THE UNIVERSITY OF ZAMBIA

SCHOOL OF ENGINEERING

DEPARTMENT OF GEOMATIC ENGINEERING

Vegetation Type Mapping Using Decision Tree Classification (DTC)

BY

Lusekelo Kasunga

A Dissertation submitted to the University of Zambia in partial fulfilment of the requirements of a **Master of Engineering degree in GeoInformatics and Geodesy**

January, 2017 **Lusaka, Zambia**

DECLARATION

I, **Lusekelo Kasunga**, hereby declare that the contents of this dissertation represent my own work and that it has not previously been submitted to this or any other university for any academic qualification.

-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.- -.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-

Signed **Date**

CERTIFICATE OF APPROVAL

This thesis by **Lusekelo Kasunga** is approved as partial fulfilment of the requirements for the award of the Master's Degree of Engineering (MEng) in GeoInformatics and Geodesy by the University of Zambia.

Examiners:	**Signature**	**Date**
1.	....	....
2.	....	....
3.	....	....

Supervised by:

Dr. A. Mulolwa	....	....

DEDICATION

To my loving Parents ….. Bernard and Lizzie Kasunga

And also to my two elder brothers…… for their constant support since my pre-school days

Table of Contents

List of Tables

List of Figures

ABBREVIATION AND ACRONYMS

AEZ	Agro-Ecological Zone
AOI	Area of Interest
AVHRR	Advanced Very High Resolution Radiometer
BIL	Binary Interleaved by Line
DEM	Digital Elevation Model
DN	Digital Number
DT	Decision Tree
DTC	Decision Tree Classification
ENVI	Environment for Visualization
EVI	Enhanced Vegetative Index
ETM	Enhanced Thematic Mapper
FLAASH	Fast Line-of-sight- Atmospheric Analysis of Spectral Hypercubes
GDP	Gross Domestic Product
GIS	Geographic Information System
GPS	Global Position System
LDA	Linear Discrimination Analysis
MDM	Minimum Distance-to-Mean
MLC	Maximum Likelihood Classification
MODTRAN	Moderate Resolution Atmospheric Transmission
NDVI	Normalized Difference Vegetative Index
NRSC	National Remote Sensing Centre
NSDI	National Spatial Data Infrastructure
WRS	World Reference System
ZARI	Zambia Agriculture Research Institute

ACKNOWLEDGMENT

I am very grateful to many individuals for their precious contributions and support during my masters study at the University of Zambia. First of all, I would like to express my sincerest thanks and appreciation to my supervisor Dr. Augustine Mulolwa for his supervision and valuable guidance during the research period. His sincere technical advice and moral support allowed me to reach this far with my work. I am thankful to the Management and staff at the National Remote Sensing Centre (NRSC) for according me the opportunity to study for my master's degree and for their continued support throughout the study period. Also, special thanks go to the lecturers in the Department of Geomatic Engineering for their unending supervision and guidance up to this far.

I would like to express my sincere gratitude to all of the friends and colleagues in the Department of Geomatic Engineering as well as those in other places especially to my "best man" Chindi Kapembwa for all the support and guidance towards the fruition of this work.

I would like to express my honest and sincere thankfulness to my parents, sisters and brothers for their constant support and good wishes. Last but not the least, my wife, Thelma Mukomba for her unlimited support and endless patience that encouraged and gave me strength to complete the study.

ABSTRACT

Remote sensing provides area coverage, and aids the mapping and classification of land cover features, such as vegetation, soil, water and forests. However, with remote sensing using medium to coarse spatial resolution imagery, some difficulties are encountered in the definition of vegetation type classes based on their spectral responses alone, thus, it is a challenging task to use such images for mapping vegetation at species level. However, when these images are integrated with other ancillary data, it becomes possible to map vegetation at species level. This research endeavored to demonstrate the possibility of integrating spectral with ancillary data in mapping vegetation types of Zambia and, for that, ENVI Decision Tree classifier was used. The nature of the vegetation in an area is determined by a complex combination of effects related to climate, soils, history, fire and human influences. Therefore, this mapping method takes advantage of the relationship that these features, vegetation types, have with their environmental factors, such as soil type and elevation. The study focused on three vegetation types, namely Miombo, Mopane and Munga, and thus the factors that influence their spatial distribution were studied and identified. Based on the literature reviewed and the GIS desktop analysis, it was found that, for the area of study, Miombo had the following factors: Ferralsols as the dominant soil type where it thrives; at 900m-1600m elevations; Band4 reflectance of 0.0 - 1.0; NDVI values of -1.0 to 1.0; maximum Band4/Band3 ratio value of 16.235294; and EVI of -1.0 to 1.0. While for Mopane: Luvisols as the dominant soil types where it thrives; at 700m-900m elevations; Band4 reflectance of 0.0 - 0.5198; NDVI value of -1.0 to 0.857818; maximum Band4/Band3 ratio value of 12.707866; EVI of -1.0 to 1.0; and also significantly occurs in Agro-ecological zone I. For Munga: Luvisols phaezom as the dominant soil types where it thrives; at 580m-1320m elevations; Band4 reflectance of 0.0 - 0.7249; NDVI value of -0.536278 to 0.872132; maximum Band4/Band3 ratio value of 14.641149; and EVI of -1.0 to 1.0. These parameters were used to develop the decision tree classifier binary rules and executed for the final produced map of the three vegetation types. With the decision tree map produced, the study demonstrated the possibility of mapping vegetation at sub-nation level by combining spectral response with other geographic parameters via the use of a decision tree classifier.

Keywords: decision tree classification; land cover; vegetation map;

1. CHAPTER 1-- INTRODUCTION

1.1. Background

Mapping of vegetation has always been an essential prerequisite to effective and efficient management and monitoring of the land resources. Many studies have been carried out to achieve more accurate vegetation maps for improved resource management. The attainment of these more accurate vegetation maps entails the application of better and more advanced mapping methodologies. The traditional methods of mapping vegetation such as field data collection were found to be not effective in mapping the vegetation as the methods were found to be very costly, time consuming and also they produced less accurate maps. These shortcomings, as well as many others, spurred the need for more advanced and effective methods of mapping the Earth surface. Thus the coming of remote sensing revolutionized, generally, the mapping of Earth's surface and specifically, for this study, the mapping of vegetation.

Remote sensing provides an important synoptic view of the Earth's surface; it provides wide areal coverage and thus aids mapping and classification of land cover features, such as vegetation, soil, water and forests (Zakaria, 2010). However, with remote sensing using medium to coarse spatial resolution imagery, some difficulties are encountered in the definition of vegetation type classes based on their spectral responses alone (Zakaria, 2010). These difficulties arise due to the common heterogeneity of the cover type and the factors affecting spectral responses (Zakaria, 2010). Due to this limitation of spatial resolution, medium to coarse resolution imagery are usually used to map vegetation at community level. It is a challenging task to use such images for mapping vegetation at species level, especially in a heterogeneous environment (Xie *et al*, 2008). However, Xie *et al*., (2008) further reported that when these images are integrated with other ancillary data, it becomes possible to map vegetation at species level. Many research studies have found that the integration of ancillary geographical data and multi-spectral satellite data can indeed improve classification results (Ozesmi & Bauer, 2002). Among the studies were those done by Xiaodong *et al*., (2009) in the integration of Landsat TM data with ancillary data to map land cover of a marsh area; Hansen *et al*., (2006) integrated AVHRR imagery with geographical ancillary data to produce a global land cover map using a DT that had a set of 41 metrics generated from five spectral channels and NDVI for input; and studies by Abdelhamid *et al*., (2010) where using a similar method as Hansen *et al.*, (2006) mapped salt- affected soils over large areas. These and many other research studies showed the usefulness and the mapping potential of integrating multi-spectral and other geographical ancillary data.

This method of integrating spectral data with other geographic ancillary datasets, in mapping vegetation types, relies on the relationship that exists between vegetation types and their geographic environmental factors. Researchers such as Zakaria, (2010) observed that the nature of the vegetation in an area is determined by a complex combination of effects related to climate, soils, history, fire and human influences. Thus a method of integrating the satellite data with these ancillary geographical data, which are related to the above mentioned environmental factors, can improve the mapping of vegetation, and for this study Decision Tree classification (DTC) is identified as one such method.

Generally, vegetation is important because it provides a basic foundation for all living beings hence its mapping is valuable. Zambia is endowed with a wide variety of vegetation and for improved management and monitoring of these resources; their mapping at sub-nation level is a primary requirement. This study endeavored to demonstrate the integration of remote sensing data with other ancillary geographical data for mapping the main vegetation types of Zambia using DTC, and the selected study area was a stretch of land covering Copperbelt to the North and Southern Province to the South of Zambia. The area covers Landsat satellite WRS2 path 172 in Zambia (between latitude 12° and 18° S and longitude 26° and 30° E); this area runs across all the three agro ecological zones of Zambia hence it encompasses a wide diversity of vegetation types for mapping.

1.2. Statement of the Problem

Zambia is endowed with a wide variety of vegetation types whose great importance to the nation should not be overlooked. Forests play a crucial role in enhancing human well-being and in sustaining the economy of Zambia. They contribute to economic growth, employment, wealth, export revenues, a stable supply of clean water, recreation and tourism opportunities, as well as essential building materials and energy for a wide range of economic sectors (Turpie *et al.*, 2015). Given the variety of vegetation types and the importance vegetation has to the country, it is of great importance that more improved cost-effective management and monitoring methods are implemented. These methods would enable the detailed management of the resources, that is, management of vegetation at sub nation level. In order to achieve these improved vegetation management and monitoring methods, advanced vegetation mapping approaches at sub national level are required.

However, the achievement of these management and monitoring methods has been hindered by a number of factors which in effect has resulted in inefficient and ineffective management of these resources.

In Zambia the first, and only, detailed vegetation map for the whole country was compiled in 1976 at a scale of 1:500,000 comprising nine tiles and for this map, a combination of aerial photos and traditional mapping methods were used. "Traditional methods (for instance, field surveys, literature reviews, map interpretation and collateral and ancillary data analysis), however, are not cost effective to acquire vegetation covers because they are time consuming, date lagged and often too expensive" (Xie *et al*., 2008). Thus "the technology of remote sensing offers a practical and economical means to study vegetation cover changes, especially over large areas" (Langley *et al*., 2001; Nordberg & Evertson, 2003). Further, classification of vegetation using remote sensing is valuable because it can determine vegetation distribution and occurrence and how the vegetation is influenced by physical soil and atmospheric factors.

However, there are some difficulties in the definition of vegetation classes based on their spectral responses alone, these are caused by the common heterogeneity of the cover type and the factors affecting their spectral responses (Zakaria, 2010). These difficulties are attributed to the limitations of medium to low spatial resolution satellite images, such as Landsat satellite images, for mapping vegetation at species level, especially in a heterogeneous environment (Xie *et al*., 2008). The foregoing has resulted in the mapping of vegetation using "broad" classes such as dense or moderate forests as opposed to vegetation "type" classes such as Miombo, Mopane and parinari forests. While mapping of the vegetation at these "broad" class levels is important for national and international reporting, it does not however guarantee an effective and efficient management of vegetation at the level where these resources are being consumed i.e. at the level of vegetation "type". A case in point is the exploitation of the Mukula tree (Pterocarpus chrysothrix); a tree of great value to the country, at "broad" class level of mapping, estimation of its distribution and volume becomes almost impossible. However, when the medium to low resolution images are integrated with other ancillary data, it becomes possible to map species (Xie *et al.*, 2008). Also due to the variations in atmospheric effects on the images, an object may be captured with varying spectral responses in different spatial locations hence the integration of spectral data with other ancillary data may minimize the misclassifications which may arise as a result of the above.

Therefore, in order to ameliorate the identified challenges, a methodology of mapping vegetation types that incorporates vegetation spectral characteristics and the environmental factors, which influence their spatial distribution such as soil type and elevation, needs to be studied and adopted as a suitable approach for sub national mapping of vegetation. Thus, in this study DTC; which has the ability to integrate spectral data with other ancillary geographical data, was studied and its application demonstrated.

1.3. Aim

The main aim of this research was to produce a vegetation type map of the study area by using medium resolution satellite images integrated with other ancillary geographical data.

1.4. Study Objectives

The main objectives of the research were to:

(i) study the environmental factors which determine the occurrence and spatial distribution of main vegetation types of Zambia and

(ii) apply DTC to mapping these vegetation types of Zambia.

1.5. Research Questions

In order to guide the flow of this research, the following research questions will be addressed:

- What are the main key environmental factors that determine the occurrence and spatial distribution of the main vegetation types of Zambia?
- Is the Decision Tree Classifier appropriate for mapping these vegetation types?
- What is the spatial distribution of the main vegetation types in the study site?

1.6. Significance of the Study

Primarily, this study is a significant endeavor to coming up with a model for mapping vegetation types in Zambia using the classes set by the 1976 vegetation map. The study takes advantage of the correlation between vegetation and environmental factors hence this makes it possible to model the vegetation based on for instance the rainfall recorded, soil type and elevation of an area. This will enable easy and timely production of vegetation maps, and also easy and timely updating of vegetation inventories. Furthermore, vegetation maps produced at sub national level are required for modern detail management and monitoring of vegetation, and thus techniques, such as the one discussed in this study, are required. Institutions such as the Forestry department may adopt such methods of mapping vegetation types as way of improving their monitoring and protection of endangered species such the Mukula tree (Pterocarpus Chrysothrix).

As vegetation types' spectral signatures/reflectance were analyzed, this study serve as a preliminary work to other future studies related to vegetation mapping using remote sensing such as vegetation water content mapping, plants' spectral characteristics mapping using multi-temporal satellite data, and crop yield estimations; such studies may help in ensuring national food security as the yield of crops such as corn may be estimated before the end of the farming season based on the rainfall projections by institution such as the Zambia National Farmers Union in conjunction with the Meteorological department. Therefore, the study may help in extending the applicability of remote sensing to different fields in Zambia.

The mapping method applied in this research required the integration of various datasets such as soil type, elevation and satellite data, from various sources. Therefore, prior to the adoption and eventual implementation of a mapping method, such as DTC, there was need for the production and acquisition of such data sets as the soil map, more accurate elevation data, and updated rainfall data. Thus, such a mapping method would further enhance the need for more research into the production and acquisition of more accurate and updated ancillary data sets as those mentioned in the preceding.

The use of such a mapping approach would enable more detail mapping of features by the use of medium to coarse resolution satellite images. Mapping of these features at such detail would ideally require the use of high resolution images; which are of high cost as compared to the medium to coarse resolution satellite images, hence making this mapping method cost effective.

2. CHAPTER 2--LITERATURE REVIEW

2.1. Overview of Remote Sensing and GIS

Lillesand *et al*., (1979) formally defined remote sensing as the science and art of obtaining information about an object, area or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation. This science encompasses a multitude of activities, which include the operation of satellite systems, image data acquisition and storage, as well as the subsequent data processing, interpretation and dissemination of the processed data and image products (Zakaria, 2010).

The evolution of remote sensing, as an applied science, has been in tandem with other technological advancement, such as the improvement in optics, sensors electronics, satellite platforms, transmission systems and computer data processing (Zakaria, 2010). The remote sensing technology has proved to be the most ideal method for acquiring information of various land cover and land use of an area and its application has found great use in mapping features on large areas. Remote sensing relies on spectral characteristics of the objects being mapped as ideally each object has its own spectral characteristic which is referred to as spectral signature and hence it can be identified and mapped. Therefore, in remote sensing the spectral characteristics of these objects are captured by satellite images of various spatial resolutions. The spatial resolution may be said to denote the pixel size of the satellite image covering the earth's surface (Mather *et al.,* 2009) in actual sense it is the smallest object that can be resolved on the ground, and satellite imagery are usually categorized in terms of their spatial resolutions with high spatial resolution imagery ranging under 2m such as Ikonos imagery, medium spatial resolution ranging between 2m and 30m these are images such as Landsat imagery, and low spatial resolution images are larger than 30m such as METEOSAT-8 imagery. Thus, the higher the spatial resolution, the more detailed the satellite image is and as a result, especially if the landscape to be mapped is highly fragmented and land cover parcels have irregular shapes, pixel mixing is reduced. However, the possibility of pixel mixing in a classification increases with lower spatial resolution imagery.

As such, in mapping objects of similar spectral characteristic, high spatial resolution imagery is able to distinguish between the objects while the medium to coarse spatial resolution imagery may result in mixing the objects' pixels. For that reason, the spatial resolution of imagery plays a significant role in determining the image classification output of remote sensing, and in the case of this study, the vegetation classification outputs.

Classifying and mapping of vegetation is an important technical task for managing natural resources as vegetation provides a base for all living beings and plays an essential role in affecting global climate change, such as influencing terrestrial CO_2 (Xiao *et al*., 2004). "Vegetation mapping also presents valuable information for understanding the natural and man-made environments through quantifying vegetation cover from local to global scales at a given point of time or over a continuous period. It is critical to obtain current states of vegetation cover in order to initiate vegetation protection and restoration programs" (Xie *et al*., 2008). However, challenges are encountered in the definition of vegetation classes based solely on their spectral responses. Zakaria, (2010) stated that these observed challenges are due to the common heterogeneity of the cover type and the factors which affect their spectral responses, this is especially the case with the use of medium to coarse spatial resolution imagery, such as Landsat imagery, to map the vegetation types.

Xie *et al*., (2008) observed that owing to its longest history and widest use for monitoring the earth from space, i.e. since 1972, Landsat imagery has proved to be very significant in the mapping of vegetation as it enables the study of the spatiotemporal changes of vegetation of an area. However, with a spatial resolution of 30m for the multispectral bands and 60m for the thermal infrared band, which makes it a medium to coarse spatial resolution imagery, Landsat images are usually used to map vegetation at regional/community level. Therefore, due to this spatial resolution, it is a challenging task to use Landsat images for mapping at vegetation species level, especially in a heterogeneous environment. However, Xie *et al*., (2008) recommended that when integrated with other ancillary data, it becomes possible to map species using Landsat imagery.

Many research studies have found that "using ancillary geographical data and multi-spectral satellite data can improve classification results" (Ozesmi & Bauer, 2002). These studies include those done by:

- Xiaodong *et al*., (2009); integrated the Landsat TM data with ancillary data to map Land cover of a marsh area;

- Hansen *et al*., (2006); integrated AVHRR imagery with geographical ancillary data to produce A global land cover map using a DT that has a set of 41 metrics generated from five spectral channels and NDVI for input;
- Abdelhamid *et al*., (2010); using a similar method as Hansen *et al.*, (2006) mapped salt-affected Soils over large areas.

Among the ancillary data that were used to assist in the identification of mapped features in most of these studies were: "soils data, topographic or elevation data, bedrock geology and landforms and climate data". Na *et al*., (2009:178)

This method of integrating remote sensing with other ancillary data takes advantage of the relationship that exist between the features to be mapped, in this case vegetation types, with their environmental factors, such as soil type and elevation. As several studies have explicitly shown and Zakaria (2010) stated that the nature and properties of vegetation are fundamental attributes of the landscapes they occur, and further that the nature of the vegetation in an area is determined by a complex combination of effects related to climate, soils, history, fire and human influences. Therefore, this complex combination of effects can be integrated with spectral data to map the vegetation species/types. This integration is usually intended to support digital classification through the use of auxiliary data as the auxiliary data is used to help improve the classification of the satellite data. Remote sensing technology has emerged as a potentially powerful tool for providing information on natural resources at various spatial and temporal resolutions (Zakaria, 2010), and the integration with the ancillary data enhances its strength further. There are many algorithms for integrating the datasets, in this study DTC, which is also known as classification tree, was used.

2.1.1. Satellite Image Classification Algorithms

Satellite image classification can simply be defined as the process of grouping image pixels to represent features such as built up areas, agricultural crop fields, waterbodies and many other feature types on the Earth's surface. There are several satellite image classification techniques or algorithms of which the main ones are supervised classification, unsupervised classification and non-parametric classification.

2.1.1.1. Supervised Classification

In this parametric image classification technique, a user/ analyst delineates sample pixels of land cover features to be mapped. This delineation is done via the use of Region of Interests (ROIs); which are created by tracing out polygons on representative pixel groups for the mapped land covers. Therefore, these ROIs form "training sites" for the classification algorithm to be applied in the classification of the entire image pixels; the classification of land cover is thus based on the spectral signature defined in the training set. Under this technique, among the most frequently used classification algorithms are the parallelepiped, minimum distance, and maximum likelihood.

As a parametric classification technique, the algorithms here assume that the observed measurement vectors Xc for each class in each spectral band during the training phase of the supervised classification are Gaussian in nature; that is, they are normally distributed (Kumar, 2013).

2.1.1.2. Unsupervised Classification

This is a type of parametric image classification technique where image spectral classes are formed based on the natural statistical grouping of image pixels. Here the analyst specifies the number of classes required for the features to be mapped and based on the specified number, the algorithm groups the pixels according to their spectral values hence the formed classes are known as spectral classes. After all the spectral classes are formed, the analyst compare the classified data with some form of reference data (such as larger scale imagery or maps or field data) to determine the identity and informational value of the spectral classes.

In this classification category there are several clustering algorithms that can be used to determine the pixel groups and the main ones here are the ISODATA (Interaction Self-Organizing Data Analysis Technique) and k-means.

2.1.1.3. Non-parametric classification

These classifiers do not make any statistical assumption about the data thus do not base their class separation calculations on statistical parameters and are especially suitable for incorporation of non-remote sensing data into a classification procedure (LU *et al.*, 2007), these classifiers enable the integration of remotely data with other ancillary datasets. The main algorithms in this category are neural networks, decision tree classifiers and support vector machine.

Of the mentioned non-parametric algorithms, many researchers such as Pooja *et al*., (2011) found decision tree classifiers to be more interpretable than other classifiers such as neural networks and support vector machines because of their ability to combine simple questions about the data in an easily understandable way. Furthermore, decision tree approach has been found to have substantial advantages for land use classification problems because of their flexibility and ability to handle non-linear relations between features and classes, hence improve the classification accuracy to a great extent. These advantages of decision tree classifier qualify it to be used in mapping vegetation types by integrating moderate spatial resolution satellite imagery with other ancillary datasets.

2.2. Decision Tree Classification (DTC)

A decision tree classifier is a type of multistage classifier that can be applied to a single image or a stack of images; it performs multistage classifications by using a series of binary decisions to place pixels into classes. The tree is composed of a starting node (root), a set of internal nodes (splits), and a set of terminal nodes (leaves) (Liu & Shi, 2008). Instances are classified starting at the root node and sorted based on their feature values. Various types of decision tree algorithms have been developed by many researchers over a period of time with enhancement in performance and ability to handle various types of data. The main algorithms in this category include the ID3 (Interactive Dichotomiser), C4.5, C5.0 and CART (Classification and regression tree). Decision tree is a structure that includes a root node, branches, and leaf nodes. For all the decision tree algorithms, each branch denotes the outcome of a test and each leaf node holds a class label. The ENVI DTC, the decision tree classifier used in this study, implements the CART algorithms in its classification. Table 1 shows a comparison of the various decision tree classifiers (Friedl & Brodley, 1997; Kumar, *et al*., 2015; ITT Visual Information Solution, 2009).

Table 1: Comparisons between different Decision Tree Algorithms (Kumar, et al., 2015)

Algorithms	ID3	C4.5	C5.0	CART
Type of data	Categorical	Continous and Categorical	Continous and Categorical, dates, time, timestamps	Continous and nominal atrributes data
Speed	Low	Faster than ID3	Highest	Average
Pruning	No	Pre-pruning	Pre-pruning	Post pruning
Boosting	Not supported	Not supported	Supported	Supported
Missing Values	Cannot deal with	Cannot deal with	Can deal with	Can deal with
Formula	Use information enropy and information Gain	Use split info and gain ratio	Same as C4.5	Use Gini diversity index

From Table 1, it can be noted that the ENVI decision tree algorithm, CART, performs relatively better than the other decision tree algorithms especially in terms of the data type handled and the execution time. Those features coupled with the readily availability of the ENVI decision tree classifier made this classifier the best option for the vegetation type mapping in this study.

However, generally, in decision tree classification, data from many different sources and files such as satellite imagery, digital elevation models and soil maps, as well as datasets of varying accuracies and spatial resolutions can be combined together to produce a single decision tree classification map showing all the resulting classes.

Such capabilities make decision tree classifiers potentially appropriate tools for mapping and modelling vegetation types as they take advantage of the correlation that exist between vegetation types and their environmental factors hence their mapping by integrating these environmental factors' datasets and files to produce vegetation maps.

In mapping vegetation types, He *et al*., (2005) reported that under some circumstances, decision tree can be very useful when vegetation types are strictly associated with other natural conditions e.g. soil type or topography. For example, some vegetation species may only grow in areas with elevation higher than a certain level. This can be integrated within decision tree to assist the classification process from imagery if such ancillary data are available (Xie *et al*., 2008). Many studies recommended the use of this technique as it has substantial advantages for remote sensing classification problems because of its flexibility, intuitive simplicity, and computational efficiency (Mustafa *et al*., 2009).

Many researchers have successfully used classification trees to integrate remote sensing imagery with ancillary geographical information for land use/cover classification (Na *et al*., 2009). Breiman *et al*., (1984) described the advantages of classification trees as possessing the versatility to integrate both numerical and categorical variables into classifications and that they make no distributional assumptions when classifying the pixels. Among the advantages of classification trees that Pal, (2005) observed were that classification trees require less training time compared to other machine learning techniques such as artificial neural networks and support vector machines, while attaining similar accuracy hence its use in the this study.

Traditionally, classification tasks are based on statistical methodologies such as Minimum Distance-to-Mean (MDM), Maximum Likelihood Classification (MLC) and Linear Discrimination Analysis (LDA). These classifiers are generally characterized by having an explicit underlying probability model, which provides a probability of being in each class rather than simply a classification. The performance of this type of classifier depends on how well the data match the predefined model. If the data are complex in structure, then to model the data in an appropriate way can become a real problem (Ghose *et al*., 2010). Thus, with the use of a decision tree classifier, such classification challenges may be overcome.

Despite the numerous advantages that it brings, such as being a cost-effective mapping methodology, integration of remote sensing (RS) with other geographical ancillary data, challenges are encountered, mainly related to ancillary data accuracy and availability. The unavailability of accurate or of any necessary ancillary data, such as soil or elevation data may result in less accurate maps. However, these problems can easily be overcome with more research work in those necessary fields.

Unlike in other regions and countries, some of which are in Africa, the use of decision tree classification in vegetation type mapping has never been explored in Zambia and thus the importance of this study cannot be overemphasized as it would serve as the basis for future studies and mapping of vegetation types using decision tree classification method.

2.3. Vegetation of Zambia

In Zambia, vegetation is classified into four major categories: closed forests, woodlands or open forests, termitaria and grasslands (Fanshawe, 1971; Storrs, 1995). These are further divided into sub-vegetation types such as Cryptosepalum, Parinari, Miombo and Mopane (Sekeli & Phiri, 2002).

Closed forests cover about 6% of the country. Cryptosepalum evergreen forests are the most extensive within this category and occur in the western part of the country while the Baikiaea forests in the south west are the second most extensive characterized by the commercially valuable *Baikiaea plurijuga* (State of the Environment in Zambia, 2000: pp 58-59).

Open forests or savannah woodlands are the dominant vegetation type covering 66 % of Zambian land. There are four types of these woodlands the most extensive being Miombo woodlands that covers about 42% of the country characterized by the *Brachystegia*, *Julbernardia* and *Isoberlinia*; followed by the Kalahari woodlands, Mopane, Munga and Termitaria (State of the Environment in Zambia, 2000). The predominant Miombo woodland (a sub-category of the savannah) is two-storied with an open and evergreen canopy 10 - 20 m high. Termitaria or anthill vegetation covers about 3 % and is present in all regions except in areas of pure sand; it is classified according to its association with other vegetation types; hence: Miombo termitaria, Kalahari termitaria, Mopane termitaria, Munga termitaria, Riparian termitaria and Grassland termitaria (State of the Environment in Zambia, 2000).

Grasslands cover 27% of the land and range from pure grassland to grassland with scattered trees. They occur in poorly drained dambos, flood plains or swamps. The dominant grasses are *Themeda triandra*, *Hyparrhenia* spp. and *Heteropogon contortus*. Having declined at 2.4% per annum, forest, mostly savannah bushveld, covers 42% of the land area. The high eastern plateau consists of open grassy plains with small trees and some marshland. Arable land comprises 7% of the total land area (Aregheore, 2009). These vegetation types are as shown in table 1:

Most of the tree species of commercial value in Zambia are used for timber production, as construction material and for making furniture products. Others are used for charcoal production, poles, soil improvement, fodder, medicines, turnery and many other uses (Aregheore, 2009). This further emphasizes the need for mapping these vegetation types. Forty-one indigenous tree species in Zambia produce edible fruit and seeds, 44 are good for animal feed, 38 for tannin production, 39 for dyes, 11 for resins and gums, and 30 for timber.

Therefore, from the few highlighted importance of forests and the forest related activities, it can be seen that forest indeed plays a crucial role in enhancing human well-being and in sustaining the economy of Zambia. They contribute to economic growth, employment, wealth, export revenues, a stable supply of clean water, recreation and tourism opportunities, as well as essential building materials and energy for a wide range of economic sectors (Turpie *et al*., 2015)

Furthermore, the summary result of an analysis carried out by Turpie *et al*., (2015) showed that the direct and indirect values of forests are estimated to make a direct contribution equivalent to about 4.7% of gross domestic product (GDP) or US$97.5 million (using 2010 figures).

With such a wide variety of vegetation and the importance that it has to the country, the need for improved methods of managing and monitoring of these resources cannot be over-emphasized. However, in order to achieve this improved management and monitoring of resources, proper and advanced methods of mapping these resources need to be adopted and applied. In this case, a method that integrates readily available medium to coarse resolution remote sensing data; which provides important coverage, mapping and classification of land cover features such as vegetation, soil and water using spectral responses only (Zakaria, 2010), with other geographical ancillary data has potential for such improved mapping and is thus investigated. From the several research studies reviewed in the literature, it can be seen that indeed this method has the potential of classifying vegetation at “type” level by using medium to coarse remote sensing data. With such details, vegetation mapping can assist improved management of these resources such as in determining the spatial distribution and estimating volumes of a vegetation type. For instance, the method may be applied in managing and monitoring vegetation species such as *Mukula* tree (Pterocarpus chrysothrix). With such detailed mapping, the spatial distribution and volume estimates of the *Mukula* tree may be determined for proper monitoring and protection from its illegal felling.

Table 2a: Vegetation of Zambia (Table complied from studies by Forestry Department, (2000);Fanshawe, (1971); Storrs, (1995); Sekeli & Phiri, (2002) and Aregheore, (2009))

Main vegetation types	**Sub-types**	Genus	**Species**(*dominant*)
Closed forests	Dry evergreen forests	- Parinari	- *Parinari excels* - *Syzygium guineense*
		- Marquesia	- *Marquesia acuminata,* - *Marquesia macroura* - *Syzygium guineense*
		- Cryptosepalum	- *Cryptosepalum pseudotaxus* - *Guibourtia coleosperma*
		- Lake basin(chipya)	
	Dry deciduous forests	- Baikiaea	- *Baikiaea plurijuga* - *Pterocarpus antunesii*
		- Itigi	
	Montane		- *Parinari excelsa* - *Podocarpus milanjanus* - *Trichilia preuriana*
	Swamp		- *Chyrosophyllummagalismontanum* - *Ilex mitis* - *Mitragynastipulosa* - *Syzygium cordatum*
	Riparian		- *Diospyrosmespiliformis* - *Khayaathothetica* - *Parinari excelsa* -*Syzygiumcordatum associated with Fauriasaligna and Raphia palms*

Table 2b; Vegetation of Zambia (Table complied from studies by Forestry Department, (2000);Fanshawe, (1971); Storrs, (1995); Sekeli & Phiri, (2002) and Aregheore, (2009))

Woodlands (or open forests)	Miombo	*It is characterized by species of the genera* - *Brachystegia* - *Isoberlinia* - *Julbernadia* - *Marquesia macroura* - *Pericopisisangolensis* -*Erythophleumafricanum* - *Parianricuratelifolia are frequent associates*
	Kalahari	*most common species are of the genera* - *Guibourtia* - *Burkea* -*Diplorhynchus* - *Parinari* - *Baikiaea plurijuga* - *Pterocarpus angolensis*
	Mopane	- *Colophospermum Mopane*
	Munga (or savanna woodland)	- *Acacia* - *Combretum* - *Terminalia spp*
Termitaria	Miombo Kalahari Mopane Munga	- *Diospyrosmespiliformis* - *Asparagus racemosus* - *Bosciaangustifolia* - *Capparistomentosa* -*Sterculiaquinqueloba* - *Maeruajuncea*
Grasslands		-*Themeda triandra* -*Hyparrhenia spp.* -*Heteropogon contortus*

In this research, only three vegetation types, namely Miombo, Mopane and Munga, were identified, due to the availability of literature on them. Initially, all the vegetation types within the area of study were to be mapped however detailed descriptive literature on most of these vegetation types is currently unavailable hence the three vegetation types of Miombo, Mopane and Munga. These vegetation types were studied and attempts were made to map them using DTC; where their spectral characteristics were integrated with their respective environmental factors.

3. CHAPTER 3--RESEARCH METHODOLOGY

3.1. Study Area Description

3.1.1. Geographical location

The selected study area is a band running from Copperbelt province in the North to the Southern province in the South of Zambia between latitude 12° and 18° S and longitude 26° and 30° E, that is, the area falls in the strip of Zambia covered by Landsat World Reference System (WRS)2 path 172. This area runs across through all the three agro ecological zones of Zambia (zones I, II & III). It covers an area of approximately 9,880,000ha and encompasses 5 provinces of Zambia. This study area was selected because of its location and stretch across the 3 agro ecological zones which results in its diverse general environmental conditions and vegetation cover, Figure. 1 shows the study area.

3.1.2. Climate conditions

The study site is located across the three agro-ecological zones of Zambia and it covers an estimated area of 9,880,000ha. These rainfall zones are distinguished as follows:

1. Zone I. Rainfall less than 800
2. Zone II. Rainfall 800-1000
3. Zone III. Rainfall above 1000

The temperature of the study area varies from about 25.9° to 30.9° (Zambia Meteorological Department, 2011).

3.1.1. Soil and geology

The dominant soil of this area has diffuse horizon boundaries, a clay assemblage dominated by low activity clays (mainly kaolinite) and a high content of sesquioxides. This is followed by the soil type having an argic horizon starting within 100cm from the soil surface, or within 200 cm from the soil surface if the argic horizon is overlain by loamy sand or coarser textures throughout (Driessen *et al*., 2001) as in Figure. 2 showing the study area/ area of interest (AOI) clipped soil map of Zambia. The altitude of the area ranges from about 400 to 1080m above sea level, as such it is a mixture of relatively high and low lands.

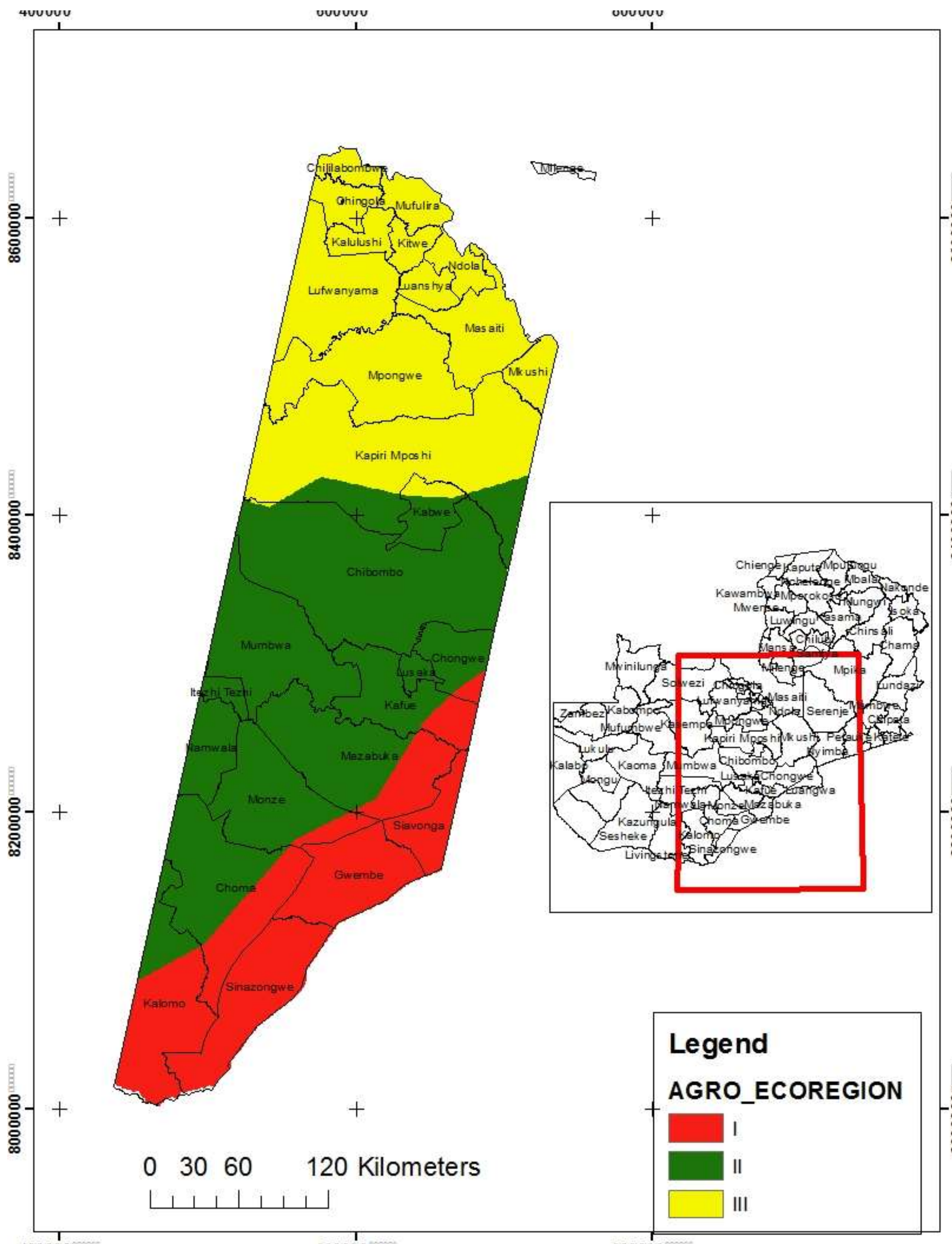

Figure 1: Location of study area(with Agro Ecological zones overlay(Zambia Meterological Department(n.d)))

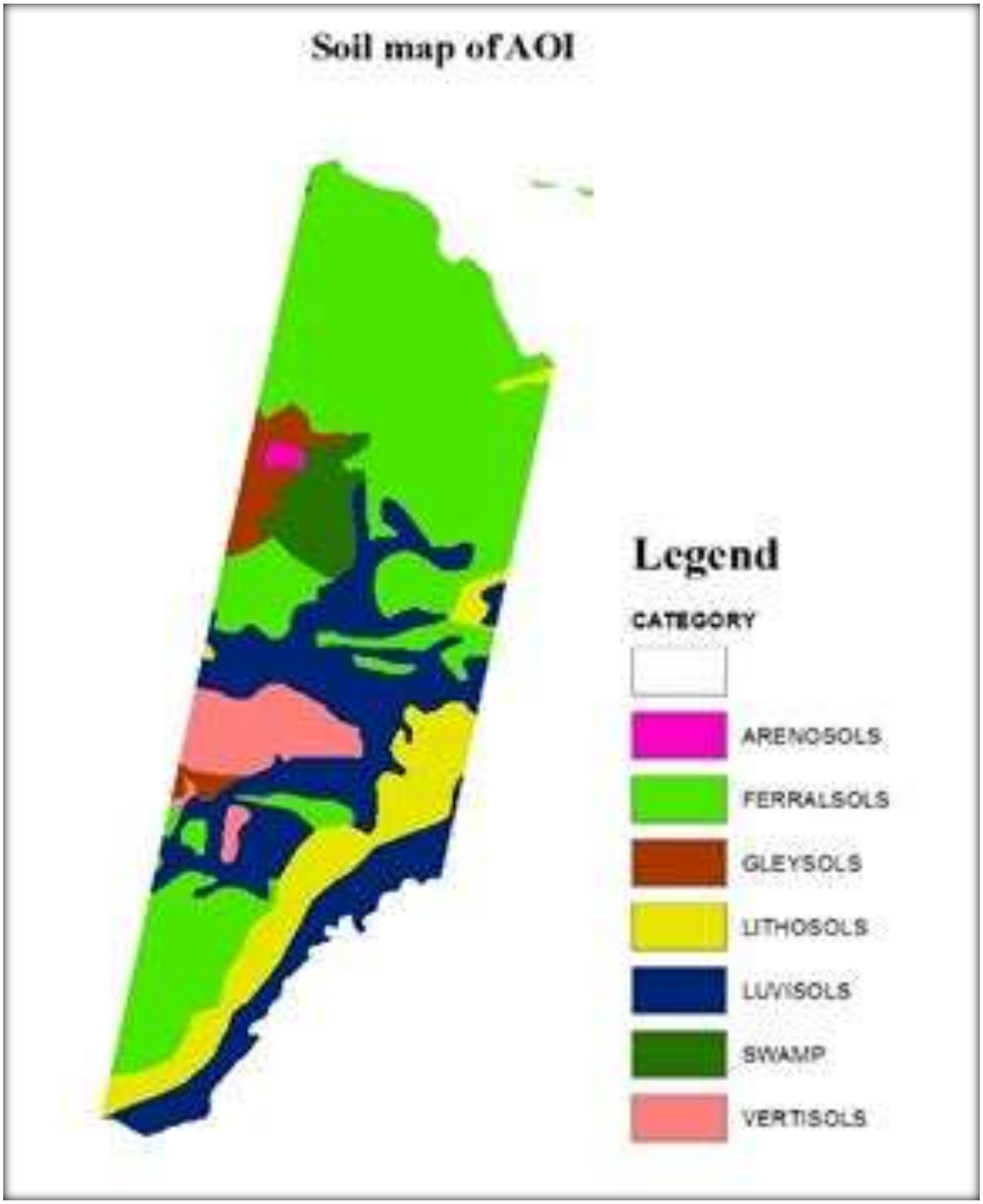

Figure 2: Soil map of study area, National Council for Scientific Research (1983)

The geology of the study area comprises of a granitic gneisses complex around the north-east of the area, some strips of alluvium colluvium laterite around the central section. The area also has some traces of shale siltstones and sandstones, and also/ meta-carbon-volcanic rocks, as shown in Figure. 3:

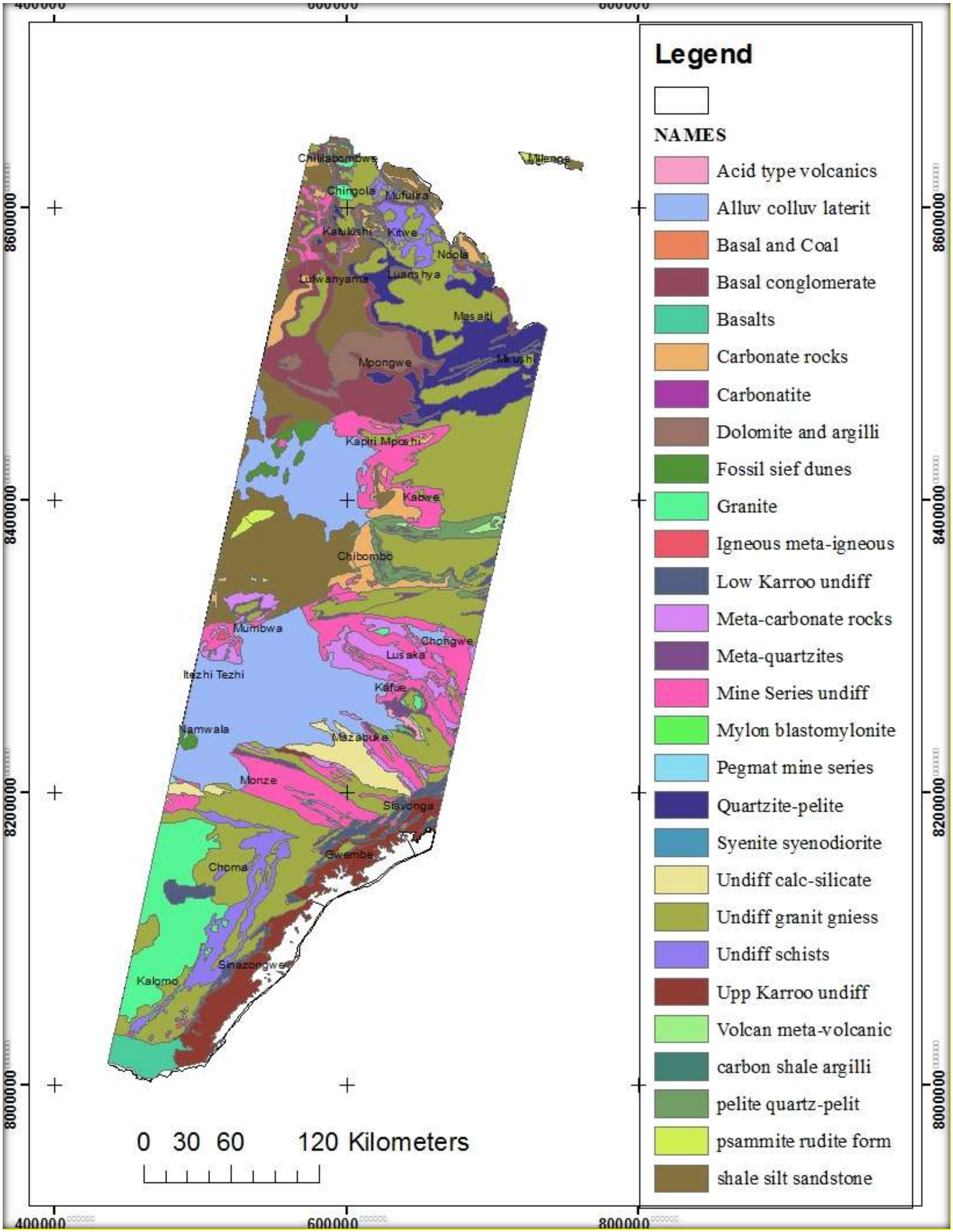

Figure 3: Geological map of study area, Zambia Geological Survey (n.d)

3.1.2. Description of Vegetation cover in study area

With the study area falling across all the three agro-ecological zones of Zambia, the area is covered by a wide variety of vegetation types/species. The vegetation types range from the most dominant, Miombo vegetation type to some few patches of other vegetation types such as Lake Basin chipya in the northern part and kalahari woodlands in the west and also a strip of Mopane in the southern margin.

3.2. Materials and method

The study primarily started with a thorough review of the available literature on similar studies carried out in other countries and regions, this was to assess the applicability of the identified methodology of mapping vegetation, that is, studying DTC from other research studies, and also reviewed literature on the vegetation type/species of Zambia as well as on the factors that determine their occurrence and spatial distribution. This literature reviewed carried out served as the basis for the practical analysis and eventual mapping of the vegetation using DTC.

The study integrated data from different sources and formats, the data comprised Landsat imageries for the years 1984, 1976 Vegetation map of Zambia, soil map of Zambia, Digital Elevation Model (DEM) and the agro-ecological zone map of Zambia.

3.2.1. Satellite data

Landsat satellite data, path: 172, rows: 069, 070, 071, 072 and 073, of the same date namely, 20th June 1984, were collected and used in this study. Initially for this study, Landsat images for year 1976 were to be used in the classification however, due to the unavailability of these images in the source archive, the 1984 Landsat satellite images were the closest alternative to the planned 1976 images hence adopted. These images were to be incorporated with the 1976 vegetation map of Zambia in the analysis of the spectral statistics of the vegetation types in the study area. The satellite data were procured from the National Remote Sensing Centre (NRSC) - Zambia satellite imagery archive.

3.2.2. Ancillary data:

The ancillary data used included the soil map of Zambia compiled by the National Council for Scientific Research; a 90m DEM; and agro-ecological zone data from NRSC archive.

The mapping approach adopted for this study was as illustrated in flowchart Figure 4:

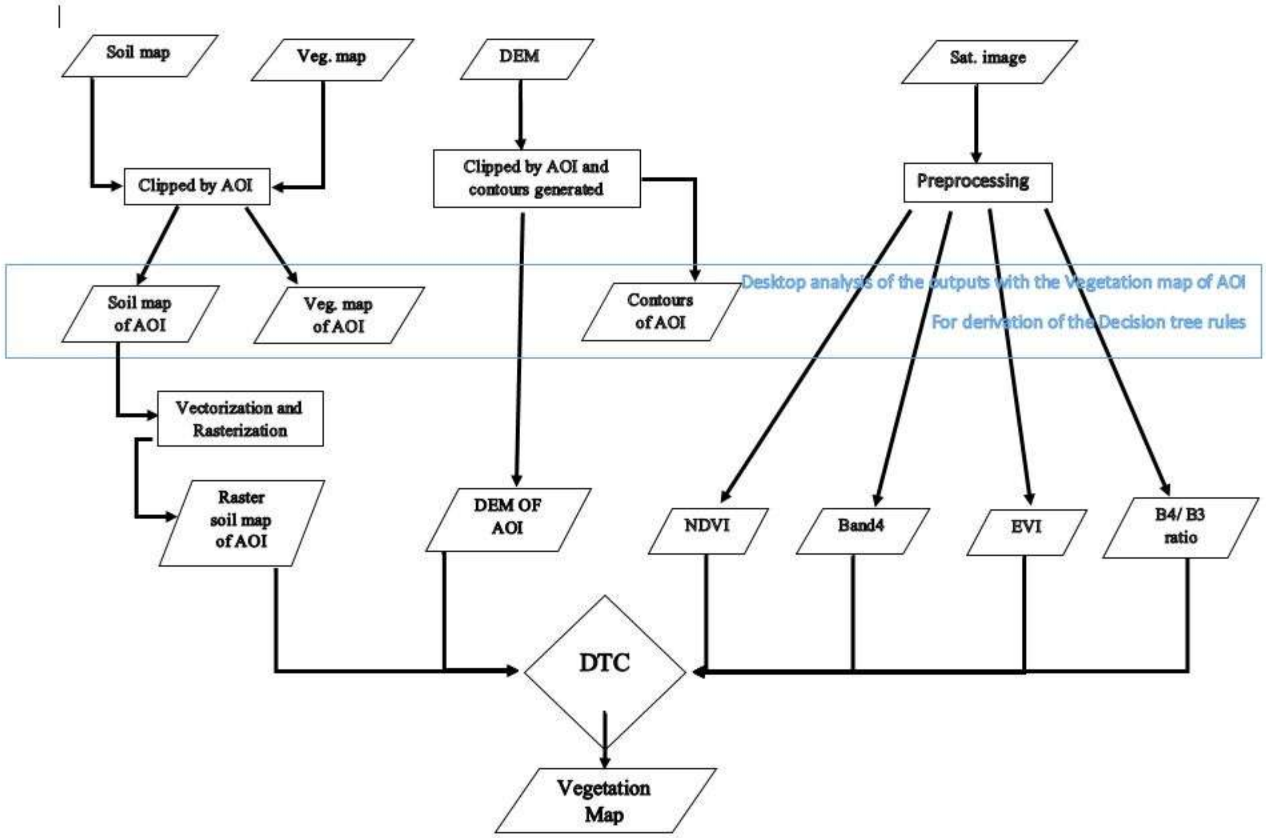

Figure 4: Diagram of methods and materials used in the research

3.2.3. Image pre-processing

Often the raw data image is not directly suitable for specific purposes and should thus be processed in some way or another to make it suitable. For that, essential steps known as image pre-processing must be done before the actual image classification. Thus in this study, the type of pre-processing carried out on the raw satellite images was atmospheric correction. Atmospheric correction is a method used to correct sensor radiance for atmospheric effects by mathematically modelling the physical behavior of radiation as is passes through the atmosphere (ITT Visual Information Solution, 2009).

Application of atmospheric correction is essential for the current study for one reason. The imagery was acquired at different times of day hence at different atmospheric conditions. Removing atmospheric effects involves calibration and atmospheric correction. Calibration adjusts the image by converting raw radiance values of each pixel to top-of-atmosphere absolute (radiance) or relative (reflectance) values. Atmospheric correction then adjusts these values to ground radiance or reflectance at each pixel based on sun-ground-sensor geometry and atmospheric composition (Zakaria, 2010).

3.2.3.1. Converting digital number to radiance:

Digital number (DN) is referring to the quantized and calibrated values for individual pixels (Stellingwerf & Hussin, 1997). For converting the DN values to radiance, in this study, ENVI Landsat Calibration module was used. This module has equations 1 and 2 inbuilt in it and prompts for the input of the satellite metadata such as the Satellite sensor, maximum and minimum radiance values, and the date of acquisition of the imagery. As the DN values of the TM and ETM+ data are geometrically correct, equation 1 was used in this study. Thus with all the necessary metadata inputted, the DN values were converted to spectral radiance Lλ (m W cm-2sr-1 μm-1) (Zakaria, 2010).

$$L_{\lambda} = \mathrm{Lmin}_{\lambda} + \left(\frac{\mathrm{Lmax}_{\lambda} - \mathrm{Lmin}_{\lambda}}{\mathrm{DNmax}}\right) * \mathrm{DN} \qquad (1)$$

$$L_{\lambda}=\mathrm{GAIN}*\mathrm{DN\_}\lambda+\mathrm{BIAS} \qquad (2)$$

Where:

L_{λ}	Spectral radiance (i: band)
Lmax_{λ}	Maximum spectral radiance (mW cm-2sr-1 μm-1)
Lmin_{λ}	Minimum spectral radiance (mW cm-2sr-1 μm-1)
DN	Absolute calibrated digital number

The obtained radiance is known as the Spectral radiance at the sensor’s aperture (at-satellite radiance). In order to finally obtain the surface reflectance of the pixels, the at-satellite radiance was further corrected for solar and atmospheric effects as in section 3.2.3.2:

3.2.3.2. Converting at-satellite radiance to surface reflectance

The at-satellite radiance was converted to surface reflectance using FLAASH (Fast Line-of-sight-Atmospheric Analysis of Spectral Hypercubes) module in ENVI 4.7 (Zakaria, 2010). FLAASH is a first-principles atmospheric correction modeling tool for retrieving spectral reflectance from hyperspectral and multispectral radiance images. It provides the most accurate means of compensating atmospheric effects considering the elevation, water vapor, and aerosol distribution properties (Adler-Golden *et al*., 1999). FLAASH accurately compensate for atmospheric effects and corrects wavelengths in the visible through near-infrared and short-wave infrared regions, up to 2.5 μm. Unlike many other atmospheric correction programs that interpolate radiation transfer properties from a pre-calculated database of modeling results, FLAASH incorporates the MODTRAN4 radiation transfer code (ITT Visual Information Solution, 2009).

Input in the FLAASH module includes the average elevation of the study area, scene centre coordinates, sensor type, flight date and time, information about aerosol distribution, visibility, and water vapor conditions (Figure. 5). The input images for FLAASH were radiometrically calibrated to radiance images in band-interleaved-by-line (BIL) format. Results showed that pixel spectral resolution is improved with FLAASH and creates an image of retrieved surface reflectance, with 16 bit instead of 8 bits (Schmidt *et al*., 2009). After applying the FLAASH on the image stack, some reflectance values are sometimes negative or greater than 1 in the corrected results. This was fixed by using band math: (b1 le 0) x 0 + (b1 ge 10000) x 1 + (b1 gt 0 and b1 lt 10000) x float(b1)/10000, on the resulting bands; where "b1" is a variable holding the preprocessed image bands inserted in the band math equation; "le" means "less or equal to"; "ge" means "greater or equal to"; "gt" means "greater than"; and "lt" means "less than". The images produced after the band math, that is, the preprocessed Landsat satellite images were then mosaicked using ENVI 4.7 for the study area.

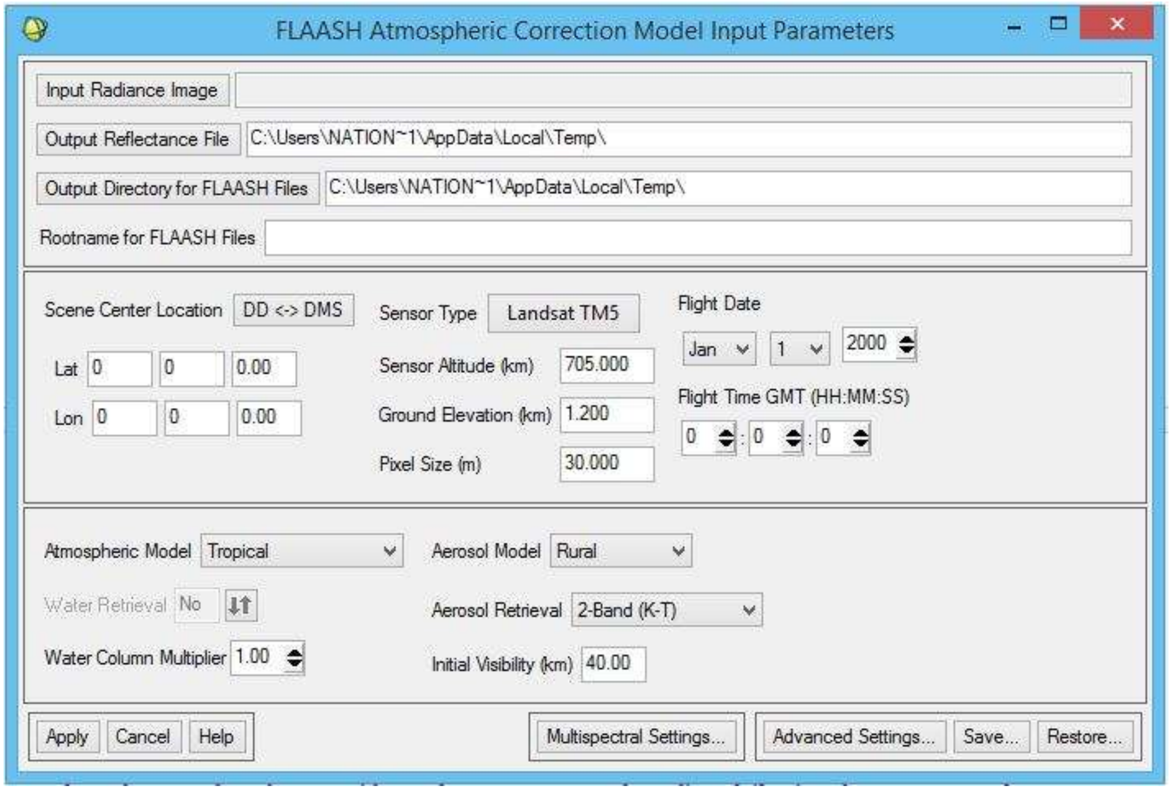

Figure 5: FLAASH input box

3.2.4. Preparation of spectral and ancillary data

The 1974 Vegetation map of Zambia was viewed in ArcGIS 10.1 where the three shapefiles, i.e. the Miombo, Mopane and Munga vegetation types, were exported into separate shapefiles. These exported shapefiles were later used to subset the preprocessed 1984 satellite imagery. This subsetting was done in order to extract and save the spectral statistics of the three vegetation types, the subsets of the input satellite imagery were viewed and the spectral statistics extracted in ENVI 4.7. Here the spectral statistics of interest were the NDVI, the band 4, band4/band3 ratio and the EVI statistics of the three respective vegetation types; all the spectral statistics of interest such as the NDVI and EVI were produced from the preprocessed Landsat imagery. Using ArcGIS 10.1, the digital soil map and the DEM were clipped using the study area boundary shapefile as the clip feature, and the subsetted soil map was vectorized for in order to add the attributes and other further preparation of the soil map. The vectorized soil map was later rasterized, in ArcGIS 10.1, so as to make it compatible for analysis in decision tree classifier. Contours were generated from the subsetted DEM using GlobalMapper 11.02; the generated contours were to be used in assessing the elevations of the study area. Thus, the generated contours were intersected with the three vegetation types respectively in order to analyze the distribution of the three vegetation types with respect to the elevations of the study area. Furthermore, the distribution of the three vegetation types was analyzed with respect to the soil types in the study area as well as with the rainfall data (agro-ecological zones data). All the rasterized datasets used in the tree were converted to 30m pixel size, to match the Landsat data, prior to the input into the tree for execution.

3.2.5. Image classification using Decision Tree Classification (DTC)

An important part of image analysis is the identification of groups of pixels that have specific spectral characteristics and to determine the various features or land cover classes represented by these groups (Lillesand *et al*., 2008) and this analysis is performed using the traditional digital image classification.

Digital image classification is the process of sorting all the pixels in an image into a finite number of individual classes based on the spectral information and characteristics of these pixels (Zakaria, 2010). However, in DTC, the sorting of the pixels does not rely on spectral information only but also on several other ancillary information such as elevation and soil information. From the literature review and data preparation analyses in the preceding sections, the spectral and ancillary information of the three respective vegetation types were gathered and prepared for use in the DTC classification. This information was eventually compiled for inputting, via binary expressions, into the ENVI 4.7 decision tree classifier for the mapping of the vegetation types namely, Miombo, Mopane and Munga in the study area.

For the accuracy assessment of the produced DTC classified map, the method adopted was the Non-site-specific-accuracy assessment. This method, despite its several shortcomings, is appropriate given the available research resources. Non-site-specific-accuracy assessment can provide some measure of agreement between a reference map and classification in terms of the areal extent of each mapped class; however it does not provide any information about the locational accuracy of the classification. Regardless of the foregoing, the method is very useful in situations where field verification is cannot be carried out due some constraints or other possible reasons. In this study, the satellite imagery used were for 1984 hence field verification was not possible.

4. CHAPTER 4--FINDINGS

4.1. Vegetation types and their Environmental factors

From the literature reviewed it can be concluded that the occurrence and spatial distribution of vegetation types is influenced and determined by a number of environmental factors such as the elevation of an area, rainfall regime and soil type.

Trapnell and Clothier (1996) observed that conditions of altitude and climate naturally play a very important part in determining the distribution of the main vegetation types, such as Miombo and Mopane, but these stated conditions are by no means the only decisive factors. From that, Trapnell and Clothier (1996) thus concluded that a broad correlation exists between the major regional soil types and the corresponding types of vegetation and that however, the correlation is not an absolute one, for the vegetation type predominantly associated with one soil class may in some regions overstep its limits and occupy marginal or outlying forms of the soils of another class. Consequently, a single vegetation type is often found on more than one type of soil. This is liable to happen more especially in a region where the climate is essentially favorable to one type of vegetation, such as the *Brachystegia-Isoberlinia* woodlands in the higher parts of the central and western territory of Zambia.

4.1.1. General Overview descriptions of Miombo, Mopane and Munga vegetation types

4.1.1.1. Overview of Miombo

The Miombo woodland is the most extensive and economically important vegetation type. It cover about 35 million hectares (47 percent of the total land area) (Aregheore, 2009). It occurs in highly weathered soils that are often more than 3 m deep on the plateau. The soils are generally freely drained (Campbell, 1996). The dominant soils in the higher rainfall zones are classified as Ferralsols and Acrisols. In the lower rainfall zones are Ferralic and Chromic Cambisols, Chromic Luvisols and Plinthic Luvisols. Miombo woodland soils are typically acid and have low cation exchange capacities (CEC) (Campbell, 1996). Driessen *et al*., (2001) described this environment as characterized by highly weathered and poor nutrient content soils with high acidity. Miombo also occurs in areas of about 700mm to over 1000mm annual rainfall (Chidumayo *et al*., 2010). This vegetation usually occurs in areas of elevation ranging from about 914m to about 1825m (Woode, 1985). From the description, the dominant soil is the Ferralsols, which actually is the most dominant soil type of Zambia.

4.1.1.2. Overview of Mopane

Whereas Miombo woodland is generally found on lighter-textured, nutrient poor, well-drained soils on the African Plateau, Mopane woodland, on the other hand, is mostly confined to lower-lying areas with clay- and nutrient-rich soils (Chidumayo *et al*., 2010) and on areas at an elevation of 200–1200m, but normally from 300–900m. Rainfall in these areas ranges from 400 to 700mm per year (Chidumayo *et al*., 2010). The Mopane is prevalent in hot river valleys of Luangwa, Kafue and Zambezi, on areas occupied by alkaline alluvial soils. This vegetation type thrives well on sodium affected clay soils (Woode, 1985) with a cation exchange capacity equal to 24 cmol (+)/kg (high nutrient level). This soil has a high clay content (argic horizon) as well as a high base saturation levels (Driessen et al., 2001). From the above description and literature, the dominant soil of such characteristics is the Luvisols.

4.1.1.3. Overview of Munga

Munga woodland is characteristic of the flood plains along the Kafue River and its tributaries. Soils vary from the loose sand of comparatively recent sand banks to impervious clays, some of which contain limestone nodules (Fanshawe, 2010). Munga vegetation is used by agriculturists as an indicator of good arable land. Munga vegetation usually occurs on flat land. In some places strips of Munga vegetation thrive along rivers and streams. The Munga soils in texture vary from clays to sand clays and have a high base exchange capacity (Woode, 1985).

The preceding environmental descriptions of where each of the three respective vegetation types occurs describe the general environments and for the study area, these general descriptions were further refined as summarized in the succeeding Section 4.1.3. This environmental description of the three vegetation types in the study area was to form the basis for the decision tree binary expressions, for the ancillary data, to be integrated with the remote sensing Landsat imagery to be used in the decision tree classification vegetation type mapping.

4.2. Vegetation types in the study area: Their spectral and ancillary information

Vegetation types or species mapping is one of the important primary prerequisites when it comes to the accurate detail monitoring, managing and planning for the natural resources of a country. These resources are of great value to all aspects of the country such as economic and social. From the literature review and the desktop analysis of the datasets, a number of observations were made regarding the vegetation types and their environmental factors within the study area. These observations were compiled into binary rules and these rules were used in the ENVI decision tree classifier to classify the three vegetation types. The following describes these observations in the study area:

Miombo

In the study area, after the analysis with the generated contours, it was found that the Miombo vegetation type occurs in areas of elevation ranging from 900m to 1600m above sea level. These areas, when intersected with the soil map, are dominated by Ferralsols followed by Lithosols soils with some traces of Miombo falling in other soil types such as Luvisols. In terms of rainfall, the Miombo occurs in all the three agro-ecological zones of the study area i.e. from less than 800m to above 1000mm rainfall area.

Regarding the spectral statistics analyzed for the Miombo in the study area, it was found that the maximum NDVI value was revealed to be 1 while the minimum value was -1. The maximum band4 reflectance was 1 while the minimum was 0 and a maximum band4/band3 ratio of 16.235294. The maximum EVI for Miombo was 1 with a minimum of -1. These spectral values were examined and evaluated on pixel-by-pixel level using ENVI 4.7 software using the Miombo-subsetted mosaicked images of the study area as shown in Figures. 6, 7 and 8, respectively:

Basic Stats	Min	Max	Mean	Stdev
Band 1	0.000000	0.543100	0.010086	0.019100
Band 2	0.000000	0.865100	0.015783	0.029589
Band 3	0.000000	0.874900	0.019482	0.038050
Band 4	0.000000	1.000000	0.051836	0.093060
Band 5	0.000000	0.906600	0.055638	0.104175
Band 6	0.000000	1.000000	0.034831	0.068718

Figure 6: Spectral statistical data for Miombo vegetation with band4 highlighted

Basic Stats	Min	Max	Mean	Stdev
Band 1	-1.000000	1.000000	0.114755	0.212491

Figure 7: Statistical data for NDVI values of Miombo vegetation

Basic Stats	Min	Max	Mean	Stdev
Band 1	-1.000000	1.000000	0.059158	0.108775

Figure 8: Statistical data for EVI values of Miombo vegetation

Mopane

In the study area, after the analysis with the generated contours, it was found that the Mopane vegetation type occurs in areas of elevation ranging from 420m to 914m above sea level. These areas, when intersected with the soil map, are dominated by Luvisols soils with some traces of Mopane falling in other soil types such as Lithosols. In terms of rainfall, the Mopane is dominant in agro-ecological zone I of the study area i.e. less than 800m rainfall area.

Regarding the spectral statistics analyzed for the Mopane in the study area, the maximum NDVI value was revealed to be 0.857 while the minimum value was -1. The maximum band4 reflectance was 0.5198 while the minimum was 0 and a maximum band4/band3 ratio of 12.707. The maximum EVI for Mopane was 1 with a minimum of -1. These values of the NDVI and band4 image were examined and evaluated on pixel-by-pixel level using ENVI 4.7 software using the Mopane-subsetted mosaicked images of the study area as shown in Figures. 9, 10 and 11, respectively:

Basic Stats	Min	Max	Mean	Stdev
Band 1	0.000000	0.540300	0.002851	0.012399
Band 2	0.000000	0.339100	0.004291	0.018645
Band 3	0.000000	0.383700	0.005492	0.024228
Band 4	0.000000	0.519800	0.012327	0.053010
Band 5	0.000000	0.906800	0.014173	0.061940
Band 6	0.000000	1.000000	0.009463	0.042299

Figure 9: Spectral statistical data for Mopane vegetation with band 4 highlighted

Basic Stats	Min	Max	Mean	Stdev
Band 1	-1.000000	0.854098	0.020548	0.090293

Figure 10: Statistical data for NDVI values of Mopane vegetation

Basic Stats	Min	Max	Mean	Stdev
Band 1	-1.000000	1.000000	0.011812	0.051979

Figure 11: Statistical data for EVI values of Mopane vegetation

Munga

In the study area, after the analysis with the generated contours, it was found that the Munga vegetation type occurs in areas of elevation ranging from 580m to 1320m above sea level. These areas, when intersected with the soil map, are dominated by Luvisols Pharzoam soils with some traces of Munga falling in other soil types such as Ferralsols. In terms of rainfall, the Munga occurs in all the three (3) agro-ecological zones of the study area.

Regarding the spectral statistics analyzed for the Munga in the study area, the maximum NDVI value was revealed to be 0.8721 while the minimum value was -0.5362. The maximum band 4 reflectance was 0.7249 with a minimum of 0 and a maximum band4/band3 ratio of 14.641. The maximum EVI for Munga was 1 with a minimum of -1. These values of the NDVI and band 4 image were examined and evaluated on pixel-by-pixel level using ENVI 4.7 software using the Munga-subsetted mosaicked images of the study area as shown in Figures. 12, 13 and 14, respectively.

Basic Stats	Min	Max	Mean	Stdev
Band 1	0.000000	0.540900	0.007142	0.018913
Band 2	0.000000	0.531700	0.010913	0.028932
Band 3	0.000000	0.539300	0.014564	0.038973
Band 4	0.000000	0.724900	0.029908	0.078877
Band 5	0.000000	0.879400	0.037523	0.099495
Band 6	0.000000	1.000000	0.025867	0.069594

Figure 12: Spectral statistical data for Munga vegetation with band 4 highlighted

Basic Stats	Min	Max	Mean	Stdev
Band 1	-0.536278	0.872132	0.045369	0.121047

Figure 13: Statistical data for NDVI values of Munga vegetation

Basic Stats	Min	Max	Mean	Stdev
Band 1	-1.000000	1.000000	0.025754	0.069486

Figure 14: Statistical data for EVI values of Munga vegetation

However, these ranges of spectral values presented above are not intended to define an absolute range for each vegetation type class, but rather to illustrate the concept of spectral analysis. The values may vary with the atmospheric conditions at time of image capture.

Using DTC, several additional environmental variables were incorporated with the spectral statistics in developing the vegetation type prediction map. DTC was carried out using the ENVI decision tree classifier algorithm. A constructed decision tree consists of nodes representing variables or attributes, branches representing attribute values, and leaves representing classes. A decision tree is built based on selecting the attribute that minimizes the amount of disorder in the sub-tree rooted at a given node (Abdelhamid *et al*., 2009). Figure. 15 illustrate the structure of the tree and the data inputted based on the spectral and environmental findings for the study area:

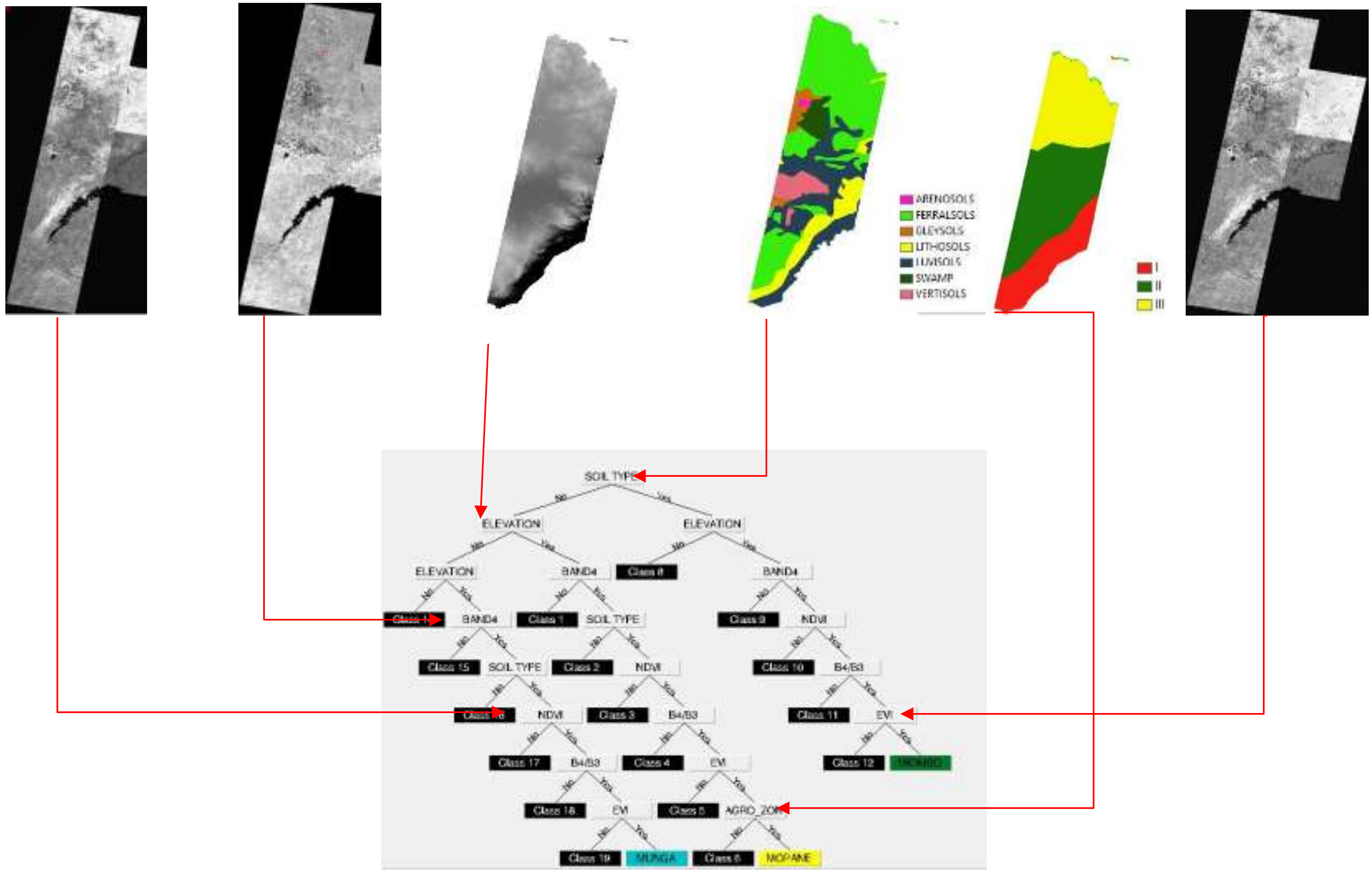

Figure 15: Decision Tree structure with Input Datasets

In this case, the input nodes are the spectral and the environmental variable datasets, that is, the NDVI and Band 4 data, and the DEM, soil type, slope and the agro-ecological zone datasets, as shown in Figure. 15 DTC illustration. Based on the characteristics of the vegetation type, noted from both the literature review and the desktop analysis, the binary expressions were constructed as follows:

Note that for the binary expressions, the nominal data such as the soil type and slope were denoted by their numerical value from the attribute table as follows:

OID	Value	Count	AGRO_REGIO
0	0	4215784	
1	1	3014632	II
2	2	1449890	I
3	3	2261696	III

OID	Value	Count	CATEGORY
0	0	96173482	
1	1	51455898	FERRALSOL
2	2	3647175	GLEYSOLS
3	3	399350	ARENOSOLS
4	4	4385562	SWAMP
5	5	28484878	LUVISOLS
6	6	12053507	LITHOSOLS
7	7	6216066	VERTISOLS

Figure 16: Attribute Tables for the nominal datasets

These findings are consistent with what prevails with the published literature.

Sample Binary Expressions used

1. For Mopane
    a. Soil Type:
       {SM[1]} eq 5
    b. Elevation:
       {d1[1]} eq 700 or {d1[1]} LT 900 OR {d1[1]} eq 900
    c. Band 4:
       {band4[1]} lt 0.5198 or {band4[1]} eq 0.5198
    d. Agro-ecological zone:
       {AZ[1]} eq 2
    e. EVI:
       {EVI[1]} eq -1 or {EVI[1]} lt 1 or {EVI[1]} eq 1
    f. NDVI:
       {N[1]} eq -1 or {N[1]} lt 0.857818 or {N[1]} eq 0.857818
    g. Band4/Band3 ratio:

{RT[1]} lt 12.707866 or {RT[1]} eq 12.70786

2. For Miombo
    a. Elevation:

       {d1[1]} eq 900 or {d1[1]} lt 1600 or {d1[1]} eq 1600
    b. Soil Type:

       {SM[1]} eq 1 or {SM[1]} eq 6
    c. Band 4:

       {band4[1]} lt 1 or {band4[1]} eq 1
    d. NDVI:

       {N[1]} eq -1 or {N[1]} lt 1 or {N[1]} eq 1
    e. Band4/Band3 ratio:

       {RT[1]} lt 16.235294 or {RT[1]} eq 16.235294
    f. EVI

       {EVI[1]} eq -1 or {EVI[1]} lt 1 or {EVI[1]} eq 1

3. For Munga
    a. Elevation:

       {d1[1]} eq 580 or {d1[1]} LT 1320 OR {d1[1]} eq 1320

       Soil Type:

       {SM[1]} eq 5
    b. Band 4:

       {band4[1]} lt 0.7249 or {band4[1]} eq 0.7249
    c. NDVI:

       {N[1]} eq -0.536278 or {N[1]} lt 0.872132 or {N[1]} eq 0.872132
    d. Band4/Band3 ratio:

       {RT[1]} lt 14.641149 or {RT[1]} eq 14.641149
    e. EVI:

       {EVI[1]} eq -1 or {EVI[1]} lt 1 or {EVI[1]} eq 1

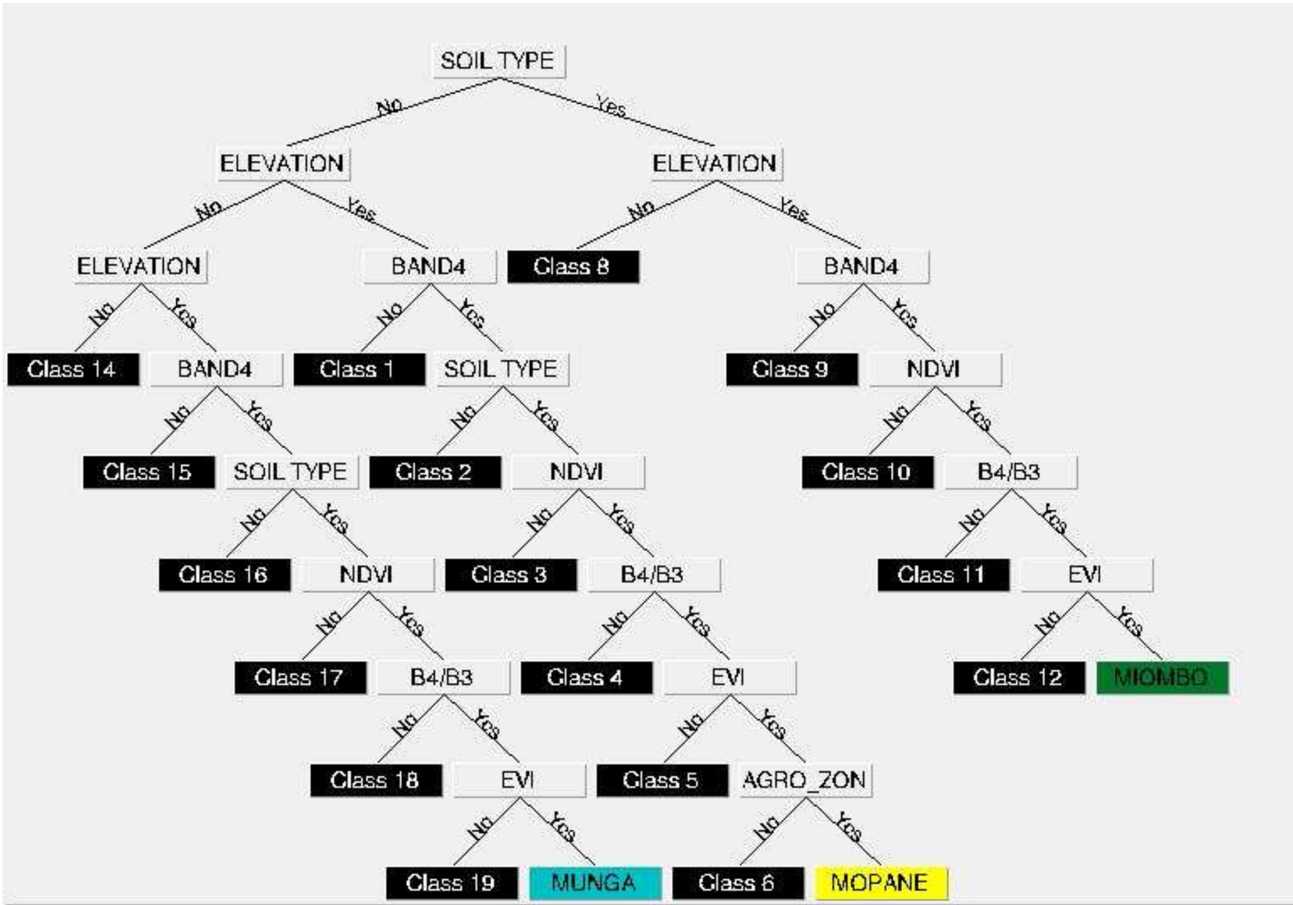

Figure 17: Decision Tree structure

4.2.1. Explanation of Decision Tree (DT) constructed

With decision tree classification of an image the pixels of the image are sorted according to the criteria defined in the binary expressions. Here there were 18 binary expressions in the structure of the decision tree (Figure. 15). The vegetation type map was produced by using these production rules.

Binary expressions were constructed based on the identified and above stated characteristics of the three vegetation types, such as the soil types where the vegetation type occur and the elevation. As the decision tree allows the integration of satellite image with other ancillary data, the binary expressions are constructed in such a way that the pixels are sorted according to the identified vegetation characteristics. For instance, for the Mopane vegetation type, the first node binary expression, i.e. node “ELEVATION”, here the pixels were split between “YES” those which fall in an elevation range of 700-900m and “NO” those out of that elevation range. The pixels falling within that elevation range were further split into “YES” those with a maximum “BAND4” reflectance of 0.5198 and “SOIL TYPE” of Luvisols. The pixels are further split into those of NDVI range of -1 to 0.857818 and of Band4/Band3 ratio maximum value of 12.707866. Furthermore, the pixels were split between “YES” those with “EVI” values range of -1 to 1 and which fall within agro ecological zone I, i.e. at node “AGRO_ZON” of the Decision tree structure above. The inclusion of the agro-ecological zone binary expression restricted the occurrence of Mopane pixels to the lower portion of the area of study.

Similarly, the binary expression used for classifying the Miombo vegetation type were constructed based on the gathered Miombo characteristics; gathered from the literature review and desktop analysis, these characteristics are such as the soil type and elevation. For the Miombo vegetation type, the sorting of the pixels started from the initial topmost “SOIL TYPE” node, as shown in figure. 17. This node had two categories of pixels i.e. those falling in Ferralsols soils (“YES” branch); which initiated the classification of Miombo and those falling in any other types of soil (“NO” branch). For this study, and area of study, it was taken that Miombo occurs in areas dominated by Ferralsols only and thus classification of Miombo took the right-hand topmost branch of the Tree structure, figure. 17.

With the pixels identified i.e. those pixels falling in Ferralsols soils; it should be noted here that Miombo, and generally all vegetation types, are naturally not restricted to specific soil types but for the sake of demonstration of DTC in this study, those restrictions are made. The elevation binary expression of the Miombo vegetation type was constructed, at node “ELEVATION”. This binary expression was based on the observed elevations where the Miombo vegetation type occurs within the area of study. Thus, in this area of study, it was observed, after analysis of the Miombo shapefile with the generated contours shapefile that Miombo occurs at an elevation range of 900m – 1600m. Further the “BAND4” and “NDVI” nodes’ binary expressions were constructed; here the BAND4 reflectance for Miombo was identified as having a maximum of 1 while with an NDVI value range of -1 to 1. Furthermore, the “BAND4/BAND3” ratio for Miombo had a maximum value of 16.235294 and also an “EVI” range of -1 to 1.

Similarly, from the analysis of the spectral statistics and the ancillary data regarding the Munga vegetation, it was observed that, in this area of study, the “ELEVATION” range of Munga was 580-1320m with a maximum “BAND4” reflectance value of 0.7249. In this study, and hence in the constructed “SOIL TYPE” node, Munga vegetation was restricted to Luvisols Pharzoam with an NDVI value range of -0.536278 to 0.872132. Further, the “BAND4/BAND3” ratio for the Munga vegetation had a maximum value of 14.641149 and also an “EVI” range of -1 to 1.

With the above mentioned nodes and their respective binary expressions constructed and inputted, the DTC structure was complete and executed for the final map output. After the execution of the above created decision tree structure, a map of the three vegetation types of the area of study was produced as shown in Figure. 18:

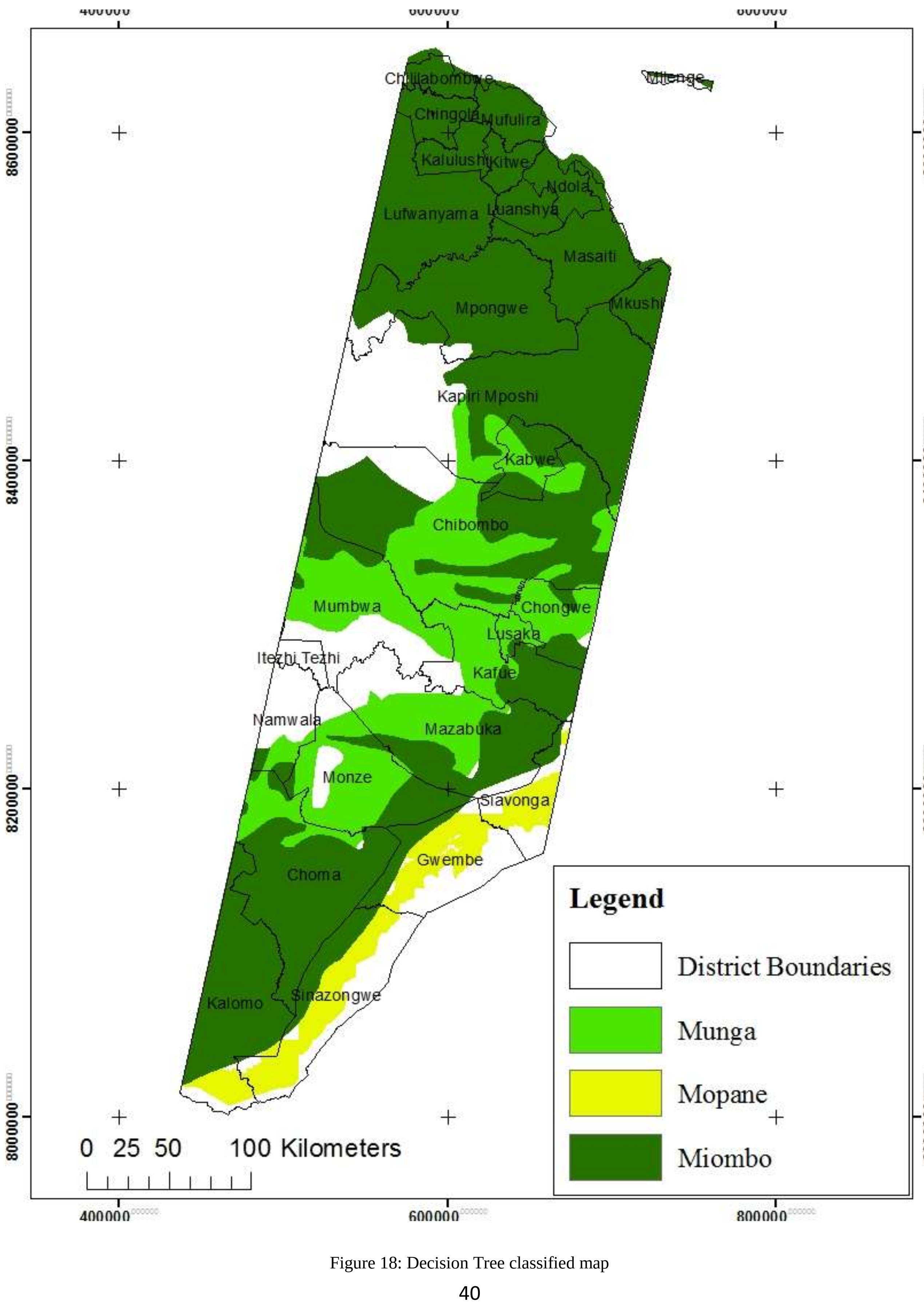

Figure 18: Decision Tree classified map

With the DTC vegetation map produced, and shown in Figure. 18, visual spatial distribution and areal comparisons between the produced DTC vegetation map of the study area and the 1976 vegetation map of the study area was performed. Therefore, for this comparison analysis, masks for the three DT mapped vegetation types were extracted from the tree in ENVI, these masks of Munga, Mopane and Miombo are as shown in Figure. 19. From these masks, the estimates for the areas occupied by the three vegetation types were determined in ArcGIS as well as calculating the areas occupied by the same vegetation types from the 1976 clipped vegetation map of the study area. Thereafter, the 1976 vegetation maps of the three vegetation types were overlaid on the extracted DTC masks of the three vegetation types. The overlay was performed so as to assess the visual spatial discrepancies between the produced DTC vegetation map and the 1976 vegetation map for the three vegetation types.

Results from these analyses, shown in Table 3, indicate that for each of the three vegetation types, the following areal estimates were obtained:

Table 4: Comparisons between the 1976 Vegetation type map with DTC Vegetation map

	1976 veg. Map	DTC veg. Map	Discrepancy	% error
Mopane	505,814 ha	525,014 ha	19,200 ha	4
Miombo	5,062,517 ha	5,715,584 ha	653,067 ha	13
Munga	1,566,647 ha	1,908,983 ha	342,336 ha	22

For the visual accuracy assessment and comparisons, in Figures 19, 20 and 21, the 1976 vegetation maps for each of the three vegetation types, in red polygons, were overlaid on their respective vegetation type masks this was done in order to cartographically show the experienced visual discrepancies. Figure 19, 20 and 21 show the overlay displays.

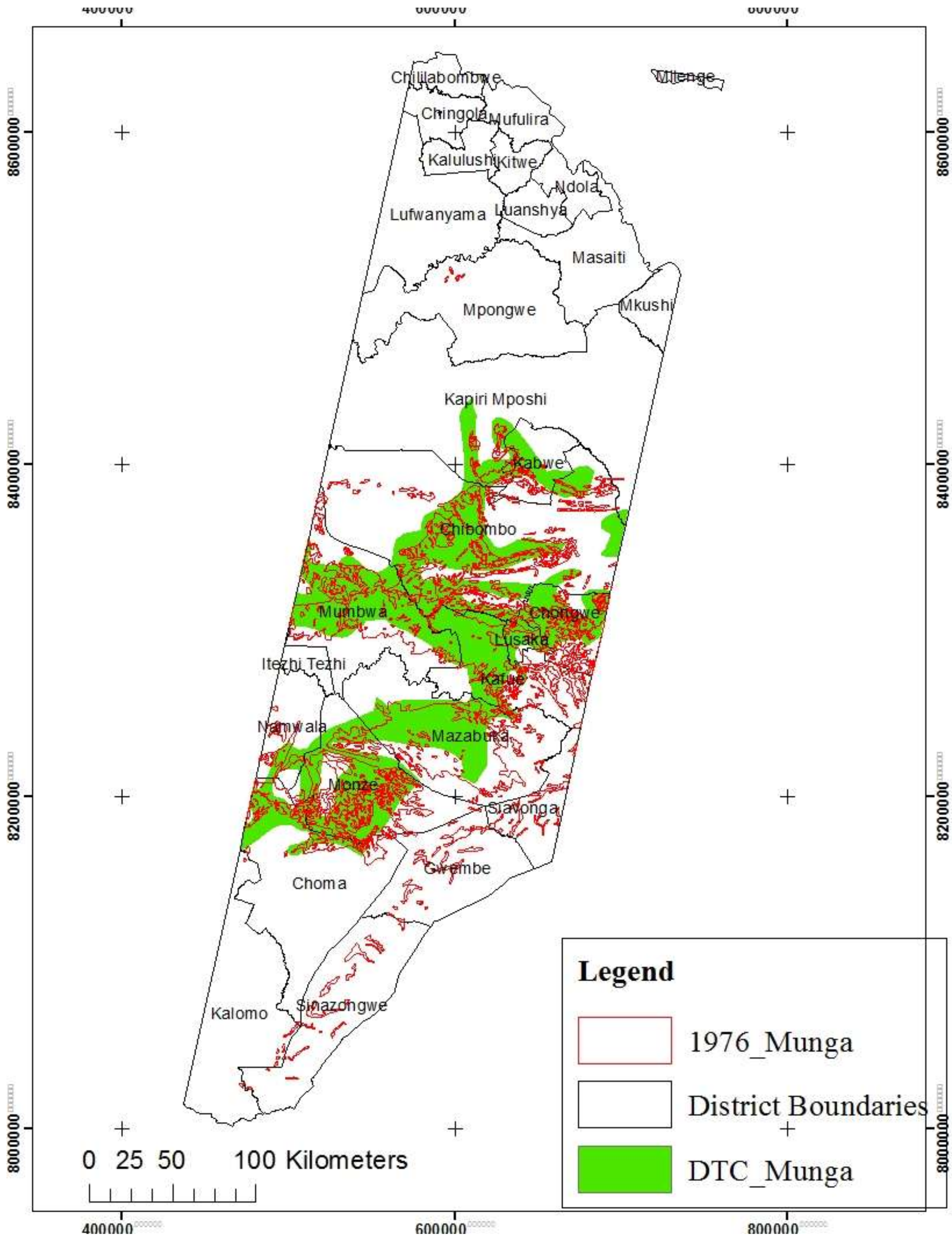

Figure 19: Visual comparison of 1976 Munga overlaid on Decision tree Munga

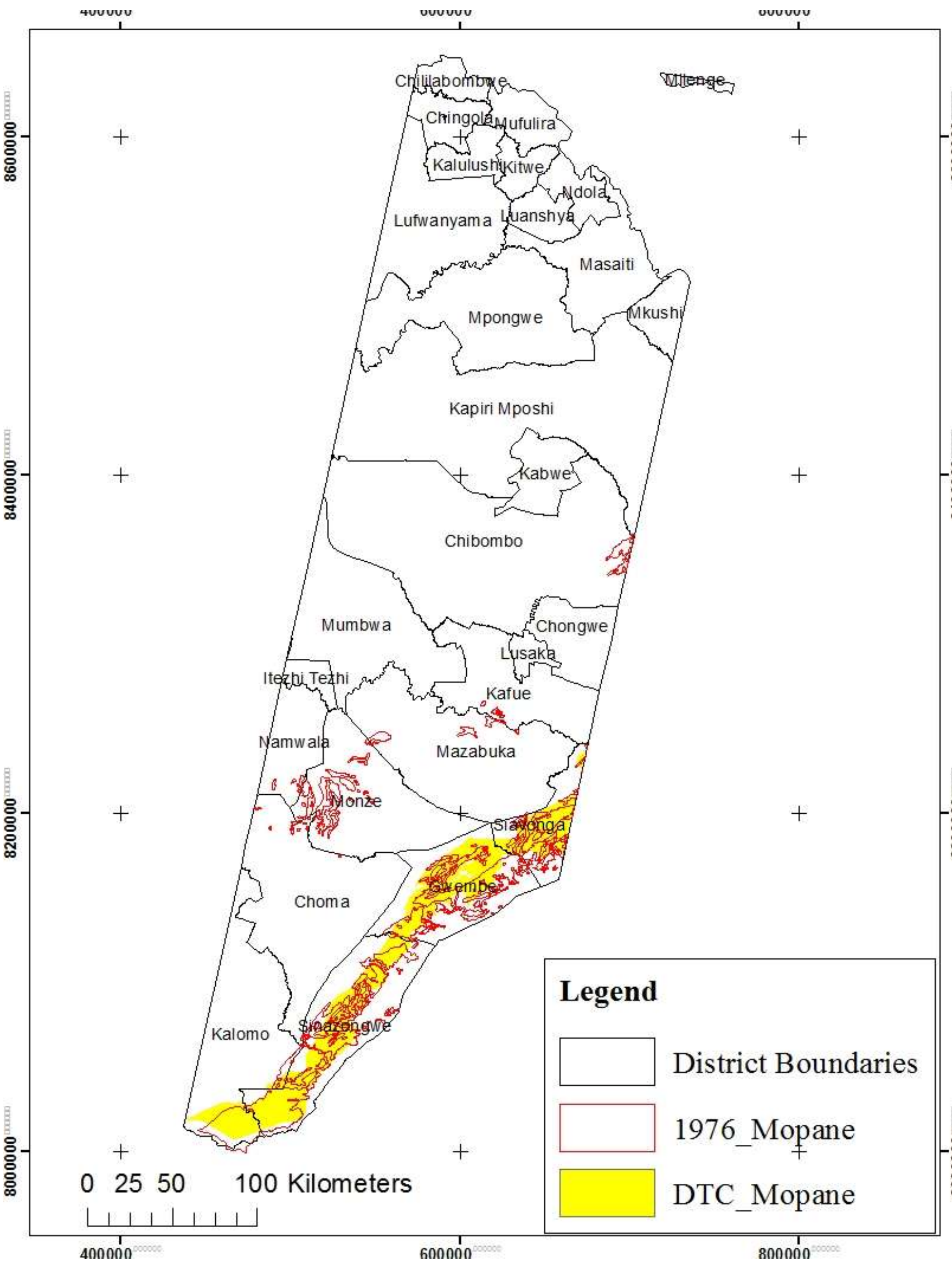

Figure 20: Visual comparison of 1976 Mopane overlaid on Decision Tree Mopane

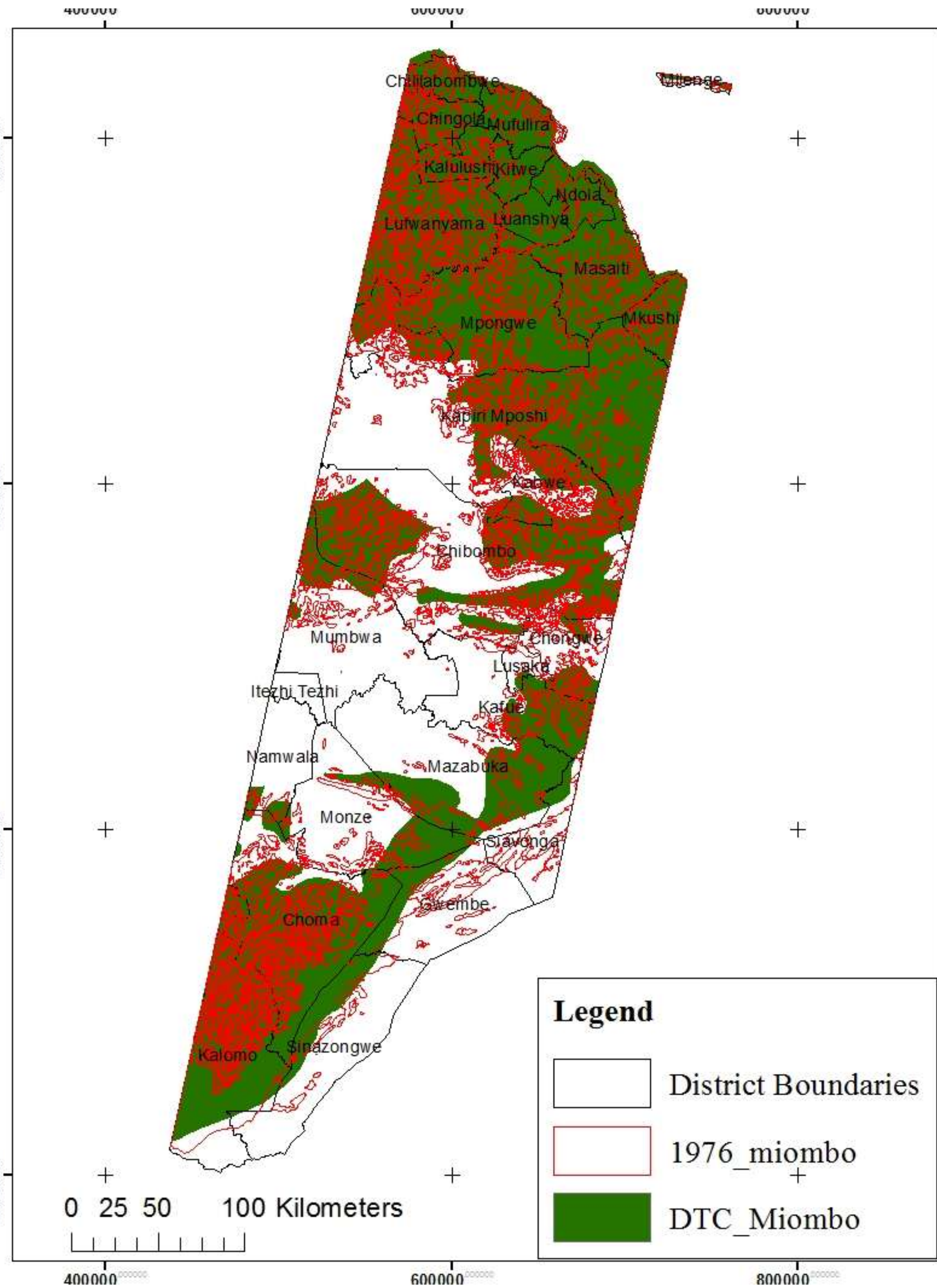

Figure 21: Visual comparison of 1976 Miombo maps overlaid on Decision tree Miombo

5. CHAPTER 5--DISCUSSION

From the produced DTC map and as shown by many research studies such as that done by Fanshawe (1972) and Brady (1990), it can be stated that the occurrence and spatial distribution of vegetation is determined by many factors which include; rainfall, temperature, elevation, human activity and time. In this research, three vegetation types were studied and the factors which affect their occurrence and distribution were noted and used in the DTC for the mapping.

The Miombo vegetation generally was found to be associated with areas which are poor in nutrients and some places which are shallow. According to the soil map of Zambia published by the National Council for Scientific Research (1983), it was observed that the highly weathered, low nutrient soils (Ferralsols and Acrisols) are largely located in agro-ecological zone III (northern half of Zambia) and partly in zones I and II of Zambia. In this study, it was also observed that the Miombo vegetation distribution followed the occurrence of the highly weathered Acrisol and Ferrasol soils.

Munga woodland occurs mainly in the Southern, Central and Eastern provinces of Zambia. They occur on flat land and in some places thrive along rivers and streams. The Munga soils vary in texture from clays to sandy-clays with a high Base Exchange capacity (Woode, 1985). The soil map of Zambia, published by the National Council for Scientific Research (1985), shows that the Munga vegetation distribution follows the occurrence of Luvisols soils; which are base rich soils with distinct clay accumulation as described by Driessen, et al., 2001.

While the Mopane woodland was found to be in areas occupied by alkaline alluvial soils which are sodium affected clay soils. The soil map of Zambia shows that the mopane vegetation occurs in areas dominated by vertisol soils such as those found in Southern and Central Zambia and partly in the Luvisols and other soil types. Vertisols are churning clay soils with 2:1 clays that are dominated by sodium in the interlayer spaces (Driessen, et al., 2001).

Using the relationship that exists between the vegetation types of Zambia and the environmental factors, this study endeavored to map the main vegetation types by integrating the environmental factors and spectral data with the use of Decision Tree Classification (DTC). Other researchers such as Xian, et al., (2002) and Mustafa, et al., (2009) used the decision tree classification to map forest groups and basal areas, and degraded lands respectively.

In this study, the map generated using the DTC was compared with the Vegetation map of Zambia (Forestry Department, 1976), using the non-site-specific-accuracy assessment and the percentage error for each vegetation type was computed. The results showed that Miombo had a percentage error of 13% while Mopane and Munga had percentage errors of 4% and 22%, respectively. These low percentage errors showed that the method was effective in mapping the vegetation types especially over a large area.

These observed discrepancies between the DTC map and the vegetation type map shapefile, from the 1976 Vegetation map of Zambia, may be attributed to the challenges encountered in the study as well as some other factors which may have affected the study. A number of challenges were encountered during the execution of this study and a few major ones were as highlighted in the preceding section.

Unavailability of more (accurate) ancillary data on which the accuracy of the DTC map output strongly relies. A report by the Japan Association of Remote Sensing (1999), also noted that the accuracy of the decision tree map depends fully on the design of the decision tree and the ancillary data hence the unavailability of these ancillary data and/ or the availability of less accurate data may hamper the use of the DTC method. Other challenges were encountered in gathering ancillary data for this mapping, in particular, the soil map of Zambia which was found in hardcopy format only. This map is currently the most accurate soil map of Zambia produced by ZARI in 1995. Due to the format of this soil map and the level of its details, digitizing the map requires a long period of time to complete. Thus, for this study, a simplified and less detailed version of the soil map was used.

Also the vegetation map of Zambia used in this study was produced in 1976 while the Landsat satellite imagery used weres acquired in 1984, which is an 8 years period gap. This 8 year gap could have resulted in mixing of pixels, due to the land cover changes that may have occurred between 1976 and 1984, which might have affected the spectral statistics analyses carried out in this study. For the elevation data, a 90m DEM was used for elevation analysis and in the Decision tree structure hence this elevation accuracy may also have contributed to the visual discrepancies observed in the accuracy assessment. Generally, all the ancillary data may have contributed some level of errors towards the output. And also the issue of scale may have further increased this error i.e. the various datasets used in the decision tree were at various mapping scales, for instance, the soil map used was at 1:1,000,000, the Landsat image can be mapped to a maximum of 1: 60,000 scale while the 1976 vegetation map was at 1:500,000 scale, varying scales of the datasets translate into varying levels of detail of the maps and hence which may have contributed to the observed discrepancies observed.

Challenges were also encountered in gathering literature for the vegetation types of Zambia. The study had set out to using DTC to map all the vegetation types within the area of study, however, at the end of the study, only three vegetation types, namely Miombo, Mopane and Munga, were mapped. This was due to the unavailability of literature on the many other vegetation types in the area of study such as Cryptosepalum, Parinari and Baikea. This kind of literature provides the descriptive information of the environmental conditions, such as the soil type, the elevation and the rainfall data, in which these vegetation types occur or thrives, and it is this kind of information that is used in the DTC for constructing the binary expressions. This information was available for only the Miombo, Mopane and Munga vegetation types while for the other vegetation types was unavailable or not descriptive enough for use here. Also the literature information used for the Miombo and Mopane may not have been specific for Zambia hence this could have contributed towards the discrepancies in the DTC map output. While also not much literature was found on Munga, this literature was only deduced from the desktop analysis using the spatial software over the study area. This may have contributed to observed percent error obtained for Munga vegetation; which was highest of the three mapped vegetation types.

The occurrence of vegetation types is not strict to specific conditions, that is, vegetation types are not exclusively confined to a specific condition for instance a vegetation type does not wholly strictly grow or occur in a certain soil type or at a certain elevation only, hence some vegetation types, such as Miombo, occur in various overlapping conditions. For instance, Miombo is not strictly confined to one specific soil type but a variety of soils. (Young, 1976; FAO-Unesco, 1977; Thompson & Purves, 1978; Purves *et al*., 1981; Nyamapfene, 1991; Anderson *et al*., 1993) described the area where Miombo occurs as; the dominant soils in the higher rainfall zones are classifed as Haplorthox and Haplustox in the USDA taxonomy (approximate FAO equivalents are Orthic, Rhodic and Xanthic Ferralsols); Paleustults and Palexerults (Ferric Acrisols). Haploxeralfs (Ferric Luvisols), Tropudalfs and Paleustults (Eutric Nitosols), and Paleudults and Tropudults (Dystric Nitosols) occur over basic rocks. The dominant soils in the lower rainfall zones are

Ustropepts (Ferralic and Chromic Cambisols), Paleustalfs and Rhodoxeralfs (Chromic Luvisols), and Plinthustalfs (Plinthic Luvisols). Psamments (Arenosols) are wide spread along the south western margins on soils derived from aeolian Kalahari sand.

However, in this study the vegetation types were exclusively mapped to specific soil types or elevation ranges hence producing an output with a less accurate representation of reality, this criterion was based on the fact that each vegetation type predominantly occurs more in certain environmental conditions such certain soil types and/ or on certain elevation than others. Hence for high accuracy DTC vegetation type maps to be achieved, all these necessary environmental conditions for each vegetation type have to be taken into consideration in constructing the binary expressions and also high accurate ancillary datasets need to be used in the DTC classification. Researchers such as Coops, et al. (2006) stated that as with most empirical approaches, the issue with the decision tree analyses is that it provides no discriminatory power outside the range of the input data, with the developed rules only suitable for datasets with equivalent ranges. And as such decision trees require input data that are completely representative of the actual data set in order to produce meaningful rules. This issue restricts the mapping of vegetation types to certain conditions only as opposed to complex prevailing scenario.

From the challenges encountered during the course of this study, it can be seen that in order to use DTC mapping of vegetation types in Zambia, more work needs to be done in researching and studying the vegetation types, other than the three mapped in this study. For instance, work needs to be done on vegetation types such as Cryptosepalum woodlands and made available prior to the DTC mapping. Such research studies would specify the kind of environmental or geographical conditions in which these vegetation types occur and that information would be used to construct the more accurate binary expression for the DTC. Also ancillary, such as detailed soil maps, need to be available in raster format so as to enable ease production of more reliable vegetation maps.

The accuracy assessment method used in this study, the non-site-specific assessment approach, has

The limitation of the Non-site-specific assessment, the accuracy assessment method used in this study, lies in its ability to only compare areal sizes of the reference map and the classified map and does not give an indication of any locational accuracy of the mapped features hence work round this limitation, a visual comparison of the DTC MAP to the "referenced" 1976 Vegetation map was done and shown in figures 19, 20 and 21 of the three respective vegetation types mapped.

However, despite, the challenges encountered, the study has demonstrated the potential and applicability of DTC in mapping of vegetation types in Zambia. From the literature reviewed it has been observed that indeed mapping vegetation types using the traditional remote sensing, with medium to coarse imagery, is difficult and also that vegetation types indeed follow trends, i.e. environmental conditions such as soil types maps.

The spectral statistics data, such as the NDVI and band reflectance, are not necessarily absolute values of the respective vegetation types but only applicable in this study as these values are relative and as such vary with varying conditions such as atmospheric conditions. Ideally, spectral information of vegetation types are supposed to be obtained from "pure" stands of the vegetation types of interest. In this case a "pure stand" of a vegetation type is identified in the fields and its coordinates collected using a GPS. Thereafter, the collected coordinates are plotted on the satellite imagery of the area of interest and a spectral analysis of the pixels is done to obtain the more accurate spectral data of the identified "pure stand". This analysis can further be enhanced with the use of field radiometers in the acquisition of these spectral statistics of the respective vegetation types and using that information in the DTC with preprocessed satellite imagery, up to surface reflectance.

The knowledge obtained from the study, with the help of the many reviewed studies, indicate that indeed with spectral responses only, for medium to coarse resolution imagery, it is difficult to distinguish and map vegetation types and thus the integration of these satellite imagery with ancillary data such as soil type data, elevation data or rainfall data via a decision tree classifier, is cited as the most appropriate approach to mapping vegetation types. Such knowledge may assist institutions such as; Forestry department to effectively manage the nation's vegetation resources, Zambia National Farmers Union to perform their crop yield projections and this may also help lead to production of more accurate ancillary data such as soil maps by institutions like ZARI.

6. CHAPTER 6--CONCLUSION AND RECOMMENDATION

6.1. Conclusion

The main objectives of the research were to study the environmental factors which determine the spatial distribution of vegetation types of Zambia and to apply DTC to mapping these vegetation types. However, the set objectives were not fully achieved as not all the vegetation types within the study area were mapped. This was due to the unavailability of the necessary literature describing the environmental characteristics of these vegetation types. Thus, for this study, decision tree mapping was demonstrated for only three vegetation types, namely Miombo, Mopane and Munga, as a good number of research studies have been carried out on the three vegetation types and hence literature on these vegetation types were quite readily available. This and several other challenges encountered negatively affected the execution of this research and the obtained results.

However, despite the challenges encountered during the research, the decision tree mapping of Zambia's vegetation types was demonstrated and used to map three vegetation types within the study area. Thus, with its ability to incorporate remote sensing with other ancillary geographic data, DTC proved to be an appropriate methodology for mapping land cover such as vegetation type while maintaining the use of medium to low spatial resolution imagery.

Given the mentioned aspect of DTC, the most important contribution of this research with respect to remote sensing may be summarized as follows. This study may save as a basis for future further studies related to vegetation and remote sensing such as in crop estimation activities e.g. harvest projections hence this may be used by, among many, Zambia National Farmers Union (ZNFU). With the atmospheric and environmental conditions requirements for a certain crop known from literature and with the availability of such ancillary data, DTC can be used to predict the amount of yield given the prevailing conditions at a given time before the end of that crop season. Furthermore, such studies, which integrate spectral data with other geographical ancillary data via decision tree classifiers, may lead to the creation of vegetation modelling systems which can be used in predicting the distribution and amounts of vegetation types in an area; for improved management and monitoring of these vegetation types e.g. managing vegetation such as Mukula tree such works may be useful for institutions such as the Zambia Forestry Department (ZFD) and Department of Wildlife. These and many other potential applications of DTC need to be explored.

6.2. Recommendation

From the reviewed literature on DTC and thus its mapping demonstrated in this study, the application potential and strength of DTC can be appreciated and thus its importance in land cover mapping. The study has demonstrated the concept of creating a DTC structure and its execution. It has also shown that for the production of a reliable DTC map, all the necessary ancillary data should be available and of acceptable accuracy furthermore literature on vegetation types must also be available for the construction of DTC binary expressions to use in the classification. Therefore, it can be recommended that more thorough research work in these vegetation types should be carried out in order to make available the descriptive information of these vegetation environmental characteristics such as the soil type or the elevation at which they occur. Such descriptive information, when displayed in form of maps, serve as primary datasets to use in the DTC mapping hence the need for it to be available prior to the mapping.

Further research should be set up to investigate the DTC mapping approach and how this approach can be improved upon and also how it can be extended to the creation of land cover mapping models. The creation of models would ease the mapping exercise of land cover e.g. vegetation types and would also enable the prediction of the land cover type at a given area given the prevailing conditions or parameter at a given time. The ability of predicting land cover from the given parameters may find great use in important national activities such as crop yield estimation projections; whose importance can never be overemphasized.

7. References

1. Adler-Golden, S. M., Matthew, M. W., Bernstein, L. S., Levine, R. Y., Berk, A., Richtsmeier, S. C., Acharya, P. K., Anderson, G. P., Felde, G., Gardner, J., Hoke, M., Jeong, L. S., Pukall, B., Ratkowski, A., and Burke, H. H. (1999). Atmospheric Correction for Short-wave Spectral Imagery Based on MODTRAN4. SPIE Proceedings on imaging spectrometry, 3753: 61-69.
2. Akhter, M. (2006). Remote Sensing for Developing an Operational Monitoring Scheme for the Sundarban Reserved Forest, Bangladesh. PhD thesis, TechnischeUniverstät Dresden, Germany.
3. Anderson, D.M.W. (1993). Some Factors Influencing the Demand for Gum Arabic (Acacia senegal, (L.) Willd.) and Other Water-soluble Tree Exudates. *Forest Ecology and Management* 58: 1-18.
4. Aregheore, E.M. (2009). Country Pasture/ Forage Resource Profiles.
5. Bolstad, P. V., and Lillesand, T. M. (1992). Rule-based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data. *Photogrammetric Engineering and Remote Sensing*, 58(7): 965–971.
6. Breiman, L., Friedman, J.H., and Olshen, R.A. (1984). Classification and Regression Trees. Boca Raton: Chapman & Hall.
7. Campbell, J. B. (2002). Introduction to Remote Sensing. (3rd edition.), Taylor and Francis.
8. Campbell, B. (ed), (1996). The Miombo in Transition: Woodlands and Welfare in Africa. Center for International Forestry Research (CIFOR). Printed in Malaysia
9. Chidumayo, E.N., and Gumbo, D.J. (2010). “The Dry Forests and Woodlands of Africa”. Managing for Products and Services
10. Chuvieco, E., and Huete, A. (2010). Fundamental of Satellite Remote Sensing, Taylor and Francis Group, New York.
11. Congalton, R. G. and Green, K. (1999). Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers.
12. Congalton, R. G. (1988). A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing*, 54: 593-600.

13. Dafalla, M. S. (2007). Mapping and Assessment of Land use/Land cover Using Remote Sensing and GIS in North Kordofan State, Sudan. PhD thesis, TechnischeUniversität Dresden, Germany.
14. Di Gregorio, A., and Jansen, L. J. M. (2005).Land Cover Classification System-Classification concepts and user manual. FAO Manual
15. Driessen, P., and Deckers, J. (2001). Lecture Notes on the Major Soils of the World
16. Egbert, S. L., Park, S., and Price, K. P. (2002) Using Conservation Reserve Program Maps Derived from Satellite Imagery to Characterize Landscape Structure. *Comput Electron Agric* 37:141–56.
17. Ernst, C. L., and Hoffer, R. M. (1979). Digital Processing of Remotely Sensed Data for Mapping Wetland Communities. West Lafayette. Laboratory for Applications of Remote Sensing, Purdue University.
18. Fanshawe, D. B. (2010). Vegetation Descriptions of the Upper Zambezi Districts of Zambia. *Occasional Publications in Biodiversity* No. 22
19. FAO. (2005-2008). Integrated Land Use Assessment (ILUA- Final Report. Zambia
20. FAO-UNESCO. (1977). Soil Map of the World, Volume VI, Africa. UNESCO, Paris.
21. Friedl, M. A., Brodley, C.E., and Strahler, A. H. (1999). Maximizing Land Cover Classification Accuracies Produced by Decision Trees at Continental to Global Scales. *IEEE Transactions on Geoscience Remote Sensing* 37(2): 969-977.
22. Ghose, M K., Pradhan, R., and Ghose, S. S. (2010). Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matrix. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 1(5).
23. Hansen, M. C., DeFries, R. S., and Townshend, J. R. (2006). Global Landcover Classification at 1 km Spatial Resolution Using a Classification Tree Approach. *International Journal of Remote Sensing*, 21:1331–6
24. He, C., Zhang, Q., and Li, Y. (2005). Zoning Grassland Protection Area Using Remote Sensing and Cellular Automata Modeling—a case study in Xilingol Steppe Grassland in Northern China. *Journal of Arid Environment* 63:814–26
25. Hinson, J. M., German, C. D., and Pulich, W. J. (1994). Accuracy Assessment and Validation of Classified Satellite Imagery of Texas Coastal Wetlands. *Marine Technology Society Journal*, 28(2): 4–9.

26. Hunter, S. C., and Paysen, T. E. (1986).Vegetation Classification System for California: User's Guide
27. Irish, R. (2004). Landsat 7 Science Data Users Handbook: Greenbelt, Maryland, Goddard Space Flight Center. http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html (accessed on 25 December 2007)
28. ITT Visual Information Solution. (2009). ITT Visual Information Solution, ENVI FLAASH Module User´s Guide.
29. Joy, S. M., Reich, R. M., and Reynolds, R. T. (2003). A Non-parametric Supervised Classification of Vegetation Types on the Kaibab National Forest using Decision Trees. *International Journal of Remote Sensing*, 24(9): 1835−1852. DOI: 10.1080/01431160210-154948
30. Kumar, S. V. K., and Kiruthika, P. (2015). An Overview of Classification Algorithm in Data Mining. *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 4, Issue 12
31. Langley, S. K., Cheshire, H. M., and Humes, K.S. (2001). A Comparison of Single Date and Multitemporal Satellite Image Classifications in a Semi-arid Grassland. *Journal of Arid Environment* 49:401–11
32. Lea, C. (2011). Vegetation Classification Guidelines: National Park Service Vegetation Inventory, version 2.0. Natural Resource Report NPS/NRPC/NRR—2011/374. National Park Service, Fort Collins, Colorado
33. Liu, K., and Shi, Xun. (2008). Monitoring Mangrove Forest Changes Using Remote Sensing and GIS Data With Decision-tree Learning. *Wetlands*, 28:336. DOI: 10.1672/06-91.1
34. Lillesand, M. T., KIefer, W. R., and Chipman, N, J. (2008). Remote Sensing and Image Interpretation (6th edition). John Wiley and Sons, Inc, New York.
35. Lu, D., and Weng, Q. (2007). A survey of Image Classification Methods and Techniques for Improving Classification Performance. *International Journal of Remote Sensing*, Vol. 28, No. 5, 823– 870.
36. Markham, B. L. and Barker, J. L. (1987). Radiometric Properties of U.S. Processed Landsat MSS Data. *Remote Sensing of Environment*, 22: 39-71.

37. Mather, P. M. and Brandt, T. (2009). Classification Methods for Remotely Sensed Data, (2nd edition). Taylor and Francis Group, LLC.
38. Minakshi Kumar. (2013). Digital Image Processing
39. Mustafa, A. A., Singh, M., Sahoo, R. N., Ahmed, N., Khanna, M., Sarangi, A., and Mishra, A. K. Mapping of Degraded Lands from Multidate Remotely Sensed Data Using Decision Tree Based Classification (DTC).
40. Na, X. D., Zhang S Q. (2009). Integrating TM and Ancillary Geographical Data with Classification Trees for Land Cover Classification of Marsh Area
41. Coops, N. C., Wulder, M. A., and White, J. C. (2006). Integrating Remotely Sensed and Ancillary Data Sources to Characterize a Mountain Pine Beetle Infestation. Mountain Pine Beetle Initiative Working Paper 2006–07
42. Nordberg, M. L., and Evertson, J. (2003). Vegetation Index Differencing and Linear Regression for Change Detection in Swedish Mountain Range Using Landsat TM and ETM+ imagery. *Land Degradation & Development* 16:139–149.
43. Nyamapfene, K. (1991). Soils of Zimbabwe. Nehanda Press, Harare.
44. Ozesmi, S. L., and Bauer, M. E. (2002). Satellite Remote Sensing of Wetlands. *Wetlands Ecology and Management*, 10(5): 381–402. DOI: 10.1023/A:1020908432489
45. Pal, M. (2005). Random Forest Classifier for Remote Sensing Classification. *International Journal of Remote Sensing*, 26(1): 217− 222. DOI: 10.1080/01431160412331269698
46. Pooja, A. P., Jayanth, J., and Koliwad, S. (2011). Classification of RS data using Decision Tree Approach. *International Journal of Computer Application* (0975-8887). Volume 23- No. 3.
47. Price, J. C. (1987). Calibration of Satellite Radiometers and the Comparison of Vegetation Indices. *Remote Sensing of the Environment*, 21 (15): 15-27.
48. Purves, W.D., Fullstone, M.J., Godwin, R.G.H., and Wells, J.D. (1981). Soils of the Banket area. *Zimbabwe Agricultural Journal Technical Handbook 4*. Government Printer, Harare.
49. Sabins, F. F. (1997). Remote Sensing, Principles and Interpretation. (3rd edition). W. H. Freeman and Company, New York.
50. Sader, S. A., Ahl, D., and Liou, W. S. (1995). Accuracy of Landsat TM and GIS rule-based Methods for Forest Wetland Classification in Maine. *Remote Sensing of Environment*, 53(3): 133–144. DOI: 10.1016/0034-4257(95)00085-F

51. Schmidt, A., Karrasch, P., and Neubert, M. (2009). Vergleichende untersuhungen der atmosphärenkorrekturprogramme ATCOR und FLAASH auf der datengrundlage des satellitensystems IKONOS. Rhombos-verlag, Berlin.
52. Sesnie, S. E., Gessler, P. E., and Finegan, B. (2008). Integrating Landsat TM and SRTM-DEM Derived Variables with Decision Trees for Habitat Classification and Change Detection in Complex NEotropical Environments. *Remote Sensing of Environment*, 112(5): 2145−2159. DOI: 10.1016/j.rse.2007.08.025
53. Stellingwerf, D. A., and Hussin, Y. A. (1997). Measurements and Estimations of Forest Stand Parameters Using Remote Sensing, VSP BV, Zeist, the Netherlands.
54. Thompson, J.G., and Purves, W.D. (1978). A Guide to the Soils of Rhodesia. Agricultural. Journal Technical Handbook No. 3. Government Printer, Salisbury.
55. Trapnell, C.G., and Clothier, J.N. (1996). The Soils, Vegetation and Traditional Agriculture of Zambia, Vol.1, (Central and Western Zambia Ecological Survey 1932-1936); and Vol.2 (by Trapnell, North Eastern Zambia, Ecological Survey 1937-1942).
56. Vohland. M., Stoffels, J., Hau, C., and Schüler, G. (2007). Remote Sensing Techniques for Forest Parameter Assessment: Multispectral Classification and Linear Spectral Mixture Analysis. Silva Fennica 41(3): 441–456
57. Wright, C., and Gallant, A. (2007). Improved Wetland Remote Sensing in Yellowstone National Park Using Classification Trees to Combine TM Imagery and Ancillary Geographical Data. *Remote Sensing of Environment* 107(4): 582−605. DOI: 10.1016/j.rse. 2006.10.019
58. Woode, P.R. (1985). Soil Taxonomy and Agro technology Transfer.
59. Xiao, X. M., Zhang, Q., and Braswell, B. (2004). Modeling Gross Primary Production of Temperate Deciduous Broadleaf Forest Using Satellite Images and Climate Data. *Remote Sensing Environment* 91:256–70.
60. Xie, Y., Sha, Z., and Yu, M. (2008). Remote Sensing Imagery in Vegetation Mapping: A Review. *Plant Ecology* 9-23 doi:10.1093/jpe/rtm005.
61. Zakaria H.E.A. (2010). Integration of Remote Sensing and GIS in Studying Vegetation Trends and Conditions in the Gum Arabic Belt in North Kordofan, Sudan. PhD thesis, Technische Universität Dresden, Germany