to produce distinct classes of coastal forests and woodlands, commercial cultivated lands and improved grasslands.

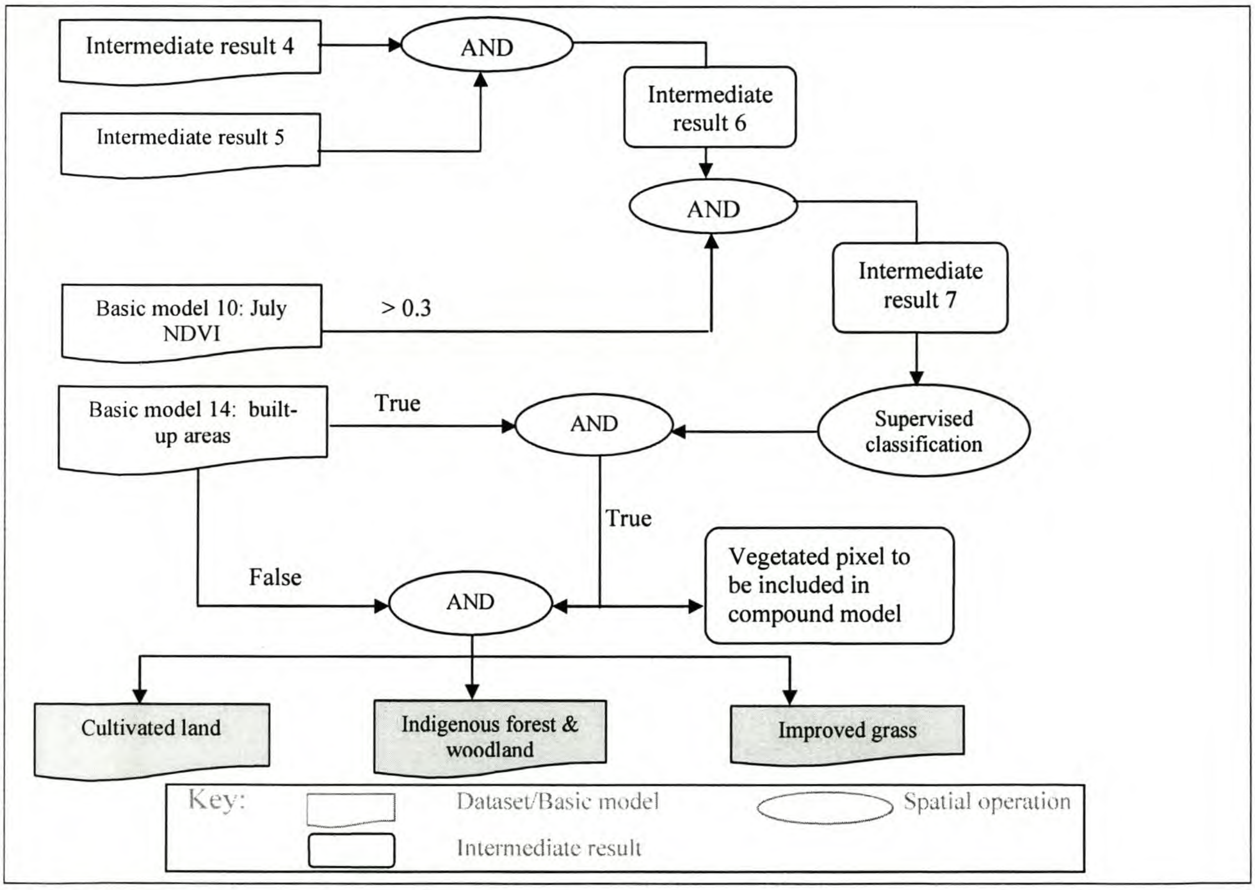


Figure 3.13: Compound model 7. Coastal forests and woodlands.

Improved grassland, which includes golf courses, racing tracks and parks, was treated as a class on its own and constitute a very small area compared to other vegetation covers. Almost all improved grasslands occur on flat to low relief areas. Improved grasslands resemble cultivated areas in many respects and there is high possibility that improved grasslands can be spectrally confused with some cultivated crops, which can lead to significant misclassification. The selection of suitable training data is therefore crucial when supervised classification is used.

Pixels identified as coastal forests and woodlands but falling within the residential boundary were excluded. The exclusion was done by applying the method used in compound model 6. The excluded pixels are included in the compound model 12 (residential areas).

In this progressive classification process, the next class to be identified is thicket and bushland forests, and compound model 3.3.8 discusses the approach taken and basic models used. Pixels classified in this compound model were excluded from further processing.

3.3.8 Compound Model 8: Thicket and bushland forests

Thicket and bushland is defined as vegetation communities (essentially indigenous species) mainly composed of tall, woody, self-supporting, single or multi-stemmed plants that branch at or near the ground, in most cases lacking definable structure. Tree heights of thicket and bushland communities are in the range of 2 – 5m and have a canopy cover of greater than 10%. Even though structural class definitions (i.e. canopy cover and height parameters) are suitable for satellite based vegetation classification (Thompson 1996)[4], the correlation of these parameters with spectral information may require a substantial amount of empirical data in order to do regression analysis.

As empirical data was not available an alternative approach was taken in this study. Thicket and bushlands occur on riverbanks, in valleys and on a variety of relief types. Thicket and bushlands therefore do not show a definite pattern in terms of local relief. In this compound model the vegetation index (NDVI) was used as a suitable explanatory variable. The sub-set vegetation layer from which forest plantations, indigenous forests, coastal and woodlands were excluded constituted a vital input.

[4] The community-type method is based on functional definition rather than conspicuous structural characteristics that can be correlated with satellite data.

To get this layer intermediate result 6 of compound model 6 was masked with intermediate result 7 of the same compound model. A suitable NDVI threshold was determined from known thicket and bushland areas to isolate potential pixels from the vegetation sub-set layer. Pixels with July NDVI values of between 0.2 and 0.3 were isolated to represent potential thicket and bushland areas. In the resulting layer commercial cultivated areas are also included. A supervised maximum-likelihood classification was used to obtain distinct classes of thicket and bushland areas and commercial cultivated lands (see figure 3.14).

Pixels identified as thicket and bushlands but falling within the residential boundary were excluded. To do so the technique applied in compound model 6 was used. These pixels are included in the residential compound model 12.

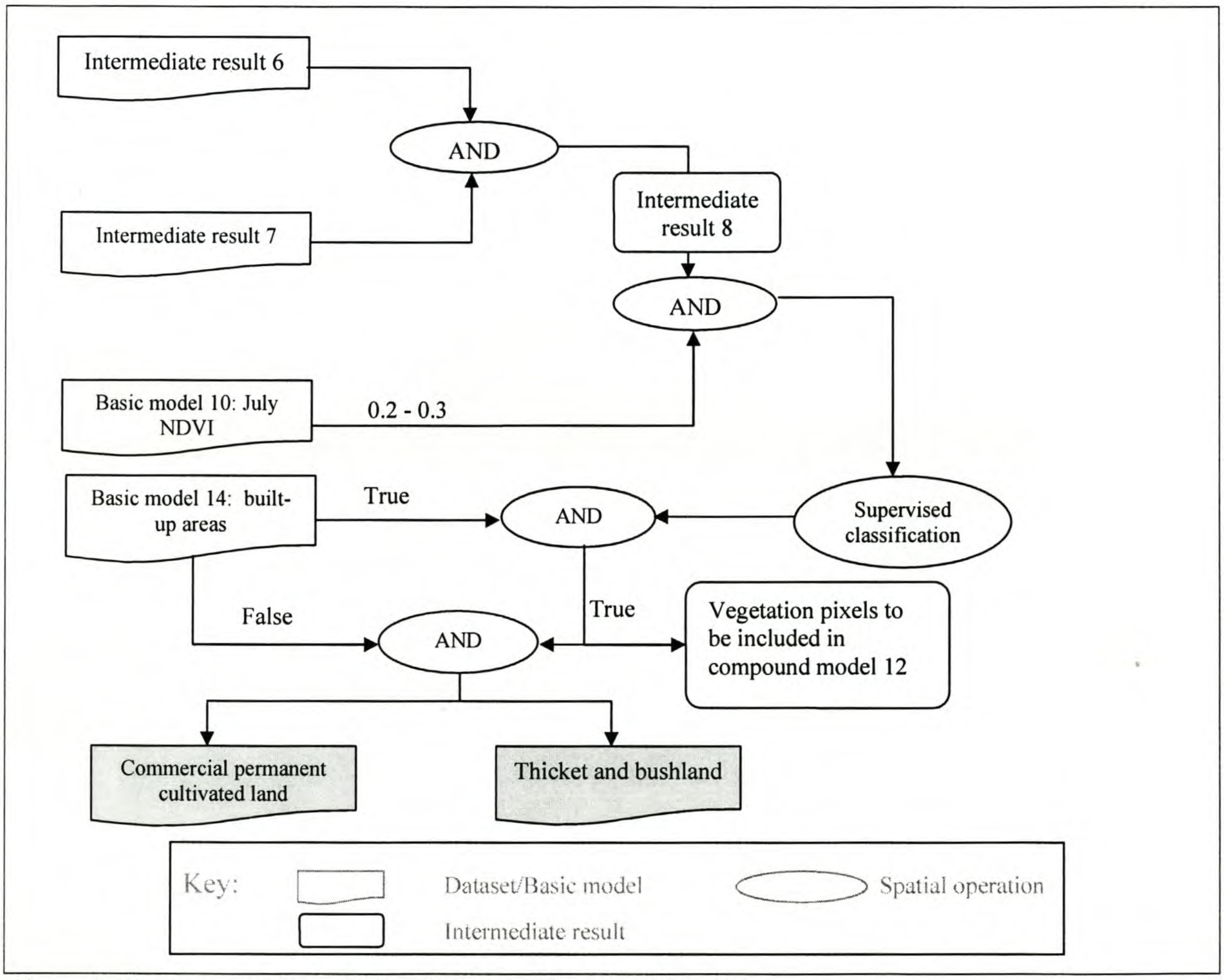


Figure 3.14: Compound model 8. Thicket and bushland forests.

3.3.9 Compound Model 9: Shrub land

As described in the *National Land Cover 2000, Illustrated Field Guide*, shrub land is vegetation dominated by low, woody, self-supporting, multi-stemmed plants, branching at or near the ground. The tree height range of this class is between 0.2 and 2 metres, and with a total tree cover of less than 0.1% (CSIR 2000). These structural parameters are poorly correlated with spectral information.

Huang *et al.* 2001 developed a technique for estimating tree canopy density using Landsat 7 ETM+ images. Although this technique is of great value, it requires empirical relationships between tree canopy density and Landsat data as well as good empirical data on the structural parameters of trees (i.e. canopy cover and tree heights) in order to apply linear and other regressions and to establish the correlation a priori. This information was not available for this study.

The approach taken for identifying shrub lands is similar to that of thicket and bushland forests. The very low total tree cover (less than 0.1 percent) of shrub lands result in very low NDVI. Vegetation that were not identified in the previous compound models were considered as potential shrub lands.

The compound model (see figure 3.15) that identifies shrub lands utilizes intermediate result 8 (from compound model 7). Pixels with a value less than or equal to 0.2 were classified as shrub land. The confused residential pixels were again excluded using the technique applied in compound model 7.

Unlike natural forests cultivated lands show significant NDVI variation. Moreover, almost all cultivations are undertaken in low local relief to flat areas. These and other properties can be used to distinguish cultivated lands and natural vegetation. The following two sections attempt to identify cultivated lands.

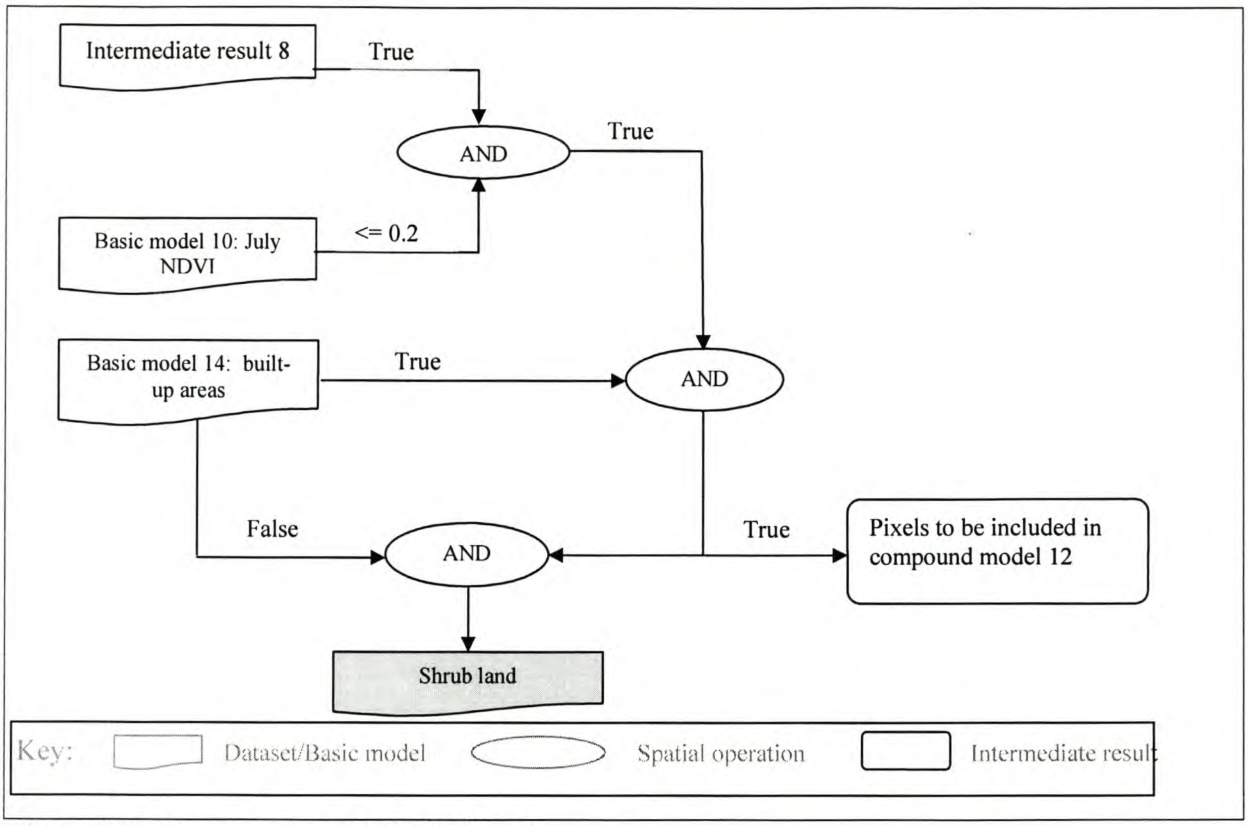


Figure 3.15: Compound model 9. Shrub land.

3.3.10 Compound Model 10: Commercial permanently irrigated cultivated lands

Of all the land cover categories in the standard classification scheme, cultivated land is perhaps the most diverse. In the *illustrated field guide for the National Land Cover 2000 project* (CSIR 2000) more than 7 subclasses are identified under the broad class 'cultivated land'. Irrigated cultivated areas are characterized by their location relative to large water resources, either in the form of rivers or dams and in exceptional cases, ground water. Cultivated maturing crops are usually associated with vigorous growth resulting in NDVI values of above 0.5. Cultivated crops have the highest variance, both spatially and temporally, and the wide variety of crops ranging from sisal, sugar cane, maize, wheat, vegetables and fruit orchards together with the range of different development stages, complicate the definition of spectral signatures (CGA 2003).

The phenological pattern of crops is a useful source of information for the identification of and discrimination between cultivated crops. Crops are seasonal and differ in their seasonal cycles of greening, flowering and ripening. Forests also show seasonality by changing leaf colouration. These are the patterns plants show in response to seasonal and climatic changes, which include duration of sunlight, precipitation and temperature. A number of studies have employed seasonal changes in vegetation indices, such as NDVI, to represent phyto-phenological differences between vegetation formations and crop types (Shoshany & Cohen 2002), while Dymond, Mladenoff & Radeloff (2002) used Tasseled Cap indices of multi-temporal imagery. The relative strength of the phenological approach depends on its interpretability by the analyst, as it requires a good knowledge of vegetation phenomenology (Shoshany & Cohen 2002). Moreover, this method requires multi-temporal data, preferably at more than two dates, to derive the phenological stages of different crops.

For the identification of permanently irrigated cultivated lands, compound model 10 (see figure 3.16) used intermediate result 3 of compound model 5. Compound model 10 classifies the zones that showed negative mean NDVI change as permanently irrigated cultivated land. This class may include non-timber plantations under drip irrigation, which can be identified from its low-hue light orange appearance in a 4-5-3 composite colour image (CGA 2003). Since some irrigated cultivated areas may not have negative mean NDVI values they may be identified as a different cultivated land sub-category in other compound models. Temporary cultivated lands that are likely to show high positive mean NDVI change values are identified in the next compound model.

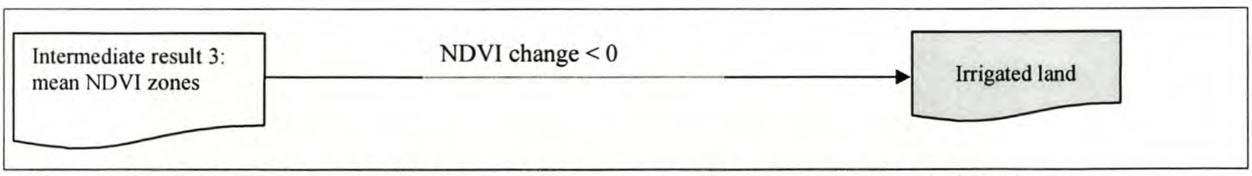


Figure 3.16: Compound model 10. Commercial permanently irrigated cultivated lands.

3.3.11 Compound Model 11: Temporary cultivated areas and unimproved grasslands

Temporary cultivated lands mainly include seasonal crops and subsistence cultivated lands (CSIR 2000; Thompson 1996). Temporary crops are described as annual crops harvested at the completion of the growing season; remaining idle until the next season. Examples include maize and Soya bean. Subsistence agriculture is characterized by small and numerous field units of usually less than 10ha in size that are found in close proximity to rural population centres. Discrimination of this category based on phenological patterns is ideal. However, as noted in the previous section, this approach requires multi-temporal image data.

For identifying temporary cultivated lands and unimproved grasslands the strategy used by this compound model (see figure 3.17) was divided into two phases. In the first phase an image was created by extracting pixels that were identified as vegetation in the February (summer) vegetation image (refer to basic model 13) and non-vegetation in the July (winter) vegetation image (refer to basic model 8). From this image, burnt areas (refer to compound model 3) and open water bodies (refer to compound model 1) were then excluded. The resulting image contained small plantations that were not identified in the winter vegetation (basic model 8). This is possibly as a result of stressed growth due to climatic and environmental conditions. Other features that can be found in this image include thicket and bushland (less dense).

Unimproved grasslands have little vegetation cover and they show significant NDVI seasonal variation, becoming more vigorous in the summer due to higher precipitation. This is even evident by visual inspection and comparison of the satellite images. Temporary non-irrigated cultivated lands show a similar pattern. Pixels in the sub-set image (described above) with February NDVI (basic model 2) values of less than 0.64 were therefore classified as unimproved natural grassland. The result obtained also contained cultivated lands, thicket, bushlands and forest plantation areas.

In the second phase of this model supervised maximum-likelihood classification, based on the training data, was performed to distinguish between different land cover types. As in previous compound models pixels that represent residential areas were separated based on the technique applied in compound model 7. These were included in the residential class as discussed in the next section.

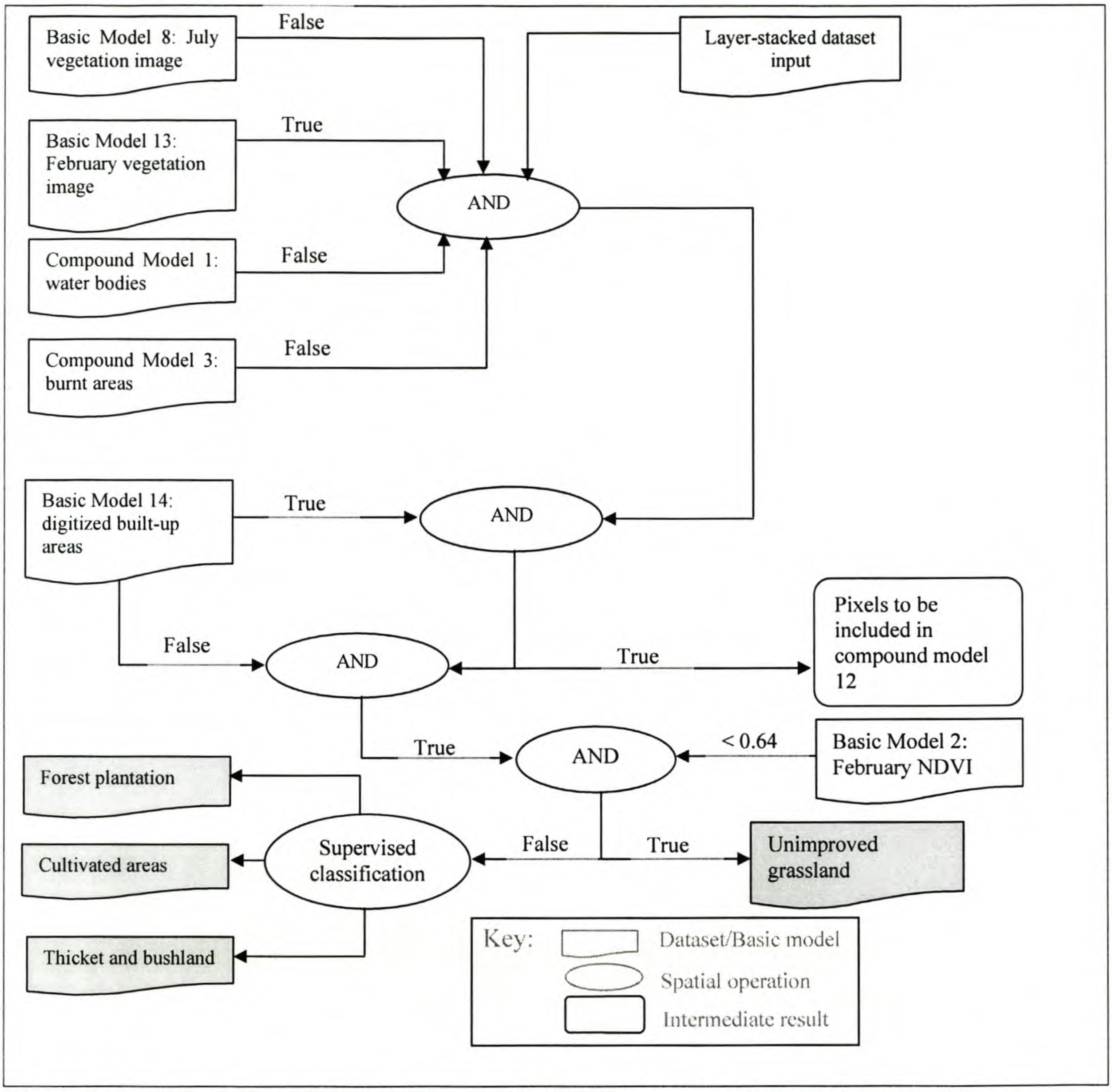


Figure 3.17: Compound model 11. Temporary cultivated areas and unimproved grasslands.

3.3.12 Compound Model 12: Residential areas

Residential areas is the most difficult class to isolate from satellite imagery. This type of built-up area is characterized by high intra-pixel spectral variability as well as inter-pixel variation. The presence of a combination of vegetation and hard surfaces makes it one of the most heterogeneous categories. Built-up areas in general have high texture, and they show explicit spatial pattern (e.g. shape) (Moller-Jensen 1998). Because of high heterogeneity the use of ancillary data is indispensable.

To delineate residential areas ancillary data was used. Street features were extracted from a roads dataset (at a scale of 1:50 000) obtained from the Chief Directorate: Surveying and Mapping (CDSM), which was used to calculate a distance layer. Urban built-up areas indicated on 1:50 000 topographical maps (also obtained from CDSM) were digitized and subsequently converted to grid format. The two grid layers were overlaid to represent residential areas. Apart from those pixels that were identified as water bodies (Compound model 1), fire scars (compound model 2), wetlands (compound model 3), forest patches and commercial permanently irrigated land (compound model 10), pixels within this boundary were classified as residential.

This technique proved to be successful in major urban areas, but was less effective in identifying formal and informal townships and villages. Compared to formal urban areas, informal settlements and rural villages are characterized by lower spectral brightness, probably due to the difference in building materials. Brightness is however higher than the surrounding areas. Rural settlements are associated with small gardens, which may result in slightly higher NDVI values compared to bare soil areas and can be easily confused with subsistence and semi-commercial agriculture. They are therefore not easily extracted from satellite images alone.

In the study area plenty of informal townships and villages are know to exist. Most of the villages appear to be distributed in the central and southern part of the study area while the informal settlements are scattered near the coastal areas such as Durban.

The compound model (see figure 3.18) to identify these land cover features is based on the basic models mentioned below as well as basic models 2 and 5 (see section 3.3.1).

Basic Model 21: Texture (contrast)

Texture models are often used for urban classifications (Moller-Jensen 1998). Since texture measures provide spatial relations of pixels, they can be used to identify land cover categories with unique spatial patterns. In this basic model grey level co-occurrence matrix texture was calculated.

The grey level co-occurrence matrix was first calculated based on ETM+ band 4 using the EASI/PACE module in PCI Geomatica V-9 (PCI/Geomatica 2003). Next the contrast texture measure was calculated on this grey level matrix. It was determined from known areas that built-up areas have texture (contrast) greater than 15.

Basic Model 22: Low to unvegetated areas layer

Degraded and unimproved grasslands images identified by compound models 4 and 10 respectively, may include rural and informal townships as they are dominated by low vegetation and have some spectral similarity with degraded and unimproved grassland. To isolate these built-up areas degraded and unimproved grasslands layers were combined.

The use of texture on the combined image of degraded and unimproved grasslands along with other ancillary data can resolve the confusion. A combination of basic model 21 (texture (contrast)), basic model 2 (February NDVI) and basic model 5 (Tasseled Cap brightness) was used. A pixel within the combined image was, based on known samples, classified as township and villages if the pixel has an NDVI value of less than 0.57, texture greater than 15 and Tasseled Cap brightness of greater than 80. The result showed that the method was able to identify most of the informal settlements and villages, although it did include some areas of bright and unvegetated soil areas that should have been classified as bares soils. The misclassified pixels account approximately for less than 1 percent of the combined layer created in this basic model.

Built-up infrastructure may also occur in other land cover types including mines and quarries. The following section discusses the identification of mines and quarries.

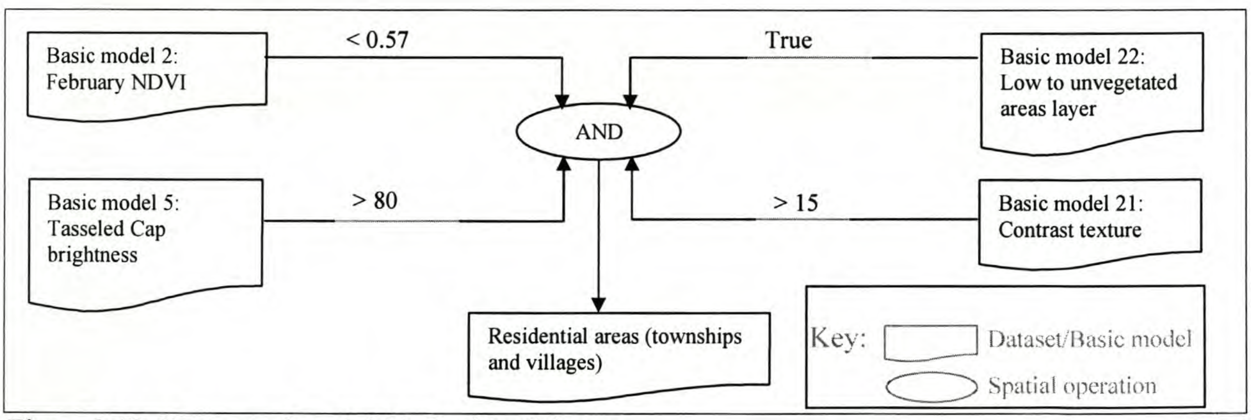


Figure 3.18: Compound model 12. Residential areas

3.3.13 Compound Model 13: Mines and quarries

Mines and quarries include active or non-active mining areas that may be underground or sub-surface. The sub-surface mining includes hardrock and sand quarries as well as opencast mining (coal) (CSIR

2000). This category is heterogeneous since it includes associated surface infrastructure. Pixels of this class can therefore be confused with other classes such as commercial and industrial areas and rock outcrops.

In the KwaZulu-Natal province a number of mining activities are known to exit and some of these are found within the study area. These include heavy mineral, coal, marble, gold, dimension stone and stone aggregate mines (Geoscience 2003).

A progressive masking approach was taken to identify mines and quarries. This method ensures that the pixels of each land cover class are mutually exclusive. Land covers that have been identified in previous compound models were excluded from being considered for this category. Besides basic model 2 (February NDVI), 8 (July vegetation) and 13 (February vegetation) and outputs of compound model 1 (water bodies), 2 (fire scars) and 3 (wetlands), a newly created basic model 23 was used.

Basic Model 23: Mines and quarries

Spectrally, mines and quarries can easily be confused with commercial, industrial and transport areas, as they are usually devoid of any vegetation. Mines and quarries are generally smaller in extent compared to commercial and industrial land covers. Mines and quarries delineated on the 1:50 000 topographic maps, obtained from Chief Directorate: Surveys and Mapping, were digitized and adjusted manually using the 4-5-3 colour composite to match the position on the satellite images. The resulting polygon shapefile was then rasterized.

In this compound model, output of compound model 1, 2 and 3 as well as basic model 8 and 13 were excluded from being considered. A suitable February NDVI threshold (< 0.05) was used to further exclude pixels that are less likely to represent mines and quarries. The resulting image is expected to contain commercial, industrial, transportation and mines and quarries. From this image, mines and quarries were extracted using basic model 23 (mines and quarries). The remaining unclassified pixels were used in compound model 14 (commercial, industrial and transportation) (see figure 3.19).

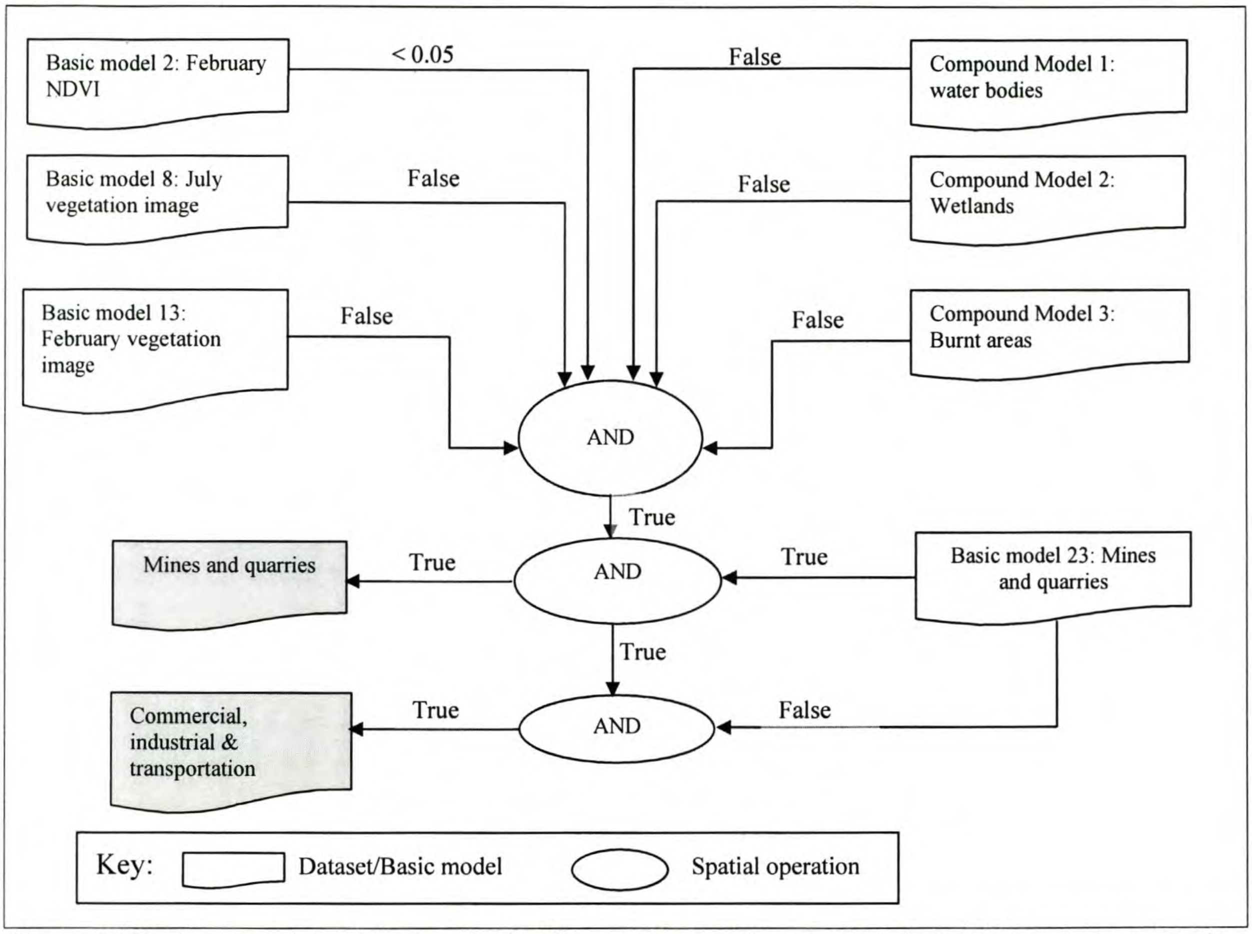


Figure 3.19: Compound Model 13 and 14: Mines, quarries, Commercial, industrial and transportation areas.

3.3.14 Compound Model 14: Commercial, industrial and transportation areas

Built-up, commercial urban areas are typically located in the central business district (CBD) of towns and cities and industrial and transportation areas are infrastructure related to major industrial and transport activities (Thompson 1996). This level 1 land cover class is characterized by bright spectral reflectance in all the ETM+ bands and high texture. It is usually devoid of vegetation and has as a result very low NDVI values. A combination of spectral and textural information is therefore ideal, but care should be taken as confusion with residential areas may occur. Because of this, both Zhang (1999) and Moller-Jensen (1998) used texture based on co-occurrence matrices and contextual spatial information to classify urban areas. Commercial, industrial and transportation areas can also be confused with mines and quarries (see previous section).

For identifying commercial, industrial and transport areas a similar process was followed as with mines and quarries. A single compound model was therefore enough to identify both of these land covers. From the image that contained commercial, industrial and transport and mines and quarries (refer to compound model 13), commercial, industrial and transportation areas were extracted using basic model 23. Pixels that are not mines and quarries were classified as commercial, industrial and transportation areas.

3.3.15 Compound Model 15: Shade

Shade is not a land cover, but was treated as a class on its own. This is because land covers in shade do not demonstrate the expected reflectance properties and shaded areas can cause confusion with water, burnt areas and other dark objects. Shaded areas should therefore be excluded from processing.

Shaded areas are usually found in areas of rough terrain and are accentuated during winter when the sun's azimuth is low. The occurrence of shade is related to aspect and slope. For example concave slopes are more likely to be shaded.

The study area includes mountains and valleys, which increase inland from the coastal areas. Shade constitutes a minority class (with 0.000063 percent of the total pixels classified) compared to other covers with only few groups of pixels scattered in the study area.

To identify shade, basic models 1 and 3 (refer to sections 3.2.1) and output from compound model 3 were used. Pixels with a slope (basic model 3) of greater than 27%, band ratio (basic model 1) between 2 and 60 and not representing fire scars were classified as shade. These discriminatory thresholds were based on shade samples identified on the satellite images.

3.4 POST-CLASSIFICATION MANIPULATION

Since the classification model developed in this study was pixel based, outlier pixels ("salt-and-pepper" effect) had to be manipulated to produce a more meaningful thematic result. A contextual classifier in the form of a low-pass filter is often used in post-classification manipulation. For this the most widely used window size is 3x3, but larger windows are often used for small-sized pixels or larger land cover entities such as in this study (Stuckens, Coppin & Bauer 2000). A 5x5 low-pass filter was therefore applied to the classification to obtain a generalized classification map. The resulting map was used in the accuracy assessment based on reference samples of the land cover classes. The following section assesses the accuracy of the classification process.

3.5 ACCURACY ASSESSMENT

The accuracy with which image classifiers identify land cover features is evaluated using accuracy assessment. Accuracy assessment is a function of training data and classifier performance and generalizes to a certain degree the information content and the accuracy of the resulting thematic map (Stuckens, Coppin & Bauer 2000). The first step toward an accuracy assessment is the collection of ground truth samples of each land cover feature. Ground truth samples may be collected by field surveying, from topographical maps, previously classified images, aerial photographs or any combination of these sources. Samples for accuracy assessment need to be different from the training samples used in the classification process.

Aalderes, in Burrough and McDonnell (1998), grouped GIS data accuracy into thematic accuracy, positional accuracy and temporal accuracy. The first two factors apply to classified images. As with GIS analyses errors can occur at various stages in the classification process and can be propagated throughout. One can therefore conclude that the performance of an image classifier depends on the algorithm employed and the accuracy (which includes image quality, image registration and consistencies and completeness) of the input layers that were used. Classification accuracy also varies with scale, showing a general trend of accuracy increase with coarser levels of spatial aggregation of pixels. Accuracy is also related to the number of land cover classes to be identified in a classification, and in addition to increasing the level of effort to create a land cover classification product, higher numbers of classes generally result in higher levels of error (Vogelmann *et al.* 2001).

The commonly employed method for assessing the accuracy of a classifier is by means of an error matrix. An error matrix reveals two basic kinds of errors namely the user's and producer's accuracy. By examining the relationship between these errors a map user can obtain an understanding of the varied reliabilities of classes on the map, while the map producer acquires insight into the performance of the process that generated the map. Measures such as these do not adjust for chance agreement between the expected and observed and they could be dependent on the samples and sampling strategy used in the analysis.

These deficiencies were to a large extent solved with the introduction of the κ (kappa) measure (formula 3.10), which measures the difference between the observed agreement between two maps and the agreement that could be achieved solely by chance matching (Campbell 1996).

$$\kappa = \frac{observed - expected}{1 - expected} \quad \text{.... (3.10)}$$

It indicates how much better a classification process performed than would be expected from chance assignment of pixels to categories. The kappa measure approaches +1.0 as the percentage of correctly identified pixels approaches 100 and the contribution of chance agreement decreases to 0. Measures near zero indicate that the chance agreement and contribution from the classification are about equal. The kappa index can assume negative values if chance agreement increases and percentage of contribution from the classification decreases (Campbell 2002).

Most of the pixels (99.66%) within the study area were classified (see Appendix B for sub-set classification images). The unclassified pixels were grouped into regions using the 'Clump' function to see if they formed substantial areas that could represent specific land cover classes. The average regions size was 3 pixels and the maximum 91. The few regions that were bigger in size were examined and found to belong to water bodies, wetlands and burnt areas. Variable constraints of water bodies, wetlands and burnt areas compound models were then relaxed to include these pixels (these changes are reflected in the compound models discussed earlier in the chapter). More than 80% of the unclassified pixels were reclassified in this manner. The remaining unclassified pixels were less than 3 pixels in size and were conflated by using a low-pass contextual filter.

Accuracy assessment was carried out on the classification result using the ERDAS Imagine (ERDAS 2001) 'accuracy assessment' facility. A total of 5062 pixels were used as input, with each class having on average 374 reference point/pixel samples. As not enough reliable reference samples were available for the shrub land cover, it was omitted from the accuracy assessment. Results of the accuracy assessment are presented in Table 3.2.

Table 3.2: Accuracy assessment result

Class Name	Reference Totals	Classified Totals	Number Correct	Producer's Accuracy	User's Accuracy	Kappa stat.
Burnt/fire-scarred areas	367	306	305	83.11%	99.67%	.9965
Water bodies	246	239	238	96.75%	99.58%	.9956
Wetlands	327	292	274	83.79%	93.84%	.9341
Forest plantations	595	555	509	85.55%	91.71%	.9061
Cultivated land	565	704	521	92.21%	74.01%	.7074
Residential	491	495	428	87.17%	86.46%	.8501
Indigenous forests	547	536	438	80.07%	81.72%	.7950
Improved grass land	193	130	130	67.36%	100.00%	1.000
Thicket bushland	333	362	265	79.58%	73.20%	.7132
Commer./indust. & trans.	370	312	312	84.32%	100.00%	1.000
Degraded land	242	267	199	82.23%	74.53%	.7325
Unimproved grass	300	428	248	82.67%	57.94%	.5529
Shade	65	63	60	92.31%	95.24%	.9518
Mines & quarries	231	190	187	80.95%	98.42%	.9835

Overall Classification Accuracy = 84.31%

Overall Kappa Statistics = 0.8294

As the result above shows 14% of the land cover classes have a producer's accuracy of less than 80% and 28% of them have a user's accuracy of less than 80%. Unimproved grasslands showed the lowest user's accuracy of 57% with a kappa measure of 0.5529. The low accuracy can be attributed to the absence of effective explanatory ancillary data. Identification of this class relied on vegetation indices that were derived from the spectral information. The confusion with agricultural fallow areas has also been a significant factor contributing to the low accuracy obtained.

Cultivated lands also produced a low kappa statistic. This land cover includes a variety of agricultural vegetation types that were not clearly distinguished by NDVI. The application of supervised classification methods on NDVI to identify cultivated lands proved to be less successful. Many researchers use phenological methods to identify cultivated crops. Unfortunately this method could not be used in this study due to unavailability of multi-temporal satellite images.

Other land cover classes that had low kappa statistics include thicket and bushlands and degraded lands. The reason for their low accuracy can to a large extent be attributed to the spectral confusion with other classes.

More than two thirds of the land cover classes were however accurately classified, with kappa statistics of 0.80 and better. Improved grasslands and commercial, industrial and transportation areas showed a user's accuracy of 100%. This means that none of the pixels belonging to other classes within the accuracy-testing sample were incorrectly classified as either of these classes. The accuracy assessment rated the overall accuracy of the expert classification system at 84.31% and kappa statistics of 0.829. This high accuracy can be attributed to the integration of ancillary, spectral information and expert knowledge.

From these results it is clear that the selection of an appropriate classifier is one of the major factors that influence classification accuracy. In the next and final chapter, accuracy and other criteria identified from the literature are used to evaluate the expert system developed in this study.

CHAPTER FOUR: DISCUSSION

4.1 EVALUATION

The primary objective of this study was to develop an expert system and to test its feasibility and suitability for the National Land Cover Project (NLC). In this chapter, the system is evaluated based on criteria identified from the literature. These are structure, complexity, accuracy, ancillary data (information requirement) and explanatory variables and rules.

4.1.1 Structure

According to Argialas & Harlow (1990), the representation of expert knowledge in unordered and unstructured sets of rules is not ideal. Logical structure and knowledge representation is therefore an important criterion that can be used to evaluate an expert classification model. Although the expert classifier developed in this study is less structured than some other models, the modular structure provides the flexibility to modify individual compound and basic models. It also allows additional models to be easily incorporated. Fine-tuning the model for a different area or time should therefore only involve minor changes.

4.1.2 Complexity

The developed expert classifier is less complex when compared to other non-parametric neural networks and fuzzy classifiers that implement complex statistical formulae and algorithms. The model is simple enough to be implemented (and modified) in most GIS or RS software packages such as ArcView, ArcGIS, IDRISI and ER mapper. The logical structure is understandable and enables the less technical user to interpret results (including intermediate results) for model tweaking.

4.1.3 Ancillary data

The developed model has the disadvantage of requiring complete and consistent ancillary information as input, which is not always available. Although the use of more ancillary data may increase accuracy, it could also have a negative effect on accuracy if the input layers are not accurate. The use of ancillary data from different sources makes the model prone to error propagation as most of input layers will have variable positional, thematic and temporal accuracy (Burrough and McDonell 1998). In contrast, artificial neural networks and fuzzy systems are more tolerant to noise within the data, limited training data, and inaccurate or missing data (Hepner *et al.* 1990; Kalogiron 2002; Campbell 2002). These systems are however much more complex (Berberoglu *et al.* 2000).

4.1.4 Explanatory variables and rules

One of the biggest challenges of this research was to identify effective explanatory ancillary data and to determine the rules based on thresholds obtained from training areas. Automated statistical techniques (e.g. CART, Classification and Regression Trees) could have been employed for this purpose, but (as mentioned in section 3.1) this technique requires suitable training samples. In addition, the different compound models needed different input explanatory variables. The inference of rules based on expert knowledge was therefore more suitable in such cases.

4.1.5 Accuracy

The main strength of expert systems is their ability to integrate spectral and ancillary information and use them in a step-wise classification process. The maximization or identification of potential pixels before they are actually assigned to a particular land cover class enables the avoidance of potential confusion that can occur among land cover classes. The identification of potential pixels is done by using ancillary information, which permits the making of contextual decisions. This contributes to the accuracy of land cover identification. For some land cover classes such as unimproved grasslands the finding of effective ancillary information may be difficult.

The overall accuracy attained by this classification model was 84.31% with a kappa statistic of 0.8294. This level of accuracy should be suitable for most earth resource management purposes. Certain classes, e.g. water and forest plantation, have higher accuracy than other classes (see Table 3.2). This may be due to the availability of appropriate and accurate variables (data) or the nature of the classes itself. While not effective for cultivated and unimproved grasslands, the use of vegetation indexes, topographic and spectral information for indigenous forests resulted in acceptable classification accuracy.

The integration of ancillary and spectral information proved to be effective for most land cover classes and resulted in higher accuracies. Although aerial photographs of the study area would have been suitable to further assess the accuracy of the model, they were not available. Visual interpretation as well as comparison with other classifications (NLC 94 and CGA (2003)) show that the model performed well in terms of its overall accuracy. The results are encouraging, but more research is needed to further improve accuracy.

4.2 DISCUSSION

This study shows that the use of expert systems for the NLC or similar projects is feasible. These systems allow users to generate local objective-specific land cover classes without having to do lengthy data preparation and manipulation prone to error generation. They are especially suitable for such projects because they will ensure that the classification process is transparent and that the data is standardized, which will improve comparability of classification results in different areas and times.

Expert systems also provide the mechanism for making all the information related to land cover/use to be significant in the classification process. This becomes more applicable for land cover/use categories at Level III. For example certain land use classes are not supposed to be located in close proximity to each other (e.g. industrial and educational built-up areas). Such rules can be used in expert systems to guide the classification process.

Considering that most decision makers often lack the necessary technical abilities to perform land use/cover classification, expert systems can be automated and customized to provide those in authority an ability to generate project specific land cover/use classes. The supply of data layers (that include spectral clusters, ancillary information and training data, which can provide a great deal of information) and metadata of the rules, both in terms of spatial and textual format, via the Internet can provide users with the flexibility to download both the database variables and rules and to even make simple for modification their own purposes (Collin *et al.* 2000). Since the Internet has revolutionized GIS, such a set-up might provide the user with a quick preview of land cover for a specified area.

4.3 FURTHER RESEARCH

Many land cover/use classes, such as cultivated lands and settlement areas are affected by topography. Topographic indices (e.g. slope and local relief) need to be explored to develop a model for use as ancillary information. Texture has also been very useful ancillary information and textural classifiers can be used in expert models. Combination of such models with other information such as vegetation indices or synthetic bands can prove to be effective. The assimilation of such information in techniques such as object-oriented classification is untapped potential. Object-oriented image classification as currently implemented only in the eCognition Imaging software package (Baatz *et al.* 2002) is algorithm is ideal for the hierarchical segmentation of a target area. In addition to speeding up the classification process, object-oriented techniques might reduce the number of variables (ancillary layers) and rules needed in a classification process. Object-oriented expert systems can prove to be superior to pixel based expert systems, as they can incorporate functional aspects of land cover classes. Moreover pixel based systems can result into more confusion of pixels; since pixel based systems operate at pixel level.

Further research in the area of building expert systems implemented in object-oriented packages is a worthy future research avenue.

4.4 CONCLUSIONS

Land cover classification is a lengthy process that starts with choosing appropriate imagery for the land cover objects in a classification scheme. There are numerous techniques available; most of which are implemented in remote sensing image processing software. Broadly, image classifiers may be divided into parametric and non-parametric. Although the non-parametric classifiers are known to be more flexible, giving the analyst the opportunity to interact with the classification process, most of them are based on complex algorithms, such as neural networks and fuzzy classifiers.

The expert classification model developed in this study is simple to implement and has an easily interpretable logical structure of sub-models that can be modified without major restructuring. In addition, its modular structure is easy to understand and robust.

The use of such a classification model within a standardized framework ensures the development of a database that will enable users to easily share classification models and results. The use of expert systems is therefore feasible and highly suitable for National Land Cover (NLC) projects.

REFERENCES

ACRES 2003. *Landsat* [online]. Canberra: Geoscience Australia. Available from: http://www.ga.gov.au/acres/prod_ser/landdata.htm [Accessed 26 April 2004].

Argialas PD & Harlow AC 1990. Computational image interpretation models: an overview and a perspective. *Photogrammetric Engineering and Remote Sensing* 56, 6: 871-886.

Baatz M, Benz U, Dehghani S, Heynen M, Höltje A, Hofmann P, Lingenfelder I, Mimler M, Sohlbach M, Weber M & Willhauck G 2002. *eCognition user guide 3*. Definiens Imaging GmbH.

Barr S & Barnsley M 2000. Reducing structural clutter in land-cover classifications of high spatial resolution remotely-sensed images for urban land use mapping. *Computers and Geosciences* 26:433-449.

Berberoglu S, Llyod CD, Atkinson PM, & Curran PJ 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. *Computer and Geosciences* 26: 385-396.

Bolstad PV & Lillesand TM 1991. Rapid maximum likelihood classification. *Photogrammetric Engineering and Remote Sensing* 57, 1:67-74.

Bolstad PV & Lillesand TM 1992. Rule-based classification models: flexible integration of satellite imagery and thematic spatial data. *Photogrammetric Engineering and Remote Sensing* 58, 7: 965-971.

Boyd DS, Foody GM & Ripple WJ 2002. Evaluation of approaches for forest cover estimation in the pacific Northwest, USA, using remote sensing. *Applied Geography* 22: 375-392

Brabyn, L 1998. *GIS analysis of Macro Landform* [online]. Hamilton, New Zealand: University of Waikato. Available from: http://divcom.otago.ac.nz/sirc/webpages/conferences/SIRC98/98Absracts/98Brabyn/Brabyn.pdf [accessed 4 May 2004].

Brown de Colstoun CE, Story HM, Thompson C, Commisso K, Smith GT & Irons RJ 2003. National park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sensing of Environment* 85: 316-327.

Buiten HJ 1993. General View of remote sensing as a source of information. In Buiten HJ & Clevers JGPW (eds) *Land observation by remote sensing: theory and applications,* pp9-26. Paris: Gordon and Breach Science Publishers.

Burrough PA 1993. Geographical information system techniques for linking remote sensing images with other spatial data. In Buiten HJ & Clevers JGPW (eds) *Land observation by remote sensing: theory and applications,* pp323-336. Paris: Gordon and Breach Science Publishers.

Burrough AP & McDonnell AR 1998. *Principles of Geographical Information Systems*. Oxford: Butler & Tanner Ltd.

Campbell BJ 1996. *Introduction to remote sensing*. 2nd ed. London: Taylor & Francis.

Campbell BJ 2002. *Introduction to remote sensing.* 3rd ed. London: Taylor & Francis.

CGA 2003. National Land Cover 2000: satellite image processing and classification guide. Stellenbosch: Centre for Geographical Analysis, University of Stellenbosch.

Chen MJ & Cihlar J 1996. Retrieving leaf area index of bored conifer forests using Landsat TM images. *Remote Sensing of Environment* 55, 2: 153 – 162.

Cohen Y & Shoshany M 2002. A national knowledge-based crop recognition in Mediterranean environment. *International Journal of Applied Earth Observation and Geoinformation* 4: 75-87.

Collin H, Huang C, Yang L & Wylie B 2000. *Development of CIRCA 2000 landcover database for the united states*. Sioux Falls: USGS/EROS data center.

Cordon O, Del Jesus JM, & Herrera F 1999. A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning* 20: 21-41.

CSIR 2000. *National Land Cover 2000: Illustrated field guide*. Working document. Pretoria. CSIR.

CSIR 2003. *The South African national land cover Database* [online]. Pretoria: CSIR. Available from: http://www.csir.co.za/plsql/ptl0002/PTL0002_PGE015_PRODUCT?PRODUCT_NO=1600858 [accessed 20 October 2003].

Debeir O, Van de Steen I, Latinne P, Van Ham P & Wolff E 2002. Textural and contextual land-cover classification using single and multiple classifier systems. *Photogrammetric Engineering and Remote sensing* 68, 6:597-605.

DeMers NM 2000. *Fundamentals of geographic information systems*. 2nd Ed. New York: John Wiley & Sons, Inc.

Dilley L, Earle J, Euston-Brown K, Keats G & Ravenscroft G 2001. *Focus on geography*. Cape Town: Maskew Miller Longman (Pty) Ltd.

Dorren KAL, Maier B & Seijmonsbergen CA 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology and Management* 183: 31-46.

Dymond CC, Mladenoff JD & Radeloff CV 2002. Phenological differences in Tasseled Cap indices improve deciduous forest classification. *Remote Sensing of Environment* 80: 460- 472.

Edwards AJ & Mumby PJ 1999. The importance of acquiring images of the appropriate scale and spatial resolution for your objectives. In Edwards AJ (ed) *Applications of satellite and airborne image data to coastal management: seventh computer-based learning module*, pp65-74. Paris: UNESCO.

ERDAS 2001. *ERDAS field guide*. 5th ed. Atlanta: ERDAS, Inc.

Ferro SJC 1998. Scale & texture in digital image classification. MA thesis. Morgantown: University West Virginia, Department of Geography.

Foody GM 2000. Estimation of sub-pixel land cover composition in the presence of untrained classes. *Computer and Geosciences* 26: 469-478.

Fuller RM, Smith GM & Devereux BJ (In press). The characterization and measurement of land cover change through remote sensing: problems in operational applications. *International Journal of Applied Earth Observation and Geoinformation.*

Gegg G, Gunther O & Riekert W 1990. Knowledge-based processing of remote sensing data: development of the expert system RESEDA. In Ehler Manfred (ed) *Advances in spatial information extraction and analysis for remote sensing*, pp 71-80. Bethesda, Maryland: American society for photogrammetry and remote sensing.

Geldenhuys CJ 1994. Berg-wind fires and the location pattern of forest patches in the southern Cape landscape, South Africa. *Journal of Biogeography* 21: 49-62.

Gerbrands JJ 1993. Digital image processing. In Buiten HJ & Clevers JGPW (eds) *Land observation by remote sensing: theory and applications,* pp323-336. Paris: Gordon and Breach Science Publishers.

Geoscience 2003. *Summary of economic geology of provinces: KwaZulu-Natal* [online]. Pretoria: Council for Geoscience. Available from: http://www.geoscience.org.za/samindaba/maps/kwazulu.htm [accessed 10 April 2004].

Gumbricht T, McCarthy J & Mahlander C 1996. Digital interpretation and management of land cover: a case study of Cyprus. *Ecological Engineering* 6: 273-279.

Hansen M, Dubayah R & DeFries R 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing* 17, 5: 1075 – 1081.

Hepner FG, Logan T, Ritter N, & Bryant N 1990. Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing* 56,4: 469-473.

Huang C, Yang L, Homer C, Coan M, Rykus R, Zhang Z, Wylie B, Hegge K, Zhu Z, Lister A, Hoppus M, Tymcio R, DeBlander L, Cooke W, McRobersts R, Wendt D & Weyermann D 2002a. *Synergistic use of FIA plot data and landsat 7 ETM+ images for large area forest mapping* [online]. Denver: USGS. Available from: http://landcover.usgs.gov/pdf/synergistic.pdf [accessed November 2003].

Huang C, Wylie B, Yang L, Homer C & Zylstra G 2002b. *Deviation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance* [online]. Denver: USGS. Available from: http://landcover.usgs.gov/pdf/tasseled.pdf [accessed July 2003].

Huang C, Yang L, Wylie B & Homer C 2001. *A strategy for estimating tree canopy density using Landsat 7 ETM+ and high resolution images over large areas* [online]. Denver: USGS. Available from: http://landcover.usgs.gov/pdf/canopy_density.pdf [accessed December 2003].

Hutchinson CF 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering and Remote Sensing* 48, 1: 123-130.

Jansen JM & Gregorio AD 2003. Land-use data collection using the "land cover classification system": results from a case study in Kenya. *Land Use Policy* 20: 131-148.

Kalogiron S 2002. Expert systems and GIS: an application of land suitability evaluation. *Computers, Environment and Urban Systems* 26: 89-112.

Langley KS, Cheshire HM & Humes SK 2001. A comparison of single date and multi-temporal satellite image classification in a semi-arid grassland. *Journal of Arid Environments* 49: 401- 411.

Lawrence LR & Wright A 2001. Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric Engineering and Remote Sensing* 67, 10: 1137-1142.

Lindgren DT 1985. *Land use planning and remote sensing.* Boston: Martinus Nijhoff publishers.

Lira J & Maletti G 2002. A supervised contextual classifier based on a region-growth algorithm. *Computers and Geosciences* 28: 951-959.

Liu Xue-Hua, Skidmore AK & Oosten VH 2002. Integration of classification methods for improvement of land-cover map accuracy. *ISPRS Journal of Photogrammety and Remote Sensing* 56: 257-268.

Thompson M 1996. A standard land-cover classification scheme for remote-sensing applications in *South Africa. South African Journal of Science* 92: 34-43.

Thompson M, Marneweck G, Bell S, Kotze D, Muller J, Cox D & Clark R 2002. *A methodology proposed for a South African national wetland inventory.* Wetlands conservation program.Pretoria. Department of Environmental Affairs and Tourism.

Tourism KwaZulu-Natal 2004. *Climate and weather in KwaZulu-Natal* [online]. Durban: Tourism KwaZulu-Natal. Available from: http://www.kzn.org.za/kzn/612.xml [accessed July 2004].

Mehldau G & Schowengerdt RA 1990. A C-extension for rule-based image classification systems. *Photogrammetric Engineering and Remote Sensing* 56, 6:887-892.

Molenaar M 1993. Digital image interpretation. In Buiten HJ & Clevers JGPW (eds) *Land observation by remote sensing: theory and applications,* pp257-272. Paris: Gordon and Breach Science Publishers.

Moller-Jensen L 1998. Classification of urban land cover based on expert systems, object models and texture. *Computer, Environment and Urban Systems* 21, 3/4: 291-302.

Mongkolsawat C & Thirangoon P 1990. *Land-cover change detection using digital analysis of remotely sensed satellite data: a methodological study* [online]. Utter Pradesh, India : GISdevelopmentAvailable from:http://www.gisdevelopment.net/aars/acrs/1990/G/lclu003.shtml [accessed 9 May 2003].

Mucina L, Geldenhuys C, Lawes M, Eeley H, Adie H, Von Maltitz G, Vink D, Flerning G & B Bailey C 2003. Classification system for South African indigenous forests. ENV-P-C2 2003-017. Pretoria: CSIR Environmentek.

Mulders MA & Jordens ER 1993. Remote sensing and Land evaluation. In Buiten HJ & Clevers JGPW (eds) *Land observation by remote sensing: theory and applications,* pp405-412. Paris: Gordon and Breach Science Publishers.

Mumby P & Clark C 2000. Radiometric correction of satellite and airborne images. In Edwards AJ (ed) *Applications of satellite and airborne image data to coastal management: seventh computer-based learning modul*e, pp65-74. Paris: UNESCO.

Oetter RD, Cohen BW, Berterretche M, Maiersperger KT & Kennedy ER 2000. Land cover mapping in an agricultural setting using multiseasonal thematic mapper data. *Remote Sensing of Environment* 79: 139 – 155.

Pal M & Mather MP 2002. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* 86: 554 – 565.

PCI/Geomatica 2003. PCI Geomatica for Windows –Version 9.0.0. Help document.

Raptis VS, Vaughan RA & Wright GG (In press). The effect of scaling on land-cover classification from satellite data. *Computer and Geosciences.*

Richards JA 1986. *Remote sensing digital image analysis.* Paris: Springer-Verlag.

Sader AS, Ahl D & Liou W 1995. Accuracy of Landsat TM and GIS rule-based methods for forest wetland classification in Maine. *Remote Sensing of Environment* 53: 133 – 144.

Shrestha PD 2000. *Image transformation and geo-statistical techniques to enhance and map soil features: an application in assessing sedimentation problem in southern Nepal* [online]. Enschede, the Netherlands: International Institute for Geo-Information Science and Earth Observation (ITC).Available from: http://www.itc.nl/library/papers/shrestha.image.ACRS.2002. pdf [accessed 21 October 2003].

Skidmore AK, Turner BJ, Brinkhof W & Knowles E 1997. Performance of a neural network: mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering and Remote Sensing* 63, 5: 501-514.

Smit GS 1993. Forestry and remote sensing. In Buiten HJ & Clevers JGPW (eds) *Land observation by remote sensing: theory and applications,* pp451-457. Paris: Gordon and Breach Science Publishers.

Smit JW 2001. A comparison of satellite remote sensing techniques for mapping fire scars in limestone fynbos. MA thesis. Stellenbosch. University of Stellenbosch, Department of Geography and Environmental Studies.

Smith L 2003. *Remote sensing technologies* [online]. Topanga, California: GIS lounge. Available from: http://gislounge.com/features/aa121900.shtml [accessed March 2004].

Stefanov WL, Ramsey MS & Christensen PR 2001. Monitoring urban land cover change: an expert system approach to land cover classification of semiarid to arid urban centres. *Remote Sensing of Environment* 77: 173-185.

Stuckens J, Coppin RP & Bauer EM 2000. Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment* 71: 282-296.

Tapiador FJ & Casanova JL 2003. Land use mapping methodology using remote sensing for the regional planning directives in Segovia, Spain. *Landscape and Urban Planning* 62: 103-115.

Tsatsoulis C 1993. Expert systems in remote sensing applications: Review article. *IEEE Geoscience and Remote Sensing Society Newsletter* 7-15.

Vogelmann EJ, Howard MS, Yang L, Larson RC, Wylie KB & Van Driel N 2001. Completion of the 1999s National Land Cover data set for the conterminous united states from landsat thematic mapper data and ancillary data sources. *Photogrammetry Engineering and Remote Sensing, Highlight*: 650-662.

Wulder M 2002. Mapping the land cover of the forest part of Canada with Landsat data. P Proceeding of 2002 International GeoScience and Remote Sensing Symposium, Toronto: Canada.

Zhang Y 1999. Optimization of building detection in satellite images by combining mulitspectral classification and texture filtering. *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 50 – 60.

Appendix A: Tasseled Cap coefficients for Landsat 7 ETM+ at-satellite reflectance.

Index	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Brightness	0.3561	0.3972	0.3904	0.6966	0.2286	0.1596
Greenness	-0.3344	-0.3544	-0.4556	0.6966	-0.0242	-0.2630
Wetness	0.2626	0.2141	0.0926	0.0656	-0.7629	-0.5388
Fourth	0.0805	-0.0498	0.1950	-0.1327	0.5752	-0.7775
Fifth	-07252	-0.0202	0.6683	0.0631	-0.1494	-0.0274
Sixth	0.4000	-0.8172	0.3832	0.0602	-0.1095	0.0985

Source: Huang 2002b: 6

Appendix B: Sub-set images of the classification map and their corresponding composite false colour (4-5-3) satellite images.

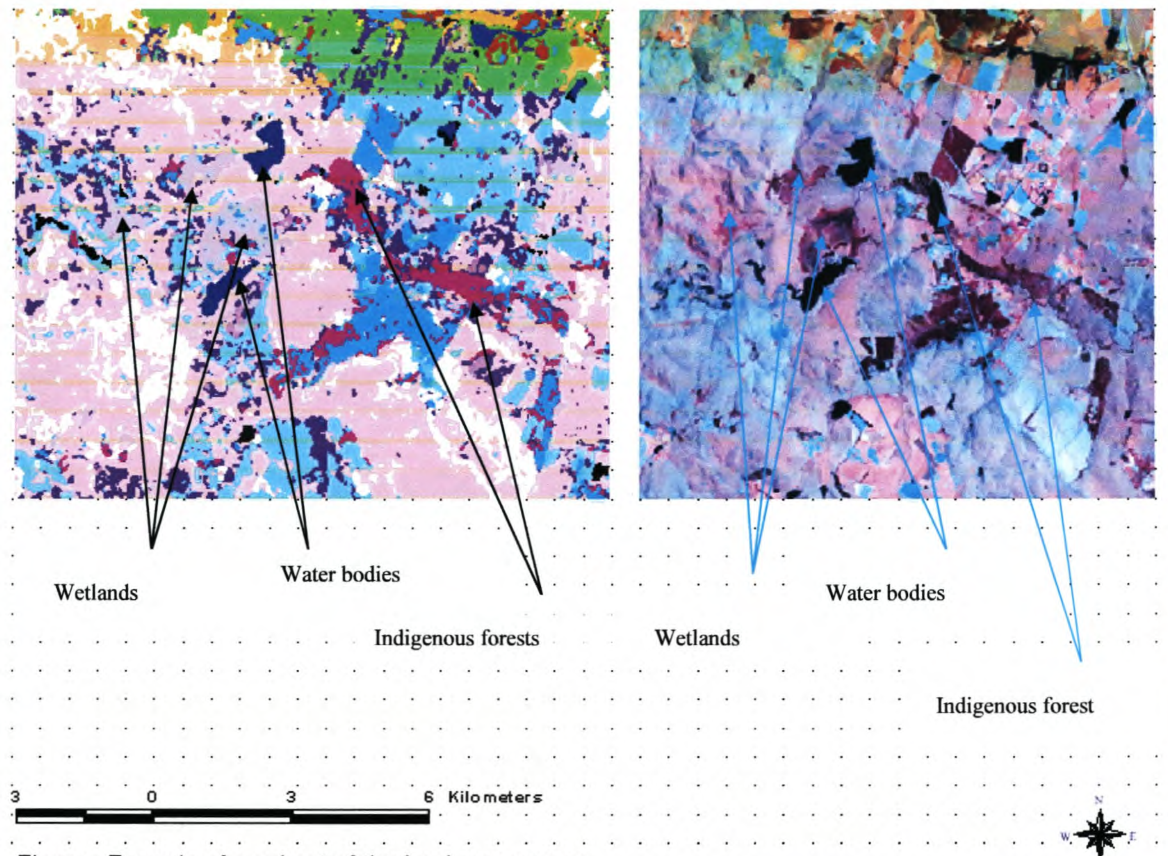
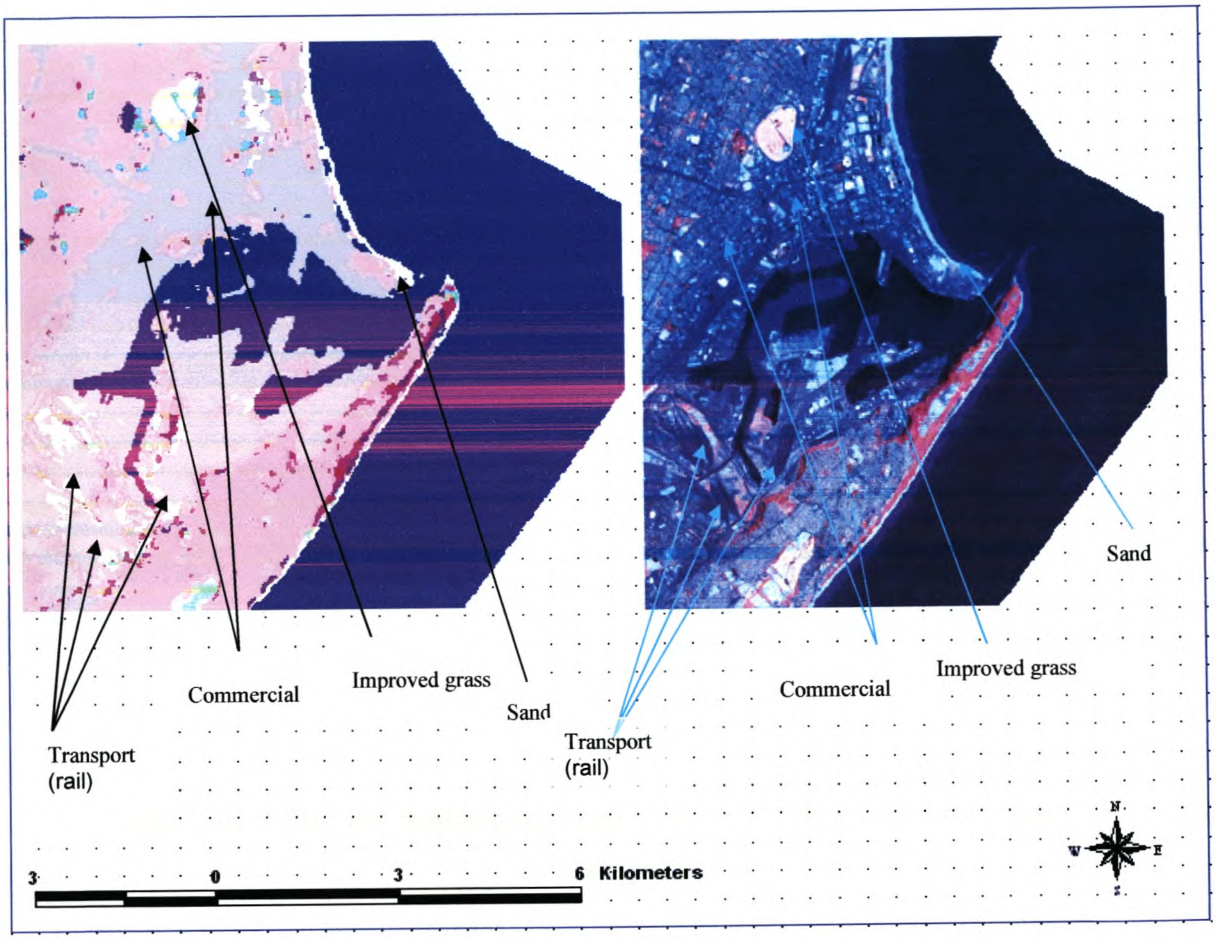
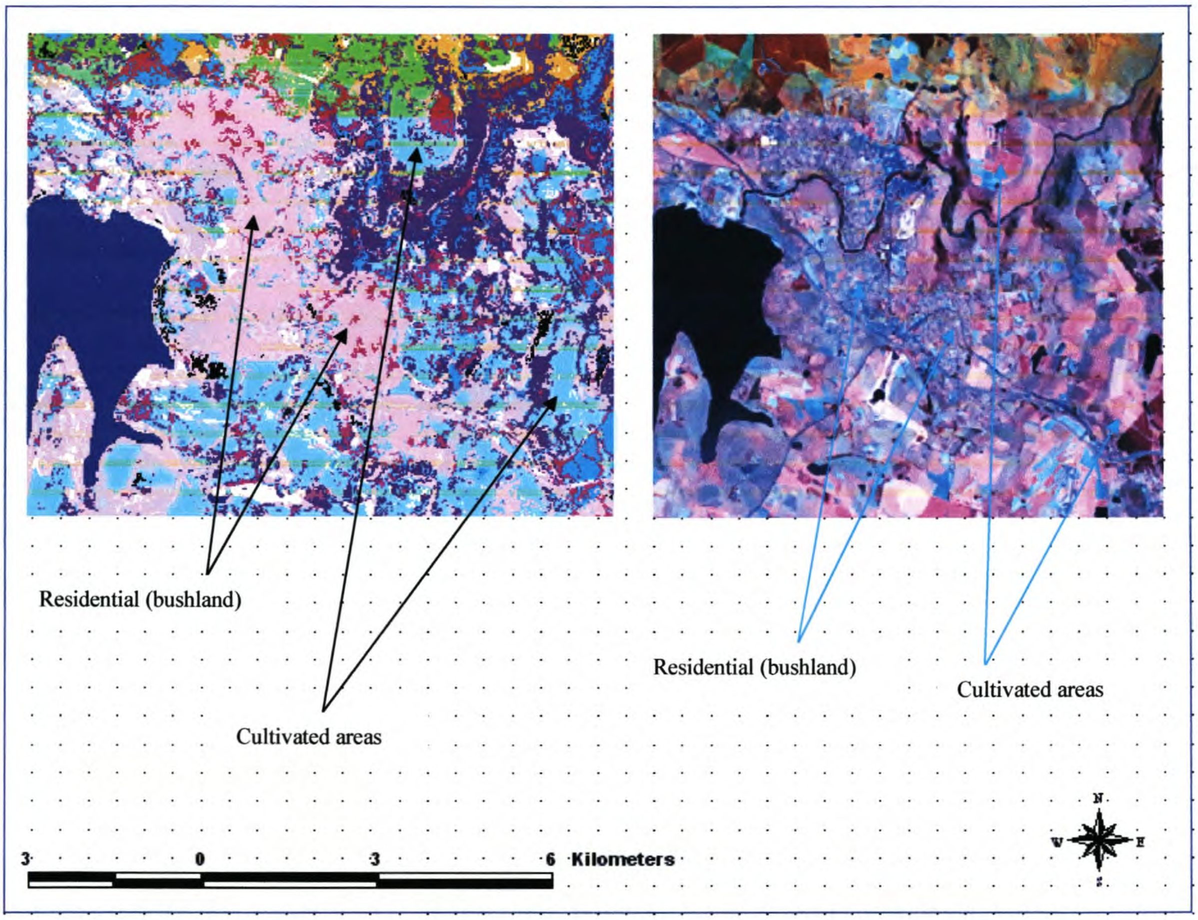
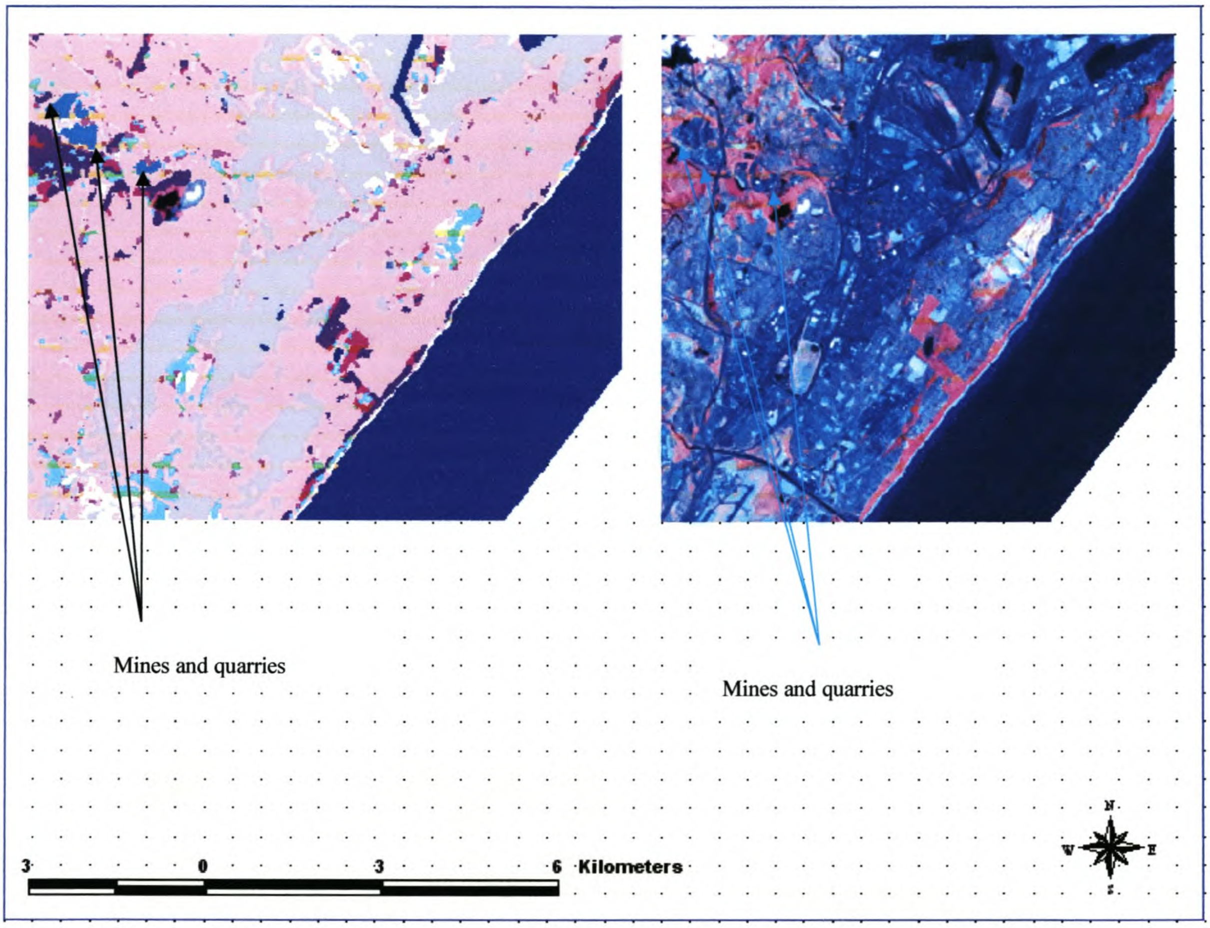
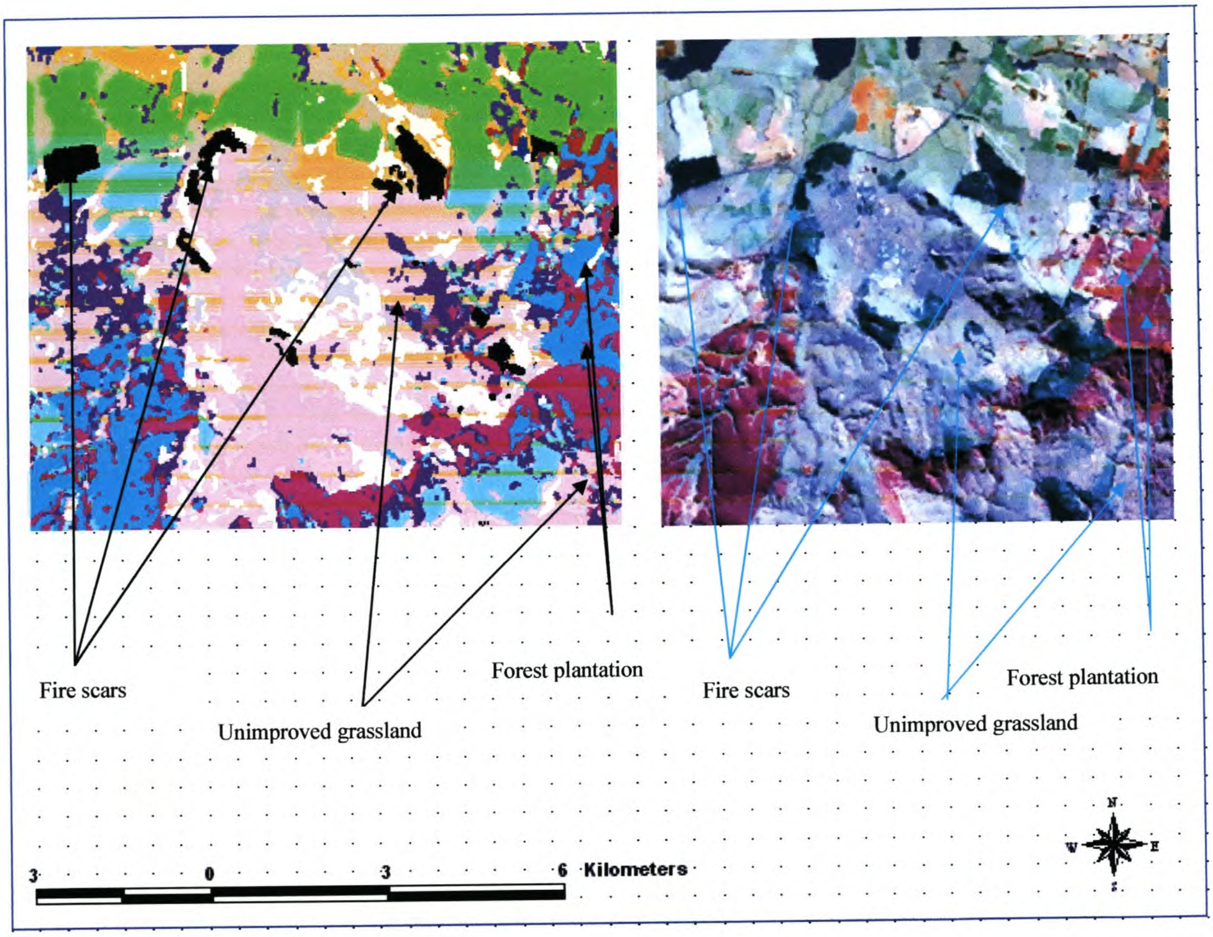


Figure : Zoom in of a subset of the land cover mage


Commercial
Improved grass
Sand
Transport
(rail)
Sand
Improved grass
Commercial
Transport
(rail)
3
0
3
6
Kilometers
N
W
E
S


Residential (bushland)
Cultivated areas
Residential (bushland)
Cultivated areas
3
0
3
6
Kilometers
N
W
E
S


Mines and quarries
Mines and quarries
3
0
3
6
Kilometers
N
W
E
S


Fire scars
Unimproved grassland
Forest plantation
Fire scars
Unimproved grassland
Forest plantation
3
0
3
6 Kilometers
N
W
E
S