UNIVERSIDADE D COIMBRA

Esther Maria Joy Holden Baptista

Methodology for Land Cover Mapping Based on Remote Sensing Data in a Subtropical Area (SW of Angola)

Tese de doutoramento em Geologia, ramo Recursos Geológicos e Ambiente, orientada pelo Professor Doutor Alcides José Sousa Castilho Pereira, e apresentada a Faculdade de Ciências e Tecnologia da Universidade de Coimbra/Departamento de Ciências da Terra

Julho 2018

Esther Maria Joy Holden Baptista

METHODOLOGY FOR LAND COVER MAPPING BASED ON REMOTE SENSING DATA IN A SUBTROPICAL AREA (SW OF ANGOLA)

Tese de doutoramento em Geologia, ramo Recursos Geológicos e Ambiente,

orientada pelo Professo Doutor Alcides José Sousa Castilho Pereira,

apresentada ao Departamento da Ciências da Terra da Faculdade de

Ciências e Tecnologia da Universidade de Coimbra

Julho 2018

UNIVERSIDADE DE COIMBRA

I dedicate this thesis to my loving husband, who has been at my side during the long journey which led up to its completion.

ACKNOWLEDGEMENTS

I would like to thank my husband for his unconditional support and encouragement in this journey we have undertaken together, and my son who became a loyal companion in a lot of my field work.

I would like to thank my colleagues and tutors in Coimbra who have been extremely helpful and patient since the beginning.

I would like to thank my mother for the constant encouragement she has provided during the whole process and for her expertise in proofreading my final text.

I would like to thank my brothers and sisters for their interest and support along the way, especially Paul for his help with brainstorming ideas and proofreading my text and Daniel for his IT expertise.

ABSTRACT

This study proposed to develop and assess a methodology for land cover mapping in a subtropical region of Southwestern Angola. The goal was not only to prepare a methodology for creating a land cover map but also an accompanying assessment procedure based on fieldwork observations to acquire quality reference data. The land cover classification was produced by training a Decision Tree Model using a high quality training set.

The proposed methodology began with the characterization of the study area to create an appropriate land cover classification system. Literature review and extensive field work led to the characterization of the area which was centered on the municipality of Humpata, confined primarily to an afro-montane forest-grassland mosaic ecoregion. The visual analysis of satellite imagery provided insight into the radiometric distinction and location of certain classes while the field work served to verify specific classes and helped to distinguishing between classes, especially vegetation classes.

Using the information acquired during the field observations a sample of 3880 pixels (combinations of individual pixels and 3x3 pixel groups) were selected for the training set. Roughly the same number of sample pixels was selected for each land cover class. The selection of the training set was limited almost exclusively to areas observed during the field work as opposed to selecting the sample using high definition images. This was done as an effort to guarantee the quality of the training set above all else. The great limitation and difficulty of this approach, as can be imagined, was the road access and time constraints. The limited information

concerning road access in the study area led to a time-consuming exploration of the area and a dependency on guides or a third person during the field work.

The training set was then used to calibrate and validate a Decision Tree Model, using one Landsat 8 image from the dry season and one from the rainy season. Validation of the model was done using 10% of the sample set. Once satisfactory results were acquired with the model, it was then used to classify the 2013/2014 Landsat 8 images thus creating a land cover map of the area at a spatial resolution of 30 m, presented at a scale of 1:100 000.

The next step of the methodology was creating an accuracy assessment procedure to verify the quality of the land cover classification. Once again, this was done almost exclusively using field work observations. A random stratified design was used to select the sample for the accuracy assessment limited to the area with road acces. Two classes, however, were assessed using Google Earth images instead of ground observation because of complete lack of access.

The classification system was made up of 16 Level II land cover classes and 8 Level I land cover classes. The overall assessment showed 72% accuracy for the Level II classes and 81% accuracy for the Level I classes. Individual 2x2 km areas were selected in the different regions of the mapped area in order to further evaluate the accuracy of the model and therefore acquire a greater understanding of the overall quality of the classification. This analysis highlighted the accuracy of the different land cover classes in the various sectors of the study area.

The accuracy of the model was also assessed in a similar area to the East of the original mapped area to verify its application to an area in which it was not calibrated. Verification in this area, because of time constraints, was done using only Google Earth images and not ground observations. As a result, it was impossible to verify the Level II land cover classes. Using Google Earth made it very difficult if not impossible to distinguish and verify the different vegetation classes. The assessment, therefore, was restricted to verifying the 8 Level I land cover classes. The assessment showed a decrease in accuracy in this area compared to the original mapped area. The overall accuracy was 71%. This decrease in accuracy could be due to the fact that verification was done using Google Earth images, instead of ground observation, or perhaps due to the fact that the model was not calibrated in this specific area. Nevertheless, these results are promising and show the possible application of the model in classifying areas with similar characteristics as the original mapped area.

A further assessment was done of the training set acquired in the first stages of the methodology, to verify its application in training models for classifying images of different/future years. This assessment was done using Landsat 8 images from 2017. Based on a visual analysis, specific pixels were removed from the training set that showed evidences of significant change since the images used for calibration (2013/2014). A decision tree model was then calibrated using the updated training set. The validation of the algorithm for the classification of the 2017 images resulted in an accuracy of 87%. These good results attest to the quality of the training set, and its potential for future classifications and as a tool for continual land cover monitoring.

In summary, the results of this study show that even with the limited access described many times before, the proposed methodology based on field work is a robust tool for studies on land cover evaluation.

The acquired training set can also be used in future support of other research and be an essential tool for monitoring land cover in southwest Africa and, thus, in evaluating the impact of climatic change applied in different contexts as opposed to being limited to creating this specific land cover map. It is hoped that this methodology would contribute to other land cover mapping studies in places with similar characteristics.

Key words: Land cover, methodology, remote sensing, numerical modeling, Angola.

RESUMO

Este estudo visa desenvolver e avaliar uma metodologia para o mapeamento da cobertura do solo numa região subtropical no Sudoeste de Angola. O objetivo não era apenas criar uma metodologia para a criação duma carta de cobertura do solo, mas também um procedimento de avaliar a sua precisão, baseado em observações de campo para a aquisição de dados de referência de qualidade. A classificação da ocupação do solo foi produzida através da calibração de um modelo de árvore de decisão usando um conjunto de treino de alta qualidade.

A metodologia proposta começou com um processo de caracterização da área de estudo, para criar um sistema de classificação de cobertura do solo apropriado. A revisão de literatura e trabalho de campo extensivo levou à caracterização da área que estava centrada no município da Humpata, confinado primordialmente a uma eco região de Mosaico Florestal Angolano de Pastagem de Montanha. A análise visual de imagens de satélite forneceu informação sobre a separação radiométrica e localização de certas classes enquanto o trabalho de campo serviu para verificar classes específicas e auxiliou a distinguir entre diferentes classes, especialmente classes de vegetação.

Utilizando a informação adquirida durante as observações do trabalho de campo uma amostra de 3380 pixéis (combinações de pixéis individuais e grupos de 3x3 pixéis) foi selecionada para o conjunto de treino. Aproximadamente o mesmo número de pixéis de amostra foram selecionados por classe de cobertura do solo. A seleção do conjunto de treino foi limitada quase exclusivamente a áreas acessíveis ao trabalho de campo, ao invés de selecionar apenas amostras recorrendo a imagens de alta resolução. Isto foi uma tentativa de garantir, acima de tudo, a qualidade do conjunto

de treino. A grande limitação e dificuldade, como pode-se imaginar, foi o acesso por estrada e as limitações de tempo. A informação disponível acerca dos acessos na área de estudo levou a longas explorações na área de estudo e a dependência em guias durante o processo de trabalho de campo.

O conjunto de treino foi então usado para calibrar e validar o modelo de árvore de decisão, utilizando uma imagem Landsat 8 do tempo de seco e uma do tempo chuvoso. A validação do modelo foi feita utilizando 10% dos pixéis do conjunto de amostras. Uma vez obtidos resultados satisfatórios com o modelo, este foi utilizado para classificar a imagem Landsat 8 referente aos anos 2013/2014, assim criando uma carta de ocupação do solo da área numa resolução espacial de 30 m, apresentada numa escala de 1:100 000.

O passo seguinte foi o de criar um procedimento de avaliação da precisão para verificar a qualidade da classificação da cobertura do solo. Mais uma vez, isto foi feito quase exclusivamente utilizando observações do trabalho de campo. Um *design* de aleatório de estratificação foi usado para selecionar a amostra para a avaliação de precisão limitada às áreas acessíveis no terreno por estrada. Duas das classes, contudo, por forte limitações de acesso, foram verificadas utilizando imagens do Google Earth em vez de observações de campo.

O sistema de classificação foi constituído por 16 classes de cobertura do solo do Nível I e 8 do Nível II. A avaliação geral da classificação mostrou 72% de precisão da classificação com 16 classes e 81% para a classificação com 8 classes. Áreas de 2x2 km foram selecionadas em diferentes zonas da área mapeada com o propósito de melhor avaliar o modelo e assim adquirir uma maior compreensão da qualidade geral da classificação. Esta análise permitiu avaliar a precisão das diferentes classes de cobertura do solo em diferentes sectores da área de estudo.

O modelo foi também testado numa área contígua situada na margem leste para verificar a sua aplicabilidade numa área não usada previamente nas áreas de treino. Devido a constrangimentos de tempo a avaliação da qualidade dos resultados do modelo, foi apenas feita com o suporte de imagens do Google Earth e não por observações no campo. Por esta razão, não foi possível verificar as classes do Nível II, difíceis de distinguir na escala das imagens do Google Earth optando-se por trabalhar apenas com as 8 classes do Nível I. Em 71% dos casos os resultados foram satisfatórios. Comparativamente à situação observada na área em estudo, este resultado indica uma pequena diminuição da qualidade do modelo definido para a área de teste. Esta diminuição poderá ser devido ao facto da avaliação ter sido feito apenas utilizando imagens do Google Earth, em vez de observações no campo, ou talvez devido ao facto de que o modelo não foi calibrado nesta área específica. De qualquer forma, estes resultados são promissores e mostram as possíveis aplicações do modelo em classificar áreas com características semelhantes à da área original.

Foi também testada aplicação do modelo à mesma área mas em anos diferentes daquele que serviu à calibração do modelo. Esta avaliação suportou-se em dados expressos em imagens de Landsat 8 do ano 2017. Baseado numa avaliação visual, foram removidos todos os pixéis do conjunto de treino que mostravam evidências de que a área que representam sofreu alterações significativas desde o ano em que foram obtidas as imagens previamente calibradas (2013/2014). Um modelo de árvore de decisão foi calibrado usando o conjunto de treino que foi objeto do refinamento descrito. Na validação do algoritmo à classificação das imagens de 2017 obteve-se uma precisão de 87%. Estes bons resultados mostram a qualidade do conjunto de treino, e o seu potencial para classificações de imagens futuras e como uma ferramenta para a monitorização contínua da cobertura do solo.

Em síntese os resultados do presente trabalho de investigação indicam que, mesmo com as limitações atrás várias vezes descritas, a metodologia proposta, baseada no trabalho de campo, é uma ferramenta robusta em estudos de avaliação da cobertura do solo.

O conjunto de treino obtido poderá também servir no futuro de apoio a outros trabalhos de investigação e ser uma ferramenta essencial para a monitorização da cobertura do solo no sudoeste de África e, assim, avaliar o impacte das alterações climáticas aplicadas em diferentes contextos ao contrário de serem limitados apenas à criação desta carta de cobertura de solo. Espera-se que esta metodologia possa contribuir para outros estudos de cobertura de solo em lugares com características semelhantes.

Palavras chaves: Cobertura do solo, methodologia, detecção remota, modelação numérica, Angola.

INDEX

FIGURE INDEX

TABLE INDEX

LIST OF ACRONYMS

AVHRR – Advanced Very High Resolution Radiometer

CARPE – Central Africa Regional Program for the Environment

CCI – Climate Change Initiative

CEB – Compressed Earth Block

DRC – Democratic Republic of Congo

DT – Decision Tree

DTC – Decision Tree Classifier

EROS – Earth Resources Observation Systems

ETM – Enhanced Thematic Mapper

ESA – European Space Agency

FAO – Food and Agriculture Organization

GAP – Gap Analysis Program

GCOS – Global Climate Observing System

GIS – Geographic Information System

GLC2000 – Global Land Cover 2000

GLCN – Global Land Cover Network

GOFC – Global Observation of Forest Cover

GOLD – Global Observation of Land Cover Dynamics

GPS – Global Positioning Service

IBR – Inverted Box Rib

IGCA – Instituto de Geodesia e Cartografia de Angola

ISA – Impervious Surface Areas

LC – Land cover

LCCS – Land Cover Classification System

MERIS – Medium Resolution Imaging Spectrometer

MLC – Maximum Likelihood Classifier

MODIS – Moderate Resolution Imaging Spectroradiometer

MRLC – Multi-Resolution Land Characteristics

NASA – National Aeronautics and Space Administration

NLCD – National Land Cover Database

OLI – Operational Land Imager

PCA – Principal Component Analysis

RGB – Red Green Blue

ROI – Regions of interest

SANLCDP – South Africa National Land-Cover Database Project

SASSCAL – Southern African Science Service Center for Climate Change and Adaptive Land Management

SPOT – Satéllite Pour l'Observation de la Terre

SRTM – Shuttle Radar Topographic Mission

TOA – Top of the Atmosphere

TM – Thematic Mapper

TIRS – Thermal Infrared Sensor

UN – United Nations

UNESCO – United Nations Educational, Scientific and Cultural Organization

USGS – United States Geological Survey

UTM – Universal Transverse Mercator

Weka – Waikato Environment for Knowledge Analysis

WGS84 – World Geodetic System 1984 USGS

WWF – World Wildlife Fund

WWF SARPO – World Wild Fund Southern Africa Programme Office

1 – INTRODUCTION

“A modern nation, as a modern business, must have adequate information on many complex interrelated aspects of its activities in order to make decisions. Land use is only one aspect, but knowledge about land use and land cover has become increasingly important … to overcome the problems of haphazard, uncontrolled development, deteriorating environmental quality, loss of prime agricultural lands, destruction of important wetlands, and loss of fish and wildlife habitat. Land use data are needed in the analysis of environmental processes and problems that must be understood if living conditions and standards are to be improved or maintained at current levels” (Anderson, 1976).

Land cover is biophysically recognizable on the Earth’s surface at any specific point in time. Land cover data sets play a significant role in a variety of geographical studies, such as natural resources management, global climate change detection, sustainable urban development and earth system simulation (Yang et al., 2017). Land cover is continually changing, whether by natural causes or by human intervention. This change has obvious implications for the environment as well as for society. The production of land cover maps is of paramount importance to a vast array of study areas, especially as a monitoring instrument. Land cover information has been used for monitoring land cover/use including assessment, management and prediction of urban growth (Aburas et al., 2017; Tuholske et al., 2017), land and water management, monitoring impervious surfaces and underground water (Mantas et al., 2016; Kumar, 2016; Vushoma, 2016), monitoring farmland and irrigation and degradation of wetlands (Finn, et al., 2017; Heinimann et al., 2017; Reyes-Gonzalez, 2017, Anule and Ujoh, 2017), for monitoring forests, desertification and forest fires

(Padonou, et al., 2017; Waser, et al., 2017; Austin, et al., 2017; Millones, et al., 2017, Remmel and Perera, 2017), as well as assessing environmental aspects affected by land cover change (Kimball et al., 2017; Mehmood, et al., 2017; Adinehvand et al., 2017). Access to land cover information can facilitate the study of many environmental and socio-economic aspects that are directly affected by or intricately related to land cover. Land cover maps contain a lot of important information that can be used in different scientific areas as well as in social areas. Monitoring land cover change caused by human activity on the environment (whether by urban development, clearing of land for farming or extraction of rock, soil or minerals) is important in the future planning for appropriate land use policies. It can be instrumental in preventing disasters like mud flows and landslides as well as other consequences caused by environmental change (Mariyappan and Perumal, 2004).

Satellite remote sensing has become an ideal technology and source of information for large scale land cover mapping as a result of numerous national, regional, continental and global land cover mapping efforts (Franklin and Wulder, 2002). Because of easy access to innumerable satellite image products, the production of land cover maps from satellite images has become commonplace. Diverse software is available which make classification of satellite images quite straightforward (Morse-McNabb, 2012). Prasad et al. (2015) gives an overview of the research done on (1) the development of advanced classification algorithms including sub pixel, per field and acknowledged based classification algorithms, (2) the use of various remote sensing features including spectral, spatial, multi-temporal and multi-sensor information; as well as (3) the incorporation of ancillary data into classification procedures that includes topography, soil, road and census data.

On the other hand, with the increased availability of different satellite imagery and the development of different classification approaches and techniques, land cover mapping using remote sensed data can become a very complex procedure. The characteristics of the end product are determined by many factors: the purpose, the thematic content, the scale, the data as well as the processing and analysis algorithms (Cihlar, 2000). Within the classification process itself there are a variety of components which must also be decided upon: a suitable classification system, selection of training samples, image processing, feature extraction, selection of suitable classification approaches, post-classification processing and accuracy assessment (Lu and Weng, 2007). Until now no ideal approach has been developed. All classifiers at some point must handle a three-way compromise between the information classes that are desired, the spectral information content of the imagery and the method of making class decisions (Franklin and Wulder, 2002).

This has led to the development of a variety of land cover mapping methodologies or protocols used by different land cover programs, created particularly for and suited to any given or defined area, with specific objectives. In the different programs the resolution varies from coarse to fine, the scale varies from global to regional, and the information acquired originates from a variety of different sensors (Franklin and Wulder, 2002).

The development of global land cover classification systems began in the 1970's with the USGS land cover classification system, followed by others in the 1990's. They include the EarthSat GeoCover Land Cover Legend, the UN/FAO Land Cover Legend, the Global Observation of Forest Cover and Global Observation of Land

Cover Dynamics (GOFC/GOLD) Land and Forest Cover Classification System and Land Cover Classification System (LCCS) (Yang et al., 2017).

In the United States, for example, there are various land cover mapping programs (regional, national and continental) such as the Gap Analysis Program (GAP), the Utah Gap Analysis Program, the Maine Gap Analysis Program, the Multi-Resolution Land Characteristics (MRLC), and the National Land Cover Database (NLCD).

Earlier mapping programs developed for the continent of Africa include: the Central Africa Regional Program for the Environment (CARPE), the South Africa National Land-Cover Database Project (SANLCDP), the Global Land Cover 2000 (continental), and the Africover (regional). The first high-resolution map classifying land cover types on the entire African continent was released in October, 2017. The map was created using a year's worth of data from the Sentinel-2A satellite. At a resolution of 20 m per pixel, the map comprises 180 000 Sentinel-2A images representing 90 terabytes captured between December 2015 and December 2016. The map was developed under ESA's Climate Change Initiative (CCI) Land Cover project (European Space Agency).

The various methodologies and protocols for mapping land cover have been developed according to the specific characteristics of each area, the availability of satellite imagery and classification technologies, and the objectives of the end product. No one methodology is perfect or suitable to all situations. The problem in producing land cover maps using remote sensed data has been first to create an appropriate methodology using effective techniques to produce the thematic map, and second to develop a procedure to adequately verify its accuracy.

In Angola, because of the civil war (1975 to 2002), there is a general lack of updated information. Only in the last decade has research resumed across the country. The general lack of updated maps in Angola is a concern that has already been clearly established and is a frequently acknowledged concern (Chisingui, 2017). Land cover maps are no exception. There is also no land cover mapping program for the Angolan context to date, although some limited studies have been done in an attempt to present solutions to the problem of land cover mapping in the Angolan context:

- *Land-cover mapping of the Angolan territory, using MODIS satellite images*, by Cabral (2007)
- *Application of remote sensing techniques to thematic mapping: the case study of Lobito region (Angola)*, (Chiquete, 2012),
- *Land-cover mapping derived from satellite images. Case study: Humpata municipality, Angola*, by Vela (2015).
- *Land cover change in Huila* (around Lubango) by Chisingui (2017).
- An on-going project by SASSCAL (2016) to map the vegetation cover in the province of Huila.

The land-cover classification by Cabral (2007) was done from MODIS images using a decision tree method. A fitogeographic map of Angola and Landsat images were used to acquire a sample set both to calibrate and validate the classification.

The case study in the Lobito area (Chiquete, 2012) was done using MODIS, TM and ETM+ images. A set of training samples, acquired from a combination of land survey and high definition images, was used in a maximum likelihood classification. Validation was done with another set also acquired from a combination of land survey and high definition images.

In the study by Vela (2015), a land cover map of the municipality of Humpata was created from Landsat 8 images taken during the dry season. The training set was chosen using visual analysis, identified from four different regions of interest (ROI) in the study area. Using three bandwidths (B3, B4, B5) a supervised classification was done using a maximum likelihood classifier. Validation was done by comparing the supervised classification with a manual (visual) classification of the same area. The area of each of the seven land cover classes in the supervised classification was compared to the area of each land cover class in the manual classification.

The resulting land cover map describes seven land cover classes of the municipality of Humpata, at a scale of 1:250.000. It lacks, however, the identification of the herbaceous wetlands present in the area, which might have been identified by using a rainy season image as well as a dry season image.

In this study by Vela (2015), both the calibration and the validation of the supervised classification depend very much on the photo interpretation of the technician. The calibration areas were chosen by visual interpretation and not by field observations. The validation process was done by comparing the supervised classification to a manual/visual classification of the same area (produced by photo interpretation). Thus, both calibration and validation of the classified map are based on the analyst's *a priori* knowledge of the study area combined with the visual interpretation of the images. It is based on the assumption that this interpretation is 100% accurate. Any conclusions resulting from the study therefore are hinged on whether the interpretation is accurate or not. This approach, on the one hand, reduces or eliminates the time and expenses necessary for field work, but on the other hand may considerably reduce the quality/accuracy of the results. Possible errors in the photo interpretation during the calibration process may go undetected and even be erroneously validated during the validation process for lack of verification on the ground.

In the study on land cover change in the area around Lubango by Chisingui (2017) the various images were classified using a supervised maximum probability classification. The training areas were selected by photointerpretation in combination with many field visits. Although the sample design is not described clearly in the study, a confusion matrix presents the accuracy results of the 2010 image classification. Thus, it is not clear, what the basis for the accuracy assessment is.

The SASSCAL (2016) project aims at producing a plant and vegetation assessment in the region and at elaborating the regional vegetation database and vegetation maps. The preliminary vegetation map was derived from MODIS data, covering the period from 2010 to 2014. The image processing produced fourteen vegetation classes. The Remote Sensing data was then combined with data from the field surveys, in order to classify the vegetation units and validate the vegetation map (SASSCAL, 2016). This study, however is purely limited to vegetation and therefore does not include any classification of urban areas, cultivated areas or barren areas as part of the maps that were produced.

Some of the previous land cover research done in Angola, has been limited by the low spatial resolution of the data used or by the kind of classifier used. MODIS images (with a 250-1000 m pixel) are not well suited to distinguishing smaller areas of land cover. Using images with that low spatial resolution can lead to significant increase in the margin of error - especially in cases of mixed land cover.

In many of the previous land cover studies done in the Angolan context a Maximum Likelihood Classifier (MLC) was used. Literature shows however, that there are many different classifiers available that can produce an increased accuracy, including the Decision Tree Classifier (DTC) (Otukei and Blaschke, 2010). A study by Asamoah et al. (2018) shows that compared to the Maximum Likelihood Classifier (MLC) the DTC produced a better representation of the LULC classes of the study area with an increase in accuracy of more than 6%.

However, a study by Li et al. (2014) shows that insufficient or less representative training samples leads to greater classification accuracy discrepancies than classification algorithms themselves. This leads to another keys issue with the studies done in the Angolan context. Because of the very limited road access in the country, samples both for training and for validation have, in the past, been primarily acquired from high definition images. This must have been in an effort to acquire sufficient samples and guarantee a sample with sufficient representation, as suggested above. However, this approach can lead to another issue – the quality of the samples.

In cases in which training samples are acquired solely from high definition images, information concerning the land cover classes is limited to bibliographic review and the analyst's *a priori* knowledge of the area. Even if isolated field work is done in the area to ascertain the general land cover, with subsequent samples being extracted from high definition images, there is no guarantee that the sample pixels selected from high definition images do in fact represent what is on the ground. The lack of specific available knowledge concerning land cover in Angola and the general lack of access makes this issue even more relevant. Suppositions as to what is being observed and interpreted in high definition images can very easily be mistaken and entire classes can even be overlooked. Such is the case in the study by Vela (2015) in which a classification of the whole municipality of Humpata, Angola is done with no mention of the existing wetlands.

Also due to lack of road access in the country, most accuracy assessments of previous land cover maps have been limited to comparisons with other maps or reference data acquired from high definition images as opposed to ground observations. In some cases, both training samples and validation samples have been primarily acquired using high definition images. Therefore, the validation and conclusions drawn from these studies rely very heavily on the photointerpretation of the analysts. The issue that arises therefore is how to acquire sufficient number of training samples with ample representation to guarantee high enough accuracy results without compromising the quality of the samples. The same goes for the acquisition of validation samples which should also be selected in order to strongly support the conclusions of any given study.

The specific limitations and/or deficiencies presented in previous studies done in the difficult Angolan context leave much space for improvement, especially in terms of the methodology used to produce land cover maps. Although it is obvious that some efforts have been made at presenting options for land cover mapping in Angola, there is still a definite need for the development of a detailed methodology (including a reliable and practical validation process) to produce consistent and repeatable land cover maps. This would include outlining the procedures from the acquisition and processing of the satellite imagery, to the training and method of classification, as well as an adequate accuracy assessment.

1.1 Objectives

The main objective of this study is to present a methodological procedure for creating land cover maps from satellite images using a decision tree algorithm – a procedure applicable in the Angolan territory but with the goal of being extended to other similar contexts.

Assessing land cover maps on the ground has proven to be one of the more difficult tasks given the Angolan context, which provides extremely limited access and large study areas. In previous research done in Angola, assessment has been limited almost exclusively to comparing the resulting land cover map with other maps or with data acquired from high definition images. One of the objectives of this study is to present an approach that highlights an accuracy assessment based on fieldwork observations to acquire quality reference data and, so, overcome the limitations described in the previous section.

Another goal is to create and validate a classification model to classify the land cover in the study area and possibly in other areas with similar characteristics. During this process, the goal is also to select a training dataset that can be used to calibrate classification model but that can also be used to calibrate other future models.

1.2 Methodology Overview

Thematic mapping from remotely sensed data is typically based on an image classification which is a method by which labels or class identifiers are attached to the pixels making up a remotely sensed image on the basis of their characteristics (Pal and Mather, 2001). Any classification begins with choosing a suitable classification system and acquiring a sufficient number of training samples which are prerequisites for a successful classification (Lu and Weng, 2007).

The primary purpose of classification is to describe the structure and relationship of groups of similar objects. Land cover classification thus requires the definition of land cover class boundaries, which should be clear, precise, possibly quantitative, and based upon a set of objective criteria (Yang et al., 2017). The classification may be one that seeks to group together cases by their relative spectral similarities (unsupervised) or that aims to allocate cases on the basis of their similarities to a set of predefined classes that have been characterized spectrally (supervised). In each case, the resulting classified image may be treated as a thematic map depicting the land cover of the region (Foody, 2002). These two broad types of classification procedures can be used as alternative procedures or combined into hybrid methodologies (Al-Doski et al., 2013). Besides supervised and unsupervised, image classification approaches can also be grouped as parametric and non-parametric, or hard and soft (fuzzy) classification, or per-pixel, subpixel and per-field (Lu and Weng, 2007).

This study features a methodology for land cover mapping of a temperate region in the southern highlands of Angola, with limited general information and very limited road access. The major steps of image classification may include determination of a suitable classification system, selection of training samples, image processing, feature extraction, selection of suitable classification approaches, post-classification processing and accuracy assessment (Lu and Weng, 2007).

2 – STUDY AREA

The study area is centered on the northern part of the municipality of Humpata, which is the most western municipality in the province of Huíla, Angola. It covers an area of about 2000 km^2. The western edge of Humpata, along the Chela Escarpment, defines part of the border between the provinces of Huíla and Namibe (Fig. 2.1), in the southwestern part of Angola. The municipality of Humpata is bordered on the northeast and southeast by the municipalities of Lubango and Chibia, respectively (Huíla Province) and on the northwest and southwest by the municipalities of Bibala and Virei, respectively (Namibe Province).

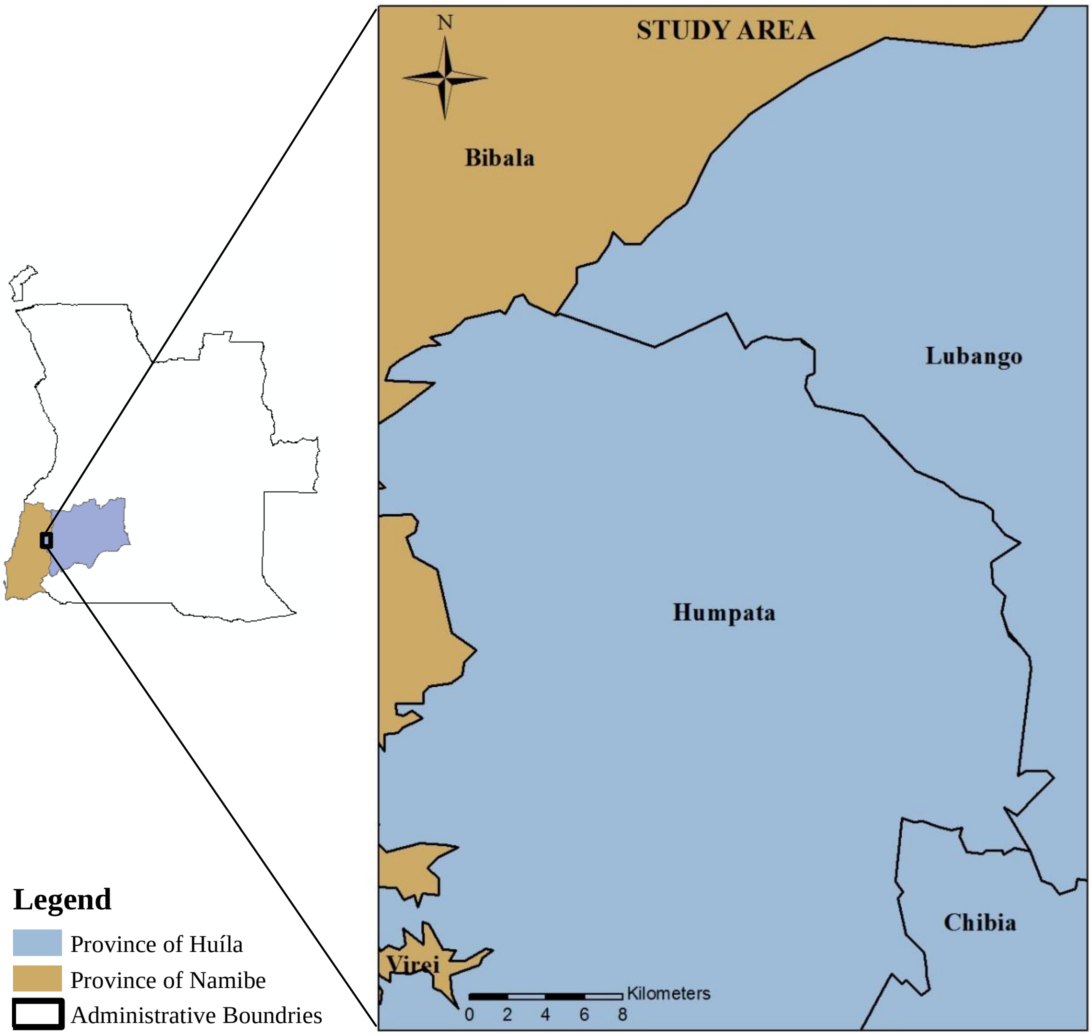


Figure 2.1 – Location of the study area with administrative boundaries (source: IGCA)

2.1 Geomorphology and Geology

The orographic system of Angola may be divided into three regions: the coastal region, attaining 500 m; the mountainous region, superior to 500 m, and sometimes surpassing 2000 m; and the highland plateau, a vast and monotonous peneplain with altitudes between 1000 and 1700 m (Hedberg e Hedberg, 1968). The *Serra da Chela*, or Chela Escarpment, rises up between the province of Namibe and the province of Huíla, forming a high plateau which is defined by an abrupt, sharp slope on its western limit and a more gradual slope on its eastern limit.

The steeper slope corresponds with the Great Escarpment of Southern Africa (Beetz, 1934). The Great Escarpment is one of the most important physical features of southern Africa (Beernaert, 1997). It is a dividing range crossing Angola, Namibia South Africa, Swaziland, Lesotho, Mozambique and Zimbabwe, which separates the high plateaus of the interior of the southern horn of Africa from the coastal lowlands (Appiah and Gates, 2005). Different sections of this U-shaped escarpment are known by different names, such as the Drakensberg in South Africa and the Kaokoveld in Namibia (Clark et al., 2011) and the *Serra da Chela* and *Serra da Mocaba* in Angola. In Angola (in the southwestern province of Cunene) the Great Escarpment becomes sharply defined in the horizontal Otavi-Chela quartzites and limestones of the *Serra da Chela*. Northward, it becomes progressively steeper and higher, attaining an altitude of 2,200 m on the edge of the Humpata Plateau, west of Lubango (Beernaert, 1997). On the eastern side, the Chela Group is formed by the Humpata Plateau with 2280 m of maximum altitude which emerges from an extensive plain to its east having an average altitude of 1800 m. On its western side, a vigorous escarpment, with slopes reaching altitudes above 1100 m, overlooks a littoral platform with an average altitude of 600 m (Pereira, et al., 2011).

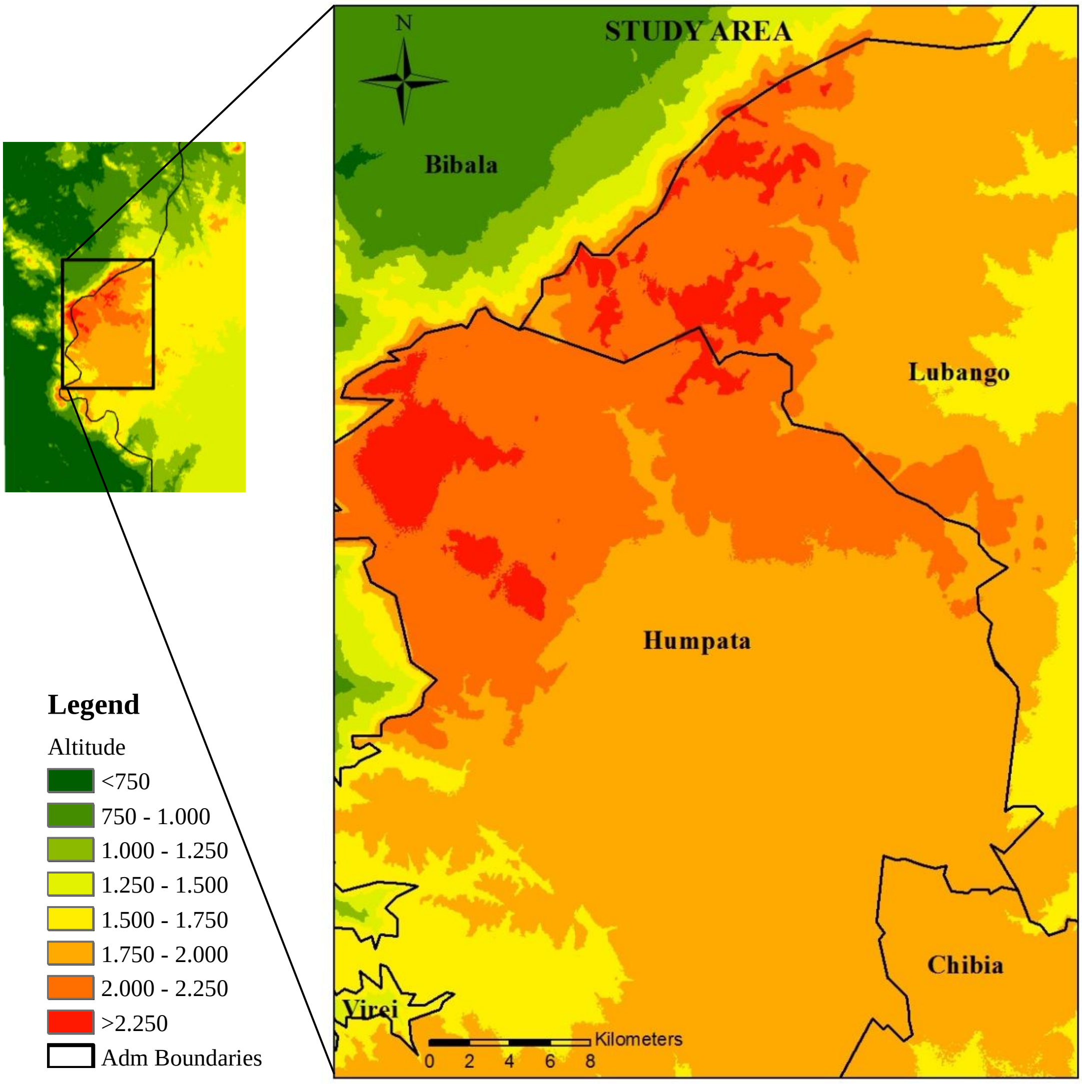


Figure 2.2 – Altimetry in study area

Although the study area includes part of the slopes of the escarpment to the northwest and west, most of the study area is located at altitudes between 1750 m and 2250 m and includes some smaller areas with an altitude exceeding 2250 found along the escarpment (Fig 2.2).

The name Chela Group is a general term for the widespread volcanic sedimentary sequence, recognized by Correia (1976) in the region of Humpata (Pereira et al., 2011). The plateau of the *Serra da Chela* is for the most part covered by fairly horizontal sediments, consisting of grits, sandstones, quartzites, breccias, conglomerates and limestones with intercalated sills and intruded dykes of dolerite. The Chela formation can be correlated with the Otavi and Nama formation of Namibia, the Transvaal system of South Africa, the Serie Schisto Calcaire of the Lower Congo, and the Lower Kundelungu system of Katanga (Beetz, 1934). The Chela Group rests in a major unconformity over a sialic basement of undifferentiated Precambrian age and is subdivided into four formal lithostratigraphic units: from the base to the top, (1) the Tundavala Formation, mainly arenitic with a local basal orrhoconglomerate; (2) the Humpata Formation with a 200-300 meters thick pile of acidic volcaniclastics; (3) the Bruco Formation mainly arenitic and (4) the Cangalongue Formation, lime-iutitic exhibiting a red face towards its upper surface (Correia, 1976).

The northwest part of the study area includes part of the Bibala municipality (formerly known as Villa Arriaga) which lies on the coastal plain at the base of the Chela Escarpment. This inselberg plain developed on Precambrian granites (Beernaert, 1997). The greater part of the study area lies on the narrow Humpata plateau (Plateau V, Jessen) near Lubango. It includes a horizontal quartzite and limestone area on the eastern edge of the marginal chain and on the western edge of the plateau is formed by the Great Escarpment, or the *Serra da Chela* (Beernaert, 1997).

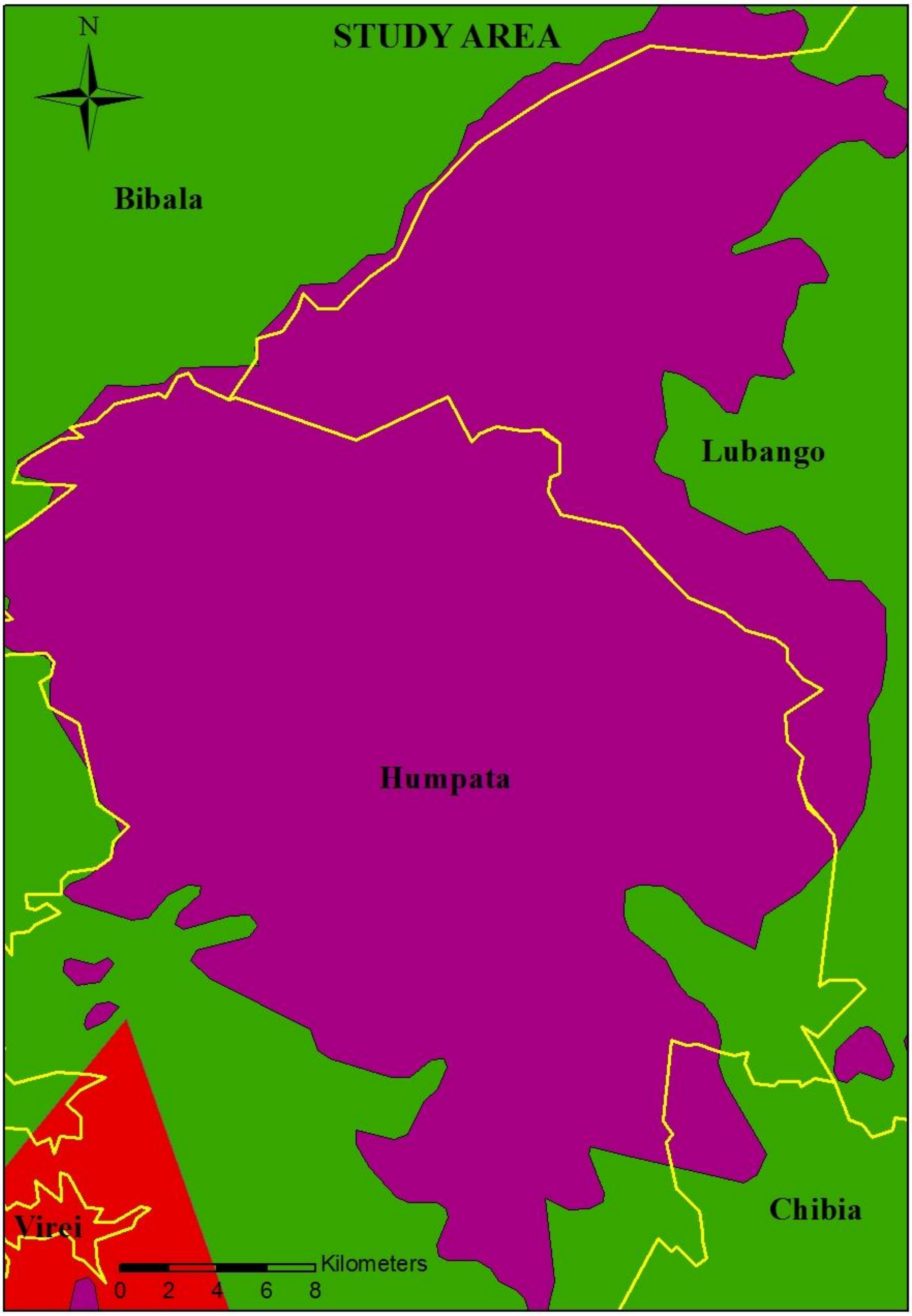
The northeastern part of the study area includes a section of the city of Lubango, which lies on the vast peneplain of granite and gabbros (Main Plateau IV, Jessen). The Main Plateau consists mainly of Paleoproterozoic igneous rocky outcroppings, associated with the Eburnean igneous cycle, which intrude an older complex unit with metamorphic and igneous rocks that may already belong to the Archean (Pereira et al., 2011).

2.2 Vegetation

The study area is included in the large Miombo Ecoregion which covers over 3.8 million km^2 in central and southern Africa, extending from the west coast in Angola to the east coast in Mozambique and Tanzania. It includes all or part of 11 countries – Angola, Namibia, Botswana, South Africa, Zimbabwe, Zambia, Democratic Republic of Congo (DRC), Mozambique, Malawi, Tanzania and Burundi (Timberlake and Chidumayo, 2011).

Much of the this extensive ecoregion is on the ancient African plateau with an altitude of 800 to 1250 m above sea level, but in the east the ecoregion transcends the escarpment and elements of the ecoregion can be found in the east African coastal zone, at 200 to 300 meters of altitude (Timberlake and Chidumayo, 2011). One of the region's main characteristics is the presence of large expanses of rolling savanna woodland on a gently undulating plain, interspersed with grassy drainage basins (Timberlake and Chidumayo, 2011).

This large ecoregion can be further divided into sub-ecoregions. In the study area the three sub-ecoregions (belonging to the larger Miombo Ecoregion) present are the Angolan Montane Forest-Grassland Mosaic, the Angolan Miombo Woodland, and the Angolan Mopane Woodland (Fig. 2.3).



Legend

Administrative Boundries

Ecoregions

Angolan Miombo Woodlands

Angolan Montane Forest-Grassland Mosaic

Angolan Mopane Woodlands

Figure 2.3 – Ecoregions in study area (source: WWF SARPO 2003)

It is an area of transition between the region of the Mopane Woodland (or Mutuati Bush) and the Miombo Woodland (or Panda Bush), but at the heart of the study area on the Humpata plateau is the Angolan Montane Forest-Grassland Mosaic (Fig 2.3).

2.2.1 Angolan Montane Forest-Grassland Mosaic

The Angolan Montane Forest-Grassland Mosaic comprises a number of small montane forest patches surrounded by grasslands and protea savanna in the west central highlands of Angola (Fig. 2.4).

Figure 2.4 – Grassland on Humpata plateau

The montane forest, sometimes referred to as “coffee forest” or “rain-and-cloud forest”, occupies a relatively small area in Angola, consisting of a narrow strip between the altitudes of 350 and 1000 meters (in the valleys and sheltered slopes of the subplateau area), running roughly parallel to the coast between 11º30 S to 7º50 S (Shaw, 1947). The forest patches are restricted to the deep ravines or remote valleys of the highest mountains in the Huambo and Cuanza Sul provinces and an area of

Afromontane forest mosaic further south, on the *Serra da Chela* (WWF, 2008a). On the slopes of the *Serra da Chela* (Fig. 2.5), at altitudes of 1600 – 2000 m, there are fragments of Montane Brushwood which belong to the Angolan Montane Forest Grassland Mosaic. The term *brushwood* is used instead of *forest* in the absence of trees more than 12 meters in height and not having more than two strata of vegetation, or only rarely of a third (Shaw, 1947).

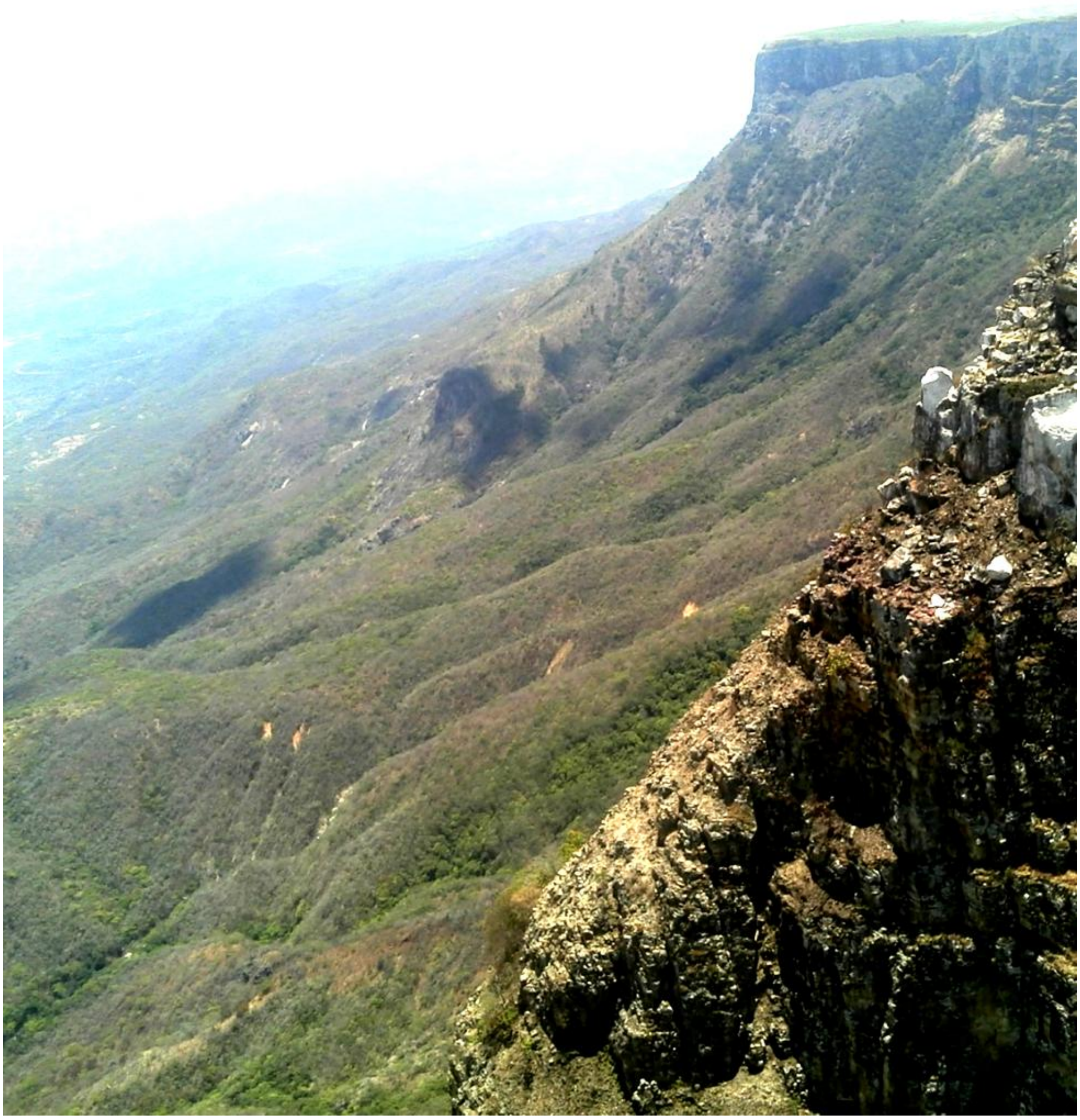


Figure 2.5 – Fragments of montane brushwood on the slope of *Serra da Chela*

The summits (Fig. 2.6) above the montane forest zone usually support a very sparse xerophilous vegetation of steppe or stony desert type (Shaw, 1947).

Figure 2.6 – Steppe vegetation on summit above *Serra da Chela* slopes

2.2.2 Angolan Miombo Woodland

About 3/5 of the total area of Angola, at 900-1700 m of altitude, is covered by woodlands and savanna-woodlands of Brachystegia spiciformis, Isoberlinia angolensis and Julbernardia paniculata, species which are dominant in the western part of the highland plateau (Hedberg e Hedberg, 1968). ‘Miombo’ is a colloquial term used to describe these central, southern and eastern African woodlands dominated by the genera *Brachystegia*, *Julbernardia* and/or *Isoberlinia;* three closely related genera from the legume family (White, 1983 apud Campbell 1996).

As mentioned above the Angolan Miombo Woodlands is one of the ecoregions in the larger Miombo Ecosystem that covers much of eastern and southern Africa.

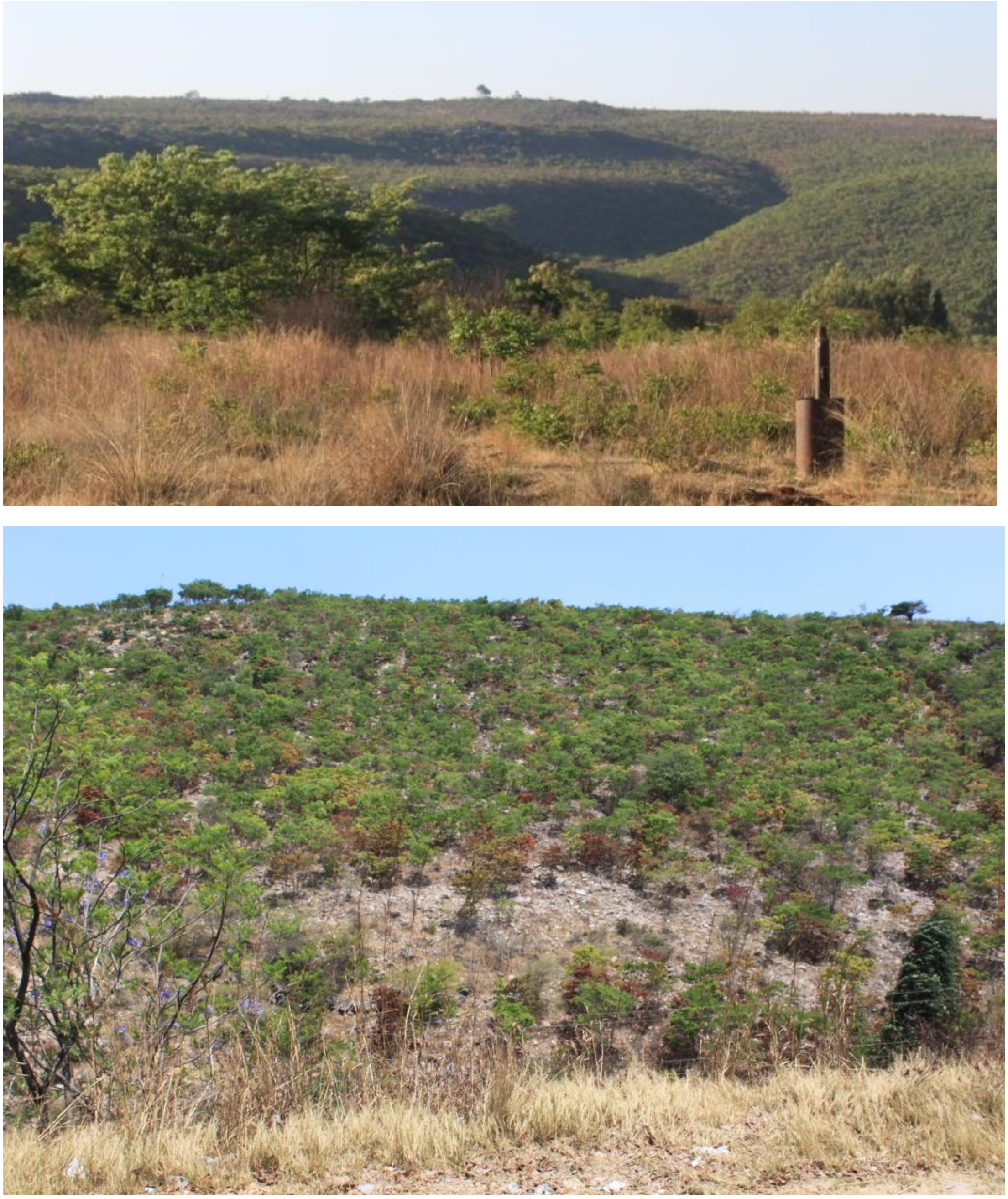


Figure 2.7 – Miombo on the hills around Lubango

Over most of its range, the mature undisturbed miombo exists as woodland, a closely growing, deciduous, nonspinescent tree cover. It generally occurs on geologically old, nutrient-poor soils in the uni-modal rainfall zone. The shrub layer is variable in density and composition. The ground cover varies from a dense coarse grass growth to a sparse cover of herbaceous grassland (Fig. 2.7). Fires are also a characteristic feature of miombo woodlands (Campbell, 1996).

2.2.3 Angolan Mopane Woodland

In contrast to the Miombo Woodland that is generally found on lighter-textured, nutrient poor, well-drained soils on the African Plateau, Mopane woodland is mostly confined to lower-lying areas containing clay and nutrient-rich soils (Chidumayo, 2010).

Figure 2.8 – Mopane Woodland with Chela Escarpment in the background

The Angolan Mopane Woodlands are located in northern Namibia and southern Angola. Mopane trees (*Colophospermum mopane*), also known as Mutuati bush, dominate the vegetation in the region (WWF, 2008b).

In Angola the Mutuati ecoregion occupies a roughly triangular tract of the country in the Namibe and Huila provinces, growing at altitudes between 200 and 1500 m (Fig.

2.8). It adjoins the coastal desert of the Namibe province, forming steppes with bushes of only 1 to 2 meters in height, and the base of the Chela Escarpment and extending to the south and south-east, where the vegetation reaches a height of 15 m (Shaw, 1947).

2.2.4 Ecological uniqueness of the area and its relevance

The Montane forest-grassland mosaic, the Miombo and the Mopani meet in the *Serra da Chela* area, a stretch of the Great Escarpment which rises 800 meters above the coastal zone. The Great Escarpment in Angola is approximately 1,000 km long, and is the least studied section in terms of biodiversity (Huntley and Matos, 1994; Dombo et al., 2002; Figueiredo 2010 apud Clark et al., 2011). In the last decade however, some research has been done in the *Serra da Chela* area.

Exploration by Bruyns (2010) in the last decade has brought to light several previously unrecorded species of *Huernia*, increasing the total number of species in this genus from 49 to 52. The most recently recorded, *Huernia humpatana Bruyns*, has only been found on the western edge of the Humpata plateau in the Chela Mountains (Bruyns, 2010). Another new species, the *Psednotrichia perennis* (*Asteraceae, Senecioneae*), was identified during a botanical excursion in 2009 in the montane grassland near *Estação Zootécnica da Humpata* (Bergh and Nordnstam, 2010). Also, a new species of African reed frog, was found in the montane forested gorge in the *Serra da Chela* Mountain range near the village of Humpata and has been identified and described by Conradie et al. (2012).

The narrow band of ‘cloud forest’ (Shaw, 1947) or ‘Angolan Escarpment Woodlands’ (Olson and Dinerstein, 1998) and the associated ‘montane brushwood’

(Clark et al., 2011) is critically endangered. In light of the ecological importance attributed to the area and the fact that it has been identified as being among the most biodiversity-rich areas on earth, it is noteworthy that it still has not been given the protection it deserves (Huntley and Matos, 1994; Dean 2001 apud Clark, 2011).

Of the natural forest ecosystems, the Afro-montane relic and the Podocarpus forests are among the most seriously threatened ecosystems in Angola and require most urgent protection (FAO, 1996; Clark, et al., 2011). Conservation challenges along the Angolan Escarpment vary although the major threats to the forests are deforestation and fires. Charcoal-making is one of the main causes of the deforestation and fires in the woodlands adjacent to larger growing populations such as the city of Lubango (Clark et al., 2011).

2.3 Climate

Angola has a tropical climate with a marked dry season being largely affected by the seasonal movements of the rain-bearing intertropical convergence zone, the northward flow of the cold Benguela Current off the coast, and elevation (Encyclopaedia Britannica, 2018). Besides these major climatic factors, local climates are influenced by altitude, latitude and the distance from the sea (Huntley and Matos, 1994).

There are two main seasons: the rainy season and the dry season. The first is the longest (from the end of September to the beginning of May); the second lasts from the middle of May until the end of September, and is characterized by the complete absence of rain (Hedberg e Hedberg, 1968).

Rainfall in Angola is seasonal throughout the country, the rainy season being the longest, in the northeast lasting from August to May, and the shortest in the south-west lasting from December to March (Huntley and Matos, 1994). In the study area the rainfall is greatest along the *Serra da Chela* and northeast of Lubango (Alvin, 1963). In Humpata the average rainfall is 817 mm while in Lubango and Chibia it is 860 mm and 714 mm, respectively. In the lower Bibala area the rainfall is similar to Humpata at 815 mm (Climate-Data).

According to the updated map of the Koppen classification, Humpata, Lubango and Chibia fall into the Cwb climate type, a subtropical temperate highland oceanic climate, with dry winters (Peel et al., 2007). The average temperature in Humpata is 16.9ºC, while in Lubango (18ºC) and Chibia (19ºC) it is slightly higher due to the lower altitude (Climate-Data). In higher regions, like Humpata, below zero temperatures are frequent in the cold season and frosts are also frequent (Hedberg e Hedberg, 1968).

A small part of the study area includes the municipality of Bibala which includes the slopes of the escarpment towards the north and northwest of Humpata. The climate in this area (most of which is below 750 m a.s.l.) is classified as Aw, tropical savanna (Peel et al., 2007) with an average temperature of 24ºC in Bibala (Climate-Data).

2.4 Urban Areas

The main urban centers present in the study area are (1) Lubango, the capital of the Province of Huíla, (2) Humpata and (3) the smaller village of Palanca (Fig. 2.9). Located at 1790 m a.s.l., on the Huíla plateau, Lubango lies surrounded in part, by the Humpata plateau which rises above it on the southern and southwestern edge. The city of Humpata (1910 m a.s.l.) and the village of Palanca (1880 m a.s.l.) are located higher up on the Humpata plateau.

Figure 2.9 – Main urban areas in study area

2.4.1 History of Urbanization

Although the native Mumwila tribe inhabited the area before any settlements were made, urbanization began with the Portuguese attempts of colonization (Jenkins et al., 2002). At that time, the Portuguese government tried to promote white settlements by attracting farmers to the southern plateau of Angola (Bastos, 2008). The first attempts at direct colonization in the area were made in 1857 with a group of 50 Germans and again in 1859 with a group of a little over a hundred Portuguese. Both attempts were considered unsuccessful (Brito, 1977).

In 1881 the Portuguese government accepted a group of Boers that founded the colony of Humpata (Brito, 1977). Two hundred and seventy Boers arrived in the Huila highlands of Angola in early 1881 and settled at Humpata, on the road leading east to the interior highlands from Namibe (formerly called Moçamedes) (Clarence Smith, 1976). They expelled large numbers of Mumwila from some of the best lands in the Huila highlands. The displaced Mumwila retreated to the mountains and forests, and carried on a protracted guerilla war with the colonialists, lasting into the 1920's. The Humpata settlement was later reinforced with a much larger group of Boers who arrived between 1892 and 1895 (Clarence Smith, 1976).

The attempts of the Portuguese government to colonize the Huila highlands continued with of a group of 250 farmers from the island of Madeira (Almeida, 1885 apud Bastos, 2008). They reached the Huila plateau in January 1885. They were enticed with the opportunity to emigrate because these island dwellers had become renowned for their adventurous spirits (Bastos, 2008), their hard working nature, and adaptability to difficult frontier-like conditions (Brito, 1977). From their perspective Angola was an opportunity as good as any other to make a new start and to make a living (Bastos, 2008). Many more Madeirans followed the original group, until the official end of the Madeiran immigration in 1892. Some of the Madeirans settled in Lubango, others in Humpata and still others founded the colony of São Pedro da Chibia (Brito, 1977).

In 1928, the Boers abandoned the plateau, and from that time on, more and more Portuguese families came to settle in the area. The greatest influx of Portuguese families in the area took place between 1940 and 1970 which increased their population from 8,521 to 31,674 (Amaral, 1978).

With the independence of Angola in 1975, and the departure of many Portuguese from the cities, the abandoned buildings were occupied *en masse* by the inhabitants of the peripheral districts and later on by the new migrants from rural areas who came to the city centre following the civil war (Rodrigues, 2009). In the 1980's rural areas were insecure and people fled to small towns and inland cities as well as to the main coastal cities (Cain, 2007). It is estimated that Lubango (one of the safer cities) had 40 to 50 times as many people in 2000 as in 1940, and ten times as many in 2000 as in 1970 (Robson and Roque, 2001).

Since the end of the war in 2002, migration patterns have become more complex (Cain, 2007). Although in some provinces there have been many internally displaced persons returning to their areas of origin, many others have left the city centers to settle in peri-urban areas around these cities (Cain, 2007).

Currently, new cities and housing districts are being built by the Chinese close to Angola's main cities (Benazeraf, 2014). These new urban centers, like Quilemba, on the outskirts of Lubango, and similar construction projects on the outskirts of other major cities, are meant to address the massive housing shortfall in the country (Benazeraf, 2014).

Within the study area, on the westward side of Humpata's main northbound road there is also one of these development areas although on a much smaller scale than those projected for the Lubango area.

3 – METHODS AND TECHNIQUES

The methodology developed in this study synthesized in figure 3.1, includes four major steps:

1. The definition of the land cover classes;
2. The procedure of acquiring high quality training data;
3. The classification procedure, using a decision tree approach; and
4. The accuracy assessment.

The first phase of the methodology includes the bibliographic research, the use of remote sensing data and the field work in order to determine the land cover classes and, thus the characterization of the study area. The second step describes the procedure of acquiring high quality data for the training set to be used in the next phase. The following phase includes the calibration and validation of the classification operator (the decision tree model) which in the end is applied in producing the land cover map. The final phase is the accuracy assessment of the resulting map, using (1) ground reference data and (2) high definition images reference data.

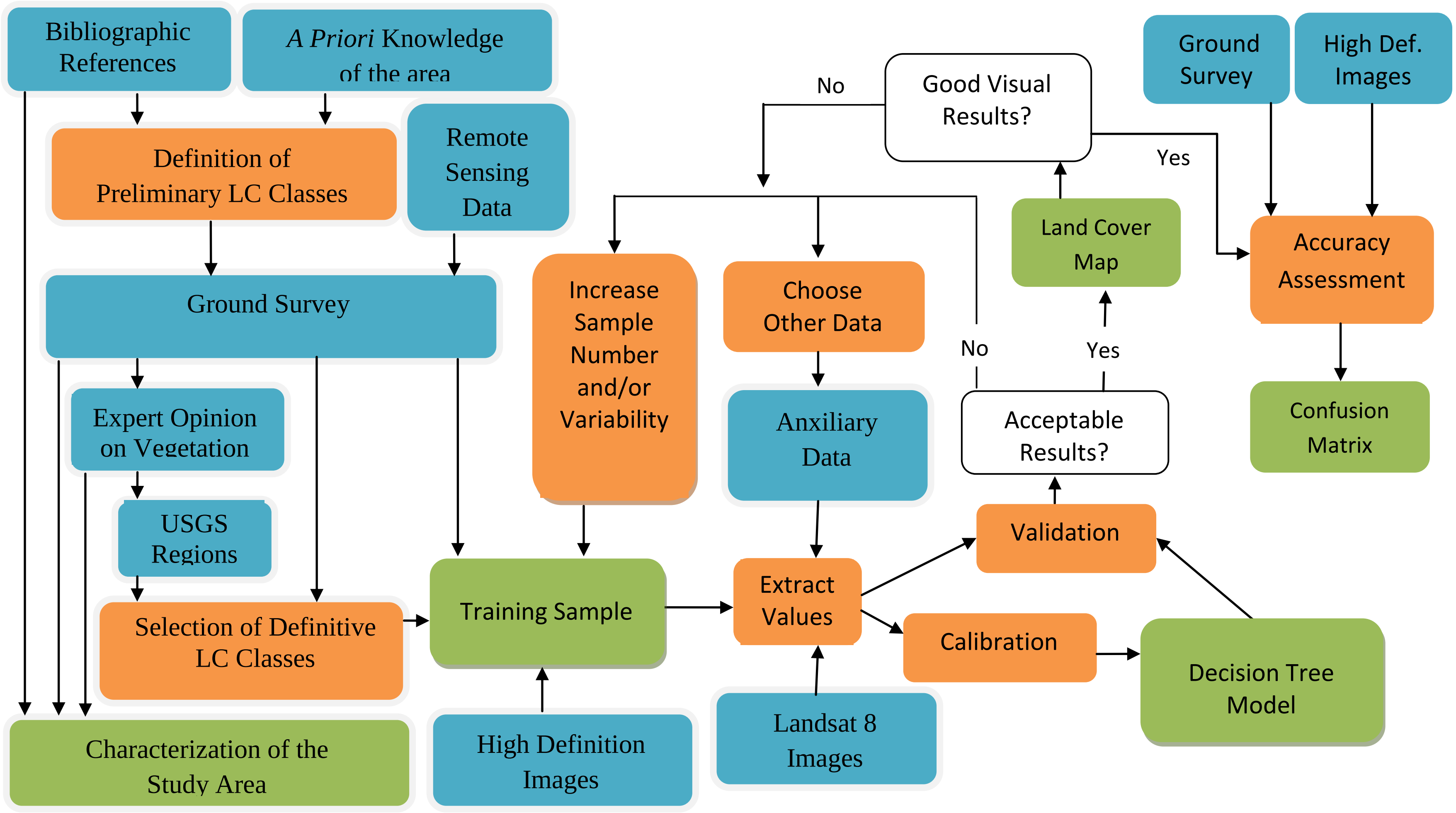


Figure 3.1 – Methodology to be developed and used during this study: - Source; - Process; - Result

3.1 Defining the Land Cover Classes

Before an attempt can be made to classify a satellite image into land cover classes, a land cover classification system or scheme must be defined. A land cover classification system (or scheme) is an abstract representation with the names, codes and definitions of the classes, the well-defined diagnostic criteria (classifiers) used to distinguish different types of land cover, and the relationship among land cover classes (Yang et al., 2017).

First of all, any classification scheme should be mutually exclusive and totally exhaustive; any area to be classified should fall into one and only one category or class (Congalton, 1991a). The failure to exhaustively define classes can result in substantial error which may also pass undetected in the assessment of classification accuracy (Foody, 2002).

Anderson (1976) also stipulated that the classification system should be applicable over extensive areas and that comparison with future land use data should also be possible. If possible, it is very advantageous to use a classification scheme that is hierarchical in nature. If such a scheme is used, certain categories within the classification scheme can be collapsed to form more general categories. This ability is especially important when trying to meet predetermined accuracy standards (Congalton, 1991a). Also according to Anderson (1976), the classification system should be suitable for use with remote sensor data obtained at different times of the year.

Since failure to try to understand the classification scheme from the very beginning results in a great loss of time and much frustration in the end (Congalton, 1991a), defining the classification scheme would be the first and most important step in the whole classification process. To guarantee that it would be informative, exhaustive and separable (Jensen, 1996 apud Lu and Weng, 2007) the definition of the classification system was divided into a series of steps: (1) analysis of bibliographic information on the area and *a priori* knowledge of the area to define preliminary land cover classes, (2) visual analysis of satellite images (Landsat 8, Google and Bing images), (3) ground survey, and consultation of expert opinion and (4) definition of final land cover classes.

3.1.1 Preliminary Land Cover Classes

Before heading out for the ground survey, preliminary land cover classes were determined using the limited *a priori* knowledge of the area and bibliography concerning the area (vegetation, urban areas and agriculture).

Based on *a priori* knowledge of the area certain land cover classes were obvious: (1) urban/built-up areas, (2) agriculture, (3) orchards, (4) inland water, (5) barren or sparsely vegetated areas, as well as the (6) eucalyptus and (7) pine groves. The woodlands/bushlands found in the literature concerning the study area were also listed as possible land cover classes from the beginning: (1) the miombo woodland/bushland, (2) the mopane woodland/bushland and (3) the montane woodland/brushwood.

These ten preliminary class labels were to be validated and possibly adapted and/or improved during the ground survey.

3.1.2 Selection, Processing and Visual Analysis of Remote Sensing Data

Nowadays there is a vast amount of readily available remote-sensing data from which land cover maps can be produced. Barnsely (1999) and Lefsky and Cohen (2003) summarize the characteristics of different remote-sensing data as spectral, radiometric, spatial and temporal resolutions, and also polarization and angularity (apud Lu and Weng, 2007). Landsat, in particular, offers several advantages to these studies, including the long-term digital archive, adequate spatial and spectral/radiometric resolution, and open data policy (Goward et al., 2006; Yang et al., 2003 apud Mantas *et al*., 2016). In this study Landsat 8 images were used in the production of the land cover map.

The LANDSAT program is made up of a group of American satellites that have been launched since the 70's, to acquire high resolution land images in order to allow observation and analysis of the Earth's surface. This experimental program began in 1972 with the launch of NASA's first satellite, ERTS-1 (Earth Resources Technology Satellite), later renamed Landsat 1 (USGS, 2001).

Images began to be commercialized in 1983 and then in 1999 Landsat 7 was launched, with the Enhanced *Thematic Mapper Plus* (ETM+) on board (Fonseca e Fernandes, 2004). This new sensor could provide images with better resolution and accuracy.

On February 11th, 2013 Landsat 8 was launched. This satellite carries two sensors: Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS), with a radiometric resolution of 12 bits. Instead of 256 tons of gray (possible using the

ETM+) it became possible to have 4096. The OLI sensor operates in nine spectral bands and TIRS in two infrared bands (Landsat).

Operational Land Imager (OLI)

- Nine spectral bands, including a pan band:
    - Band 1 Visible (0.43 - 0.45 µm) 30 m
    - Band 2 Visible (0.450 - 0.51 µm) 30 m
    - Band 3 Visible (0.53 - 0.59 µm) 30 m
    - Band 4 Red (0.64 - 0.67 µm) 30 m
    - Band 5 Near-Infrared (0.85 - 0.88 µm) 30 m
    - Band 6 SWIR 1 (1.57 - 1.65 µm) 30 m
    - Band 7 SWIR 2 (2.11 - 2.29 µm) 30 m
    - Band 8 Panchromatic (PAN) (0.50 - 0.68 µm) 15 m
    - Band 9 Cirrus (1.36 - 1.38 µm) 30 m

Thermal Infrared Sensor (TIRS)

- Two spectral bands:
    - Band 10 TIRS 1 (10.6 - 11.19 µm) 100 m
    - Band 11 TIRS 2 (11.5 - 12.51 µm) 100 m

Source: https://landsat.usgs.gov/landsat-8-history

The OLI sensor has a new bandwidth (0.443 µm, band 1) which can be used in combination with other bands to improve research of coastal waters and calculation of aerosols in the atmosphere. Another new bandwidth is band 9 (1.375 µm), that allows better detection of cirrus clouds. TIRS contains two thermal infrared bands:

10.3-11.3 µm e 11.5-12.5 µm. Figure 3.2 shows the bands available from ETM+ (Landsat 7) and from the OLI and TIRS sensors (Landsat 8):

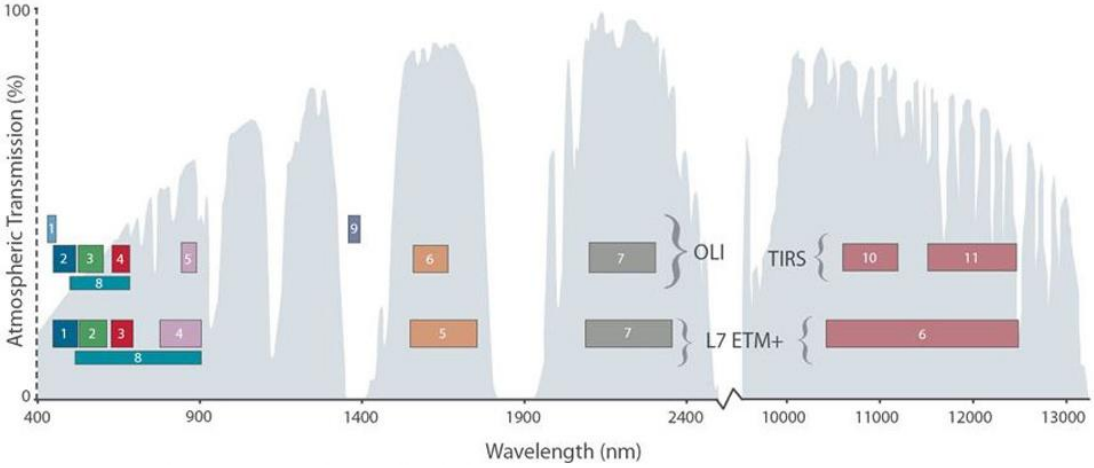


Figure 3.2 – Landsat 8 Spectral Bands and Wavelengths compared to Landsat 7 ETM+ (LANDSAT 8 (L8) DATA USERS HANDBOOK Version 2.0)

Since the study area has two very distinct seasons, two images were acquired (one from each season) to make it possible to identify and enhance the differences between the land cover classes that vary considerably from one season to another. The two most cloud free Landsat 8 scenes (Path 181, Row 70) were selected for the study: one taken during the dry season, on August 22nd, 2013, and one taken during the rainy season, on February 14th, 2014.

The products were delivered in 16-bit unsigned integer format then rescaled to the Top Of Atmosphere (TOA) reflectance and radiance using radiometric rescaling coefficients provided in the product metadata file (MTL file) using:

Radiance: $L_\lambda = M_L Q_{cal} + A_L$ (1), where:

L_λ = TOA spectral radiance (Watts/(m2 * srad * μm))

M_L = Band-specific multiplicative rescaling factor from the metadata (RADIANCE_MULT_BAND_x, where x is the band number)

A_L = Band-specific additive rescaling factor from the metadata (RADIANCE_ADD_BAND_x, where x is the band number)

Q_{cal} = Quantized and calibrated standard product pixel values (DN)

Reflectance $\rho\lambda' = M_\rho Q_{cal} + A_\rho$ (2), where:

$\rho\lambda'$ = TOA planetary reflectance, without correction for solar angle. Note that $\rho\lambda'$ does not contain a correction for the sun's angle.

M_ρ = Band-specific multiplicative rescaling factor from the metadata (REFLECTANCE_MULT_BAND_x, where x is the band number)

A_ρ = Band-specific additive rescaling factor from the metadata (REFLECTANCE_ADD_BAND_x, where x is the band number)

Q_{cal} = Quantized and calibrated standard product pixel values (DN)

TOA reflectance with a correction for the sun angle is then:

$\rho\lambda = \frac{\rho\lambda'}{\cos(\theta_{SZ})} = \frac{\rho\lambda'}{\sin(\theta_{SE})}$ (3), where:

$\rho\lambda$ = TOA planetary reflectance

θ_{SE} = Local sun elevation angle. The scene center sun elevation angle in degrees is provided in the metadata (SUN_ELEVATION).

θ_{SZ} = Local solar zenith angle; $\theta_{SZ} = 90° - \theta_{SE}$

The multispectral bands (2, 3, 4, 5, 6 and 7) from both the rainy season image and the dry season image were used to compute several products including Principal Component Analysis (PCA), Normalized Difference Vegetation Index (NDVI).

Using the colour composites and NDVI's of the Landsat 8 images from the dry season and rainy season, a visual analysis was done to identify homogenous pixel areas which could potentially be tied to specific land cover classes across the study area. Different color composites of each image were used to highlight different aspects of land cover. Similarities and differences from one season to the other were identified by visually comparing the two images. This variability once again permitted the identification of different homogenous pixel areas, some more easily identified in the rainy season image, and others more easily identified in the dry season image.

Other commercial satellite images, such as Bing images and Google Earth images, with lower spectral resolution and higher spatial resolution were also used for visual analysis. Using these high definition images of the same region and limited *a priori* knowledge of the area, a preliminary and tentative identification of some land cover classes of the different homogenous pixels areas was made. Specific areas of more difficult or uncertain visual identification were also identified during this process and access roads to these areas were considered for the ground survey.

3.1.3 Ground Survey and Expert Consultation

The initial ground survey or field work served a dual purpose: firstly, to validate the preliminary land cover classes and, secondly, to collect reference data for training the classification operator. The first step in the ground survey was to locate accessible/passable roads in the general study area, and more specifically, access roads to the areas of interest previously identified in the images.

The challenge would be to acquire sufficient information across the whole study area that has, generally speaking, limited road access, to identify all the land covers and to guarantee that the reference data would adequately represent the population of the whole study area. The major limiting factor would prove to be the road access. Some areas which originally seemed to be accessible (or close enough to accessible roads) proved to be much harder to get to from the access roads or, the roads themselves proved to be less passable than anticipated. Besides the lack of actual roads, another difficulty was the bad conditions of the existing roads and lack of bridges at river crossings. Having little previous knowledge of the road conditions and little available information concerning the roads themselves (lack of maps and signs), an incredible amount of time was spent “exploring” and discovering where roads led and if they were even passable.

Within the city of Lubango, there are many good roads, around the city (dirt and asphalt), as well as, along the base of the mountains surrounding the city. The main road leaving Lubango heading northwest to the Tundavala area and the escarpment has actually been completely paved in the years since the Landsat images were taken.

Crossing the general study area is one main asphalt road (Estrada Nacional 280) that begins in the city of Lubango, climbs up to the Humpata plateau and continues roughly south-east towards the *Leba* escarpment. This same road then winds down the famous *Serra da Leba* switchback heading west into the province of Namibe and finally reaches the coast. The dirt road heading northward out of the city of Humpata branches into two: one heading north-northeast to the Neves Dam area and the other heading north-northwest to the "*Estação Zootecnica da Humpata*" and then east towards the *Alto Bimbi* and the escarpment (Fig. 3.3).

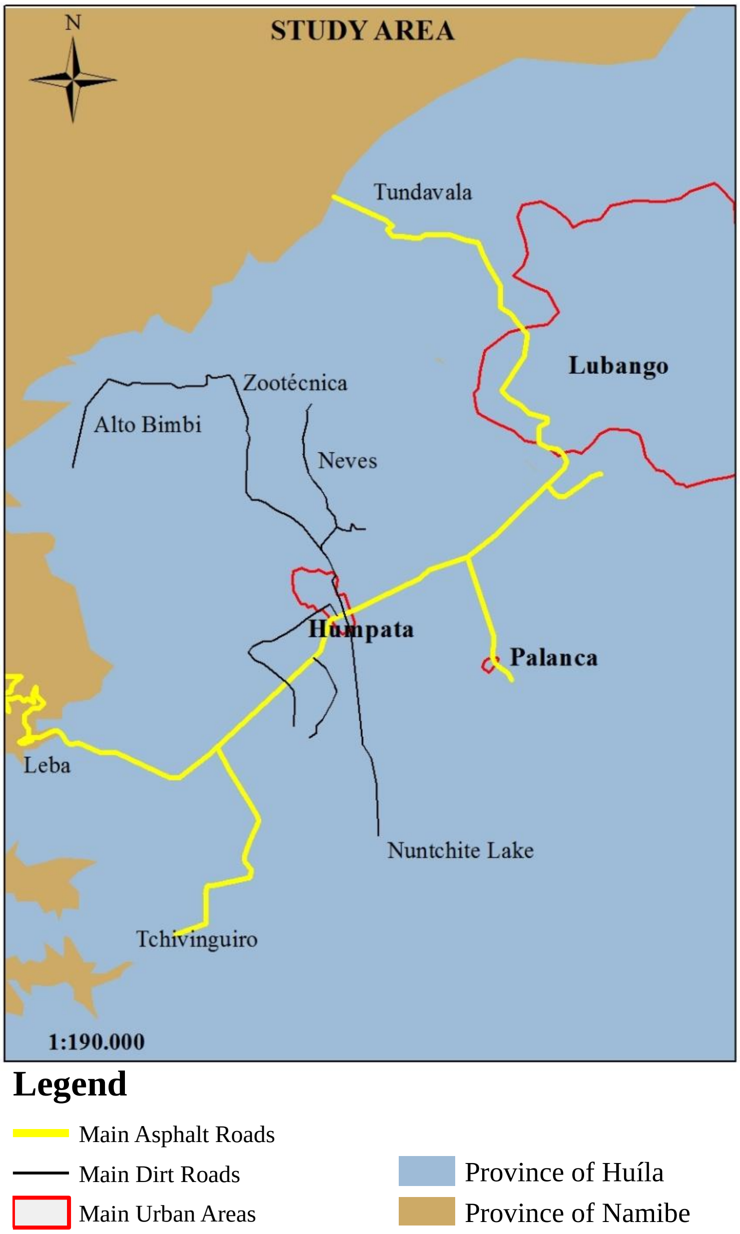


Figure 3.3 – Main roads in study area

To the south of Humpata, there is a network of dirt roads which provides access to agricultural lands, as well as one main dirt road that heads south to Nuntchite Lake and further on towards the township of Jau (in the municipality of Chibia). About halfway between the cities of Humpata and Lubango, along the main asphalt road, a recently repaired asphalt road branches off roughly south towards the rural village of Palanca. Another asphalted, but extremely potholed, road heads southeast from the main road roughly halfway between Humpata and the *Leba escarpment* and later turns southwest to Tchivinguiro (Fig. 3.3).

There are many other tracks and trails on the Humpata plateau besides the main dirt roads and many other asphalt roads within the city of Lubango. There are also many dirt roads in and around the city of Humpata, especially bordering agricultural plots, many tracks or trails that crisscross the area south of the main asphalt road (all dirt roads). The ones presented in Fig. 3.3 however, are the main ones that can be accessed without a guide.

During the rainy season the dirt roads can become impassible because of the mud and flooded rivers (Fig. 3.4).

Figure 3.4 – Flooded areas during rainy season

Even during the dry season most f the dirt roads are only passable using a 4x4 vehicle. Much of the study area was simply inaccessible because of the extremely bad condition of the roads or lack of roads altogether. Many of the dirt roads were extremely potholed and some had deep ravines across them (Fig. 3.5); other roads have no bridges at the river crossings.

Figure 3.5 – Roads in Neves Dam area

Some of the rivers however are intermittent and they are either dry or at least reduced to enough of a trickle during the dry season that four wheel drive vehicles can drive across without any difficulty. In the rainy season, however, without bridges these rivers can become a risk to cross. Other areas along the dirt roads become swampy causing vehicles to get stuck in the mud.

On one occasion during the field work, a river was crossed in the morning and, even though it didn't rain in that exact area, the river became flooded and impassable by the afternoon. It was then necessary to find a local guide who knew of an alternative route back.

Another problem in many of the more remote areas is the lack of cell phone coverage. Therefore, besides the fact that some areas can become impassable, in the event of becoming stranded, there is no way of getting help. For these reasons, most of the field work along the dirt roads could only be done during the dry season months.

So, in much of the study area, there are no defined roads. The local people use a network of unidentified tracks (Fig. 3.6) that crisscross the landscape.

Figure 3.6 – (1) Unmarked track across the grassland, (2) log river-crossing

Again, these can become impassable during the rainy season because of swampy areas and the lack of bridges (Fig. 3.7). However, independent of the season or month of the year, the most often encountered problem along these tracks is that of getting lost. Without a local guide, it is impossible to know where you are going or how to get back. There is also no way of knowing which tracks are actually passable.

Figure 3.7 – (1) Swampy area along dirt road and (2) dry river bed crossing

Another issue that was encountered during the field work was actual safety in remote locations. In general, the Angolan population is very welcoming, but on certain occasions situations arose that made it evident that visiting such remote locations alone may not be such a wise idea (especially for a woman). From then on field work in the more remote areas had to be done with at least one additional person (preferably male).

All these conditions and limitations considerably reduced the geographic area that could realistically or safely be visited, as well as limited the months of the year in which field work could be done without major risks to the safety of self or vehicle.

Although the time spent “exploring” was considerable it was well worth the effort. This process made it possible to validate and more accurately define the preliminary land cover classes by comparing previous/preliminary knowledge with the reality on the ground. Three hundred and seventeen sites were visited and catalogued along the passable routes. These selected sites were points at boundaries between different land cover features (Fig. 3.8) and points in the middle of a given feature. At each site a record was made of the GPS coordinates, photographs were taken (including date and direction) and a short description was made of the land cover observed and a land cover classification (label) was attributed.

Figure 3.8 – Different land covers observed near the *Leba* escarpment: miombo, agriculture and grassland

Whenever possible, sites were visited in both the dry season and the rainy season in order to register information showing significant seasonal differences, especially where vegetation was an issue.

Some vegetation areas were visited and discussed with experts to determine the specific classification of vegetation that should be attributed. This expert opinion helped to make the connection between the vegetation studied in the literature and that observed on the ground. It may have been possible to suspect that there were different vegetation types and perhaps, even visually observe some of the different reflectance in the images, but without this expert opinion it would have been impossible to make an adequate assessment of the different vegetation types.

At this point the goal was to get an overview of the existing land cover in the study area, to validate and better define the preliminary classes and tie them to the homogenous areas of pixels previously identified in the Landsat 8 images. Although a preliminary visual analysis was done, the visual analysis of the images was an ongoing and continuous process, comparing the satellite images with the ground observations. A deeper understanding of the visual analysis was only possible as more detailed and extensive observations of land cover on the ground were made.

Being able to connect (or not) the homogenous pixel areas to specific land cover classes observed in the field was the starting point to understanding which land cover classes were obviously radiometrically distinguishable (even from simple visual observation) and which ones would require more information to make distinction possible.

Each point registered during the ground survey was entered into Geographic Information System (ArcGIS 10.1) and overlaid on the dry season and rainy season images (Fig. 3.9). Using the information registered at each point it was possible to connect the correspondent homogenous pixel areas of the images to a specific land cover class observed during the ground survey.

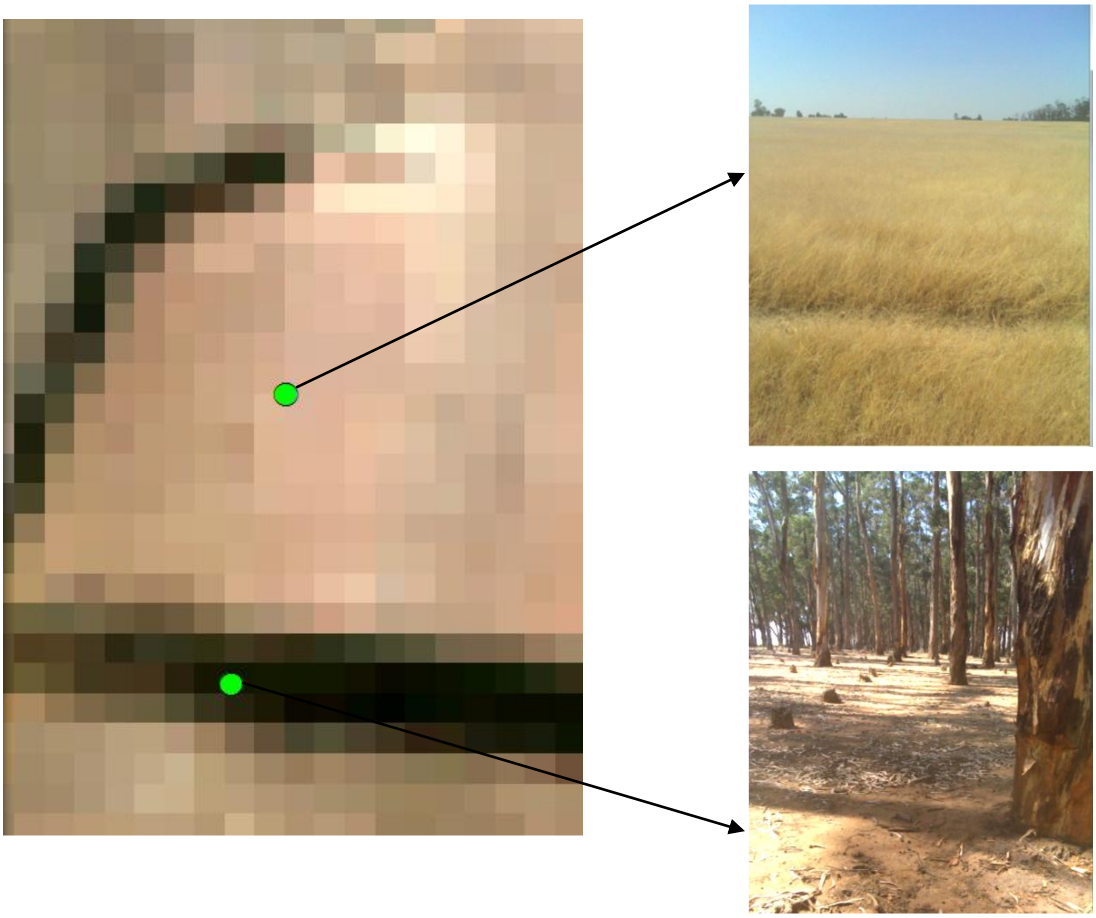


Figure 3.9 – Two points on Landsat 8 dry season image visited in *Alto Bimbi*

A second visual analysis of the images was done after incorporating the ground information. At this point it was possible to observe, with more certainty, the pixel variation within the different land cover classes.

3.1.4. Final Land Cover Classes

The information acquired from all the different sources, from (1) the literature, (2) *a priori* knowledge, (3) the ground survey and from (4) the personal contact with vegetation experts, was compared to the visual analysis of the images. Using this analysis, the land cover observed during the ground survey was grouped into 16 land cover classes.

These in turn were compared to the USGS Africa Seasonal Land Cover Regions Legend which resulted in the final classification scheme containing 8 Level I land cover classes and 16 Level II land cover labels to be used in the classification of the Landsat 8 image.

3.2 Training Data

Once the land cover classes are defined a sample of these classes must be selected to train the classification. Any supervised classification requires site data to calibrate (train) and validate (test) classification algorithms (Muchoney e Strahler, 2002). In the same way, the classification operator used in this study requires high-quality training data from which relations among features and classes present within the data are “learned”. Therefore, a set of training samples representative of the population to be classified must be available to construct an accurate decision tree (Friedl and Brodley, 1997).

In some studies, training samples are acquired from some kind of pre-existing information or data set. In the study done by Zhao et al. (2016) doing land mapping in Chile, researchers had access to the Chilean Forestry Services Database, which contains generalized polygons, labeled with major land cover classes. The training sample was selected using this database in combination with different remote sensing techniques.

In the 2001 NLCD (Homer et al., 2007) training data, was collected from a variety of sources including high-resolution orthoimagery, local datasets, field collected points, and Forest Inventory Analysis (FIA) plot data. In many mapping zones, training data collection took advantage of existing regional land cover maps (e.g. NLCD 1992, Gap Analysis Program (GAP), and National Agricultural Statistics Service (NASS) cropland data) to improve classification efficiency.

In a study by Xian et al. (2013) very high spatial resolution imagery is used in a combination with ground observations. WorldView-2 (WV-2) satellite was used to create the first phase of training dataset. This satellite imagery, with a spatial resolution of 1.8 m, makes it possible to discern different fine scale land cover features. In the second phase, considering the coverage of the remote sensing data and public road access, three areas were chosen for field sample collection.

No matter the method of acquisition the goal continues to be *acquiring training samples representative of the population to be classified*. The source of the training samples however is usually limited by the availability of funding, resources and time as well as, accessibility to sampling areas.

3.2.1 Quality of Training Set

As mentioned above the set of training samples can be acquired or selected by simple visual analysis of high definition images or by ground observations. Some studies also resort to pre-existing reference datasets which are used to assess specific maps (Tsendbazar et al., 2015).

Originally, the accuracy of photointerpretation was accepted as correct without any confirmation (Congalton, 1991a). Many digital classifications are still created (calibrated) and/or assessed (validated) with reference to photointerpretation, as in the case of Vela (2015) and Chisingui (2017). However, the assumption that the photointerpretation is 100% correct is rarely valid and can lead to a rather poor and unfair assessment of the digital classification (Biging and Congalton, 1989 apud Congalton 1991a).

Nevertheless, there has been increased improvement in the spatial resolution of high resolution data available to analysts depending on the area to be mapped. In a study done in Portugal (Mantas et al., 2016) on mapping Impervious Surface Area (ISA) the training data was acquired using aerial photographs with a spatial resolution of 0.5 m as well as high resolution satellite imagery from multiple sources including QuickBird 2 and WorldView 1/2, with spatial resolution between 0.46 and 0.65 m. The samples were manually mapped by the authors upon interpretation of the high resolution satellite and aerial imagery. However, the high resolution imagery used in this modeling may not be available depending on the area (Mantas et al., 2016).

In the present study area, very little updated knowledge concerning the land cover is available and there is no pre-existing dataset that could be relied upon to acquire any kind of training data. Therefore, it was considered that doing field work, would be the optimum way to acquire high-quality ground reference data (even considering the limited access). Field observations were also considered to be most important because of the very little available information concerning the actual land cover.

Visual analysis of available high definition commercial images was used to acquire training data, however, but exclusively in situations which made field observation absolutely impossible.

In an effort to guarantee a higher quality training sample, most of the sampling was limited to the areas accessed and observed by onsite visits during the field work. Two classes however, one because of its temporary nature and one for complete lack of road access had to be sampled mostly using of the high definition images.

3.2.2 Representativeness of training data

To correctly predict an output class, training data must be representative of classes and their subclasses and be sufficient in number to allow pattern recognition to occur (Muchoney and Strahler, 2002). In an effort to guarantee the representativeness of the sample, classes that were identified in different regions of the study areas were sampled in each region. This was an attempt to guarantee not only an extensive geographic spread of the sample for each class but also a representation of different geological formations underlying them and different altitudes characterizing each class. This however, was not possible for classes that were limited to specific areas. Samples were also taken in the middle and on the edges of class areas to guarantee that they would be representative of the different land cover classes as well as intra-class variability.

3.2.3 Number of Sample Pixels (Training Data)

A sufficient number of training samples and their representativeness are critical for image classification (Hubert-Moy et al., 2001, Chen and Stow 2002, Landgrebe 2003, Mather 2004, apud Lu and Weng, 2007). If sufficiently representative training samples are used, most algorithms perform reasonably well. However insufficient (less representative) training can cause large accuracy drops in all supervised algorithms (Li et al., 2014).

When using a stratified sampling design, some studies have used a number of samples proportional to those in the land cover class, other studies have used and equal number of samples per land cover class. There are advantages to each approach. A study by Zhu et al. (2016) showed better results when extracting data proportionally to the occurrence of land cover classes than an equal distribution of training data per class. The problem of unbalanced training data was alleviated by extracting between 600 and 800 training pixels per class. In a study by Rodriguez-Galiano et al. (2012) they used an equal distribution by land cover class. This in general is to guarantee the representation of each land cover class.

In this study it was proposed to select the same number of sample pixels per land cover class to ensure that each class was sufficiently well represented in the calibration of the model. As mentioned above sampling was primarily restricted to the areas accessed and observed by onsite visits during the field work. This to a certain extent also influenced the number of pixels obtainable to include in the training set.

According to Li et al. (2014) most supervised algorithms produce satisfactory results when the training samples are sufficient (more than 200 samples per class). The goal in this study was to obtain 250 samples pixels for each land cover class, aiming at a total of roughly 4000 pixels for the complete training set. For some classes however, it was not possible to reach the desired number of pixels. For this reason, a couple of classes were slightly underrepresented in the sample set.

3.3 Decision Tree Approach

As mentioned before, there are many different approaches for classifying satellite images and different pros and cons for each method (Lu and Weng, 2007). In this study a decision tree classifier is used. A decision tree (DT) is defined as a classification procedure that recursively partitions a data set into smaller subdivisions on the basis of a set of tests defined at each branch (or node) in the tree (Friedl and Brodley, 1997).

Decision Tree Classifiers (DTC) have a simple form which can be compactly stored and that efficiently classifies new data (Pal and Mather, 2001). They use a multi-stage or sequential approach to the problem of label assignment. The labeling process is considered to be a chain of simple decisions based on the results of sequential tests rather than a single, complex decision (Pal and Mather, 2003). Sets of decision sequences form the branches of the DT, with tests being applied at the nodes. The leaves (or branch termini) represent labels (Pal and Mather, 2003).

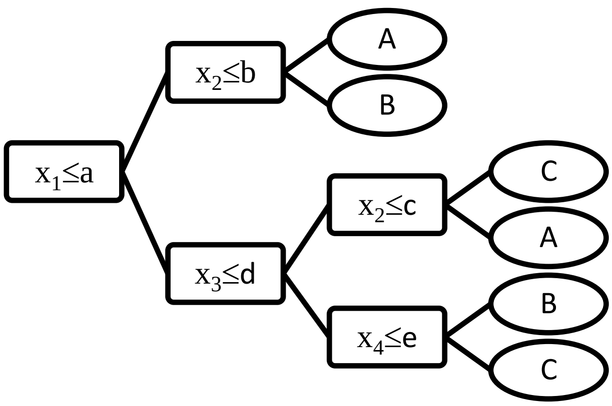


Figure 3.10 – A classification tree with four dimensional feature space and three classes (A, B, C) (Pal and Mather, 2003)

As one of a method of data mining, a decision tree learns from a given data set and formulates explicit rules to classify, segment or make predictions about a target variable (Kuching, 2007). This approach, as all supervised classification approaches, requires a dataset used to train the decision tree. This dataset is selected as representative of the different classes either by doing field work or using high definition images. The quality of the result of the classification algorithm depends crucially on the adequacy of the training data to represent the classes of interest and on the number and nature of the classes relative to the spatial scale of the imagery (Pal and Mather, 2001).

A DT is a non-parametric, per-pixel classifier. Classifiers can be characterized in different ways including whether they use parameters such as mean vector and covariance matrix or not, thus defining whether they are parametric classifiers (such as Maximum Likelihood and Linear Discriminant Analysis) or non-parametric classifiers (such as Artificial Neural Network, Decision Tree Classifiers (DTC), Evidential Reasoning, Support Vector Machine and Expert System) (Prasad et al., 2015). In parametric classifiers a Gaussian distribution is assumed, while in non-parametric classifiers, such as the DTC, no assumption about the data is required (Lu and Weng, 2007).

Parametric algorithms have the advantages of ease-of-use and widespread availability; however, statistical classifiers do not always perform well when the data display non-normal distributions (Franklin and Wulder, 2002). When landscape is complex, parametric classifiers often produce 'noisy' results (Prasad et al., 2015) whereas non-parametric classifiers can handle non-normal, non-homogenous and noisy data sets, as well as non-linear relations between features and classes, missing values and both numeric and categorical inputs (Quinlan, 1993 apud Pal and Mather, 2001) making them especially suitable for incorporation of non-remote-sensing data into a classification procedure (Prasad et al., 2015).

Different studies have compared various classifiers in relation to different aspects of functionality and accuracy (Kuching, 2007; Otukei and Blaschke, 2010; Li et al., 2014; Asamoah et al., 2018). Each classification method has its own merits and demerits (Prasad et al., 2015). However, no classifier is overall advantageous or accurate. Some are more appropriate in certain contexts with given specifications while others perform better in other situations. However, DT techniques have substantial advantages for remote sensing classification problems because of their flexibility, intuitive simplicity and computational efficiency (Friedl and Brodley, 1997). Compared to ANN classifiers (also a non-parametric classifier), DT classifiers are much simpler, involving only the decision of what attributes and pruning methods to use; while ANN classifiers involve decision concerning the type of network, the network architecture, and the initial values of various parameters (Pal and Mather, 2003). Some studies have shown that compared to ML classifiers (a parametric classifier), DT classifiers have higher accuracy results (Pal and Mather, 2003; Otukei and Blaschke, 2010; Asamoah et al., 2018). The question of choosing a classification method continues to be ambiguous, however, because many factors such as spatial resolution of RD, multi-sensor data, availability of different classification software are involved (Prasad et al., 2015).

DT classifiers have been used in a variety of land cover studies with a wide range of spatial resolution, using different satellite imagery. They have been used in producing global land cover maps with 8km and 1 km resolution (De Fries et al., 1998; Hansen et al., 2000), national land cover maps with 250 m (Cabral, 2007) and 30 m resolution (Homer et al., 2015) as well as in many other more specific land cover mapping such as Impervious Surface Areas (ISA), shrubland, crops, and wetlands (Mantas et al., 2016; Xian et al., 2013; Danielson et al., 2016; Hui et al., 2009). The DT algorithm has also been used as an integral part of the NLCD mapping module since 2001 (Homer et al., 2015). The NLCD 2001, 2006 and 2011 were all based primarily on a decision tree classification of Landsat image and several geospatial ancillary datasets.

3.3.1 Calibration and Validation

Once a training set has been selected to be used in the decision tree procedure, it is necessary to define the variables to be used. Many potential variables may be used in image classification, including spectral signatures, vegetation indices, transformed images, textural or contextual information, multitemporal images, multisensor images and ancillary data (Lu and Weng, 2007).

In this study a classification tree is used, in which the target variable is categorical. The decision tree therefore, is used to identify the class. The DT model relies on high quality training samples which in this case, are used as the dependent (target) variable. The satellite imagery becomes the independent or explanatory variables during model development (Mantas et al., 2016). The imagery used in this study includes Landsat-8 OLI multi-spectral bands and the 6 derived PCA's of both the dry season image and rainy season image. Values from band 2, 3, 4, 5, 6, and 7 of the Landsat 8 images from the dry season and rainy season were extracted for each sample pixel. The model was built on Weka using the J48 classifier (Quinlan, 1992). Waikato Environment for Knowledge Analysis (Weka) is a comprehensive suite of

Java class libraries that implement many state-of-the-art machine learning and data mining algorithms (Witten et al., 1999).

The validation of the model outputs relied on traditional accuracy metrics (Janssen and Van der Wel, 1994) applied to independent reference data (equivalent to 10% of the calibration set) (Mantas, et al. 2016). At this point, the validation process showed the accuracy of the resulting DT model. To improve the accuracy of the DT model, the initial training set was adjusted to increase variability and balance the number of sample pixels per land cover class. Another adjustment experimented with was adding the PCA's to the DT model.

3.3.2 Classification

Once adequate results were reached in the validation process, the final DT model was translated into GeoTIFF rasters (one per rule) using a custom python script developed by the authors of Mantas et al. (2016) for ArcGIS. The rasters were then combined into a single image file containing the classified land cover of the area. The product describes the land cover of the area, at a spatial resolution of 30 m for the reference year of 2013/2014.

3.4 Accuracy Assessment

A classification is not complete until it has been assessed. Then and only then can the decisions made based on that information have any validity (Congalton, 1991a). Classification accuracy is typically taken to mean the degree to which the derived image classification agrees with reality or conforms to the 'truth' (Campbell, 1996; Janssen & Van der Wel, 1994; Maling, 1989; Smits et al., 1999; apud Foody, 2002). An accuracy assessment may be undertaken for different reasons: to provide an overall measure of the quality of a map, to form the basis of an evaluation of different classification algorithms or in an attempt to help gain an understanding of errors (Congalton et al., 1998; Hay, 1979; Richards, 1996; apud Foody, 2002).

In thematic mapping from remotely sensed data, accuracy is typically expressed by comparing samples of the classified map with "ground truth" or reference data (Foody, 2002). In the history of accuracy assessments both photointerpretation and ground data have been used to provide this reference data, although photointerpretation is in general considered less accurate than ground data (Congalton, 1991a).

For these assessments, the most commonly used method to represent the classification accuracy of remotely sensed data is in the form of an error or confusion matrix (Congalton, 1991b). The confusion matrix is currently at the core of the accuracy assessment literature (Foody, 2002). The square array of numbers set out in rows and columns express the number of sample units (i.e., pixels, clusters of pixels, or polygons) assigned to a particular class label relative to the actual category as verified on the ground. The columns usually represent the reference data while the rows indicate the classification generated from the remotely sensed data (Congalton, 1991b). The simple cross-tabulation of the mapped class label against that observed on the ground or reference data for a sample of cases at specified locations provides an obvious foundation for accuracy assessment (Campbell, 1996; Canters, 1997; apud Foody, 2002).

Many measures of classification accuracy can be derived from a confusion matrix (Foody, 2002; Stehman and Czaplewski, 1998). One of the most popular is the percentage of cases correctly allocated (Foody, 2002).

If attention focuses on the accuracy of individual classes, it can be achieved from two standpoints (Foody, 2002). Firstly, the "user's accuracy" is the conditional probability of correctly classifying a location given that it has been mapped as class *i* (rows). Secondly the "producer's accuracy" is the conditional probability of having correctly mapped a location given that it is truly class *j* (columns) (Stehman and Czaplewski, 1998). According to Anderson (1976), the minimum level of interpretation accuracy in the identification of land cover categories from remote sensor data should be at least 85 percent and the accuracy interpretation for several categories should be about equal.

Another coefficient that is commonly used in accuracy assessments, and has been suggested as a standard measure, is Cohen's kappa coefficient (Smits et al., 1999 apud Foody, 2002). This measure makes some compensation for chance agreement; the fact that some cases may be allocated the correct class purely by chance (Congalton, 1991; Pontius, 2000; Rosenfield and Fitzpatrick-Lins, 1986; Turk, 1979; apud Foody, 2002)

However, not only the error matrix itself has to be evaluated but also the whole procedure of data collection for the accuracy assessment (Banko, 1998). According to Stehman and Czaplewski (1998), there are three basic components in an accuracy assessment: the response design, sampling design and analysis. Since it is too expensive and difficult to obtain the reference land cover classification for the entire region of interest, statistical sampling has become a critical component of accuracy assessment (Stehman, 2009).

3.4.1 Sampling Design

A ‘sample’ is a subset or portion of the region mapped (Stehman, 2009) and the sampling design is “the protocol by which reference sample units are selected” (Stehman and Czaplewski, 1998). The major issues in sample design are (1) how to select the sample so that it presents an unbiased view of the true population, and (2) how to draw conclusions about the true population from the results of the sample (Cochran & Cox, 1957; Snedecor & Cochran, 1989, apud Holmes et al., 2006). It is important that the sampling design is specified as it can significantly influence the results of an analysis (Friedl et al., 2000; Green, Strawderman, & Airola, 1993; Stehman, 1995, apud Foody, 2002). Selection of the proper scheme or design is critical to generating an error matrix that is representative of the entire classified image (Congalton, 1991b). “Indeed the confusion matrix cannot be properly interpreted without knowledge of the sampling design used in its construction (Maling, 1989; Stehman, 1995; apud Foody, 2002).”

“A statistically rigorous accuracy assessment is one in which the sampling design satisfies probability sampling protocol and the estimates are statistically consistent (Stehman, 2001).” In the overall scheme of an accuracy assessment, the objectives determine the analysis, and the analysis strongly influences the choice of sampling design (Stehman, 2009). Opinions concerning the appropriate sampling design vary greatly and include everything from simple random sampling to stratified systematic unaligned sampling (Congalton, 1991b). The practical reality is that limited resources will require focusing the design on priority objectives, so the key is to choose an adequate design, not necessarily the perfect design, recognizing the strengths and weaknesses of different designs and understanding the trade-offs among objectives and desirable design criteria (Stehman, 2009).

The sampling design chosen for this study was stratification by land cover class. Stratifying by map land cover class and allocating approximately equal sample sizes to each stratum is a relatively common practice in accuracy assessments (Strahler et al., 2006). Strata are groups of pixels constructed such that each pixel belongs to exactly one stratum and the strata form a partition of the population of all pixels. Strata are most often constructed based on the map class of each pixel or based on the spatial location of each pixel (Stehman, 2009). This approach is designed to provide approximately equal precision for estimated user's accuracy of each class. Without stratification, the sample size representing a rare class or small region may be insufficient to precisely estimate accuracy (Strahler et al., 2006).

Stratified random sampling using map land cover to form the strata is one of the most commonly employed sampling designs for several reasons: (1) it is practical, (2) it provides precise estimates of class-specific accuracy, (3) it has the ability to estimate standard errors and (4) it has flexibility to change in sample size (Stehman, 2009). The disadvantage is having to collect ground information for the accuracy assessment at random locations that can only be selected after the classification has been performed. This has a limitation: the accuracy assessment data can only be collected late in the project instead of in conjunction with the training data collection. This causes tremendous increases to the cost of the project (Congalton, 1991a).
The strata to be used in the accuracy assessment of this study were the 16 previously defined land cover classes, which were later on used to classify the satellite image.

3.4.2 Population

An accuracy assessment is based on comparing the map depiction of land cover to the true land cover condition. Typically, ‘ground truth’ is not easily attained. As a consequence, accuracy assessments evaluate the map land cover relative to some higher quality determination of land cover. This higher quality data, referred to as ‘reference data’, is used to produce a ‘reference land cover classification’ that is compared to the map land cover classification (Stehman, 2009). Although no reference data set may be completely accurate, it is important that the reference data have high accuracy or else it is not a fair assessment. Therefore, it is critical that the ground or reference data collection be carefully considered in any accuracy assessment (Congalton, 1991a).

In some studies, researchers choose practicality over the quality of the reference data. In these cases, reference data is obtained from a visual analysis of high definition imagery instead of accessing areas on the ground. The advantage of this approach is that the whole population can be used in the sampling process because it is not limited by accessibility and/or time constraints. The disadvantage however is the uncertainty of the reference data, which in this case depends solely on the interpretation of the technician.

Originally, the accuracy of photointerpretation had been accepted as correct without any confirmation. Many digital classifications are still assessed with reference to photointerpretation (such as Cabral, 2007; Vela, 2015; and Chisingui, 2017). An obvious assumption made here is that the photointerpretation is 100% correct. This assumption, however, is rarely valid and can lead to a rather poor and unfair assessment of the digital classification (Biging and Congalton, 1989 apud Congalton 1991a).

Another approach is to use a reduced population which permits the reference data to be collected on the ground, in accessible areas (Stehman, 2001). In such cases the assumption made is that the accuracy acquired from the reduced population is representative of the accuracy of the full population (Stehman, 2001). Again, this may not be entirely true. The advantage of this approach is the higher quality reference data. The disadvantage is the increased financial cost and time required for sample collection as well as the limitation in terms of accessibility.

Some studies are assessed using reference data acquired from high definition images in combination with field work (Chiquete, 2012) in an effort to guarantee higher quality reference data while also taking into account financial, time and access constraints.

In this study it was decided that all accessible land cover classes would be verified using ground visits to collect the reference data, while the inaccessible land cover classes would be verified using reference data acquired from high definition satellite images.

In this case population representation is reduced because of lack of access (Stehman, 2001). The target population (entire classified image) and sampled population (population accessible for sampling) differ; therefore, any generalization to the target population requires assumptions (Stehman, 2001). It is assumed therefore, in this study, that the accuracy of the target population is the same as the sampled population.

As stated by Stehman (2001), in a probability sampling protocol the inclusion probability (π_u) must be known for all elements in the sample, and the inclusion probabilities must be non-zero for all elements of the population. Therefore, the rigorous statistical inference will apply only to the population in which $\pi_u > 0$.

The reduced population (population to be sampled), used for accessible land cover classes, would be defined by a buffer of 100 meters on either side of the accessible routes within the final map area. In the case of the inaccessible classes the entire population (entire mapped area) would be used.

The final map area was defined at a scale of 1:100.000 to be compatible and comparable to other available maps of the area the topographic and geologic maps. Within this defined area the passable roads were traced, totaling 163.8 km of road in the mapped area of approximately 500 km^2. This included the main paved and unpaved routes (with the condition that they were passable at least during the dry season). Once the accessible roads in the final map areas were traced, a buffer was created of 100 meters on either side of the roads.

The final map area (entire population) is 546,790,419 m^2 equivalent to 6,075,449 pixels. The reduced population (sampled population) has an area of 30,270,906 m^2 equivalent to 33,634 pixels (roughly 5.54% of the entire population).

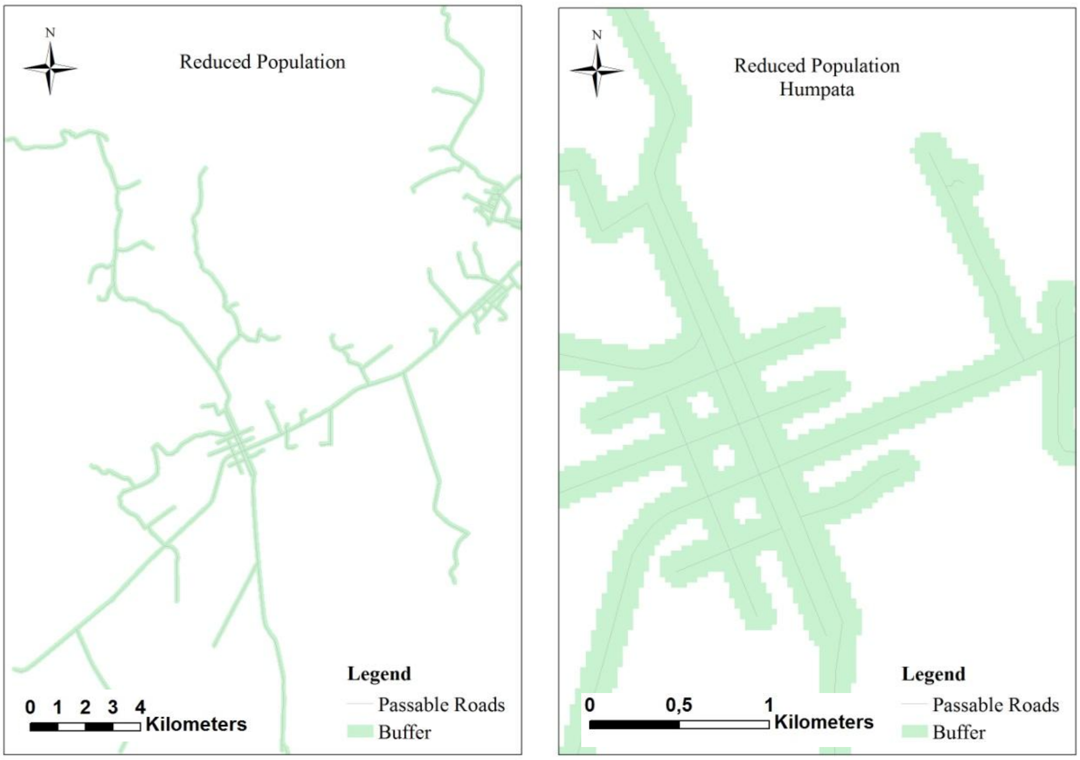


Figure 3.11 – Entire reduced population and reduced population in the city of Humpata

This buffer was used to extract the corresponding pixels from the classified image. This resulted in a raster of classified pixels, 100 meters on both sides along the passable roads. These pixels represent the population in which the sample of reference points (pixels) would be selected, roughly the same amount in each strata or land cover class.

The classified raster of the reduced population was then converted to a polygon shapefile (in the ArcGis project). Using the 16 class attributes the many individual polygons created belonging to each class were reduced to one polygon per class.

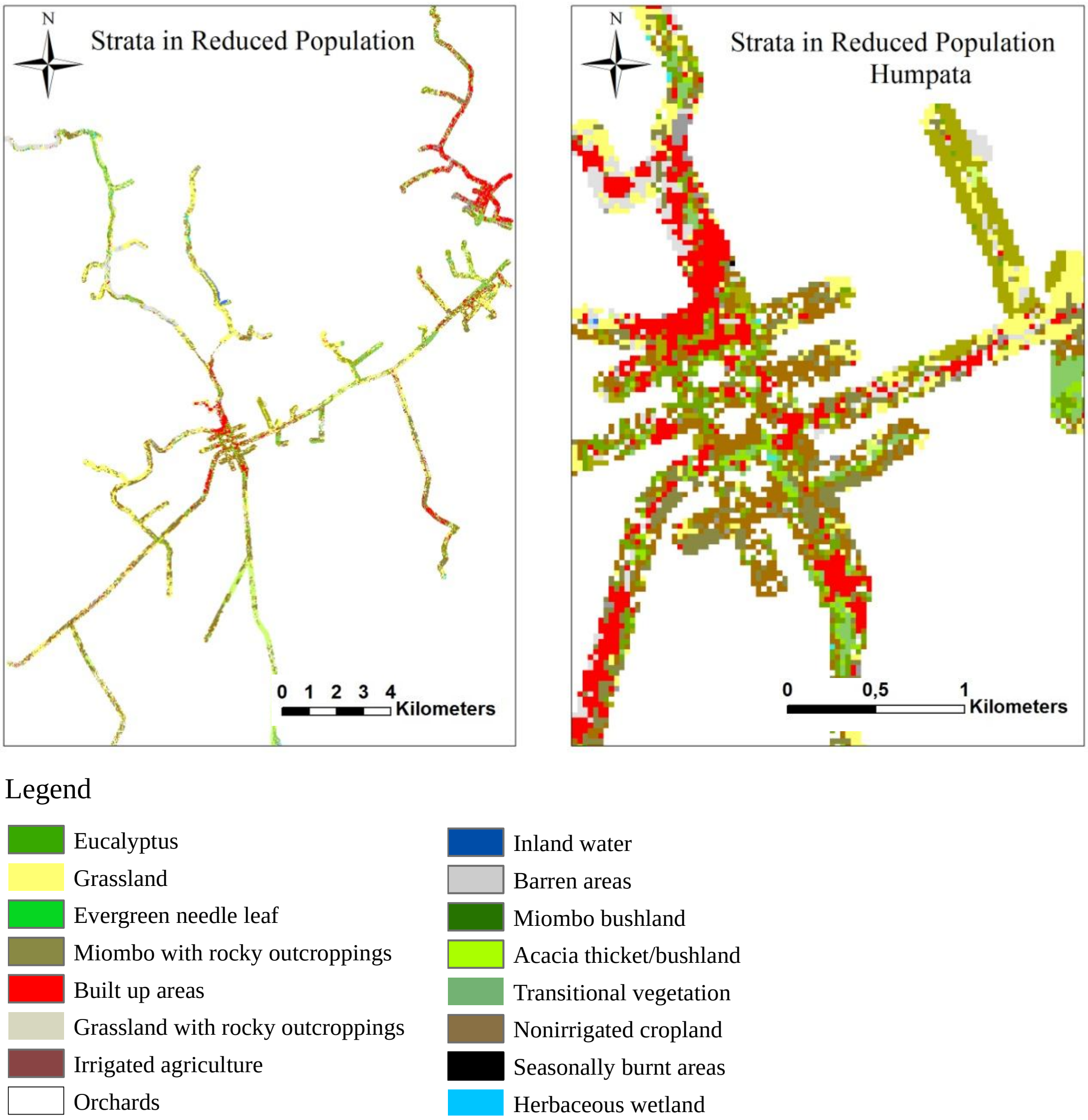


Figure 3.12 – Strata in entire reduced population and strata reduced population in the city of Humpata

3.4.3 Response Design

Stehman (2001) describes the response design as "the protocol for determining the reference classification recorded on each sampling unit," including the recording of primary and perhaps secondary land-cover classes as well as the choosing of the spatial unit on which the assessment is based. In the majority of accuracy assessment applications, both the map and reference classifications are crisp and the descriptive accuracy analyses employ an error matrix to organize the data (Stehman, 2009).

A sampling unit should be at least as large as the minimum mapping unit (Aronoff, 1985 apud Banko, 1998). In this study the map and reference classifications are crisp; that is, they are done based on an individual pixel area of 30x30 m. Therefore, the land cover class was recorded based on what was observed on the ground (for 14 of the land cover classes) and in the high definition images (for the 2 remaining land cover classes) based on their accessibility.

3.4.4 Sample

The number of samples to be used is a compromise between the effort to minimize the costs of field sampling and the requirement of a minimum sample size to be representative and statistically sound (Banko, 1998). The sample size must be selected with care and be sufficient to provide a representative and meaningful basis for accuracy assessment (Hay, 1979 apud Foody, 2002). Each sample point collected is expensive and, therefore, sample size must be kept to a minimum; yet it is important to maintain a large enough sample size so that any analysis performed is statistically valid (Congalton, 1991b).

A balance between what is statistically sound and what is practically attainable must be found (Congalton 1991a). Some authors suggest a minimum of 50 samples per category as a rule of thumb (Congalton 1991a) or 75 - 100 samples (Bank, 1998) while others suggest using a multinomial distribution to calculate the sample size (Rosenfield, 1982 apud Banko 1998). These suggestions vary depending on the size of the area, the number of land cover classes and the objectives of the project.

It was decided that, for this accuracy assessment, 25 sample pixels would be randomly selected per land cover class. However, a total of only 391 sample pixels (1.16% of the reduced population, and 0.006% of the entire population) ended up being used for lack of more pixels in certain land cover classes within the reduced population area. Two similar approaches were used: the first approach was used for the 14 classes that were accessible for verification within the reduced population. For the 2 inaccessible classes a variation of the first approach was used. In each stratum approximately the same number of random sample pixels were allocated, from which reference data would be collected either on the ground or in high definition images.

3.4.4.1 First Approach (accessible land cover classes using reduced population area)

Since the accuracy assessment data used for validating the land cover product must be independent of the data used to train the classification (Stehman, 2009), all the pixels within this reduced population that had been previously used for training the decision tree were eliminated from this population area.

Each sample pixel used for training the decision tree was identified by a point. Using these points a raster of the sample pixels was extracted, converted to polygons, reduced to a single attribute and then converted into one whole polygon. Overlaying this polygon of sample pixels on the road and buffer polygons, it was possible to split and then delete the sample pixels from the reduced population of classified pixels.

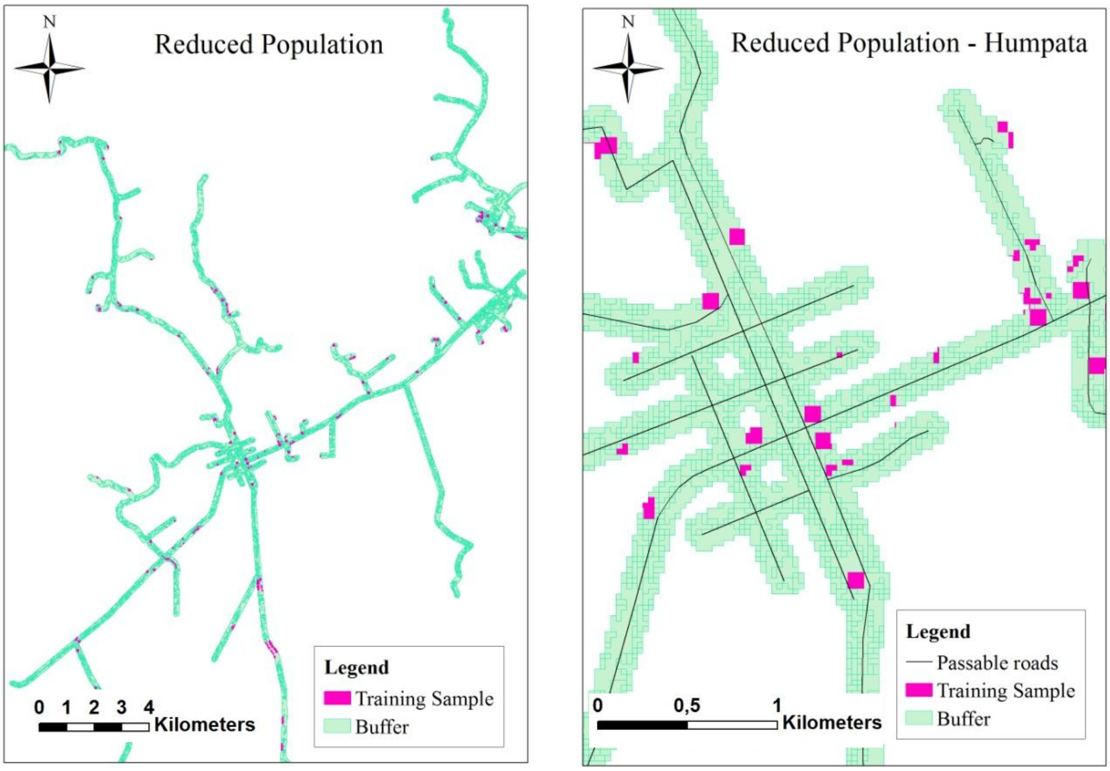


Figure 3.13 – Training samples eliminated in the entire reduced population and in reduced population in the city of Humpata

Thus, the final reduced population used in the accuracy assessment was the pixels along passable roads within the 100 meter buffer on either side, from which pixels previously used for training the decision tree had been excluded.

In the final reduced population twenty-five points (10% of the number used for the training) were randomly selected in each of the 14 accessible land cover classes (polygons). To guarantee that no two points were selected within any given pixel a minimum distance of 50 meters was established between any two given points.

In the reduced population it was possible to randomly select 25 points for each of the 14 accessible land cover classes, except for one in which only 16 were available for selection. The coordinates of the random points within the sample pixels were then extracted along with the respective classified label. These sites were then visited on the ground for verification.

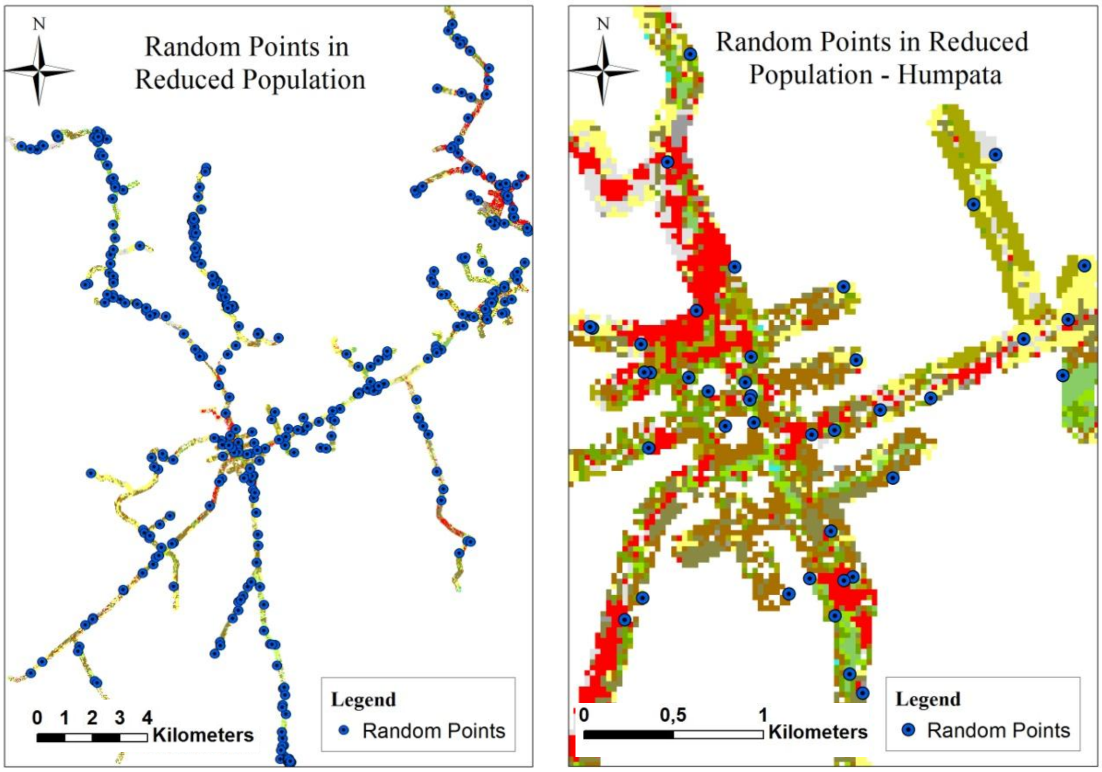


Figure 3.14 – Random points selected in entire reduced population and random points selected in reduced population in the city of Humpata

3.4.4.2 Second Approach (inaccessible land cover using entire population)

Two of the sixteen classes were not included in the first approach of verification: the seasonally burnt areas and the transitional vegetation on the escarpment slope. The seasonally burnt areas, as suggested by the name, are a temporary class which could only be verified in the same year as the Landsat images were taken. Many burnt areas were observed in the different phases of the application of the methodology (from 2015 to 2017) as well as while verifying the other land cover classes (done in July/August, 2017), however these would not necessarily, or likely be, in the same locations as the ones in the classified images (taken in 2013). Therefore, the above approach was considered inadequate for this specific class.

The verification of transitional vegetation on the escarpment slope could also not be done using the above approach because of the road access to the area. The sections of this class accessed during the observation phase and sampling for training the classification were at the Alto Bimbi and Tundavala lookouts, areas which were not included in the final mapped area. The area of the escarpment within the final mapped area is inaccessible by road. Therefore, any pixels in the reduced population (area defined by the passable roads and the buffer) that happened to be classified as this kind of vegetation would be considered most likely to be misclassifications. Had the same approach been used to verify the accuracy of this class, most likely 100% of the selection would be incorrect.

Therefore, a variation of the first approach was used to evaluate the accuracy of these two classes using high definition satellite images. Since verification would be done using high definition satellite images, there would be no limitation in terms of access, and therefore the whole mapped area could be used as the population. So, this area was defined by a rectangular polygon which was then used to extract the pixels of the classified image within the final map area. The resulting raster (of the entire population) was then converted to a polygon shapefile.

Using the 16 class attributes, the many individual polygons created were reduced to one polygon per class. This resulted in one polygon per class, covering the area within the final mapped area. Again, pixels within the population that were previously used for training the decision tree were eliminated from the population area.

In the resulting area twenty-five random pixels (10% of the number used for the training) were selected in the two land cover classes (polygons), the seasonally burnt areas and the transitional vegetation on the escarpment.

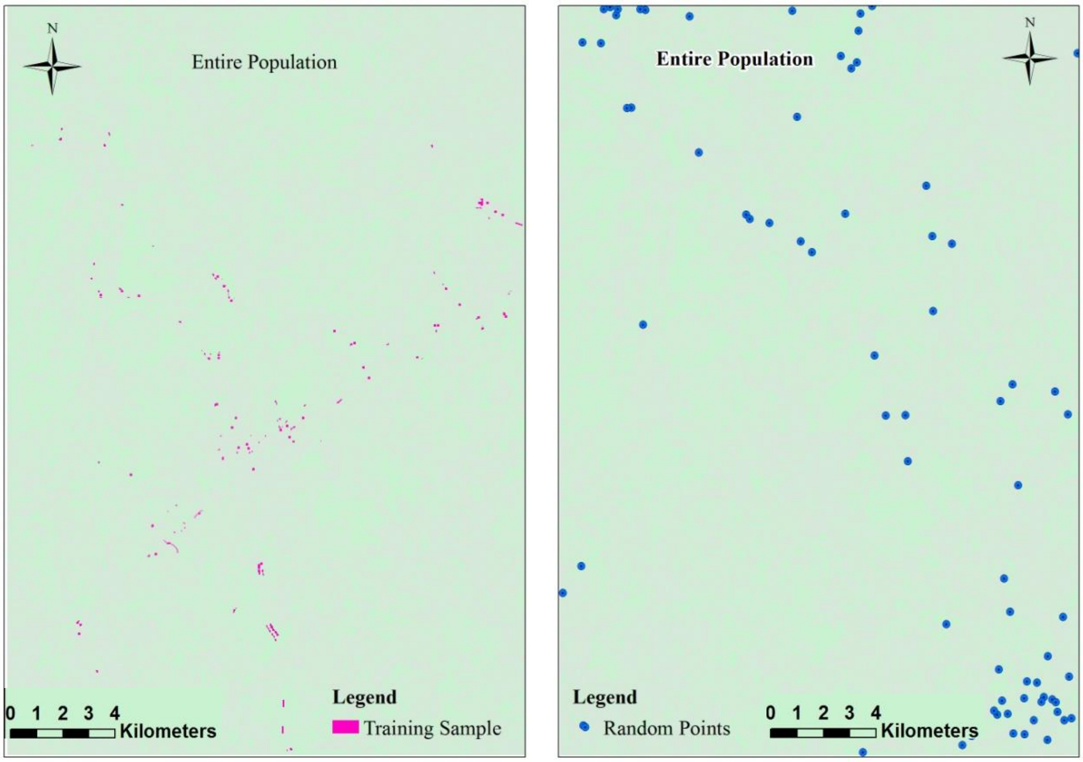


Figure 3.15 – Training samples excluded from population and random points selected from the entire population area

Again, to guarantee that no two points were attributed to the same pixel a distance of 50 meters was established between points. The coordinates of the random points within the sample pixels were extracted along with the respective classification label for verification using high definition images.

3.4.5 Verification of Sample Pixels

The sample pixels of the 14 classes accessible by road were visited on the ground. The land cover within the pixel area was observed and compared to land cover label attributed by the classification. Each sample pixel was then marked as correctly classified or incorrectly classified. In the case of misclassification, the correct classification was registered. During this process other information considered relevant was also noted especially in misclassified pixel areas.

In a few cases additional information was necessary to verify the accuracy of certain pixels. When visited on the ground, some pixels classified as barren/sparsely vegetated areas (in the 2013 image), were found to be urban/Built-up areas (in 2017). Most likely at the time the image was taken these areas had been cleared of original vegetation to begin construction of what has now become an urban/Built-up area.

This problem occurs when the time lapse between the start of the project and the execution of the accuracy assessment is so extensive. This length of time allows for possible changes on the ground, which affect ground reference data. In other words, the ground may change between the time the project starts and the accuracy assessment begins (Congalton, 1991a). This happens especially when using a stratified sampling design, as the ground reference data can only be collected after the actual classification is done.

To verify this situation, the same coordinate was located on Google images of the same year as the Landsat image. Using this extra information the pixels were then marked as correct or incorrect and the extra information was also noted.

The two remaining classes (transitional woodland and seasonally burnt areas) were verified using only Google Earth images; the first because of the complete lack of access and the second because of its temporary/changeable nature. Using extracted coordinates, each sample pixel was located on the Google images. The land cover within the pixel area was observed and then compared to the land cover label attributed by the classification. Each sample pixel was then marked as correctly classified or incorrectly classified. In the case of misclassification, a possible correct classification was attributed and registered. Using Google images however, some classes are difficult to distinguish, (especially between nonirrigated crops and grassland, as well as the different kinds of trees).

Since the seasonally burnt areas are a temporary class, the sample pixels belonging to this class were verified using the Google images (taken on the 13th of October 2013) with the closest date after the Landsat image (taken on the 23rd of August 2013).

4. RESULTS AND DISCUSSION

The key objectives of the methodology described in Chapter III were (1) to enable the overall characterization of the study area by dividing it into clearly defined land cover classes, (2) to acquire a high quality training set to calibrate the DT model, (3) to validate and apply the DT model in classifying the study area, and (4) to validate the resulting classification map. This section begins with the representation of the results reached in each of the four processes.

4.1 Final Land Cover Classes

The initial, and perhaps one of the most fundamental steps, is the careful determination of a mutually exclusive and totally exhaustive classification scheme. The information acquired from bibliographic references and field work, as described in section 3.1, culminated in the creation of an adequate classification scheme for the study area, containing 8 Level I land cover classes and 16 Level II land cover labels.

As mentioned in section 3.1.4, before finalizing the classification system the land cover classes were compared to the already existing USGS Africa Seasonal Land Cover Regions. Three classes were included in the scheme which were not in the USGS Africa Seasonal Land Cover Regions but were thought to be relevant in this study area: the Grassland with Rocky Outcroppings, the Miombo with Rocky Outcroppings and the Seasonally Burnt Areas. Another class, the Montane Brushwood and Transitional Woodland, is a variation/adaptation of the *Montane Forest Class* contained in the USGS Africa Seasonal Land Cover Regions, due to the fact that the vegetation on the escarpment slopes is not exclusively Montane Forest.

For all the other classes an appropriate equivalent class was found in the USGS Africa Seasonal Land Cover Regions.

Table 4.1 – Classification system for study area

Level I Classes	**Level II Classes**
Water	Inland Water
Built-up Areas	Built-up Areas
Barren or Sparsely Vegetated Areas	Barren or Sparsely Vegetated Areas
Herbaceous Vegetation	Grassland
	Grassland with Rocky Outcroppings
Woodland/Bushland	Miombo Bushland
	Miombo Bushland with Rocky Outcroppings
	Montane Brushland and Transitional Woodland
	Acacia Bushland/Thicket
	Evergreen Needle Leaf Plantation
	Eucalyptus Plantations
Wetlands	Herbaceous Wetlands
Cultivated Areas	Irrigated Agriculture
	Nonirrigated Cropland
	Orchards
Seasonally Burnt Areas	Seasonally Burnt Areas

The following table includes a description of each Level II land cover class, which is then followed by a section with a more detailed characterization of each LC class.

Table 4.2 – Classification system with land cover class description: adapted from NLCD 2006

Level I	**Level II**	**Description**
Water	Inland water	Areas of open water, generally with less than 25% cover of vegetation or soil.
Built-up areas	Built-up areas	Included are lands of low, medium, and high intensity with a mixture of constructed areas and vegetation, such as single-family housing units, multifamily housing units, and areas of retail, commercial, and industrial uses.
Barren or Sparsely Vegetated Areas	Barren or Sparsely Vegetated Areas	Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, strip mines, gravel pits, and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.
Herbaceous Vegetation	Grassland	Areas dominated by graminoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.
	Grassland with Rocky Outcropping	Areas dominated by graminoid or herbaceous vegetation, in which rocky outcroppings account for 25% to 50% of the area.

Level I	Level II	Description
Woodland/Bushland	Miombo Woodland/Bushland	Areas dominated by broad-leaved deciduous woodlands, dominated by species of *Brachystegia*, *Julbernardia* and *Isoberlinia* generally accounting for more than 80% of the total vegetation
	Miombo Woodland/Bushland with Rocky Outcroppings	Areas dominated by broad-leaved deciduous woodlands, dominated by species of *Brachystegia*, *Julbernardia* and *Isoberlinia*, with rocky outcroppings accounting for 25% to 50% of the area
	Montane Brushland and Transitional Woodland	Small areas of evergreen montane brushland amongst a greater area of miombo woodland/bushland transitioning to mopane woodland/bushland
	Acacia Bushland/Thicket	Areas dominated by acacia thicket generally accounting for more than 80% of the vegetation
	Evergreen Needle Leaf Plantations	Exotic evergreen needle leaf plantations mostly of pine and cedar
	Eucalyptus Plantations	Exotic eucalyptus plantations

Level I	Level II	Description
Wetlands	Herbaceous wetland	Areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
Cultivated Areas	Irrigated agriculture	Irrigated agriculture are described as areas typically cultivated year-round using some kind of irrigation, especially during the dry season
	Nonirrigated Cropland	Nonirrigated crops are described as areas used for the production of annual crops, such as corn, sorghum, and millet, typically cultivated on a perennial cycle. This class also includes all seasonally tilled land
	Orchards	Areas dominated by the plantation of a variety of fruit bearing trees (pears, oranges, apples, etc), although sections between the rows of trees may also be used for irrigated agriculture.
Seasonally Burnt Areas	Seasonally Burnt Areas	Seasonally burnt areas are described as areas of grassland and/or woodland/bushland burnt during the dry season

4.1.1 Inland Water

The largest water reservoirs in the study area are manmade lakes created by built-up dams. The Neves Dam and the Tundavala Dam are the two dams of notable size. A series of three smaller ones is located along the road from *Estação Zootécnica da Humpata* to *Alto Bimbi*. All of these provide water for irrigation, especially during the dry season.

The water level in the lakes varies from dry season to rainy season and from year to year according to the amount of rainfall. This variation is possible to visualize and quantify in the remote sensing data.

The Neves Dam (*Barragem das Neves*) is located roughly four and half kilometers north of the city of Humpata, at about 1,972 m a.s.l. (Fig. 4.1). The lake occupies about 1 km^2 and provides water for year-round irrigation to the area in and around the city of Humpata.

Figure 4.1 – Lake formed by Neves Dam near the city of Humpata

The water is distributed to local farms by a system of canals allowing for irrigation of crops and orchards year-round (Fig. 4.2).

Figure 4.2 – Canals in the Humpata area

The lake created by the Tundavala Dam (*Barragem da Tundavala*) is located about eleven and half kilometers from downtown Lubango at about 1,994 m a.s.l. (Fig. 4.3). Smaller than the Neves Lake, the Tundavala Lake occupies an area of about 200,000 m^2.

The lake originally provided irrigation for the agricultural lands in the Mapunda, an area in the northwestern part of Lubango. More recently it has also been channeled for use in the construction of the new centrality about 10 km to the east of the lake.

Figure 4.3 – Tundavala Dam

To the west of *Estação Zootécnica da Humpata*, there is a series of three dams at approximately 2,126 m a.s.l., 2,133 m a.s.l., and 2,153 m a.s.l. The relatively small lakes formed by the series of dams (approximately 7,000m^2, 23,000m^2, and 16,000m^2, respectively) provide irrigation to the *Estação Zootécnica da Humpata* located lower down at about 2,075 m a.s.l. (Fig. 4.4).

Figure 4.4 – One of the dams above *Estação Zootécnica da Humpata*

4.1.2 Built-up Areas

Lubango, the capital city of the Province of Huíla, is the largest built-up area in the province (Fig. 4.5). It became a city in 1923, with its first official Urban Plan in 1957. Demographic growth was rapid in the 1960s and 1970s, caused by strong transcontinental immigration, which caused the growth of the urbanized area (Rodrigues, 2009).

Figure 4.5 – Buildings from the colonial period in downtown Lubango

The city was a point of entry for the population coming primarily from Portugal, which was designed by the colonial policies of settlement and development of the area (Jenkins et al., 2002).

Figure 4.6 – One of the main streets in Lubango

Even though most the original built-up areas from the pre-independence era and even many of the more recently projected areas are divided into organized plots, including paved or unpaved projected streets, much of the built-up area surrounding the urbanized central core consists of temporary style adobe homes haphazardly built on whatever unoccupied land that was “considered” or “deemed” as suitable. The strong economic growth of the area during the last decades of the colonial period (60’s and 70’s) made the city a centre of development for the whole area (Fig. 4.6). This became a strong attraction for great numbers of people of a rural origin who moved into the peripheral zones of the city with subsequent development of new shanty towns (Jenkins et al., 2002) most commonly known as the *musseque*.

Although the word *musseque*, comes from the Kimbundo language, indicating “the red sanded areas of the Luanda plateau, different from the fishing villages (. . .) and from the ‘cement city’ of the Portuguese . . .” (Kasack, 1996 apud Rodrigues, 2009) it is now a commonly used and accepted terminology referring to these kinds of squatter settlements around major cities throughout the country (Fig. 4.7).

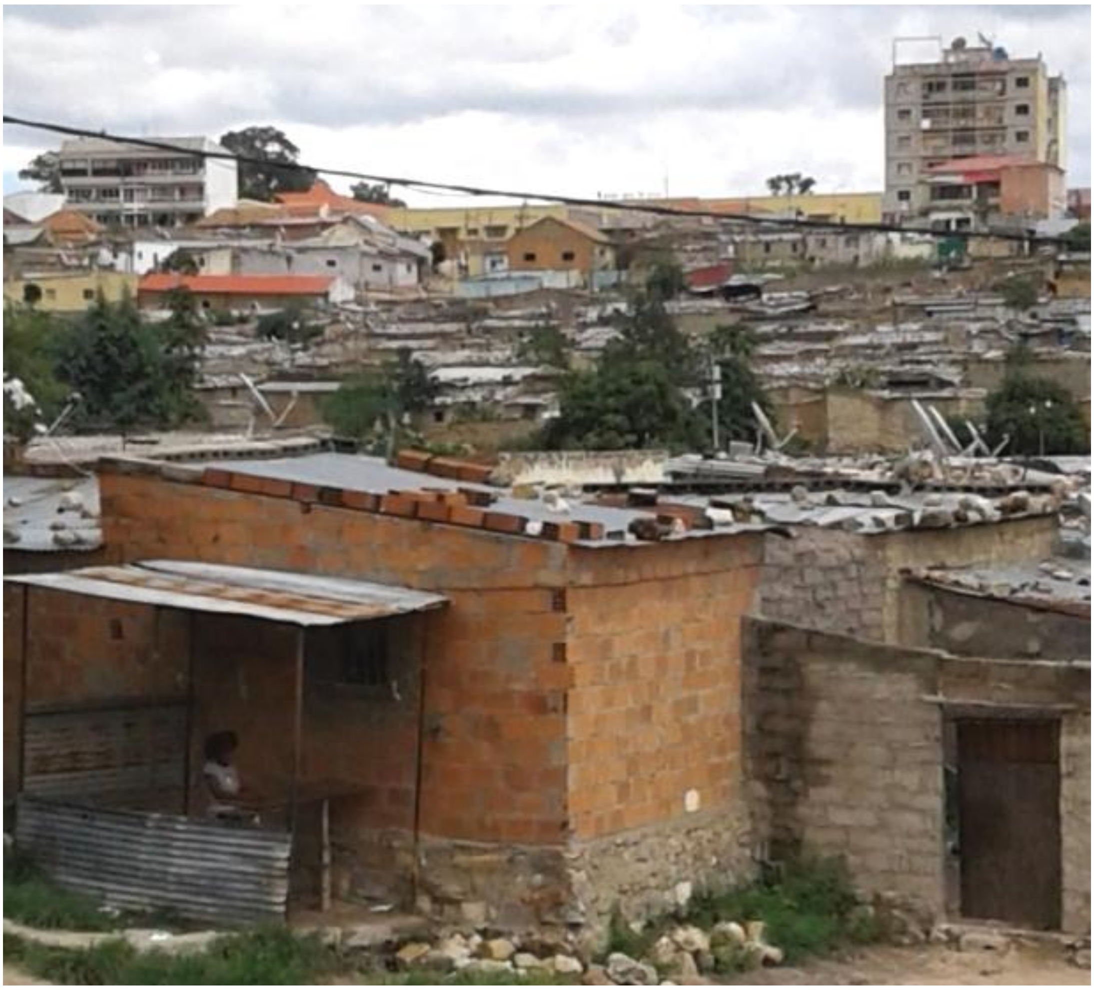


Figure 4.7 – *Musseque* near downtown Lubango

This type of urban planning was actually an organization that started during the colonial period – urbanized centers, where services and the city administration were located and where the majority of the population of European origin lived. The peripheral zones of the cities, were typically a more precarious construction that was “permitted” or tolerated and prevalent, and where asphalt and public services were absent; it was occupied largely by the native population and migrants of rural origin (Rodrigues, 2009). After independence, the increase in precarious construction in the city center and the growth of the peripheral districts became more significant with the migration of the population fleeing the war with South Africa, in 1983. There was an agglomeration of anarchical constructions in the central district that “grew up” on any available ground and along the riverbanks and at road junctions (Rodrigues, 2009).

By 2003 it was estimated that in Lubango, in general terms, unconsolidated construction was prevalent in all districts – 84.8 percent were made up of anarchical constructions and only 15.2 percent of solid buildings in areas with infrastructures (Governo da Provincia da Huila, 2003 apud Rodrigues, 2009). A study on land cover and land use by Holden (2015) showed an increase of 1658% from 1978 to 2002 and another 17% from 2002 to 2010, in the Lubango area.

Figure 4.8 – The city of Lubango in 2017

The population of Lubango increased from 104,847 in 1984 (Robson and Roque, 2001) to 731,575 in 2014 (2014 census) (Fig. 4.8). The increase in population up until 2002 was largely due to the displacement of people fleeing areas of the country more affected by war. After the war, people migrated to the city in search of better living conditions (Cain, 2007).

The increase of population in urban areas and a consequent lack of adequate housing and basic sanitation services throughout cities in Angola, led Angolan government to set the goal of constructing “1 million houses’: 685,000 through self-construction, 185,000 by government contracts, 120,000 by the private sector, and 80,000 by co-operatives (Croese, 2012). As part of this resolution, a total of 100,000 hectares of land around Luanda, Benguela, Namibe, Lubango and Malange were reserved for the program in order to build satellite towns, called ‘new centralities’ (*novas centralidades*) or ‘new cities’ (*novas cidades*) (Benazeraf, 2014). The largest development of this kind in Lubango is the new satellite town in Quilemba with 8,000 new homes built by government contracts with Chinese construction companies (Angop, 2016). There are also some smaller urbanized areas being built in Humpata (Fig. 4.9).

Figure 4.9 – New urbanized area in Humpata

The city of Humpata and the village of Palanca, located up on the Humpata plateau were historically made up of small farm holdings. Since the end of the war (in 2002) there has been a lot of new urban development in both communities. Both Humpata and Palanca have one or two paved streets running through them; the rest are dirt roads. Originally both were agricultural villages, consisting mostly of small farms irrigated by rivers and a man-made canal system. More recently, however, there has been increasing "do-it-yourself" construction, projects by private individuals wanting to move out of Lubango, as well as government-initiated construction projects related to the government funded program of new centralities. This has increased the built-up area in both Humpata and Palanca.

There has also been a continual increase of houses in the surrounding rural areas, mostly rudimentary adobe homes as well as homes built from logs and plastered with clay or mud (Fig. 4.10). Most of these do not have any road access, although some can be reached along dirt tracks.

Figure 4.10 – Rural construction near Humpata

Construction in raw earth is evident on a large scale, in periurban and rural areas (Fig. 4.11). The construction methods follow the ancestral standards, distributed throughout the region of Huila, being built by those belonging to the various representative ethnic groups in the area. Most notable among the construction techniques in earth, are: the adobe, wattle-and-daub and more recently construction of CEB (Compressed Earth Block). The type of soil used to make the adobes is mainly silty-clayed sand (Wachilala et al., 2016).

Figure 4.11 – Rural homes: adobe houses with metal sheeting and grass thatch

These houses may be covered with grass thatch (mostly in the rural areas) or with the same corrugated metal roofing sheets used extensively in the urban areas. Rural homes are usually found in small clusters of two to three huts surrounded by the family's farm (or cultivated land) and often bordered by sisal plants or temporary "fences" made of thorn branches cut from indigenous bushes. The small number of mechanized farms that exist have clusters of buildings that have used what are considered more "permanent" construction materials and engineering methods, according to standards enforced/required of construction in nearby urban centers.

The various types of roofing materials used in the urban areas include baked clay tiles, undulated asbestos roofing sheets and corrugated metal roofing sheets (the most frequently used roofing material) (Fig. 4.12). The corrugated metal roofing sheets are mostly unpainted but a small number better quality IBR metal roofing sheets have a factory coat of green or a dark rust brown colored paint.

Figure 3.4.12 – Kinds of roofing: (1) new clay tiles, (2) pink coated metal sheeting, (3) old colonial clay tiles, and (4) asbestos roofing sheets

4.1.3 Barren or Sparsely Vegetated Areas

There are some very sparsely vegetated areas that are of natural occurrence on the Humpata plateau and along the escarpment (mainly quartzite rocks) (Fig. 4.13).

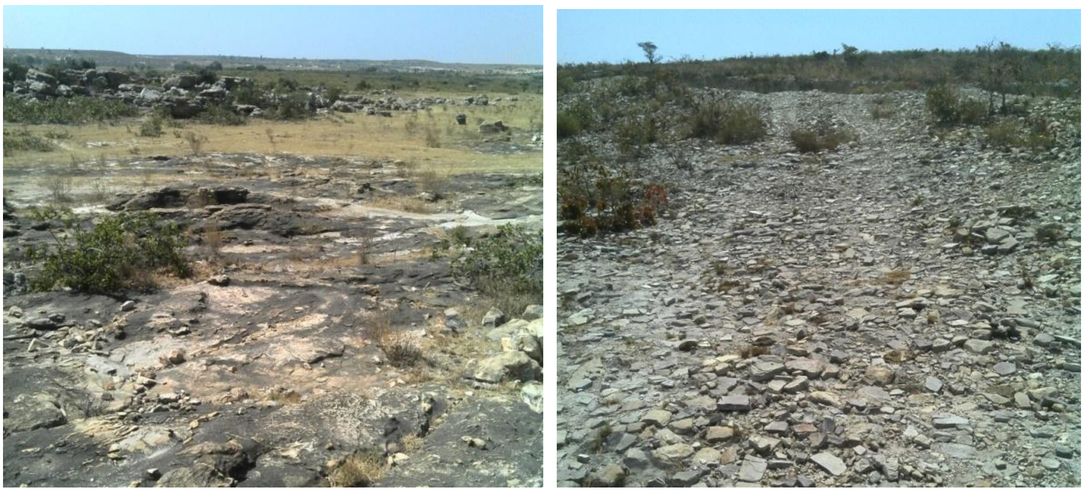


Figure 4.13 – Natural sparsely vegetated areas

However, most of the areas identified during the ground survey were those related to anthropologic activities such as quarries and construction. Some of the areas for extracting stone and gravel are worked originally by the local people with pickaxes (Fig. 4.14) and shovels while larger more industrialized sites are excavated using machinery (Fig. 4.15).

Figure 4.14 – Hand dug quarry stone (rose quartzite) extraction near “Cristo Rei”

The study by Holden (2015) shows that open pit mineral extraction, in the Lubango area, increased by 398% from 2002 to 2010, mostly due to the growing need for construction materials such as stone, crushed rock, gravel and sand.

Figure 4.15 – Google image of industrial mineral extraction near the *Leba* escarpment (limestone)

Other sparsely covered areas caused by human intervention include those areas that have been cleared of vegetation for construction purposes.

4.1.4 Grassland

Angola is one of the two countries in Sub-Saharan Africa with more than 1 million km^2 of grassland (White et al., 2000). It is estimated that about a third of the total area of Angola is covered with herbaceous vegetation – mainly savannah and steppe communities (Shaw, 1947).

Figure 4.16 – Grassland in *Alto Bimbi* (the observed outcrops are of quartzite rocks)

The Montane ecoregion is made up mostly of grasslands with widely scattered trees and shrubs, covering large areas of the highland plateau at altitudes above 1,600 m (WWF, 2008a) (Fig. 4.16). In well-drained areas, this vegetation is generally fire-prone and includes shrub species such as Philippia benguelensis, Erica spp., Protea trichophylla, Stoebe vulgaris, and Cliffortia sp., and grasses, such as Themeda triandra, Tristachya inamoena, T. bequertii, Hyparrhenia andogensis, H. quarrei, Festuca spp., and Monocymbium ceresiiforme (Huntley and Matos, 1994).

Figure 4.17 – Grassland in the Humpata area

Beernaert (1997) uses three terms to describe the grasslands of the Humpata plateau: grass/herb-land, dry savannah with trees or shrubs, and semi-arid steppe vegetation. These are described by Barbosa (1970) as (1) the “savanna das baixas”, the grassland on flat valley floors and on lower valley slopes of the plateau (Fig. 4.17) and (2) the “anharas do alto”, a steppe-like herb-grasslands, which can be found on the hills and Table Mountains of higher altitude (Fig. 4.18 and 4.19).

Figure 4.18 – Grassland above *Estação Zootécnica da Humpata*

He describes the savannahs across the plateau valleys as being herbaceous vegetation dominated primarily by Gramineae and Cyperaceae.

The “anharas do alto”, herb-grasslands on the plateau, were observed by Barbosa in the general Humpata area, near to the Estação Zootécnica da Humpata, from Humpata (city) to Tchivinguiro, as well as, in Tundavala. Besides other invasive European species, the main species grouped into this classification by Barbosa are *Protea, Parinari, Syzygium, Stoebe, Helichrysum Phillippia, Ctenuim* and *Fimbristylis.*

Figure 4.19 – Semi-arid steppe vegetation in *Alto Bimbi*

In this study all these herb-grasslands were grouped and classified as Grassland, except for areas in which presented a great percentage of visible rocky outcroppings interspersed among the vegetation. Such areas present a distinct spectral signature visually evident in the satellite images and were classified as grassland with rocky outcroppings.

4.1.5 Grassland with Rocky Outcroppings

The summits above the montane forest zone usually support a very sparse xerophilous vegetation of steppe - or stony desert type, with a narrow but well defined zone of grass savannah between it and the forest (Shaw, 1947). In many of the areas of higher altitude the herbaceous grasslands grow on visibly rocky ground and in some places amongst rocky outcroppings. Because of the obvious difference in reflectance from the grasslands on less visibly rocky soils and/or with no outcroppings, these grasslands were grouped into a separate class. Most of these areas were observed in the higher altitudes surrounding *Alto Bimbi* and on the hills in the Tundavala area (geological formations composed mainly by quartzite rocks) (Fig. 4.20). The visible rock, in these areas is identified as being between 25 to 50% of the covering.

Figure 4.20 – Rocky grassland areas, with quartzite, on the way to *Estação Zootécnica da Humpata*

4.1.6 Miombo Woodland/Bushland

To avoid any confusion with previously used terminology that identified the "Miombo" as an ecoregion, it must be emphasized that, in this section, the term "Miombo" is limited in use to describe the more specific land cover identified as occurring within the study area (Fig. 4.21).

The miombo is the most extensive vegetation unit in Angola (Shaw, 1947). These broad-leaved deciduous woodlands or shrublands, dominated by species of *Brachystegia*, *Julbernardia* and *Isoberlinia*, are found in areas of over 700 mm annual rainfall on nutrient-poor soils where there is distinct seasonality (Chidumayo, 2010). There are 21 species of *Brachystegia* in miombo woodland and three species of each of the related genera (White, 1983 apud Campbell, 1996).

Figure 4.21 – Miombo on the hills near Lubango

In Angola, this vegetation is commonly referred to as “mata de panda” or “panda bush”.

Figure 4.22 – Miombo on hillside, overlooking the city of Lubango

Barbosa (1970) describes the Miombo subtype that occurs in the study area as a dwarf Miombo measuring between 2 to 5 meters in height, which is most common in regions of high altitude (between 1900 m and 2200 m) (Fig. 4.22). This vegetation is common on the Humpata plateau, on the hills surrounding Lubango (Fig. 4.23), and in Tchivinguiro. Barbosa also describes the most important geological sub-formations in which this subtype of Miombo has been identified as being the Chela quartzite, contacts with quartzite, patches of dolomite limestone, and small patches of dolerites.

Figure 4.23 – Miombo on the hillside surrounding Lubango (taken looking east)

4.1.7 Miombo Bushland with Rocky Outcroppings

Another class decided upon during the field work, also visually evident in the Landsat 8 images, consists of miombo bushland with rocky outcroppings. These areas seem to occur mostly in the transition from lower plains or valleys of herbaceous grasslands and/or seasonal farming to geological formations that rise up above them.

The bush vegetation accompanied by visible rocky outcroppings is distinct from the grassland and/or seasonal farming on the flatter plains or valley below, and many times from more rocky grassland on the flatter tops of these formations. The difference between this class and the miombo is not so much the existing plant species present but the high occurrence of rocky outcroppings which occupy between 25 to 50% of the surface area and typically are not covered by the any vegetation (Fig. 4.24).

Figure 4.24 – Bushland with rocky outcropping (quartzite) on the hills east of Humpata

The combination of vegetation and exposed rock (Fig. 4.25) creates a distinct spectral signature, distinguishing it both from the Miombo bushland as well as from the barren or sparsely covered area (which has less than 15% vegetation cover).

Figure 4.25 – Miombo bushland with rocky outcropping (granites) in Heva

4.1.8 Montane Bushland and Transitional Vegetation

The area on the escarpment slope northward to Bibala is inaccessible, except by trails, but can be observed looking down the escarpment slope from its sharp “drop-off” at Alto Bimbi and Tundavala lookouts (Fig. 4.26).

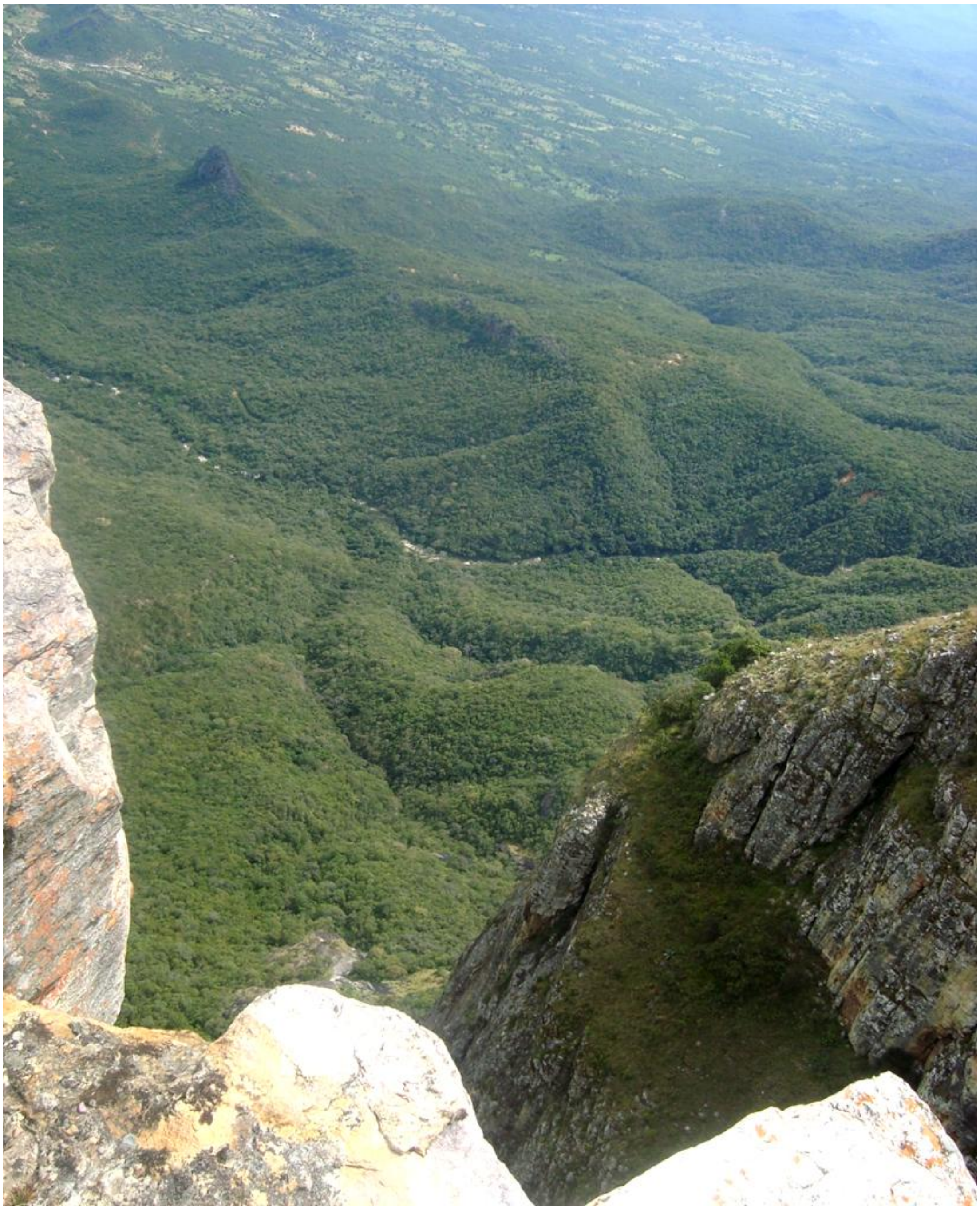


Figure 4.26 – Vegetation on the escarpment, looking down the slope from Tundavala (the outcropping rocks are quartzite)

Barbosa (1970) mentions two types of vegetation in this area: small populations of Afro-Montane forest and a much greater population of transitional or contact vegetation between the Miombo woodland and the Mopane woodland.

Although scarce in this area and relatively scarce in Angola, the Afro-Montane forest vegetation is mentioned by various authors especially because of its great phytogeographic importance in Africa (Shaw, 1947; Barbosa, 1970; Clark et al., 2011). The dominant species of this evergreen forest is the *Podocarpus milanjianus*, also known as "pinho de muxito". This forest or brushwood grows to about 5 or 6 meters in the Tundavala area, at an elevation of between 2000 m a.s.l. and 2500 m a.s.l., and in protected crevices lower down (Fig. 4.27).

Figure 4.27 – Montane bushwood in crevices looking down from Tundavala

Small patches of this forest survive mainly in deep, humid ravines and on isolated peaks higher than 1,800 m a.s.l. (Huntley and Matos, 1994). Shaw (1947) describes the vegetation in this area as "montane brushwood" as opposed to the "montane forests", having only two or rarely, three strata of vegetation, and under 12 m in height.

The greater population of the vegetation, between Bibala (formerly called Vila Arriaga) and the base of the Chela mountain range, is described by Barbosa (1970), as a contact or transitional vegetation between the Miombo and the Mopane. This kind of vegetation appears in the transition between the region of higher altitude and humidity and the drier region of lower altitude.

According to Barbosa (1970), this contact vegetation has an abundance of *Pterocarpus antunesii, Croton zambesiacus*, two more species of *Croton, Ziziphus abyssinica, Combretum mossambicensis, C. oxystanchyum, C. zeyheri, Dalbergia nitidula, Spirotachys africana, Acacia erubescens, Berchemia (Phyllogeiton) discolor*.

4.1.9 Acacia Bushland/Thicket

This land cover class was found to the south of the city of Humpata along the road heading towards Jau (Chibia municipality) (Fig. 4.28), and in the area surrounding Nuntechite Lake. This thicket, also known as "balcedo" or "mato cerrado" is described by Vasconcelos (2015) as appearing in southern regions of the country where the climate becomes semiarid and in which the soil is less deep or is excessively permeable. Vasconcelos mentions them being in the area of Chibia and further south, made up primarily of Combretum e Grewia, as well as acacias.

Figure 4.28 – Acacia bushland/thicket on the way to Jau

4.1.10 Evergreen Needle Leaf Plantations

Angola is one of the African countries with many hectares of land planted mainly in Eucalyptus and Pinus species. These plantations were established in the 30’s and enlarged between the 60’s and 70’s, mostly located on the central plateau (FAO, 2010). Since the conditions found on the plateau were good enough for some of the *Pinus* and *Eucalyptus* species many large plantations were established by Portuguese colonialists (Pinheiro, 1972 apud Domingos, 2014). Many of these plantations exist until the present although, sadly enough, harvesting these forests is presently without measures to replant.

At the time of independence (1975) forest plantations in the Angolan Highlands covered a total of 140,000 ha, 40,000 ha of which were *P. patula* (FAO, 2003 apud Delgado-Matas and Pukkala, 2012).

Figure 4.29 – Pine trees along agricultural plots in Humpata

On the Humpata plateau it is very common to find rows of non indigenous pine and cedar trees bordering agricultural plots and along the dirt roads parallel to the water canals, especially in the area southwest of the city of Humpata (Fig. 4.29). Other groves can be found on farms in the Humpata area as well as in some agricultural areas in Lubango that were formerly devoted to agricultural endeavors (Fig. 4.30).

Figure 4.30 – Grove of pine trees along the *Estrada Nacional 280* (Humpata)

4.1.11 Eucalyptus Plantations

Eucalyptus is the other prominent non indigenous species in the area. The country presents about 140,000 ha of plantations including the eucalyptus, established by the private sector during the colonial times with the purpose of pulpwood and wood fuel production for locomotive machines (FAO, 2010) (Fig. 4.31). Eucalyptus plantations in the Angolan highlands were started in the early 1910's following the construction of the *Caminho de Ferro de Benguela* (the Benguela Railway) in order to provide wood fuel for locomotives after the original miombo woodland was depleted. Later, their management objective was realigned to provide pulp and sawn timber to feed the growing Angolan colonial economy (Delgado-Matas and Pukkala, 2011).

Figure 4.31 – Eucalyptus grove along Estrada Nacional 280

The eucalyptus can be grown to a useful size within 7 to 8 years, depending on its intended use (Domingos, 2014), and thus can provide many benefits very quickly. It has been used for industrial wood and fibre, poles and posts, fuel wood and timber for household use, as well as for nectar, oils and tannins. Several species are also used for windbreaks and shelter (FAO, 2010). In the early days of eucalyptus plantations in Angola, *E. camaldulensis* was the favored species. However, based on the field experience, the dominance soon changed to the *saligna-grandis* complex (Silva, 1969 apud Delgado-Matas and Pukkala, 2011). Today, the plantations play an important role in supplying poles and firewood to local rural communities and local markets, and there is a growing demand for logs to provide prepared boards to size for construction projects and also for logs to be used as a pulpwood source for the paper industry (Domingos, 2014).

Although, most of these plantations were obviously planted decades ago, there are some farms that are producing new plantations, to supply the on-going demand.

Figure 4.32 – Eucalyptus groves in *Alto Bimbi*

Many groves of this fast-growing type of forest can be found in the Humpata area and in the city of Lubango. On the Humpata plateau some of the most obvious plantations are seen along the roads in *Alto Bimbi*, in and around the *Estação Zootécnica da Humpata*, in the *Polígono Florestal da Humpata*, bordering Nuntechite Lake on the northern edge, as well as along the roads heading north and south out of the city of Humpata (Fig. 4.32). Other considerable groves can be found on farmsteads and in the areas surrounding Humpata. In Lubango, one of the most evident areas that continues to have many eucalyptus trees is the *Nossa Senhora do Monte* Park (Fig. 4.33). Many other groves have been cut for urbanization purposes.

Figure 4.33 – Eucalyptus in *Nossa Senhora do Monte* Park, Lubango

4.1.12 Irrigated Agriculture

Agriculture is a major economic activity in Angola. In spite of the war, 73% of the workforce was employed in agriculture between 1990 and 1992 (FAO, 1996). The climatic conditions in the Humpata region allow year-round agriculture for those who have access to a constant source of water for irrigation. These farms produce a variety of vegetables year-round.

In this area farms that cultivate in both rainy season and dry season depend on one of the following water sources:

1. Dams in the area provide irrigation water through a system of manmade canals. Besides the Neves dam, the Tundavala dam and the dams above

Estação Zootécnica da Humpata (previously mentioned in this study) there are also some significantly smaller dams in the study area. They are also used for irrigation purposes (Fig. 4.34).

Figure 4.34 – Agriculture irrigated from the Neves Dam

2. Privately owned boreholes supply water to individual farms; and
3. The perennial streams and year-round springs: a lot of year-round agriculture is done along the riverbanks and in areas with year-round springs (Fig. 4.35).

Figure 4.35 – Agriculture in the riverbed (Tchivinguiro)

4.1.13 Nonirrigated Cropland

Agriculture in Angola today continues to be in the majority of cases a strongly subsistence-based agriculture, with the average size of plot cultivated per family at around 2 ha (Fig. 4.36). Over half of the small scale farmers use some form of shifting agriculture (FAO, 1996).

Figure 4.36 – Nonirrigated cropland (millet)

In the following satellite images the difference between agriculture during the dry season and the rainy season is clearly visible (Fig. 4.37). Most of the lower region (riverbed) is cultivated throughout the year, including in the dry season. In these cases irrigation can be done from the river. In the higher areas (to the east), the cultivated crops depend solely on the rains and are cultivated only in the rainy season.

Figure 4.37 – Google images of irrigated riverbed and higher nonirrigated areas (southeast of Humpata): (1) dry season and (2) rainy season

The agricultural lands used for this kind of seasonal agriculture were classified as nonirrigated cropland. The most commonly cultivated crop in these areas is maize. The planting season starts with the first rains which normally are expected by the end of September. During this season these lands are ploughed, planted and harvested. In the dry season they are either covered with the dry remains of the previous season's crops or are cleared (Fig. 4.38). This class also includes all actively/seasonally tilled land.

The boundaries of these areas used for seasonal agriculture are not always visible: lands are demarcated by locally known boundaries (controlled by the "sobas" the local traditional authorities) and sometimes by thorn fences or aloe plants, although these are generally used more for keeping out cattle than demarcation purposes. Many times the area used for subsistence farming by a family is located around or close to their huts.

Figure 4.38 – Corn fields in (1) rainy season and (2) dry season

Grassland and cropland typically occur side by side or may occur surrounded by the other. In the Landsat 8 images the visual distinction between them is not always obvious. In the higher definition Bing images, although not clearly distinguishable by the colour, they can some times be distinguished by their texture. The cropland that is fenced in with makeshift thornbush fences or by sisal plants (Fig. 4.39) is clearly visible with a defining contrast in the Bing images. The distinction is also observable in these images if the plots have recently been tilled.

Figure 4.39 – Borders around crops: (1) makeshift thornbush fence and (2) sisal plants

4.1.14 Orchards

Apple, pear, plum and orange orchards are plentiful, mostly in areas irrigated by the canals, but also along the riverbeds. Although most of them are small scale orchards, there are two orange orchards of considerable size: one to the north and the other to the south of the city of Humpata (Fig. 4.40).

Figure 4.40 – Orchards in Humpata: (1) a pear orchard with seasonal crops planted between rows and (2) an orange orchard

In Lubango, there also are some orchards in the Mapunda area. Despite the great urban growth some orchards remain, whilst many others have been cleared and the land sold to provide plots for increasing urbanization demands/development.

4.1.15 Herbaceous Wetland

The herbaceous flora of the permanently swampy locations, where water comes to the surface but does not form pools, is extraordinarily rich and varied. The general aspect of the vegetation is that of grass meadow, but with monocotyledons forming a large percentage of the total. Endemism is higher in this type of vegetation, in Angola, than in any other (Shaw, 1947). In many countries of Southern Africa, the word “dambo” is used to describe these seasonally waterlogged, predominantly grass-covered, shallow depressions in the headwater zone of rivers, generally less than 5 km^2 (FAO, 1998).

Figure 4.41 – Herbaceous wetland in Humpata area

As described by Sebadduka (2014), “the term ‘dambo’ was introduced to scientific literature by Ackermann (Mäckel, 1974) being described as “periodically inundated grass-covered depressions on the headward ends of a drainage system in a region of dry forest or bush vegetation.”

Dambos are distinctive features of the Miombo region (Campbell, 1996). These damp spots, which occur commonly both on the higher plateau and on the less elevated territory of Lunda, Moxico and the Cuanza and Cubango basins, are given a variety of names by. These include such names as “mulola” (south of Angola), “omuramba” (Cuenene basin), “neaca” (Benguela plateau), and “tenga” (Lunda). “Mulola” and “omuramba” are traditional words that both designate depressions or very open valleys, more or less clayey, which become partially or completely dry at the height of the dry season (Shaw, 1947). In the Humpata region, the locals refer to them as “etalas” (Fig. 4.41).

In the study area, Nuntechite Lake (also referred to as Lake Outite), is the most evident and accessible area with this type of land cover (Fig. 4.42). Measuring 6.3 km^2, (4.5 km long, 2 km wide) located roughly 11 km south of the village of Humpata, Lake Nuntechite (15º7'S; 13º25'E) is a seasonal lake that outflows into the Chibia River (Vanden Bossche and Bernacsek, 1990).

Figure 4.42 – Nuntechite lake in the (1) rainy season and (2) dry season, looking north toward eucalyptus border

Although it is obvious when looking at the satellite images that there are quite a few other such areas, Lake Nuntechite is the largest in the study area and was fairly easily accessed using the road that goes south from Humpata towards Jau (Fig. 4.43).

Figure 4.43 – Google images of Nuntechite lake in (1) dry season and (2) rainy season

The northwestern border of the lake is defined by a considerable grove of eucalyptus and, further north on the western edge by the same kind of Acacia bushland/thicket that borders the inbound road from Humpata.

4.1.16 Seasonally Burnt Areas

According to Clark et al. (2011) the major threats to the forests are deforestation and fires, as well as charcoal-making in the woodlands adjacent to large cities such as Lubango. About 1.3 million km^2 of fire adapted savannah and grassland burn annually in Africa (FAO, 2001 apud Malmer, 2007). Whether these seasonally burnt areas are a manmade problem or a naturally instituted process continues to be debated. Reliable studies of fire frequencies are scarce and it can be debated what is “natural” (Malmer, 2007).

Some authors such as Shaw (1947) explain how this is a natural process which stimulates renewed growth of the vegetation. With the onset of the dry winds from the east and north-east in June and July, and the ensuing fall of humidity, the herbaceous vegetation rapidly withers and dries up and thus the annual fires occur (Shaw, 1947).

According to the Miombo Ecoregion Vision Report (2009) fires in the Miombo Ecoregion also originate from people preparing land for cultivation, collecting honey or making charcoal. Others are purposefully/intentionally caused by hunters, either to drive out animals or to attract them later to the grass re-growth areas that were burnt, and by livestock herders to provide a green flush for their livestock and to control pests, such as ticks. This seems to be primarily what occurs in the grasslands in the Humpata region (Fig. 4.44).

Figure 4.44 – Area near Humpata (1) before and (2) after burning

In Angola, there is a strong incentive for charcoal production because of its huge demand as a fuel/cooking energy source. The alarming rate at which forests are being burned can put the sustainability of these resources at risk (Bahu, 2015). Charcoal as the largest source of energy (56.8% of the total consumption) in the country is used for cooking, lighting, heating and industry. The greatest production of charcoal happens in areas surrounding the large urban centers causing rapidly expanding rings of deforestation (Bahu, 2015). This is seen on a yearly basis on the hills surrounding the city of Lubango (Fig. 4.45). According to Vasconcelos (2015) the species most used for the production of charcoal in and around Lubango are *Brachystegia spiciformis* Benth (*Fabaceae*), *Julbernardia paniculata* (Benth) Troupin (*Fabaceae*), and *Pteleopsis anisoptera* (Welw. Ex. M.A, Lawson) Engl. & Diels (*Combretaceae*) (Vasconcelos, 2015).

Figure 4.45 – Burning in the dry season (area around Lubango)

The grassland areas and the Miombo scrubland areas are at greatest risk of being burned during the dry season. Although this burnt vegetation cannot be defined as a land cover in and of itself, in this study it was considered as one of the land cover classes for two reasons:

a) Since large areas were easily identified in the Landsat 8 dry season images, it was considered that not identifying it as a special case scenario in the sample and classification could cause these pixels to be misclassified in the final map.

b) Since bush fires are commonplace in Africa and monitoring them has become of increasing concern, it was decided that, although not one of the original objectives of the current study, it would be worthwhile to consider it as a "special" classification.

4.2 Training Data

Using the ground data acquired during the field work, in combination with the high definition Bing images, a total of 3880 homogenous pixels was selected to be randomly split: 90% to be used for the calibration of the decision tree rules and 10% for validation. Most of the sampling was limited to the areas accessed and observed during the field work. The seasonally burnt areas and part of the vegetation on the escarpment slope, however, had to be sampled mostly using the high definition images; the former, due to the temporary character of the class and the latter due to the limited road access.

The 3880 pixels are 0.63% of the number of pixels in the final mapped area. The limited road access in the general study area resulted in a limited accessible area from which to take samples. Therefore the sample pixels used for training the classification were taken from a larger area than that of the final mapped area. The total study area is about 2000 km^2 (from which training samples were selected) while the final mapped and assessed area is only approximately 500 km^2 (classification area which would later be assessed). This is the central part of the study area that represents, at least a small portion, of every land cover class described in section 4.1 and was mapped at a scale of 1:100 000.

For 14 of the 16 classes 252 pixels were selected per land cover class (Table 4.3). The remaining two classes (Evergreen Needle leaf and Mambo Bush land with Rocky Outcropping) in the accessible area for field work lacked sample pixels due to its small dimension. Instead of the desired 252 pixels selected for the other classes, only 115 and 237 pixels, respectively, were selected for each of these land cover classes.

For 8 out of the 16 classes, in 28 cases a square matrix of 3x3 pixels was the selected geometry. In the remaining classes this geometry was more difficult to achieve in part due to access problems. So, a combination geometry of 3x3 pixels and individual pixels, or using only this last case, was selected depending on the area of the class. Individual pixels were selected in situations or areas where a sample of 3x3 pixels was not possible. The sample pixels of the evergreen needle leaf trees, for example, was limited to individual pixels instead of 3x3 pixels because in the study area they appear mostly along roadsides not forming more than 1 or 2 pixel wide rows.

Table 4.3 presents the number of 3x3 pixels and individual pixels sampled in each class as well as the distribution of the sample in the different areas.

Table 4.3 – Distribution of training sample

	SAMPLE PIXELS			PIXELS PER AREA													
	3x3 pixels	Single pixels	Total	Lubango	Tundavala	Heva	Humpata City	Alto Bimbi	Escarpment	Road to Tchivinguiro	Area Southwest of Humpata	South of Humpata	Neves Dam Area	Area North of Humpata	Leba	Estação Zootécnica Area	Total
Acacia Bushland/Thicket	28	0	**252**	0	0	0	0	0	0	0	0	252	0	0	0	0	**252**
Bushland with outcropping	17	84	**237**	0	0	149	77	0	0	0	0	0	0	0	0	11	**237**
Miombo	26	18	**252**	171	0	18	0	0	0	0	0	0	45	0	0	18	**252**
Transitional Vegetation	28	0	**252**	0	0	0	0	0	252	0	0	0	0	0	0	0	**252**
Eucaliptus	24	36	**252**	27	0	90	45	54	0	0	0	18	0	0	0	18	**252**
Evergreen Needleleaf	0	115	**115**	0	0	0	12	0	0	0	103	0	0	0	0	0	**115**
Grassland	28	0	**252**	0	0	81	45	18	0	9	45	9	18	0	9	18	**252**
Grassland with rock	23	45	**252**	0	54	0	90	6	0	0	0	0	0	0	0	102	**252**
Herbaceous wetland	28	0	**252**	0	0	0	0	0	0	0	0	252	0	0	0	0	**252**
Built-up areas	27	9	**252**	165	0	4	74	0	0	0	9	0	0	0	0	0	**252**
Seasonally Burnt areas	28	0	**252**	0	0	0	0	0	0	252	0	0	0	0	0	0	**252**
Barren areas	28	0	**252**	81	54	36	18	0	0	0	0	0	0	9	45	9	**252**
Nonirrigated agriculture	27	9	**252**	0	0	99	27	18	0	27	18	36	9	0	9	9	**252**
Irrigated Cropland	28	0	**252**	0	0	0	162	0	0	0	36	0	0	0	54	0	**252**
Orchards	23	45	**252**	0	0	0	45	0	0	0	9	81	0	117	0	0	**252**
Inland water	28	0	**252**	0	72	0	0	0	0	0	0	0	171	0	0	9	**252**
Total	391	361	**3880**	444	180	477	595	96	252	288	220	648	243	126	117	194	**3880**

The training area selected for each class represents 6.5% of the total 3880 sample pixels, except for the Evergreen Needle leaf and the Miombo Bushland with Rocky Outcropping, which represent 2.9% and 6.1% respectively.

4.2.1 Modeling Results for the Training Sample

The J48 decision tree algorithm available with Weka was used for training the Landsat datasets. Of the 3880 pixels processed using Weka 3.6.13, J48 classifier, to create the decision tree, as stated before, 90% were used for calibration and 10% for validation. The result of this analysis, using the 6 bands described in section 3.3.1, indicated that 89.9% of the pixels were correctly with a kappa coefficient of 0.89.

In an effort to improve the accuracy and remove redundancy of the data, experiments were done adding a Principal Components Analysis (PCA) for both images. The values of the 12 resulting PCA images were also added to the model. The values extracted from the 3880 sample pixels from the 6 bands of the rainy season image, 6 bands of the dry season image and the 12 PCA images were then processed as a whole. The result of this analysis was a decision tree made up of a set of 155 rules with 91% correctly classified instances (an increase of 1.8%) with a kappa coefficient of 0.91 (an increase of 0.019).

4.2.2 Validation of the Model for Training Samples

The following confusion matrix shows the accuracies of each individual class using the 10% of the sample pixels selected for validation (a total of 388 pixels) (Table 4.4).

Only the nonirrigated cropland and grassland were slightly below 80%; in all other classes the accuracy values vary from 80 to 100 (Table 4.4). Given the good results of this model created using the training samples, it was now possible to use the model to produce a classification of the Landsat 8 images.

Table 4.4 – Results from training the model

Reference Data

Classified Data	Herbaceous wetlands	Seasonally burnt	Nonirrigated crops	Acacia thicket	Miombo bushland	Irrigated Agriculture	Barren areas	Miombo with rock	Built-up areas	Grassland with rock	Grassland	Orchards	Evergreen needleleaf	Eucalyptus	Inland water	Transitional veg	Total	User's Accuracy
Herbaceous wetlands	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	23	95.7
Seasonally burnt	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	100
Nonirrigated crops	0	0	18	0	0	0	0	0	0	0	5	0	0	0	0	0	23	78.3
Acacia thicket	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	33	100
Miombo bushland	0	0	0	0	17	0	0	0	0	0	0	1	1	0	0	1	20	85
Irrigated agriculture	0	0	1	0	0	22	0	0	0	0	0	2	0	1	0	0	26	84.6
Barren areas	0	0	1	0	0	0	24	0	2	0	0	0	0	0	0	0	27	88.9
Miombo with rock	0	0	0	0	0	0	0	22	0	1	0	1	0	0	0	0	24	91.7
Built-up areas	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	18	100
Grassland with rock	0	0	0	0	0	0	0	1	0	32	1	0	0	0	0	0	34	94.1
Grassland	0	0	3	0	0	0	0	1	0	1	18	0	0	0	0	0	23	78.3
Orchards	0	0	1	0	0	2	0	0	0	0	1	20	0	0	0	0	24	83.3
Evergreen needle leaf	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	5	100
Eucalyptus	0	0	0	0	0	0	0	0	0	0	0	1	0	25	0	0	26	96.2
Inland water	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	35	100
Transitional veg	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	22	24	91.7
Total	22	23	24	33	19	24	24	24	20	34	25	25	6	26	35	24	388	
Producer's Accuracy	100	100	75	100	89.5	91.7	100	91.7	90	94.1	72	80	83.3	96.2	100	91.7		91.8

4.3 Modeling of the Land Cover in the Study Area

The DT model containing 155 rules was applied using bands 2, 3, 4, 5, 6 and 7 from both seasons, as well as the PCA's. Next a mosaic was created using the 155 resulting images, representing the 16 Level II land cover classes (Fig. 4.47), which can, in turn, be collapsed into the 8 Level I land cover classes (Fig. 4.48).

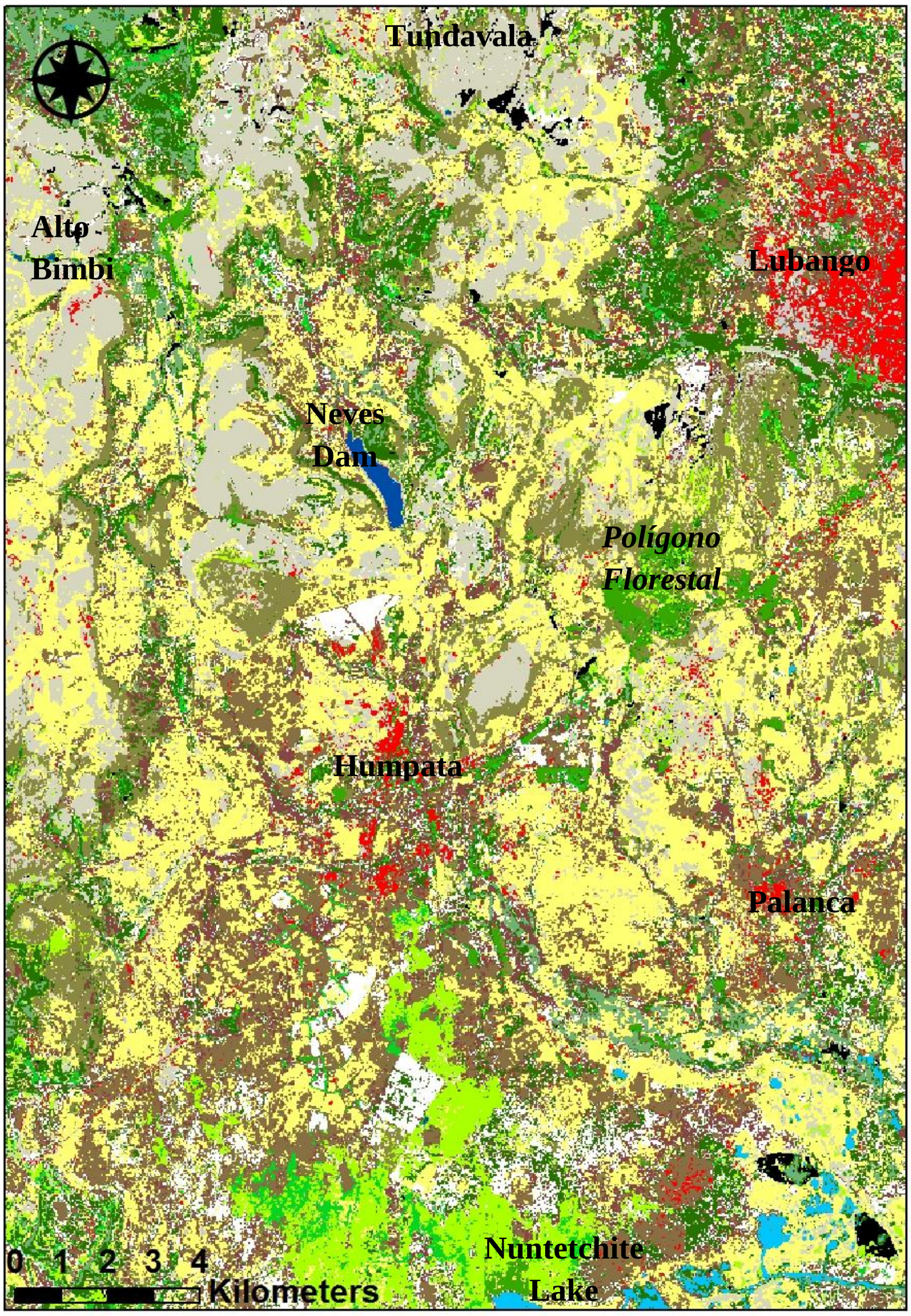


Figure 4.46 – Classified map and areas described (Legend as Annex 1)

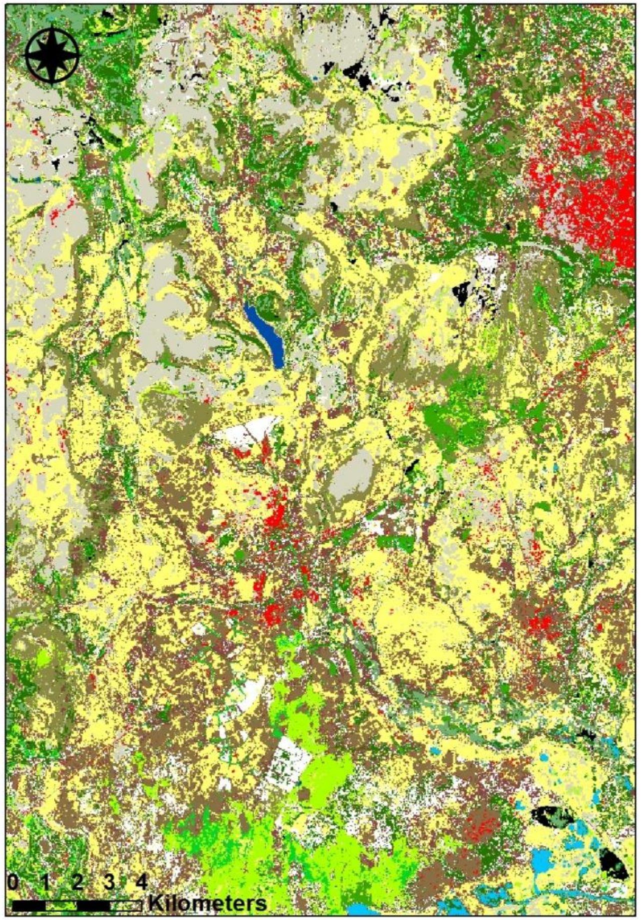


Figure 4.47 – Classified map for the 16 Level II classes
(Legend as Annex 1)

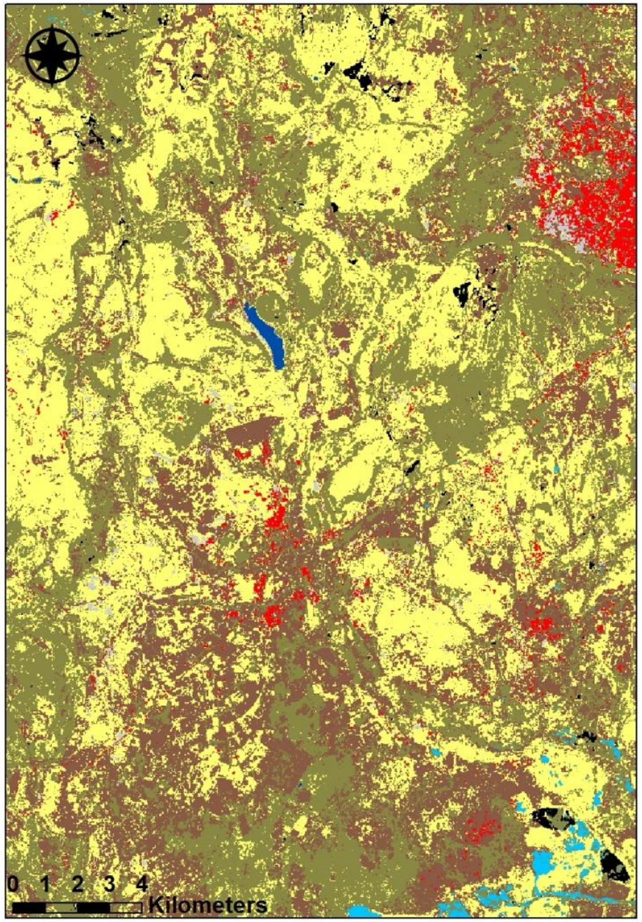


Figure 4.48 – Classified map for the 8 Level I classes
(Legend as Annex 1)

A preliminary visual analysis of the resulting classification showed very positive results, in general terms, based on the overall knowledge of the area. However, it was still possible to identify areas that were most likely misclassified.

The northern part of the mapped area (Fig. 4.46) (in the highest part of the Humpata plateau) is clearly dominated by grassland with and without rocky outcroppings. Miombo appears along the hillsides, while the transitional vegetation appears on the escarpment slopes visible in the northwest corner of the map.

On the other hand, the southern region of the mapped area (in the lower region of the Humpata plateau) is predominantly covered by agricultural lands, orchards and acacia bushland/thicket with the herbaceous wetlands in the southeast corner. The acacia bushland/thicket occupies a large area in the southern part of the mapped area.

The urban areas of Lubango, Humpata and Palanca are clearly visible in bright red (Fig. 4.46). The large body of water in the central part of the mapped area (dark blue) represents the lake created by the Neves Dam. The smaller lakes en route to Alto Bimbi are also visible in the northwest corner of the mapped area.

In the northern part of the mapped area (Fig. 4.46), in the Tundavala and Alto Bimbi regions, the summits of the mountains have mostly been classified as grassland with rocky outcroppings (light grey) while the grasslands (yellow) appear mostly in the surrounding valleys.

Miombo bushland (green) is visible on the hillsides surrounding the city of Lubango (in the northeast corner) as well as in some areas of Humpata (near the center of the mapped areas). In the Humpata area, however, there is a much greater amount of miombo with rocky outcroppings surrounding the hills.

The plantations of eucalyptus and evergreen needle leaf can be seen mostly in the central part of the mapped area. A large portion of eucalyptus is evident in the area of the *Polígono Florestal* while smaller patches can be seen in and around Humpata as well as on the northern border of Nuntechite Lake (Fig. 4.46). Although clearly evident as borders along the agricultural lands in Humpata and some patches on larger farms, the evergreen needle leaf plantation is definitely one of the classes with the smallest representation.

Many seasonally burnt areas (black) can be seen across the plateau. The largest ones are in areas surrounded by grassland and some smaller sections are surrounded by miombo bushland.

There are some pixel areas that, at least from a visual analysis, based on knowledge acquired during onsite visits, would appear to be misclassified. The most evident are some pixels near and around the region of the acacia thicket that have been classified as (1) orchards and (2) evergreen needle leaf plantations. From a visual analysis of their location and distribution it could be supposed that these pixels have been misclassified. Orchards and the evergreen needle leaf plantation occur mostly in agricultural areas with some kind of road access and their distribution is generally somewhat uniform, in rows or clusters. These areas are scattered amongst the acacia bushland/thicket in areas where there is little or no road access. Compared to the surroundings and the general knowledge of the area, the most logical conclusion would be that these pixel areas are covered by a variation of the acacia thicket, or another kind of indigenous vegetation, that was not sampled during the calibration process. One suggestion may be that it is the same kind of vegetation, for example, but with less density.

The other very likely case of misclassified pixel areas is the transitional vegetation that appears along the rivers. The transitional vegetation is characteristic of the escarpment slopes or in crevices on the mountainside and not on the plateau along the rivers. Therefore, these pixels along the rivers have most likely been misclassified. It is likely that these areas are covered by some kind of indigenous vegetation which has a constant source of water making it lush and dense year-round similar to the transitional vegetation. This may be the reason for having a spectral signature similar to that of the transitional vegetation. The other possibility is that these areas are cultivated lands, irrigated year-round by the perennial rivers.

4.3.1 Accuracy Assessment of the Model

As mentioned previously, no land cover classification is complete without an accuracy assessment. The results acquired during the accuracy assessment described in section 3.6 are presented in the following tables: the first (Table 4.5) with the 16 Level II land cover classes and the second with the 8 Level I land cover classes collapsed from the 16 Level II classes.

Table 4.5 integrates the 14 classes verified on the ground (in black) and the two classes verified using only Google images (highlighted in red). Table 4.6 shows the accuracy of the classes collapsed into 8 Level I classes.

Table 4.5 – Accuracy results for Level I land cover classes (%)

Reference Data

Classified Data	**Eucalyptus**	**Grassland**	**Evergreen needleleaf**	**Miombo with rock**	**Miombo bushland**	**Built-up areas**	**Grassland with rock**	**Irrgiated agriculture**	**Nonirrigated crops**	**Orchards**	**Barren areas**	**Inland Water**	**Acacia thicket**	**Herbacious wetland**	**Transitional veg**	**Seasonally burnt**	**Total**	**User's Accuracy**
Eucalyptus	24	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	25	96
Grassland	0	24	0	0	0	0	0	0	1	0	0	0	0	0	0	0	25	96
Evergreen needleleaf	7	0	10	1	0	1	0	0	0	2	0	0	4	0	0	0	25	40
Miombo with rock	3	4	0	15	3	0	0	0	0	0	0	0	0	0	0	0	25	60
Miombo bushland	6	4	0	0	11	0	0	2	1	1	0	0	0	0	0	0	25	44
Built-up areas	0	2	0	0	0	23	0	0	0	0	0	0	0	0	0	0	25	92
Grassland with rock	0	4	0	0	0	0	21	0	0	0	0	0	0	0	0	0	25	84
Irrigated agriculture	1	4	0	1	2	0	0	10	0	1	0	1	5	0	0	0	25	40
Nonirrigated crops	0	7	0	1	0	1	0	0	14	0	1	0	1	0	0	0	25	56
Orchards	5	5	0	0	0	0	0	0	1	12	0	0	2	0	0	0	25	48
Barren areas	0	0	0	0	0	0	0	1	1	0	23	0	0	0	0	0	25	92
Inland Water	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	25	100
Acacia thicket	3	2	0	0	0	0	1	0	0	1	0	0	18	0	0	0	25	72
Herbaceous wetlands	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	16	100
Transitional veg	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**9**	**0**	**0**	**0**	**1**	**1**	**0**	**14**	**0**	**25**	**56**
Seasonally burnt	**0**	**4**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**21**	**25**	**84**
Total	49	60	11	18	16	25	22	22	18	17	24	27	31	16	14	21	391	
Producer's Accuracy	49	40	91	83	69	92	95	45	78	71	96	93	58	100	100	100		72

Table 4.6 – Accuracy results for Level II land cover classes (%)

Reference Data

Classified Data	Bushland/Woodland	Herbaceous Vegetation	Urban Areas	Cultivated Areas	Barren Areas	Inland Water	Wetlands	Seasonally Burnt Areas	Total	User's Accuracy
Bushland/Woodland	121	11	1	16	0	1	0	0	150	81
Herbaceous Vegetation	0	49	0	1	0	0	0	0	50	98
Urban Areas	0	2	23	0	0	0	0	0	25	92
Cultivated Areas	18	16	1	38	1	1	0	0	75	51
Barren Areas	0	0	0	2	23	0	0	0	25	92
Inland Water	0	0	0	0	0	25	0	0	25	100
Wetlands	0	0	0	0	0	0	16	0	16	100
Seasonally Burnt Areas	0	4	0	0	0	0	0	21	25	84
Total	139	82	25	57	24	27	16	21	391	
Producer's Accuracy	87	60	92	67	96	93	100	100		81

4.3.1.1 Level II Land Cover Classes

The overall accuracy of the first classified map (with 16 Level II classes) is 72% (Table 4.5). The classes with best accuracies (all above 90%) are the Herbaceous Wetlands, Water, Barren areas, and Urban areas followed by the the Seasonally Burnt Areas and the Grassland with rocky outcroppings (above 80%). The class with the lowest user and producer accuracies, when comparing the observed and the calculated (both under 46%) is the irrigated agriculture. The Eucalyptus and Grassland have very high user’s accuracies (96%) but very low producer’s accuracies (both under 50%).

The table shows that there is considerable confusion between the Transitional Vegetation (on the escarpment slopes) and the Irrigated Agriculture. This may be due to the fact that irrigated agricultural areas are cultivated and receive water year-round making them lush year-round, even as the vegetation on the escarpment is, in part, made up of evergreen forest (*Podocarpus milanjianus*), also watered year round by groundwater naturally springing from crevices and rock faces along the escarpment slope.

Although not evidenced by the confusion matrix, the same kind of confusion is visible in the classified image along the rivers on the Humpata plateau. As mentioned before some vegetation along the rivers, was classified as transitional vegetation, the same as on the escarpment slope. From field work data obtained along the rivers it is known that these are not the transitional vegetation found on the escarpment slopes. Some of vegetation along the rivers is probably irrigated agriculture done year-round along the riverbanks. The rest is probably different types of vegetation requiring a constant source of water.

The seasonally burnt areas were verified using the Google images (taken on the 13th of October 2013) with the closest date after the Landsat image (taken on the 23rd of August 2013). Even so, the time lapse (seven weeks) is considerable enough for the evidence of the burning to become imperceptible, especially if the area's vegetation is already sparse to begin with or if the rains come soon after the burning and cause immediate new growth. During the process of field work, in certain burned grassy areas, it was observed that new growth was clearly visible within three weeks from the time of burning. This would mean that, depending on the kind of vegetation and the rain, it is possible that evidence of burning could be completely imperceptible within seven weeks. Therefore, the seasonally burnt areas could potentially have had higher accuracy, had there been Google images with less of a time difference as compared to the Landsat image.

4.3.1.2 Level I Land Cover Classes

Table 4.6 shows the accuracy of the 8 Level I land cover classes collapsed from the 16 Level II land cover classes. When a hierarchical classification is used, reference land cover classes can be collapsed into broader land cover categories which correspond to categories on the map being assessed (Nusser and Klaas, 2003). This is commonly done when trying to achieve a certain level of accuracy. Collapsing more detailed classes into more generic ones eliminates confusion that may exist between classes that are very close in spectral signature.

In the second image all the different bushland/woodland (including the exotic plantations) were collapsed into one class. Originally it was thought that they could be collapsed into two classes, one indigenous and one exotic. However, the confusion matrix of table 4.5 showed considerable confusion between indigenous and exotic vegetation. It may be that these classes' spectral signatures are too similar to separate them adequately, or there may be a great internal variation in the classes (for example height and density of vegetation) which makes it difficult to create a very specific spectral signature. Therefore, it was decided that bushland/woodland (including the exotic plantations) would be grouped into only one class. This is also in agreement with the NLCD 2006 which also only presents one forest class in Level I.

The nonirrigated cropland, irrigated agriculture and orchards were grouped into one class: cultivated areas. The grassland and grassland with rocky outcroppings were also grouped into one class: herbaceous vegetation.

The urban areas, inland water, sparsely covered areas, herbaceous wetlands and seasonally burnt areas were all maintained as individual classes in the second image. After collapsing some of the identified land cover, the second image continued with a total of only 8 land cover classes. The overall accuracy (correctly classified instances) of the second classified image is 81%. The one class that continues to have a very low accuracy rate is the cultivated areas. All other classes are considered as having a relatively reasonable accuracy rate.

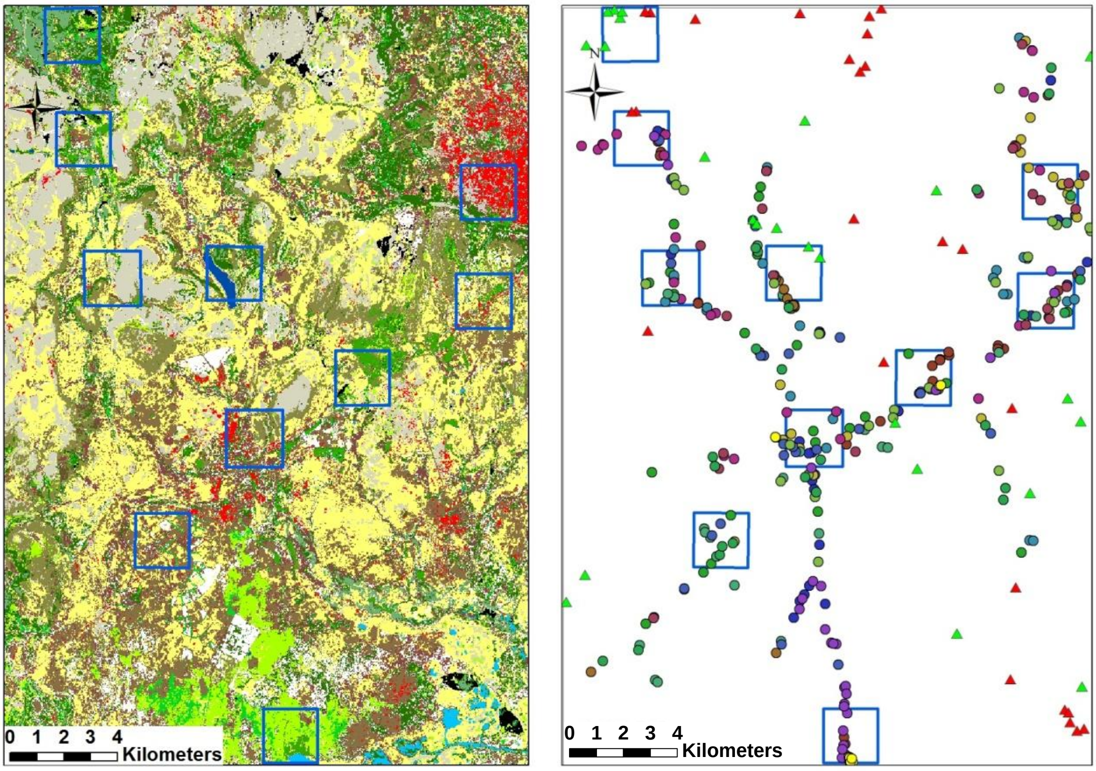
4.3.2 Detailed Analysis of the Accuracy of the Model Using Selected Areas

In order to refine the accuracy of the model and therefore acquire a greater understanding of the overall quality of the classification, 10 areas of 2x2 km (Fig. 4.49), were selected in the different regions of the mapped area, as described:

1. Escarpment area
2. *Estação Zootécnica da Humpata*
3. Lubango West
4. Area on route to *Estação Zootécnica da Humpata*
5. Neves Dam area
6. Heva area
7. *Polígono Florestal* area
8. Humpata City East
9. Agricultural lands southeast of Humpata City
10. Herbaceous Wetland area on the way to Jau

These areas were selected by taking into account the overall distribution of the land cover classes in the mapped area and also the number of the random sample points available for analysis in each area. The same sample pixels set used in the previous overall analysis are used to evaluate and analyse the smaller selected areas. These ten areas are basically smaller samples of the general area representing in more detail each land cover class (16 Level II classes).

Each 2x2 km area of the classified image (with 16 Level II classes) was enlarged to a scale of 1:12,500 and is presented side by side with a high definition image at the same scale for easy visual comparison (Fig 4.50). The same sample pixels described in section 3.6.4, used in the previous overall accuracy assessment, were used to analyze the accuracy in each individual area (Fig. 4.49). For each area, a confusion matrix was prepared using the sample pixels present in the area.



Legend

2x2km Areas

Random Points in Entire Population

- Transitional Vegetation
- Seasonally Burnt Areas

Random Points in Reduced Population

- Eucalyptus
- Grassland
- Evergreen Needleleaf
- Bushland with Rock Outcropping
- Built up areas
- Grassland with Rock Outcropping
- Irrigated Agriculture
- Orchards
- Water
- Sparsely Vegetated Land
- Miombo Bushland
- Acacia Thicket
- Unirrigated Cropland
- Herbacious Wetland

Figure 4.49 – Selected 2x2 km areas representing different land cover areas and clusters of samples

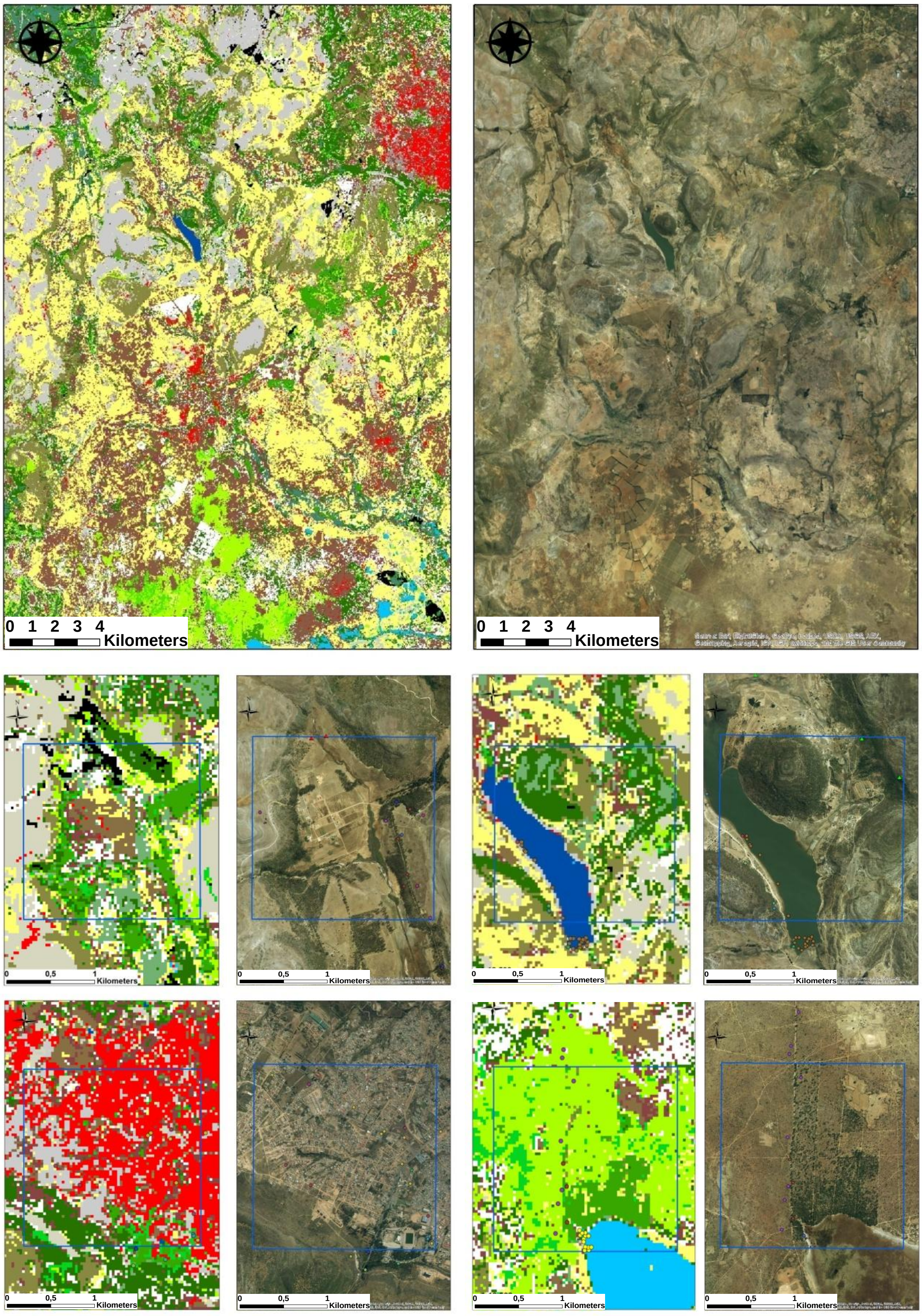


Figure 4.50 – Classified map and 2x2 km areas compared to high definition Bing images

4.3.2.1 Escarpment Area

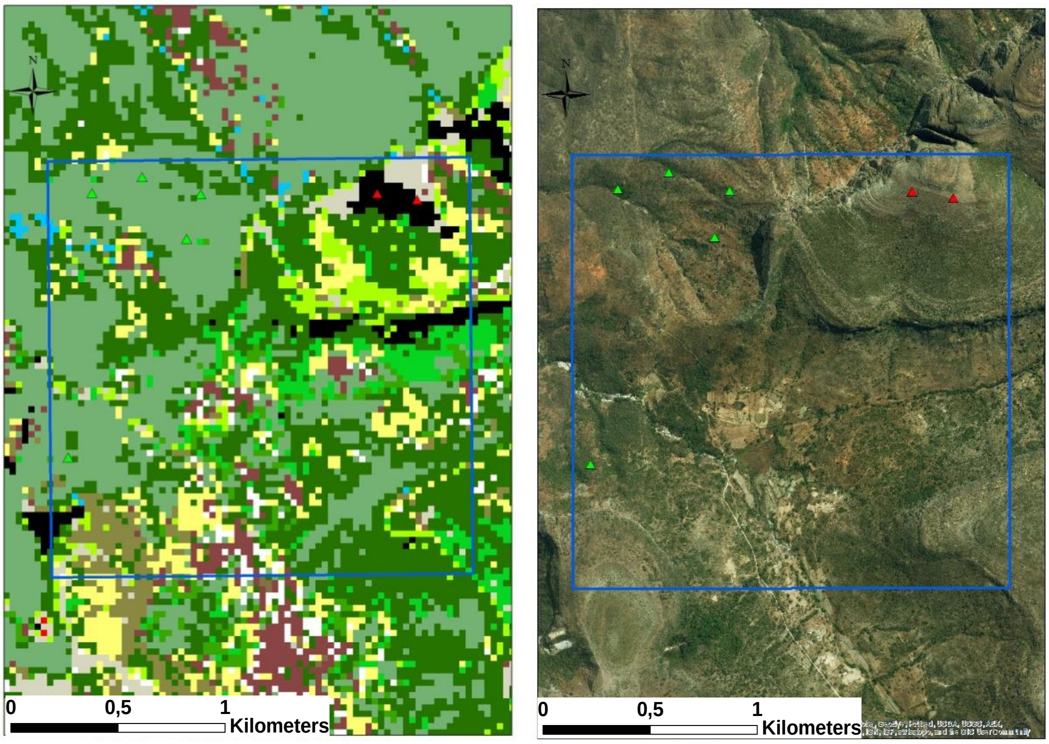


Figure 4.51 – Escarpment 2x2 km classified area compared to high definition image: transitional vegetation (green triangles) and seasonally burnt area (red triangles) samples

The 2x2 km area at the edge of the escarpment was chosen for the obvious reason that it is the main location in the mapped area in which there is representation of the vegetation specific to the escarpment slope – the transitional vegetation between Miombo and Mutuati and the Montane Bushwood. It also includes two sample points for the seasonally burnt areas (Fig. 4.51).

Table 4.7 – Confusion matrix for escarpment area

Reference Data

Classified Data		Grasslandand with rock	Transitional Vegetation	Seasonally Burnt Areas	Total	User's Accuracy
	Grassland with rock				0	
	Transitional Vegetation		5		5	100
	Seasonally Burnt Areas	2			2	0
	Total	2	5	0	7	
	Producer's Accuracy	0	100			71

All the sample points in this area were verified solely by photointerpretation of Google Earth images because of inaccessibility by road to the area. Based on this interpretation all the sample points classified as Transitional Vegetation were correctly classified (Table 4.7). It can be assumed that these results would be similar in most of the area along the escarpment slopes, especially since this vegetation is specific to this kind of environment. It might however be difficult to visually distinguish between the Miombo (which was identified around Lubango) and the Transitional Vegetation (on the escarpment slopes) using photointerpretation. Once again, the sole dependence on this data has its limitation and may lead to over optimistic results. Nevertheless, most of the Transitional Vegetation classification does occur along the escarpment slopes, leading to conclusion that in most cases it can be considered accurate.

The seasonally burnt areas were also verified solely by photointerpretation, using the available Google Earth images with less time difference to the Landsat 8 image. The sample points in this area were verified as being grassland with rocky outcroppings and not burnt areas. This however may be erroneous given the 7-week time lapse between the Google Earth images and Landsat 8 images. It could very well be that obvious visual evidences of burning may have been eliminated during this period. From the visual analysis of the Landsat 8 dry season image, and based on the results of most other sample points which were positively verified, it is very likely that these in fact were burnt areas at the time the Landsat 8 image was taken but that evidences of the fact were not visible in the Google Earth images due to immediate regrowth of the vegetation and/or due to the fact that there may have been very little vegetation previous to the burning.

In the classified image the surrounding areas was classified as grassland with rocky outcroppings. If this is accurate, the amount of actual vegetation to be burned (and the subsequent evidences of burning) would be considerably less than in grassland areas without rocky outcroppings or in Miombo bushland areas (which may have been the case for other areas that were positively verified). This could be the reason that other areas could be verified within the 7-week period and this one could not.

The other possibility is that the burning in this area had occurred earlier on than in the other areas and thus evidences of the burning would have already become imperceptible by the time the Google Earth images were taken, while in other areas (burned later on, perhaps closer to the date of the Landsat 8 image) it would still be quite evident.

Had there been a more precise way of verifying this situation, the accuracy of this 2x2 km area would probably be closer to 100%.

4.3.2.2 *Estação Zootécnica da Humpata*

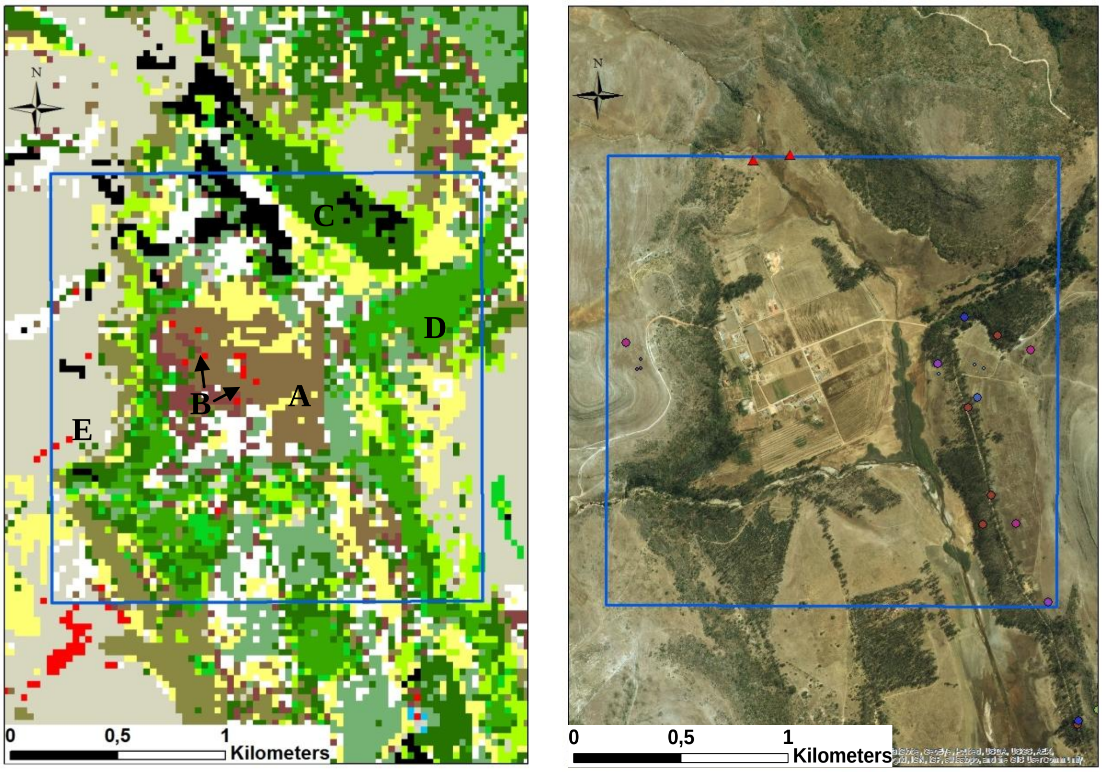


Figure 4.52 – *Estação Zootécnica da Humpata* 2x2 km classified area compared to high definition Bing image

At the center of this 2x2 km area is the *Estação Zootécnica da Humpata* which is used primarily for agriculture and animal breeding. It has scattered buildings surrounded mostly by cultivated fields with many groves of eucalyptus around its perimeter (Fig. 4.52). A section of Miombo bushland/woodland can be found on one of the hillsides in the northeast and southwest corner of the areas, while most of the summits of the surrounding hills are covered by grassland with rocky outcroppings.

Table 4.8 – Confusion matrix for *Estação Zootécnica da Humpata*

		Reference Data					
		Eucalyptus	**Grassland**	**Grassland with rock**	**Seasonally Burnt Areas**	**Total**	**User's Accuracy**
Classified Data	Eucalyptus	4				4	100
	Evergreen needle leaf	1				1	0
	Grassland with rock			3		3	100
	Orchards		1			1	0
	Acacia Thicket	1		1		2	0
	Seasonally Burnt Areas				2	2	100
	Total	6	1	4	2	13	
	Producer's Accuracy	67	0	75	100		69

The sample pixels in this area were classified as eucalyptus, evergreen needle leaf, grassland with rocky outcroppings, orchards, acacia thicket and seasonally burnt areas. The pixels classified as eucalyptus, grassland with rocky outcroppings and seasonally burnt areas were all correctly classified (Table 4.8). However the four pixels classified as orchards (1 pixel), evergreen needle leaf (1 pixel) and acacia thicket (2 pixels) were all misclassifications.

This is actually in agreement with the overall accuracy assessment in which the eucalyptus, grassland with rocky outcroppings and seasonally burnt areas have quite high user's accuracy (96%, 84% and 84% respectively) and the orchards, evergreen needle leaf and acacia thicket much lower user's accuracy (48%, 40% and 72% respectively).

Nevertheless, from a visual perspective the large sectors of land cover are obviously well identified in the classified image (Fig. 4.52): (A) the cropland of the *Estação Zootécnica da Humpata*, (B) several scattered Built-up areas, (C) the Miombo along one of the hillsides in the northeast and southeast corner, (D) the large eucalyptus groves and the summits of the surrounding hills covered by (E) grassland with rocky outcroppings. This can be observed by comparing the classified image with the high definition image and is also obvious in the photos below (Fig. 4.53 and 4.54).

Figure 4.53 – *Estação Zootécnica da Humpata* with cropland (A), scattered buildings (B) and Eucalyptus (D)

Figure 4.54 – Eucalyptus (D) and grassland surrounding *Estação Zootécnica da Humpata*

There is however one very clear misclassification in this area that can be observed by visual analysis although it is not evidenced by the confusion matrix. There is one major sector that is incorrectly classified as “transitional vegetation”, which is specifically related to the escarpment slopes. For someone unfamiliar with the area it probably would not be obvious from a simple visual analysis of the high definition image. But a person with general knowledge of the area through multiple onsite visits would realize this as an obvious error.

The transitional vegetation (including the Montane Bushwood) occurs primarily along the escarpment slopes and in crevices on the mountainsides and not in the valleys or along the rivers. The pixel areas classified as transitional vegetation in this 2x2 km area are most likely some kind of vegetation receiving year-round irrigation from a constant water source (perhaps also agriculture) and not the transitional vegetation characteristic of the escarpment slopes.

4.3.2.3 Lubango West

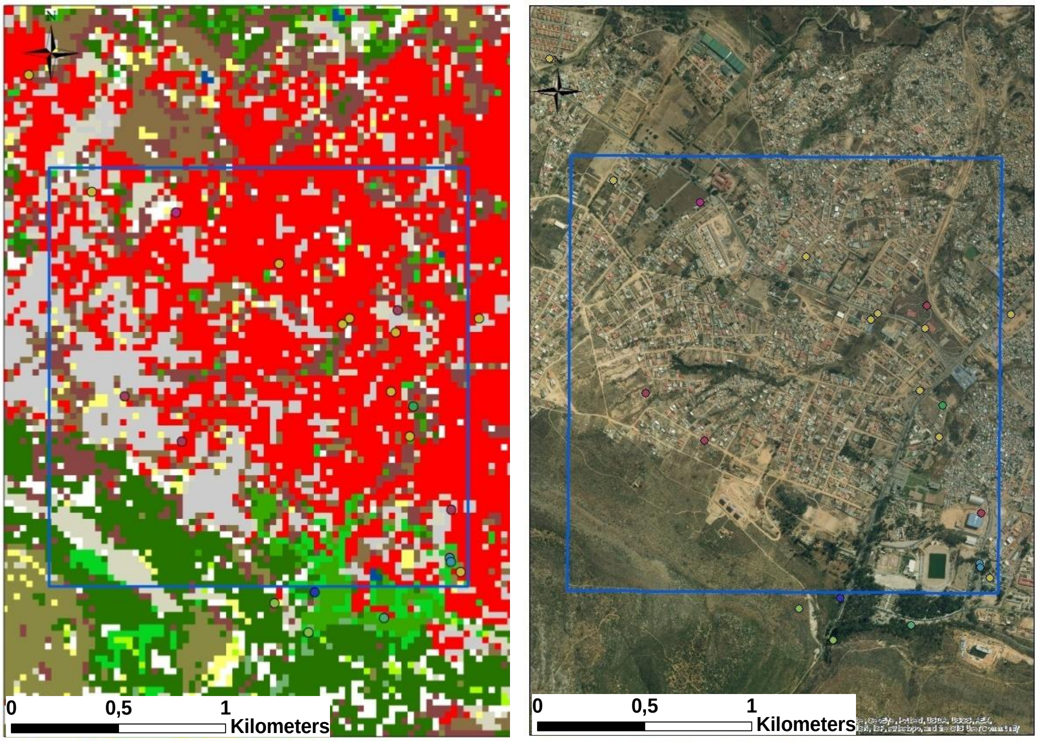


Figure 4.55 – Lubango West 2x2 km classified area compared to high definition Bing image

This 2x2 km area includes a large sector of urban/Built-up area of Lubango, near the mountainside beginning the ascent to the Humpata plateau. Between the urban area and the mountainside there are various sectors of sparsely covered land, most probably cleared for construction purposes (Fig. 4.55). In the southeast corner of the area is the *Nossa Senhora do Monte* park with a considerable amount of eucalyptus and evergreen needle leaf trees (pine and cedar) (Fig. 4.56).

Figure 4.56 – Eucalyptus in *Sra do Monte* park

In the southwest corner of the area the mountainside is covered primarily by Miombo bushland (Fig. 4.57).

Figure 4.57 – Miombo in foreground, looking east across Lubango

These large sectors can be easily identified in the classified image and compared to the high definition Bing image (Fig.4.55). There is however one group of pixels in the southwest corner classified as nonirrigated crops which is in reality more likely Miombo bushland.

Table 4.9 – Confusion matrix for Lubango West

Reference Data

Clasified Data	**Eucalyptus**	**Grassland**	**Built-up areas**	**Grassland with rock**	**Nonirrigated crops**	**Acacia Thicket**	**Total**	**User's Accuracy**
Miombo with rock	2						2	0
Built-up areas			9				9	100
Grassland with rock				1			1	100
Nonirrigated crops		1			4	1	6	67
Total	2	1	9	1	4	1	18	
Producer's Accuracy	0	0	100	100	100	0		78

The sample pixels verified in this area have relatively good accuracy results (Table 4.9). The user's accuracies for the sample pixels representing built-up areas and grassland with rocky outcroppings were both 100%. Again, this is not surprising as the overall user's accuracy of both these classes is also very high. This also accounts for the higher overall accuracy of this area seeing as half of the sample pixels are built-up areas.

Although, the nonirrigated crops have a higher user's accuracy in this area, compared to the overall assessment, the matrix continues to show the existing confusion between nonirrigated crops, grassland and acacia thicket.

4.3.2.4 Area on Route to *Estação Zootécnica da Humpata*

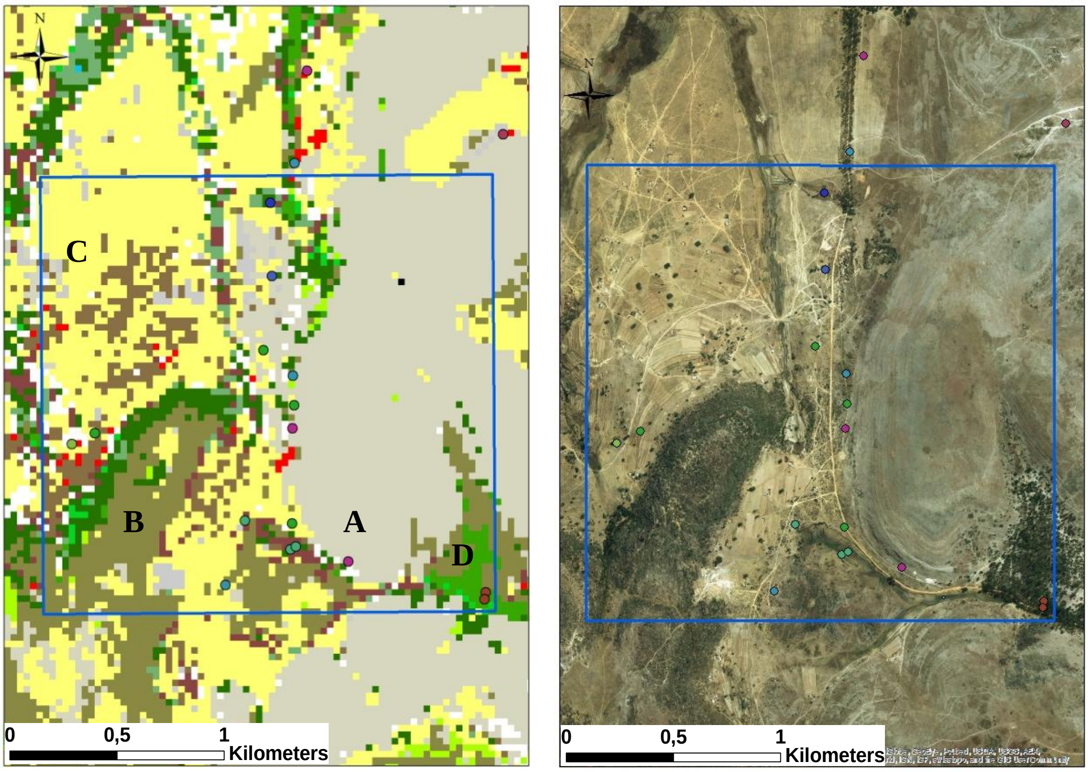


Figure 4.58 – On route to *Estação Zootécnica da Humpata*, 2x2 km classified area compared to high definition Bing image

This 2x2 km area on the way to the *Estação Zootécnica da Humpata*, shows a clear distinction between a hill covered by (A) grassland with rocky outcroppings and a hill covered with (B) Miombo and Miombo with rocky outcroppings (towards the west), as well as (C) the grassland in the valleys below the hills (Fig. 4.58 and Fig.59). These are clearly distinguishable in the classified image and can be easily compared to the high definition image (Fig. 4.58).

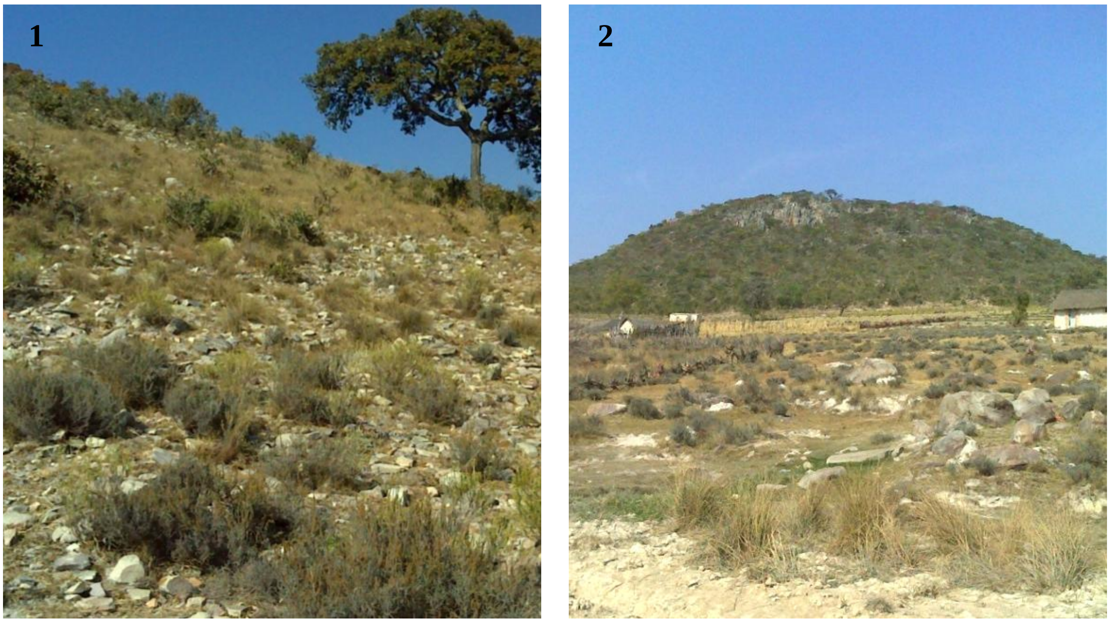


Figure 4.59 – (1) grassland with rocky outcropping (A), (2) Miombo bushland (B)

Also in the valleys below the hills are some areas of nonirrigated crops (especially in the northwest corner of the area) (Fig. 4.58 and 4.60) and some irrigated agricultural areas along the river.

Figure 4.60 – Grassland and nonirrigated crops in the valley below the hills

A considerable grove of eucalyptus (D) is also quite evident in the southeast corner of the area (Fig. 4.58).

Table 4.10 – Confusion matrix for area on route to *Estação Zootécnica da Humpata*

Classified Data \ Reference Data	Eucalyptus	Grassland	Evergreen	Miombo with rock	Grassland with rock	Irrgiated agriculture	Total	User's Accuracy
Eucalyptus	2						2	100
Grassland		4					4	100
Evergreen			1				1	100
Miombo with rock				2			2	100
Miombo Bushland		1					1	0
Grassland with rock					2		2	100
Irrigated agriculture				1		2	3	67
Orchards	1						1	0
Total	3	5	1	3	2	2	16	
Producer's Accuracy	67	80	100	67	100	100		81

The overall accuracy of the sample pixels in the area is relatively high (81%), with only 3 out of 16 pixels misclassified, confirming the visual comparison (Fig.4.10). However the confusion between the irrigated agriculture and the Miombo with rock is evidenced in this area, as well as between grassland and Miombo bushland, and orchards and eucalyptus.

4.3.2.5 Neves Dam Area

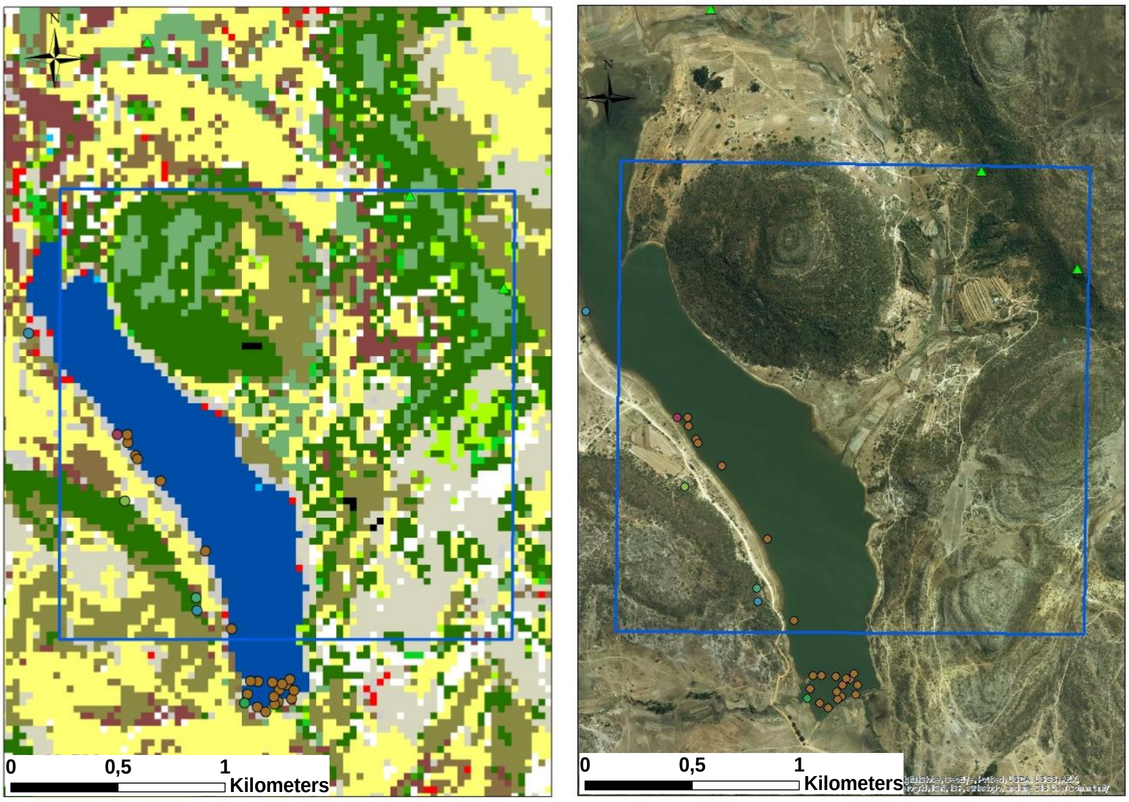


Figure 4.61 – Neves Dam 2x2 km classified area compared to high definition Bing image

This area was chosen for its obvious representation of inland water, this lake being the largest body of inland water in the mapped area (Fig. 4.61 and Fig. 4.62).

Figure 4.62 – Neves Dam lake and surrounding hill to the East

Fortunately, the area also includes some sample pixels classified as Miombo and others as Transitional Vegetation, located on the hills surrounding the lake area.

The classified area is easily compared to the high definition image (Fig. 4.61) with the lake at the center and the surrounding hills covered mostly by bushland/woodland especially on the slopes and in the crevices, as well as the grassland in the valleys surrounding the lake and rocky grasslands on the summits of some of the hills (Fig. 4.63).

Figure 4.63– Lake with hills in the background

The overall accuracy of the pixels in this area reflects the accuracy of areas mapped under the class of inland water in the overall assessment (%100 user’s accuracy and 93% producer’s accuracy) (Table 4.11). Interestingly enough the accuracy of the woodland surrounding the body of water is also high.

Table 4.11 - Confusion matrix for Neves Dam area

Reference Data

Classified Data		Miombo with rock	Miombo Bushland	Barren areas	Inland water	Transitional Vegetation	Total	User's Accuracy
	Miombo with rock	1					1	100
	Miombo Bushland		1				1	100
	Irrigated agriculture		1				1	0
	Barren areas			1			1	100
	Inland water				7		7	100
	Transitional vegetation					2	2	100
	Total	1	2	1	7	2	13	
	Producer's Accuracy	100	50	100	100	100		92

As in overall verification, the transitional vegetation was verified using only visual analysis of Google Earth images. Based on the description the transitional vegetation and the necessary conditions for its growth, the location seems to be quite characteristic of this kind of vegetation. Again, the distinction between the Miombo Woodland/Bushland and Transitional Woodland/Bushland in this area is difficult to assess, therefore the assessment is based primarily on the visual analysis of the density and geophysical nature of the area.

4.3.2.6 Heva Area

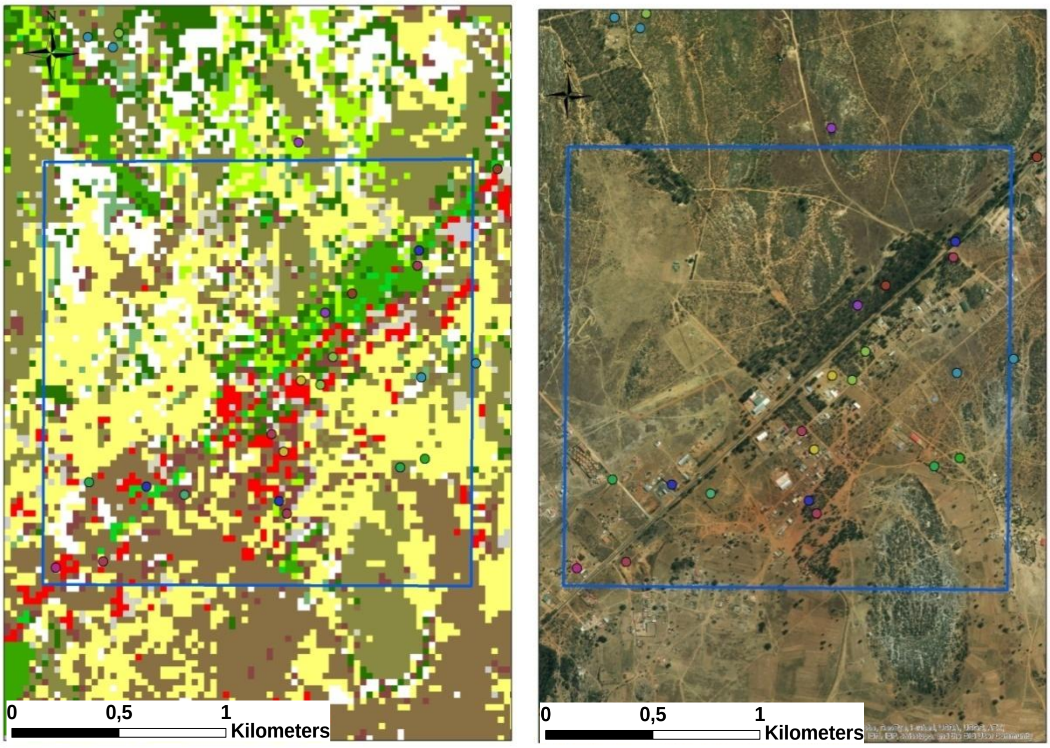


Figure 4.64 – Heva 2x2 km classified area compared to high definition Bing image

The area of Heva is located on the Humpata plateau just outside (to the southwest of) the city of Lubango. There is a developing built-up area located mostly along the southern side of the *Estrada Nacional 120*. On the northern side (in a protected government reserve) there is a large grove of eucalyptus and a lesser amount of pine trees. Most of the area (Heva), however, is covered by grassland and nonirrigated crops with a few areas identified as Miombo with rocky outcroppings (Fig. 4.64).

This built-up area is a lot less dense than in the city of Lubango appearing in the classified image as scattered groups of red pixels stretched out along the road. These groupings of buildings are also visible in the high definition image (Fig. 4.64).

Based on a visual comparison of the classified image and the high definition image, the grassland, the nonirrigated crops and the Miombo bushland with outcroppings seem to be quite accurate. In this area there are many small hills covered in Miombo bushland with rocky outcroppings that rise slightly above the grassland and cropland areas. These features are quite noticeable in both classified and high definition image as in the images below (Fig. 4.65).

Figure 4.65 – (1) Grassland with eucalyptus in background, (2) Miombo with rocky outcroppings

The eucalyptus grove is also evident in the classified image although with some possible pixels misclassified as evergreen needle leaf trees as evident also in the confusion matrix below (Fig. 4.64).

Table 4.12 – Confusion matrix for Heva area

Reference Data

Classified Data	**Eucalyptus**	**Grassland**	**Evergreen**	**Miombo with rock**	**Miombo bushland**	**Built-up areas**	**Grassland with rock**	**Barren areas**	**Total**	**User's Accuracy**
Eucalyptus	1								1	100
Grassland		1							1	100
Evergreen	1		2						3	67
Miombo with rock				1	1				2	50
Miombo bushland	1								1	0
Built-up areas						2			2	100
Grassland with rock							1		1	100
Irrigated agriculture		1							1	0
Nonirrigated crops				1					1	0
Sparsely covered land								4	4	100
Acacia thicket	1								1	0
Total	4	2	2	2	1	2	1	4	18	
Producer's Accuracy	25	50	100	50	0	100	100	100		67

Although the main eucalyptus grove can be identified in the classified image it seems as though many other areas of eucalyptus were misclassified as acacia thicket, Miombo bushland and/or evergreen needle leaf trees (Table 4.12). This occurred in line with what was already observed in the overall confusion matrix. This may be due to the different density of the eucalyptus groves. These misclassifications of the eucalyptus were a major contributing factor to the general accuracy of this area being considerably low (67%). The other pixels were misclassified grassland and Miombo.

4.3.2.7 *Polígono Florestal* Area

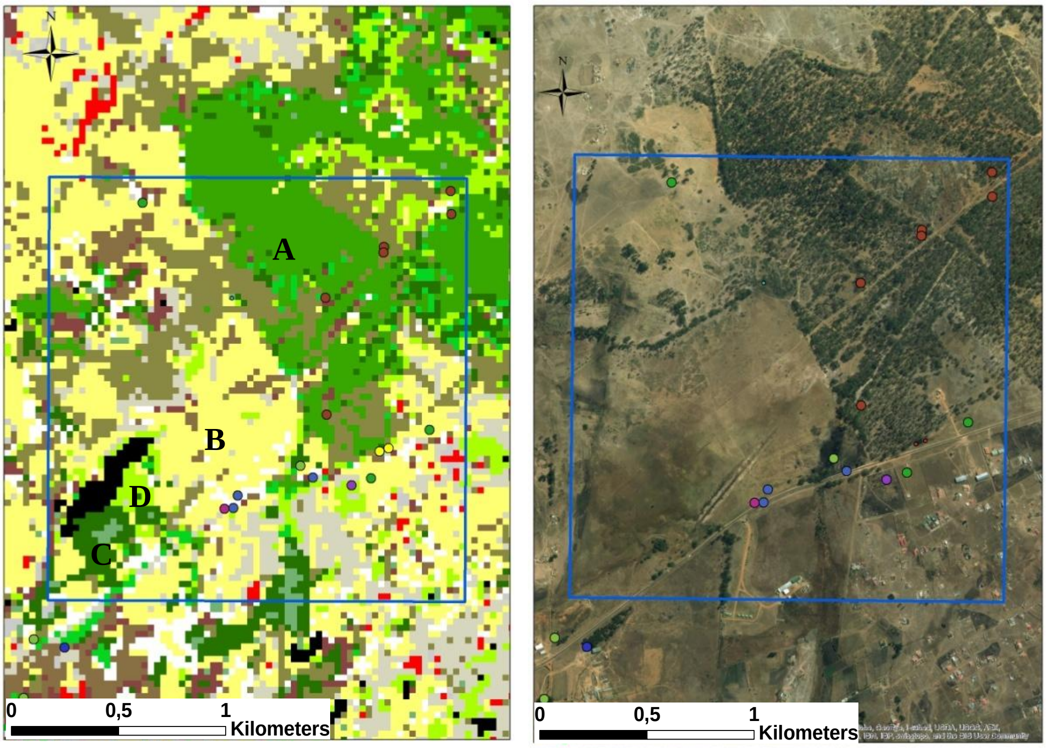


Figure 4.66 – *Polígono florestal* 2x2 km classified area compared to high definition Bing image

This area includes a section of the *Polígono Florestal*, a great part of which is occupied by (A) eucalyptus trees. To the southwest of the eucalyptus plantation is a great expanse of (B) grassland (Fig. 4.67). Based on a visual comparison between the classified image and the high definition image, the border between the grassland and woodland seems to be well defined and relatively accurate (Fig. 4.66)

Most of the surrounding grassland seems to be correctly classified except for a section in the southwest corner of the area where there are two large portions which seem to be misclassified as (C) Miombo bushland and (D) Acacia thicket. Based on the many onsite visits to this area, it can be concluded that these last two classes are most likely misclassifications.

Figure 4.67 – Eucalyptus (A)/Grassland (B) border in *Polígono Florestal* area

The classification of the woodland is much less accurate than the grassland however. Based on ground visits to this area (Fig. 68), it was confirmed that these wooded areas belonging to the *Polígono Florestal* are almost completely made up of eucalyptus trees, some areas more densely covered than others.

Figure 4.68 – Eucalyptus trees in the *Polígono Florestal* area

The Miombo bushland with rock outcroppings is the most clearly evident misclassification based on visual observation. In reality, this land cover class is non--existent in this area. In the overall assessment this was also an issue – 12% of the pixels classified as Miombo bushland with rock outcroppings were in fact areas covered by eucalyptus trees. This misclassification seems to occur in areas where the eucalyptus trees are less dense.

Table 4.13 – Confusion matrix for the *Poligono Florestal* area

Reference Data

Classified Data		**Eucalyptus**	**Grassland**	**Grassland with rock**	**Herbacious wetland**	**Total**	**User's Accuracy**
	Eucalyptus	6				6	100
	Grassland		3			3	100
	Miombo bushland	1				1	0
	Grassland with rock			1		1	100
	Orchards	2	1			3	0
	Acacia thicket		1			1	0
	Herbacious wetlands				2	2	100
	Total	9	5	1	2	17	
	Producer's Accuracy	67	60	100	100		71

The confusion matrix also presents some other misclassifications (Table 4.13). Although the user's accuracy of the eucalyptus is 100%, the confusion between the orchards, Miombo bushland and the eucalyptus is evident once more in the producer's accuracy (67%). This also happens with the grassland in this area. The user's accuracy is 100% while the producer's accuracy is only 60% because of the confusion between the grassland and the Acacia thicket and the orchards. The very small area of herbaceous wetland is correctly classified, which is also in agreement with the overall accuracy of the herbaceous wetland (100%).

4.3.2.8 Humpata City East

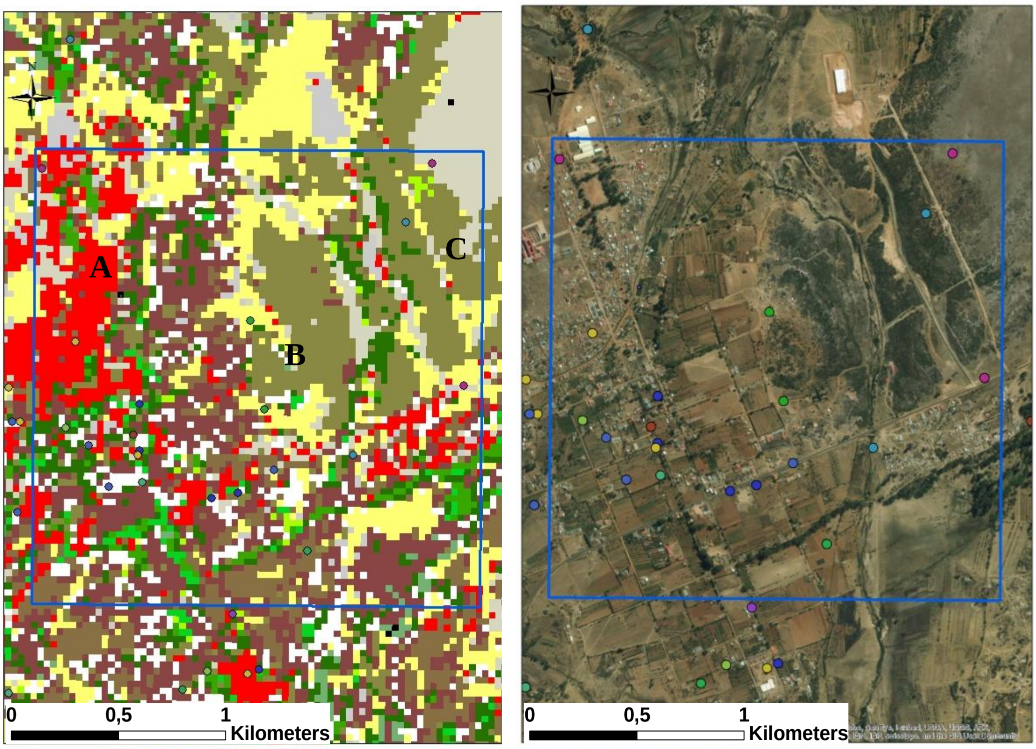


Figure 4.69 – Humpata City East 2x2 km classified area compared to high definition Bing image

This area includes the eastern side of the city of Humpata which besides built-up areas also has a lot of agricultural lands. Due to the fact that historically it was an agricultural village, most of the built-up area of Humpata City is not as dense as in Lubango and tends to be interspersed with agricultural lands and orchards. Nevertheless some of the *musseque* areas developing around the central part of the city have become quite dense as is evident in the northeast corner of the area (A) (Fig. 4.69).

In the eastern part of this 2x2 km area there are a couple hilly areas covered mostly by (B) Miombo bushland with rocky outcroppings as well as by (C) grassland with rocky outcroppings along the plateau summit. These land covers are also quite easily identified by comparing the classified image and the high definition Bing image.

Eucalyptus and evergreen needles leaf trees (D) line some of the streets as can be observed in both the classified and high definition images (Fig. 4.69). Other areas of eucalyptus stand out in definite contrast to the grassland and agricultural (Fig. 4.70).

Figure 4.70 – Grassland in foreground and eucalyptus in background

In Table 4.14 it can be seen that, in this area, the accuracy of the orchards and nonirrigated crops were both 100%. The user’s accuracy for the built-up areas and the irrigated agriculture was also 100%. These four land cover classes probably account for most of the land cover on the eastern side of the city of Humpata.

Table 4.14 – Confusion matrix for Humpata City East

Reference Data

Classified Data	Grassland	Evergreen	Miombo with rock	Built-up areas	Grassland with rock	Irrgiated agriculture	Nonirrigated crops	Orchards	Total	User's Accuracy
Eucalyptus		1							1	0
Grassland	2								2	100
Evergreen		3		1					4	75
Miombo with rock	1		1						2	50
Miombo bushland						1			1	0
Built-up areas				2					2	100
Grassland with rock	1				2				3	67
Irrigated agriculture						1			1	100
Nonirrigated crops							1		1	100
Orchards								3	3	100
Total	4	4	1	3	2	2	1	3	20	
Producer's Accuracy	50	75	100	67	100	50	100	100		75

The evergreen needle leaf user's accuracy in this area is much higher (75%) than in the overall assessment (40%). This is most likely due to the fact that this 2x2 km area is made up of mostly built-up areas and agricultural lands, which often accompanied by a greater concentration of evergreen needle leaf plantations than in the larger mapped area. One pixel area was classified was misclassified as eucalyptus while 3 others were correctly classified. This situation is in agreement with the overall assessment in which 9% of the evergreen needle leaf pixels observed on the ground were misclassified as eucalyptus. The one pixel classified as evergreen needle leaf (verified to be in fact urban/built-up area) was actually an area beside another evergreen needle leaf pixel that was correctly classified. The most likely explanation in this case, would be that the classification was strongly influenced by the neighboring pixel.

4.3.2.9 Agricultural Lands Southwest of Humpata City

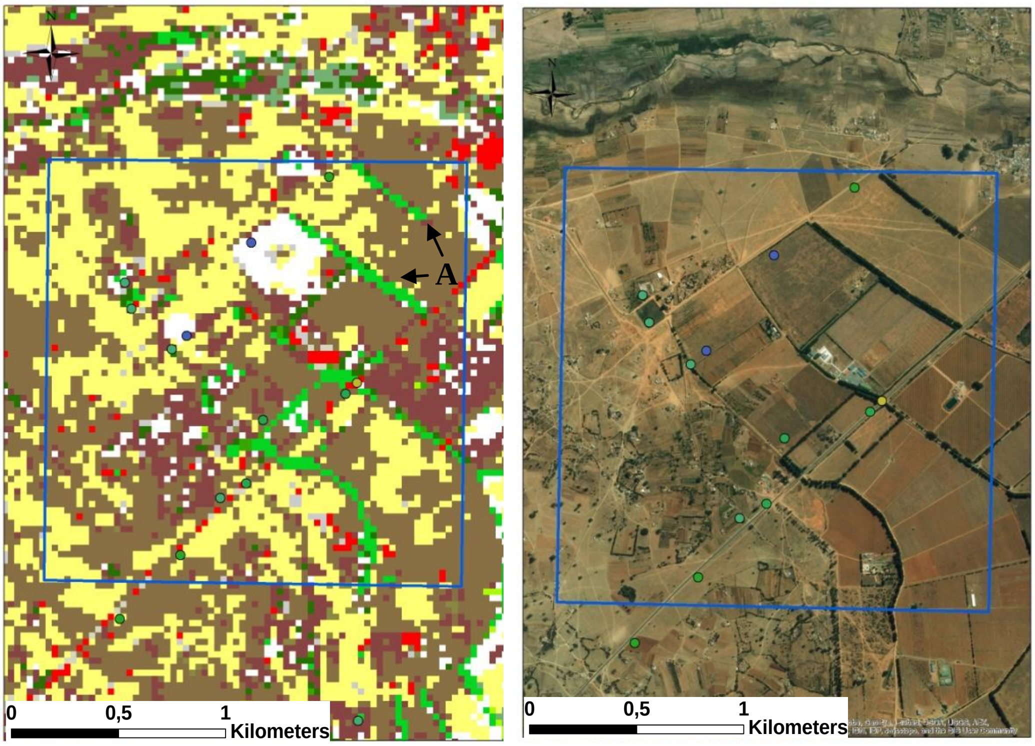


Figure 4.71 – Agricultural lands southwest of Humpata City 2x2 km classified area compared to high definition Bing image

This agricultural area is located southwest of the city of Humpata. It is a mosaic of agricultural plots and natural grassland with scattered built-up areas. Many of the agricultural plots are bordered by (A) evergreen needle leaf trees (primarily pine), which are clearly distinguishable in both the high definition image and the classified image (Fig. 4.71).

Some of the plots are cultivated year-round taking advantage of the canal system for irrigation while others are limited to cultivation during the rainy season only. During the accuracy assessment (done during the dry season) it was possible to verify which plots were being cultivated in both rainy season and dry season using irrigation from the canals (Fig.72), and those not having access to the canal system and thus limited to cultivation during the rainy season only.

In general terms the use of the plots seems to be quite consistent (from observations done between 2013 and 2017). However, there could be some variation that could easily go undetected. For instance, it is possible that in the year the satellite images were taken (2014) a given plot was cultivated in both seasons and in the year the accuracy assessment was done (2017) it was cultivated only in the rainy season, or vice-versa. This, of course, would only be an issue in areas like this one, in which access to a source of irrigation is possible.

Figure 4.72 – Irrigated agriculture

In this agricultural area there are also many small orchards, most of which also receive irrigation from the canal system (Fig. 4.73). Some plots with orchards are visibly perceptible in the high definition image (especially those with very well defined rows of trees) and can be compared to the classified image. Others, however, are more difficult to identify or distinguish from plots with other kinds of agriculture. This happens especially with orchards.

Figure 4.73 – Orchards with cultivation of seasonal crops amongst the trees and irrigated

Amongst the cultivated plots are areas of natural grassland which some cattle owners use for grazing (Fig. 4.74). Even some plots which would appear to be agricultural plots (in the high definition image) are in fact grassland. They may be “plots” in the sense that they are demarcated, but they are not actively cultivated (or at least have not been actively under cultivation since the time that the field work of this particular study started – 2014).

Figure 4.74 – Grassland in foreground and evergreen needle leaf trees in background

Table 4.15 – Confusion matrix for agricultural land southwest of Humpata City

Reference data

Classified Data	Grassland	Built-up areas	Irrgiated agriculture	Nonirrigated crops	Orchards	Inland water	Total	User's Accuracy
Grassland	2						2	100
Built-up areas		1					1	100
Irrigated agriculture			3			1	4	75
Nonirrigated crops				3			3	100
Orchards	1				1		2	50
Total	3	1	3	3	1	1	12	
Producer's Accuracy	67	100	100	100	100	0		83

The overall accuracy for this area is considerably high (83%) (Table 4.15). Only two pixels (out of twelve) were incorrectly classified. One pixel area classified as orchards was verified to be grassland. This seems to be due to the influence of neighboring pixels seeing as the misclassified pixel is on the border between an area of orchards on one side and an area of grassland on the other. In the case of the other misclassified pixel, the area was classified as an irrigated agricultural area when in fact it was an area occupied by some kind of water reservoir. This error may be due to the vegetation growth inside the actual reservoir seeing as it is not extremely deep and may have enough vegetation growth to cause this type of misclassification.

4.3.2.10 Herbaceous Wetland Area on the Way to Jau

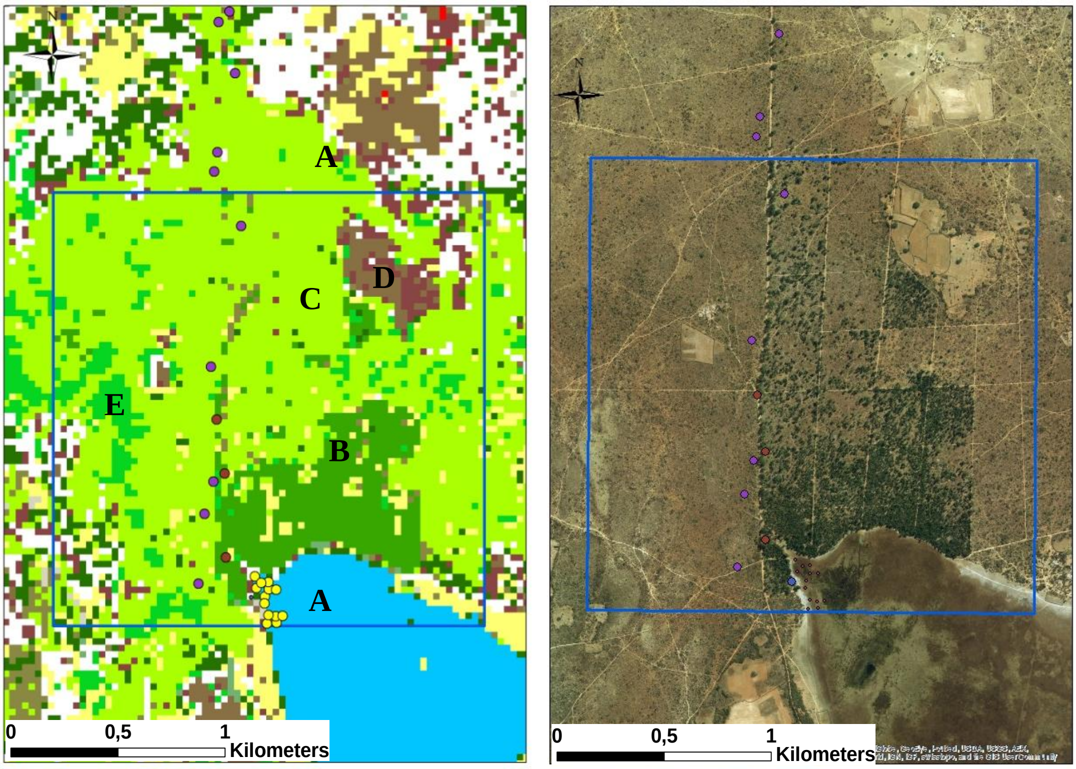


Figure 4.75 – Herbaceous wetland 2x2 km classified area compared to high definition Bing image

This area has the largest accessible representation of the (A) herbaceous wetland in the mapped area bordered to the north by a large area of (B) eucalyptus (Fig. 4.75). The rest of the 2x2 km area is covered largely by (C) acacia thicket as well as one smaller area of (D) agricultural lands. These features are easily identified in the classified image and can easily be compared to the high definition image (Fig. 4.75). However, it is also possible to observe that many pixels in the general acacia thicket region have been classified as (E) evergreen needle leaf trees. This classification is highly questionable given the location, distribution and number of pixels.

Most plantations of the evergreen needle leaf trees observed during the field work were observed on or around agricultural lands and not out in the middle of the acacia thickets. Also the plantations observed during onsite ground visits occupy much smaller and organized areas than the ones classified in this area. This leads to the deduction that these pixels have been misclassified. This conclusion is also congruent with the fact that in the overall assessment 16% of the pixels classified as evergreen needle leaf plantations were in reality observed to be acacia thicket. As previously mentioned in section 4.3, the most logical conclusion would be that these pixel areas are covered by a variation of the acacia thicket, or another kind of indigenous vegetation, that was not sampled during the calibration process. Another possibility is that it is the same kind of vegetation, for example, but with less density than the areas that were sampled. This could explain why its spectral signature would be closer to the orchards which in general have more open space among the trees.

Despite this discrepancy, the accuracy assessment of the sample pixels in this area is the highest of the ten selected areas (Table 4.16). This is most likely due to high overall accuracy of the three main land covers in this area (the herbaceous wetlands, the eucalyptus and the acacia thicket). It is probably also be due to the high concentration of sample pixels representing the herbaceous wetlands. Besides the fact that this class is 100% accurate in the overall assessment, in this area the sample pixels of this class account for 59% of the sample pixels.

Table 4.16 – Confusion matrix for herbaceous wetland area on the way to Jau

Reference Data

Classified Data		Eucalyptus	Acacia thicket	Herbacious wetland	Total	User's Accuracy
	Eucalyptus	3			3	100
	Orchards		1		1	0
	Acacia thicket		5		5	100
	Herbaceous wetlands			13	13	100
	Total	3	6	13	22	
	Producer's Accuracy	100	83	100		95

Another possible reason that the accuracy of this area is so high is that the sampled areas of the land cover in this specific area are quite homogenous. This can be observed in the high definition image as well as in the photos below (Fig. 4.76 and 4.77).

Figure 4.76 – (1) Looking south, eucalyptus (B) on one side of the road and acacia thicket (E) on the other; (2) looking north across herbaceous wetland (A) (dry season) with eucalyptus on the northern edge

Figure 4.77 – Looking north across herbaceous wetland (A) (rainy season) with eucalyptus (B) (northwestern edge) and acacia bushland/thicket (northeastern edge) in the background

4.4 Further Applications of DT Model and Training Set

The application of the decision tree model up to this point has been tested exclusively in the production of the classified map of the study area. The above results clearly show this capacity. However, it is hoped that, once the model is acquired, a broader application of this tool would be possible. At this point, the model has been proven useful in classifying the specific area in which it was calibrated using images of a specific time frame (2013/2014).

Anderson (1976) stipulated that a classification system should be applicable over extensive areas and that comparison with future land use data should also be possible. At this point in the study, the issue is not only if the *classification system* is well suited to more extensive areas and future land cover data. The question is also whether the *decision tree model* (which includes the classification system) can also have a broader application in terms of space or if its application is limited to the area which was used to calibrate it. Another issue is whether the training dataset used to calibrate the model can be used to calibrate models to classify the same area using images of other years. To verify these issues, a further assessment was done to evaluate the capacity of the model to classify (1) an area with similar characteristics (land cover classes and terrain) located in the plateau region of southern Angola and (2) if the training dataset can be used to calibrate a model for classifying the same area using images of a more recent date (2017).

4.4.1 Classification of a Similar Area Using Images of 2013/3014

The area used to calibrate the decision tree model was centered on the Humpata municipality, in the plateau region. An area to the east (Lubango and Chibia municipalities) was selected to assess the capacity of the model to classify a similar area, which was not involved in the calibration of the model. Images of this area (Lubango/Chibia) were then classified using the decision tree model that had been originally calibrated in the Humpata area.

The classified image is presented for visual comparison beside a high definition Bing image (Fig. 4.78). It represents the 2013/1014 land cover (8 Level I classes) of the Lubango/Chibia area at a spatial resolution of 30 m, at a scale of 1:100 000.

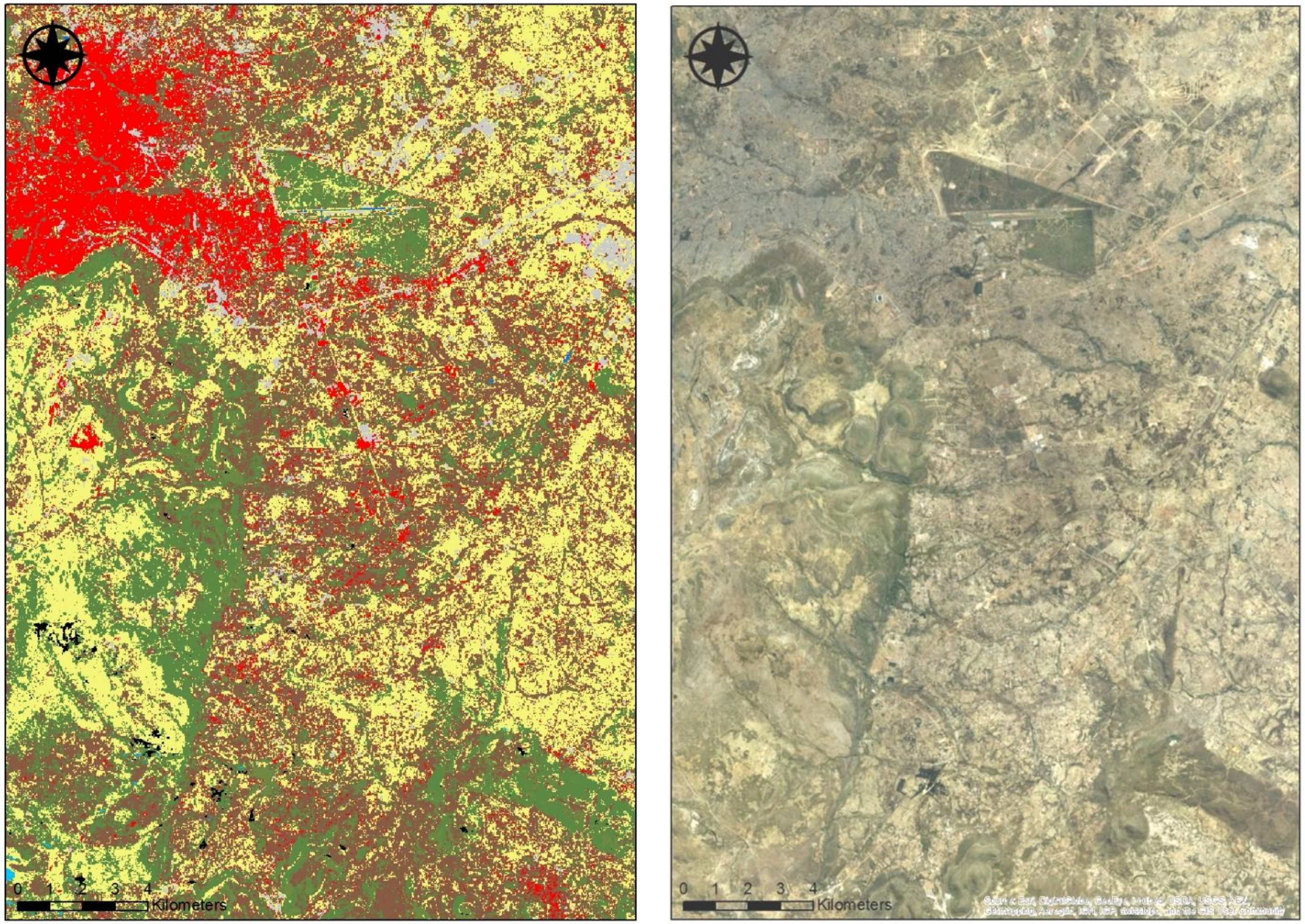


Figure 4.78 – Classified area to the east (Lubango and Chibia) with 8 Level I classes (legend as Annex 1), and high definition Bing image

To ensure an adequate comparison of the originally classified area with the area chosen to the east the same sampling design was used in the assessment, with a couple small variations. Because of time constraints at this point in the study, the verification of the accuracy of the pixels of this classified area would not be done by onsite observations but exclusively using Google Earth images. In this case, road access would not be an issue and therefore the whole area would be used as the population from which to select the sample pixels.

An assessment using only Google Earth images would make it difficult to verify the 16 Level II classes, especially in questions related to clearly distinguishing between many of the vegetation classes. It was therefore decided that the verification would be done using a stratified sampling design in which the 8 Level I classes would define the strata.

Therefore, the classification of the area east of the original area to be assessed would contain the 8 Level I classes instead of the more detailed Level II classes. The results would then be comparable to the assessment of the first classified area with the collapsed 8 Level I classes (section 4.1.3).

The plan was to attribute the same number of sample pixels in each of the 8 collapsed classes as the sum of the sample pixels attributed in the 16 original classes (section 4.2). According to the numbers in Table 4.17, the pixels would be randomly selected in each of the 8 Level I classes (Fig. 4.79), a total of 391 (the same number as in the assessment of the first classified area, section 4.2):

Table 4.17 – Number of samples for Level I land cover classes

Level I	Eucalyptus	Evergreen needleleaf	Miombo with rock	Miombo bushland	Transitional Veg	Acacia thicket	Built-up areas	Grassland	Grassland with rock	Irrgiated agriculture	Nonirrigated crops	Orchards	Barren areas	Inland water	Herbacious wetland	Seasonally burnt	Total
Woodland/ Bushland	25	25	25	25	25	25											150
Urban areas							25										25
Herbaceous vegetation								25	25								50
Cultivated areas										25	25	25					75
Barren areas													25				25
Inland water														25			25
Wetland															16		16
Seasonally burnt areas																25	25

However, when the pixels were randomly selected there were insufficient pixels available in some strata: only 23 seasonally burnt pixels, 4 water pixels and 5 wetland pixels were available for sampling. Therefore, the total number of pixels was 357, instead of the desired 391.

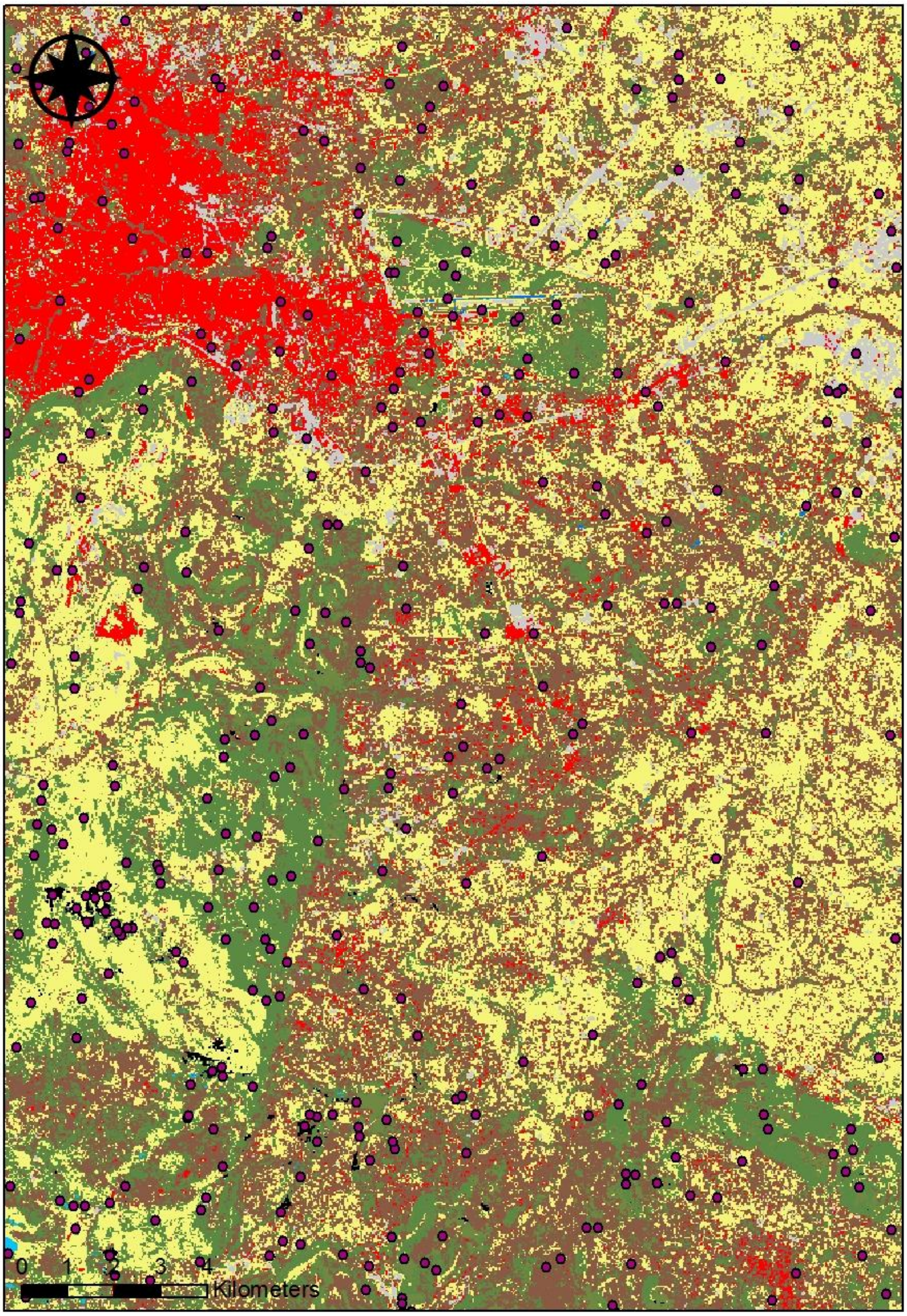


Figure 4.79 – Sample points in the area to the east

These pixels were then verified using Google Earth images. The verification was done using the procedure described in section 3.6.4.2; the same procedure used in the assessment of the original mapped area, for verifying the transitional vegetation on the escarpment and the seasonally burnt areas. Table 4.18 shows the results.

Table 4.18 – Confusion matrix of the Lubango/Chibia classification assessment

Reference Data

Classified Data		Bushland/Woodland	Herbaceous vegetation	Built-up areas	Cultivated areas	Barren areas	Inland water	Wetlands	Seasonally burnt areas	Total	User's Accuracy
	Bushland/Woodland	113	13	2	22	0	0	0	0	150	75
	Herbaceous vegetation	2	46	1	1	0	0	0	0	50	92
	Built-up areas	0	5	18	1	1	0	0	0	25	72
	Cultivated areas	32	14	1	28	0	0	0	0	75	37
	Barren areas	0	6	1	1	17	0	0	0	25	68
	Inland water	0	0	1	0	0	3	0	0	4	75
	Wetlands	0	0	0	0	0	0	5	0	5	100
	Seasonally burnt areas	0	1	0	0	0	0	0	22	23	96
	Total	147	85	24	53	18	3	5	22	357	
	Producer's Accuracy	77	54	75	53	094	100	100	100		71

The overall accuracy and that of most of the classes is lower in this area (Lubango/Chibia) than in the first mapped area (Humpata). This may be due to the fact that the verification was done using Google Earth images instead of ground data. One major conclusion is that there are difficulties in distinguishing with certainty between some classes without onsite observations. For example, the Google Earth images show some areas that are cultivated only in the rainy season as looking very similar to grasslands.

The other reason for the lower accuracy results may be that the calibration of the model was done at higher altitudes than the area being classified. Most of the training samples used to calibrate the model were located on the Humpata plateau (at altitudes higher than 1,750 m a.s.l.) whereas most of the area to the east that is being assessed varies between 1,250 and 1,750 m a.s.l.

The one class with a higher accuracy was the seasonally burnt areas (96% user's accuracy and 100% producer's accuracy). This may be explained by the simple fact that the evidences of burning were still visible in the Google Earth images (more so in this area than in the original mapped area). This could be due to the fact that in the Lubango/Chibia area there is not so much grassland with rocky outcroppings (which would have less vegetation to evidence burning) compared to the first mapped area. Thus the burnt areas sampled in this area probably have more vegetation overall than the ones in the first classified area (especially those with rocky outcroppings) and therefore burnt areas would be more visible in Google Earth images. A logical conclusion would be that these seasonally burnt areas are easier to verify in regions with less rocky outcroppings and more vegetation, causing longer lasting evidences of burning. This again would be true in this situation (and other similar ones) in which the only available Google image was taken 7 weeks after the Landsat 8 images.

If this is the case, it is once again related to the difference in altitude. The higher altitudes would probably have more grassland with rocky outcroppings and the lower altitudes would probably have more grassland without rocky outcroppings.

The water and wetland pixels in this area were very few (4 and 5 respectively). The accuracy of the wetland classification continues to be 100% as in the first mapped area, while the water is lower, 75%. This is probably due to the fact that there were very few points to assess.

The accuracy of the cultivated areas, which was already low in the first mapped area, dropped even further in this area: 37% user's accuracy and 53% producer's accuracy. Once again, some classes are clearly distinguishable using the Google Earth images, while others are not so visible. Therefore, it is difficult to judge what percentage of these results is due to real misclassification.

Again, this difference in accuracy may also be due to the difference in altitude. According to Campbell (1996) the shrub layer of the miombo is variable in density and composition. It may be that the kind of bushland/woodland vegetation in the lesser altitudes is slightly different from that which was sampled in the higher altitudes and perhaps closer in spectral signature to the cultivated areas. The confusion between the bushland/woodland and the cultivated areas was however already evident in the accuracy assessment of the first mapped area (although at a lower percentage). Based on the numbers, the confusion seems to reside primarily between the grassland and nonirrigated crops, and between the bushland/woodland and irrigated agriculture.

4.4.1.1 Statistical Comparison Between the Two Images

Table 4.19 shows a comparison between the overall accuracy of the decision tree analysis, the assessment of the original mapped area (Humpata) and the assessment of the contiguous eastside area (Lubango/Chibia).

Table 4.19 – Statistics comparison between the data of the studied and that used for testing the model

	Percentage Correctly Classified Instances	Kappa Coefficient
DT Model	91,7	0,91
Humpata	81,0	0,76
Lubango/Chibia	71,0	0,61

The DT model results are based on the limited set of sample pixel related to areas controlled on the ground in the original mapped area. The results in the overall original area (Humpata) are representative of a limited set of sample pixel areas visited as well as, certain areas that were not accessed for calibration sampling. Finally, the results of the area to the east (Lubango/Chibia) of the original mapped area are representative of a region which was not sampled at all, but which is considered to be somewhat similar to the original mapped area.

The general results show a likely decrease in accuracy from the DT accuracy (using only visited sample pixels) to the classified area to the east which was *not* sampled at all. Figure 4.80 and 4.81 compare the accuracies by land cover class:

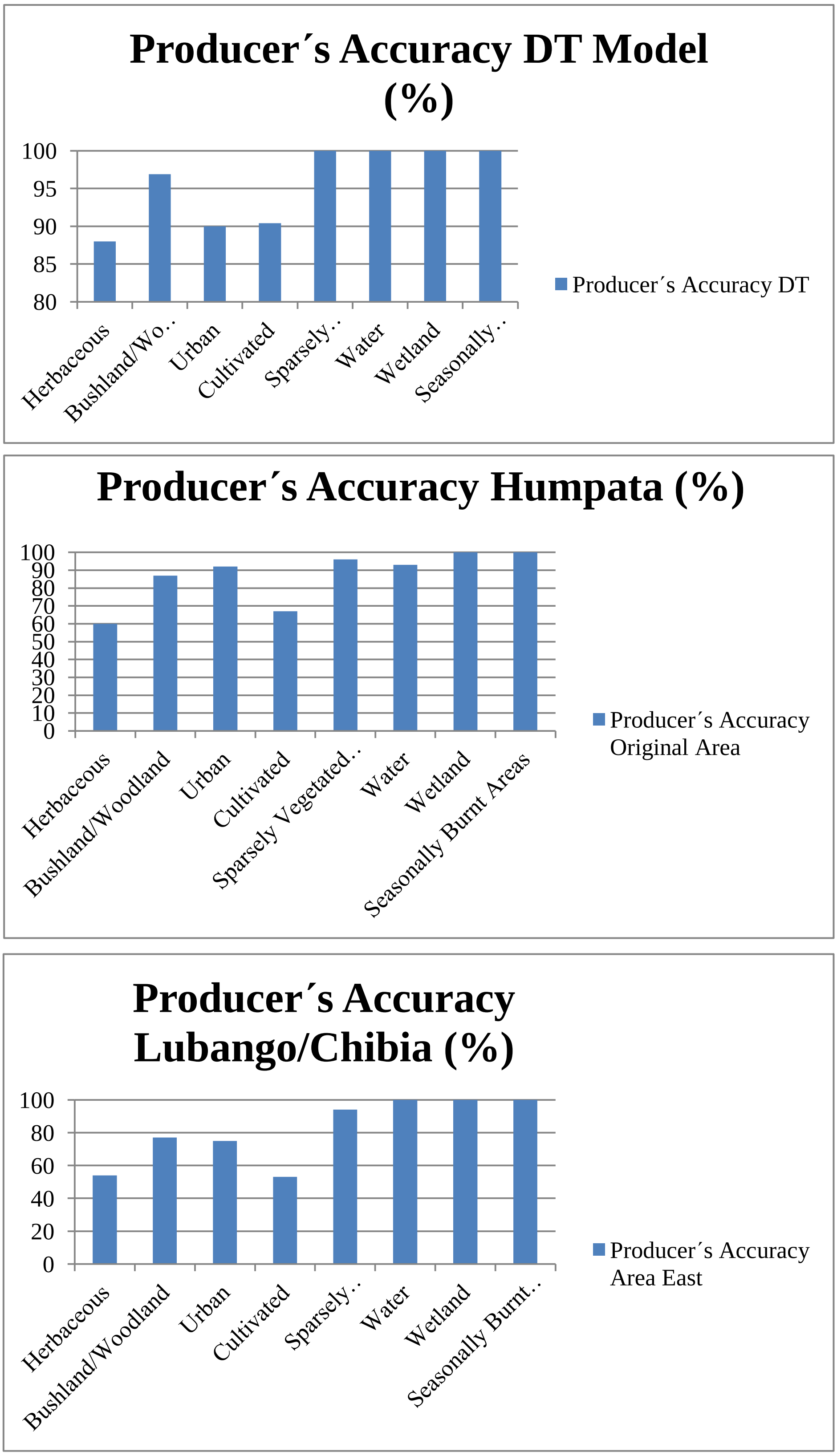
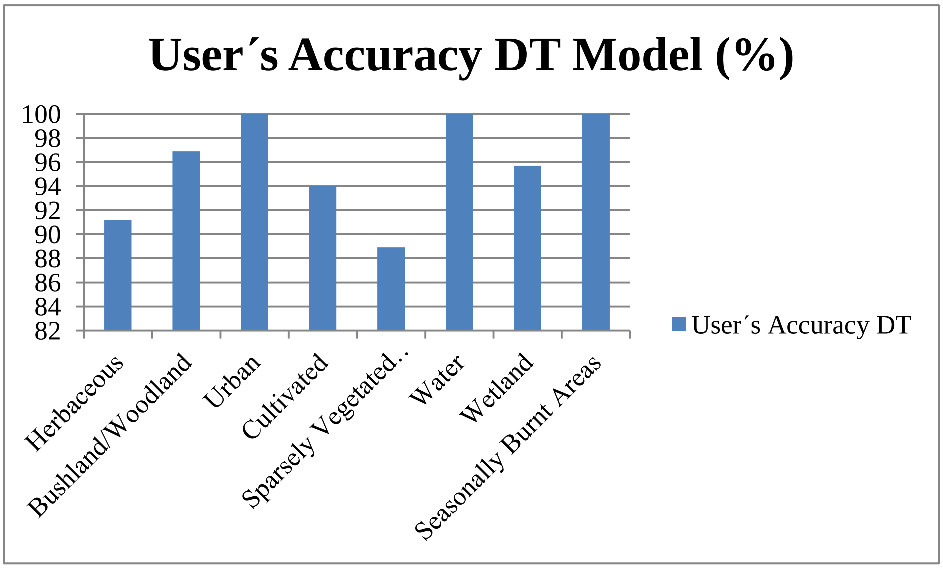
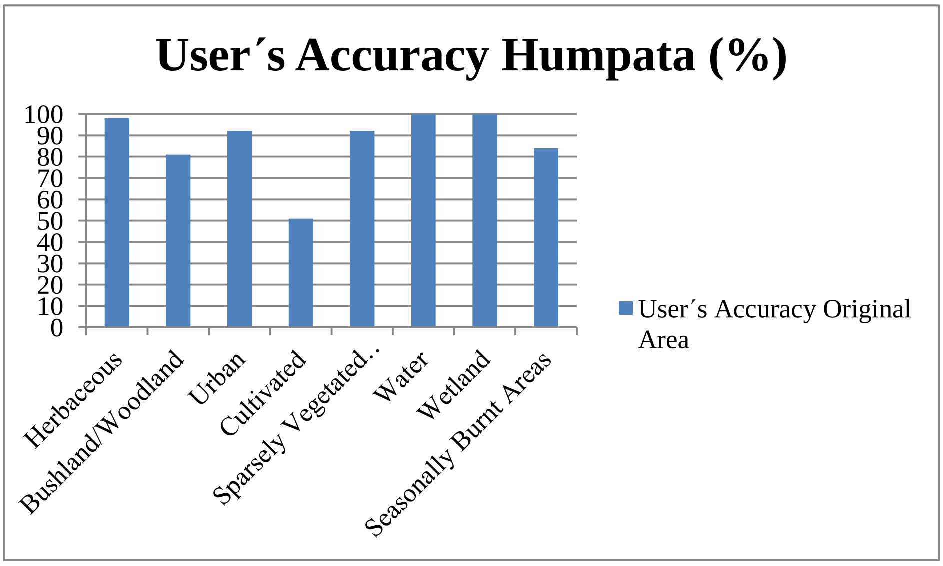


Figure 4.80 – Producer’s Accuracy (DT model, Humpata, Lubango/Chibia)





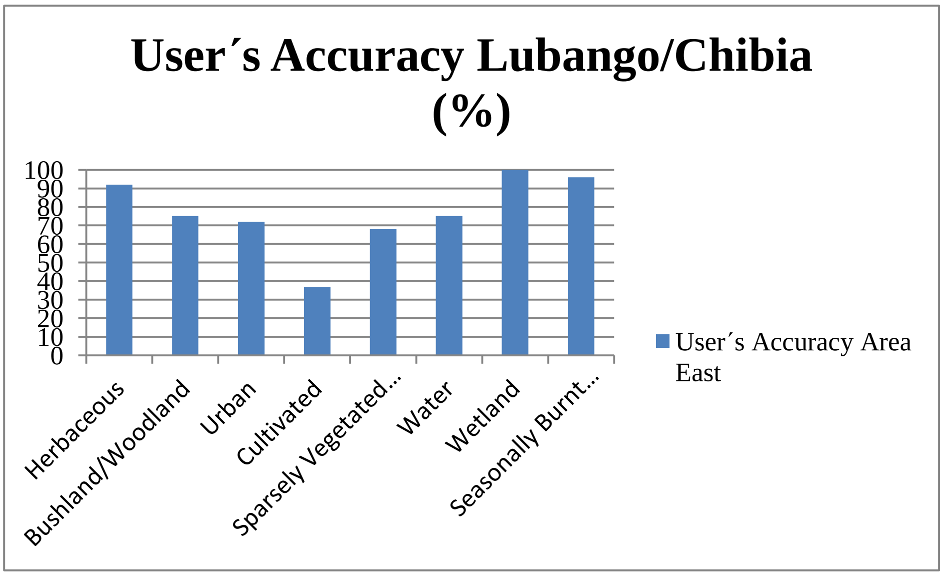


Figure 4.81 – User’s Accuracy (DT model, Humpata, Lubango/Chibia)

Although the water, wetland and seasonally burnt areas stand out as classes having high accuracy levels across the board, the others show, as in the overall results, a general decrease in accuracy from the DT accuracy (using only visited sample pixels) to the original mapped area, to the area east (which was *not* sampled at all). This is not without exceptions however.

Some classes actually show an increase in accuracy even from the decision tree to the mapped area, as are the case of the herbaceous vegetation, the sparsely vegetated areas and the wetland classes which improved in user's accuracy from the decision tree analysis to the original mapped area.

4.4.2 Application of the Training Dataset for 2017 Images

The DT model used to classify the original area was created using images from 2013/2014, one from the dry season and one from the rainy season, using the training dataset described in section 4.2. In order to verify the usefulness of this dataset to calibrate other models for classifying images of different years, more recent Landsat 8 images (Feb 6, 2017 and Aug 17, 2017) of the same area were acquired and processed in the same way as the original images (from 2013/2014).

The dataset presented in section 4.2 was adapted to fit the conditions of the 2017 images. First of all, since the burnt areas are a temporary class (specific to dry season of 2013), these pixels were eliminated from the dataset. Secondly, a visual analysis was made of the rest of the pixels comparing the 2013/2014 images to the 2017 images. Pixels that represented areas that had changed were removed from the data set. The updated dataset was then processed in Weka following the same procedure as described in section 4.3. The results DT analysis, using the 6 bands of the 2017 dry season and rainy season images, and corresponding PCA's, was 87.3% correctly classified instances and a kappa coefficient of 0.86.

These values are lower than in the original procedure but still represent good results. One reason for the lower results is that the burnt areas (which were eliminated) with a very high accuracy rate (100%) also represented part of the overall accuracy rate. Another probable reason they are lower is that some of the sample pixels may represent areas that have changed from 2013/2014 to 2017 and were not detected in the visual analysis. One way of improving these results would be to do a change vector analysis instead of relying solely on a visual analysis to remove pixels representing changed land cover.

Nevertheless, this shows that the training dataset is viable to be used as a baseline for future land cover projects and for calibrating images from different dates.

4.5 Conclusions and Recommendations

This study presents a methodological procedure for creating land cover maps (using Landsat 8 images) including adequate assessment procedures, applicable in the Angolan context with its particular limitations. Even given the time constraints, limited road information and access, as well as the limited overall information concerning the area, the study shows that it is feasible to identify, define and characterize the land cover classes in the study area and create an appropriate and detailed classification system. The results show that using visual analysis of satellite imagery in combination with field work it is possible to train a DT model to satisfactorily classify this kind of study area. The results of this study show, that even with limited access, field work is an important tool in determining land cover and selecting a training set, especially in areas where there is limited available data.

The accuracy assessment procedure developed for the validation of the land cover classification has proven to be adequate for this kind of study area and, even with its unique and unusual set of limitations, has shown statistically sound results.

The on-site verification shows to have provided high quality information for the validation process, which could not have been achieved using only high definition images. This is especially evidenced by the use of the similar validation process performed in an area to the East, using only high definition images for validation. It quickly becomes evident that the sole use of high definition images in the verification process of this area results in serious issues in differentiation and clear identification of vegetation types.

As has been mentioned before, in this kind of study there are always trade-offs. The unique and unusual limitations of the sampling process, i.e. accessibility due to terrain issues and intermittent accessibility by ground transport, reduce the possible area and time available for the training and validation process. This in turn creates the need for the assumption that the land cover observed in the visited areas also represents the land cover of the remaining unvisited areas (due to their inaccessibility). On the other hand, it permits the identification and verification of more detailed classes (especially concerning the vegetation). This would have been impossible using only high definition images either in the training process or in the validation process.

The decision tree model produced for classifying the area has shown itself adequate for classifying the study area with 72% overall accuracy (correctly classified instances) for the 16 Level II classes, and 81% for the 8 Level I classes.

Although some of the specific class accuracies of the Level II class left much to be desired (e.i. the irrigated agriculture, the miombo bushland and the acacia thicket) this detailed information can be used as a basis for further studies into the separation of these land cover classes, especially concerning the different kinds of vegetation. It is however, thanks to the extensive field work (both in the training as well as in the validation process) that it has been possible to include the great variety of bushland/woodland as well as some of the other classes in this land cover mapping.

Analysis of the Level I classes showed that the largest confusion in classification is between the cultivated areas and herbaceous and bushland vegetation. Other Level I classes had quite high individual accuracies varying from 81% to 100%. Using an image from the dry season and from the rainy season made it possible to clearly identify classes that have characteristics which are particular to each season: the seasonally burnt areas (84% UA and 100% PA) and the herbaceous wetlands (both 100% UA and PA).

Experimentation using the same DT model to classify another area to the East produced a decrease in accuracy to 71% correctly classified instances, for the Level I classes. Although this may be partly due to the fact that verification in this case was not done using ground observations, it is probably more due to the fact that the model was not trained in this area and there may be slight variations of certain classes due to specific characteristics of the area (like altitude). Nevertheless this shows that the DT model can also be used to classify other similar areas, albeit with slightly less accuracy.

The verification of the training dataset in calibrating models for classifying images of other years also produced encouraging results. This shows that it is possible to use it in calibrating future models which can then be used as instruments of monitoring the dynamic land cover of the area. This dataset can also serve as a starting point for future studies and perhaps even larger dataset.

This study has produced a land cover map of the study area for 2013/2014. This map in and of itself offers quality information concerning the land cover of the area that can be used as a base for other studies and as a reference for decision makers who work in various government sectors.

Besides this land cover map, the training dataset that has been produced during the study, can be used to produce models for classifying images for different years (using Landsat 8 images). One very obvious application of these tools is in monitoring the development of urban areas. It is possible to monitor the development of urban areas (using images since the launch of Landsat 8). This is extremely relevant in a country like Angola, where this kind of information is basically non-existent. The disorderly and unchecked growth of the peri-urban areas has been a continuous issue in Angola and has had innumerous consequences with predictable environmental and ecological issues for the future.

This dataset can be used to calibrate models to produce relatively quick and continual access to information concerning the quantitative development as well as the directionality of the development of the urban areas in the Lubango and Humpata.

Updated knowledge concerning the gradual increase and development of the urban areas could help in monitoring this growth and more importantly in forecasting tendencies, facilitating proactive planning and organizing future urban development. Using the training dataset to produce models for future images, would make it possible to update the land cover maps on a regular, yearly basis providing continually updated information, without effort having to collect further training samples. The only requirement would be to update the sample set according to the year of the images to be classified.

The advance and growth of urban areas leads to deforestation of the surrounding natural vegetation. In Angola this is also accompanied by burning for charcoal-making. Both these aspects could also be monitored on a yearly basis. This would be vital information for policy-makers, in protecting specific areas. Using these tools, the burning, for example, could be monitored all through the dry season, using more than one image of the dry season. This would provide information of the tendencies of burning not only throughout the dry season but also from one year to the next.

Having tested the DT model in another area in which it was not calibrated, and having been successful to a certain extent, this same model can been used to classify similar areas with the same expected results. Even with a lower accuracy, this tool can still be extremely useful to local authorities. In a context in which there are simply no available land cover maps having a tool that can be used to produce land cover maps (even with a lower accuracy) of areas that have no information would be extremely useful. This model however should be used in areas that have at least similar conditions to the original study area in which it was produced, with similar vegetation and in the plateau area.

It is recommended however, that further experimentation with the model be done in other areas of the plateau to assess its accuracy in more diverse situations. The results could clarify its limitations or determine a broader capacity. Another recommendation that can be made would be to do further studies in the area using time series to increase the accuracy especially of the vegetation and agriculture classes (Khatami et al., 2016) as in the study by Massey (2017).

Although the DT model is limited to the specific land cover described in the study area, the methodology could be repeated to create other models for other regions. This is obviously a much more arduous process than simply applying an existing model. However, having a defined methodological procedure is half the work. An interesting prospect would be to apply the methodology in an area towards the coast below the escarpment.

5 – BIBLIOGRAPHY

Aburas, M. M., Abdullah, S. H., Ramli, M. F., & Asha'ari, Z. H. (2017). Land Suitability Analysis of Urban Growth in Seremban Malaysia, Using GIS Based Analytical Hierarchy Process. *Procedia Engineering*, *198,* 1128-1136.

Al-doski, J., Mansorl, S. B., & Shafri, H. Z. M. (2013). Image classification in remote sensing. *Department of Civil Engineering, Faculty of Engineering, University Putra, Malaysia*.

Alvin W. Urquhart. (1963). *Patterns of settlement and subsistence in southwestern Angola* (No. 18). National Academies.

Amaral, I. (1978). Contribuição para o conhecimento do fenómeno de urbanização em Angola. *Finisterra*, *13*(25).

Anderson, J. R. (1976). *A land use and land cover classification system for use with remote sensor data* (Vol. 964). US Government Printing Office.

Angop, 2016 [WebPage] http://www.angop.ao/angola/pt_pt/noticias/reconstrucao-nacional/2016/0/4/Huila-Centralidade-Quilemba-pronta-para-ser-habitada,b985b449-3f3e-4a72-85cc-7b334c667857.html. Accessed 06-05-18.

Anule, P., & Ujoh, F. (2017). GEOSPATIAL ANALYSIS OF WETLANDS DEGRADATION IN MAKURDI, NIGERIA. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*.

Appiah, A., & Gates, H. L. (Eds.). (2005). *Africana: The encyclopedia of the African and African American experience*. Oxford University Press, USA.

Asamoah, J. N., Jnr, E. M. O., Acquah, P. C., & Amoah, A. S. (2018, March). Comparison of Decision Tree and Maximum Likelihood Using a Landsat Image of Ejisu-Juaben Municipality. In *International Conference on Applied Science and Technology Conference Proceedings* (Vol. 4, No. 1, pp. 200-210).

Austin, K. G., Mosnier, A., Pirker, J., McCallum, I., Fritz, S., & Kasibhatla, P. S. (2017). Shifting patterns of oil palm driven deforestation in Indonesia and implications for zero-deforestation commitments. *Land Use Policy*, *69*, 41-48.

Bahu, A. M. (2015). *A comercialização de carvão vegetal versus insustentabilidade das florestas naturais baseado num estudo de caso das aldeias de Nazaré e Calombo nos Municípios de Lomgonjo e Caála, Província do Huambo, Angola* (Doctoral dissertation, Universidade José Eduardo dos Santos).

Banko, G. (1998). A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory.

Barbosa, L. A. (1970). Carta Fitogeográfica de Angola. Instituto de Investigação Científica de Angola. *Luanda, Angola*.

Bastos, C. (2008). Migrants, settlers and colonists: the biopolitics of displaced bodies. *International Migration*, *46*(5), 27-54.

Beernaert, F. (1997). Development of a soil and terrain map/database for Angola. *Institute for Land and Water Management Catholic University of Leuven (Belgium)*.

Beetz, P. F. W. (1934). Geology of south west Angola, between Cunene and Lunda axis. *South African Journal of Geology*, *37*(Transactions 1934), 137-176.

Bergh, N. G., & Nordenstam, B. (2010). Psednotrichia perennis (Asteraceae, Senecioneae): A new species from the Huila plateau, Angola. *South African Journal of Botany*, *76*(2), 369-374.

Benazeraf, D. (2014). 'Oil for Housing': Chinese-built New Towns in Angola.

Brito, R. S. (1977). Acerca da colonização das Terras Altas da Huíla (Angola). *Finisterra*, *12*(24).

Bruyns, P. V. (2010). Huernia humpatana (Apocynaceae), a new species from southern Angola. *South African Journal of Botany*, *76*(3), 585-587.

Cabral, A. I. (2007). cartografia de coberto do solo para o território angolano utilizando imagens de satélite modis. *Revista Portuguesa de Estudos Regionais*, (15).

Cain, A. (2007). Housing microfinance in post-conflict Angola. Overcoming socioeconomic exclusion through land tenure and access to credit. *Environment and urbanization*, *19*(2), 361-390.

Campbell, B. M. (Ed.). (1996). *The Miombo in transition: woodlands and welfare in Africa*. Cifor.

Censo 2014 [WebPage] http://**censo**.ine.gov.ao/xportal. Accessed on 06-05-18.

Chidumayo, E. N. (2010). *The dry forests and woodlands of Africa: managing for products and services*. Earthscan.

Cihlar, J. (2000). Land cover mapping of large areas from satellites: status and research priorities. *International journal of remote sensing*, *21*(6-7), 1093-1114.

Chiquete, P. S. dos S. (2012). Aplicação de técnicas de detecção remota à cartografia temática: O caso da região do Lobito (Angola). Tese de Mestrado em Geociências – Ramo Ambiente e Ordenamento. Faculdade de Ciências e Tecnologia – Universidade de Coimbra.

Chisingui, A. V. (2017). Análise da paisagem e das alterações de uso/ocupação do solo no Lubango e arredores. Tese apresentada à Universidade de Évora
para obtenção do Grau de Doutor em Ciências do Ambiente

Clarence-Smith, W. G. (1976). The thirstland trekkers in Angola-Some reflections on a frontier society. In *Collected Seminar Papers. Institute of Commonwealth Studies* (Vol. 20, pp. 42-51). Institute of Commonwealth Studies.

Clark, V. R., Barker, N. P., & Mucina, L. (2011). The Great Escarpment of southern Africa: a new frontier for biodiversity exploration. *Biodiversity and Conservation*, *20*(12), 2543.

Climate-Data [WebPage] https://en.climate-data.org/region/1387/ Accessed 06-05-18

Colditz, R. R. (2015). An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. *Remote Sensing*, *7*(8), 9655-9681.

Congalton, R. G. (1991a). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, *37*(1), 35-46.

Congalton, R. G. (1991b). Remote sensing and geographic information system data integration: Error sources and. *Photogrammetric Engineering & Remote Sensing*, *57*(6), 677-687.

Conradie, W., Branch, W. R., Measey, G. J., & Tolley, K. A. (2012). A new species of Hyperolius Rapp, 1842 (Anura: Hyperoliidae) from the Serra da Chela mountains, south-western Angola. *Zootaxa*, *3269*(1), 1-17.

Correia, H. (1976). O Grupo da Chela e Formação da Leba como novas unidades litoestratigráficas resultantes da redefinição da Formação da Chela na região do Planalto da Humpata (Sudoeste de Angola). *Boletim da sociedade Geológica de Portugal* 20: 65-130.

Croese, S. (2012). One million houses? Chinese engagement in Angola's national reconstruction. *China and Angola: a marriage of convenience*, 124-138.

Danielson, P., Yang, L., Jin, S., Homer, C., & Napton, D. (2016). An Assessment of the Cultivated Cropland Class of NLCD 2006 Using a Multi-Source and Multi-Criteria Approach. *Remote Sensing*, *8*(2), 101.

De Fries, R. S., Hansen, M., Townshend, J. R. G., & Sohlberg, R. (1998). Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, *19*(16), 3141-3168.

Delgado-Matas, C., & Pukkala, T. (2011). Comparison of the growth of six Eucalyptus species in Angola. *International Journal of Forestry Research*, *2011*.

Delgado-Matas, C., & Pukkala, T. (2012). Growth and yield of nine pine species in Angola. *Journal of forestry research*, *23*(2), 197-204.

Domingos, J. C. (2014). *A importância da floresta indígena e exótica no desenvolvimento económico e social de Angola: situação actual e potencialidades* (Master's thesis, Universidade de Évora).

European Space Agency [WebPage] http://www.esa.int/Our_Activities/Observing_the_Earth/Space_for_our_climate/Africa_classified. Accessed on 05-05-18

Encyclopedia of Earth. Environmental Information Coalition, National Council for Science and the Environment, Washington. [WebPage] http://www.eoearth.org./article/Angolanmontane_forest-grassland_mosaic. Accessed 18-01-18 2018.

Encyclopedia Britannica Online. [WebPage] https://www.britannica.com/place/Angola. Accessed 18-01-18

Finn, J. A., et al. (2017). Predicting the distribution of High Nature Value farmland in the Republic of Ireland. In: *Grassland resources for extensive farming systems in marginal lands: major drivers and future scenarios. Proceedings of the 19th Symposium of the European Grassland Federation, Alghero, Italy, 7-10 May 2017*. CNR-ISPAAM. p. 274-276.

Food and Agriculture Organization of United Nation (FAO). 1996. Contribuição do sector florestal no desenvolvimento rural, in Angola: Agricultural recovery and development options review.

Food and Agriculture Organization of the United Nations (FAO). (1998). Wetland characterization and classification for sustainable agricultural development. *Harare: Sub-Regional Office for East and Southern Africa (SAFR).*

Food and Agriculture Organization of the United Nations (FAO). (2010). Evaluation Des Ressources Forestières Mondiales. Rapport National Angola. http://www.fao.org/forest-resources-assessment/current-assessment/country-reports/en. Accessed 15-06-18.

Fonseca, A. D. and Fernandes, J. C. (2004). Detecção Remota. 1ª Edição. Lisboa: Lidel

Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment*, *80*(1), pp.185-201.

Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography*, *26*(2), 173-205.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, *61*(3), 399-409.

Hansen, M. C., DeFries, R. S., Townshend, J. R., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International journal of remote sensing*, *21*(6-7), 1331-1364.

Hedberg, I., & Hedberg, O. (1968). Conservation of vegetation in Africa south of the Sahara: proceedings of a symposium held at the 6th plenary meeting of the" Association pour l'étude taxonomique de la flore d'Afrique tropicale"(AETFAT) in Uppsala, Sept. 12th-16th, 1966. Sv. växtgeografiska sällsk.

Heinimann, A., Mertz, O., Frolking, S., Christensen, A. E., Hurni, K., Sedano, F., ... & Hurtt, G. (2017). A global view of shifting cultivation: Recent, current, and future extent. *PloS one*, *12*(9), e0184479.

Holden, D. (2015). Transformação do uso e ocupação do solo na cidade do Lubango no período de 1978 a 2010 – Relações físicas e societais. *Tese de Mestrado em Geociências – Ramo Ambiente e Ordenamento. Faculdade de Ciências e Tecnologia – Universidade de Coimbra.*

Holmes, K. W., Van Niel, K., Kendrick, G., & Baxter, K. (2006). Designs for remote sampling: review, discussion, examples of sampling methods and layout of scaling issues. *Technical report, Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management, Australia*.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., ... & Wickham, J. (2007). Completion of the 2001 national land cover database for the counterminous United States. *Photogrammetric engineering and remote sensing*, *73*(4), 337.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., ... & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, *81*(5), 345-354.

Hui, Y., Zhang, R., Li, X. (2009). Classification of wetland from TM imageries based on
decision tree. WSEAS Trans. Info. Sci. and App 6 (7), 1155–1164.

Huntley, B. J. (1974). Outlines of wildlife conservation in Angola. *South African Journal of Wildlife Research-24-month delayed open access*, *4*(3), 157-166.

Huntley, B. J., & Matos, E. M. (1994). Botanical diversity and its conservation in Angola. *Botanical diversity in southern Africa.((Strelitzia*, *1*, 53-74.

Janssen, L. L. F., & Van der Wel, F. J. M. (1994). Accuracy assessment of satellite derived land-cover data: a review. Photpgrammeric and Remote Sensing, 60: 419-426.

Jenkins, P., Robson, P., & Cain, A. (2002). Local responses to globalization and peripheralization in Luanda, Angola. *Environment and Urbanization*, *14*(1), 115-127.

Khatami, R., Mountrakis, G., & Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, *177*, 89-100.

Kimball, H. L., Selmants, P. C., Moreno, A., Running, S. W., & Giardina, C. P. (2017). Evaluating the role of land cover and climate uncertainties in computing gross primary production in Hawaiian Island ecosystems. *PloS one*, *12*(9), e0184466. , 12.9:

Kumar, K. (2016). *Land and water resource management in Gagas River Valley Watershed using remote sensing and GIS* (Doctoral dissertation, GB Pant University of Agriculture and Technology, Pantnagar-263145 (Uttarakhand)).

Li, C., Wang, J., Wang, L., Hu, L., & Gong, P. (2014). Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sensing*, *6*(2), 964-983.

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, *28*(5), 823-870.

Malmer, Anders (2007, February). General ecological features of miombo woodlands and considerations for utilization and management. In *MITMIOMBO–Management of Indigenous Tree Species for Ecosystem Restoration and Wood Production in Semi-Arid Miombo Woodlands in Eastern Africa. Proceedings of the First MITMIOMBO Project Workshop held in Morogoro, Tanzania* (pp. 6-12).

Mantas, V. M., Marques, J. C., & Pereira, A. J. (2016). A geospatial approach to monitoring impervious surfaces in watersheds using Landsat data (the Mondego Basin, Portugal as a case study). *Ecological indicators*, *71*, 449-466.

Mariyappan, R., & Perumal, S. (2004). Land use changes: The initiator of environmental problem.

Massey, R. (2017). *Satellite Remote Sensing of Cropland Characteristics in 30m Resolution: The First North American Continental-Scale Classification on High Performance Computing Platforms* (Doctoral dissertation, Northern Arizona University).

Mehmood, R., Butt, M. A., Mahmood, S. A., & Ali, F. (2017). Appraisal of Urban Heat Island and Its Impacts on Environment Using Landsat TM in Peshawar, Pakistan. *Advances in Remote Sensing, 6*(03), 192.

Millones, M., Rogan, J., II, B., Parmentier, B., Harris, R. C., & Griffith, D. A. (2017). Fire Data as Proxy for Anthropogenic Landscape Change in the Yucatán. *Land*, *6*(3), 61.

Morse-McNabb, E., Sheffield, K., Clark, R., Robson, S., Lewis, H., & Morse, E. (2012). Calibration and Validation of State Wide Land Cover Mapping. In *GSR*.

Muchoney, D. M., & Strahler, A. H. (2002). Pixel-and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data. *Remote Sensing of Environment*, *81*(2-3), 290-299.

Nusser, S. M., & Klaas, E. E. (2003). Survey methods for assessing land cover map accuracy. *Environmental and Ecological Statistics, 10*(3), 309-331.

Olson, D. M., & Dinerstein, E. (1998). The Global 200: a representation approach to conserving the Earth's most biologically valuable ecoregions. *Conservation Biology*, *12*(3), 502-515.

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, *12,* S27-S31.

Padonou, E. A., Lykke, A. M., Bachmann, Y., Idohou, R., & Sinsin, B. (2017). Mapping changes in land use/land cover and prediction of future extension of bowé in Benin, West Africa. *Land Use Policy, 69*, 85-92.

Pal, M. and Mather, P.M. (2001). Decision tree based classification of remotely sensed data. In *Paper presented at the 22nd Asian Conference on Remote Sensing* (Vol. 5, p. 9).

Pal, M. and Mather, P.M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment*, *86*(4), pp.554-565.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences discussions*, *4*(2), 439-473.

Pereira, E., Tassinari, C. C., Rodrigues, J. F., & Van-Dúnem, M. V. (2011). New data on the deposition age of the volcano-sedimentary Chela Group and its Eburnean basement: implications to post-Eburnean crustal evolution of the SW of Angola. *Comunicações Geológicas*.

Prasad, M. S., Savithri, T. S., & Krishna, I. V. M. (2015). Techniques iIn Image Classification; A Survey. *Global Journal of Research In Engineering*.

Quinlan, J.R., (1992). Learning with Continuous Classes. In: 5th Australian Joint Conference on Artificial Intelligence. World Scientific, Singapore, pp. 343–348.

Remmel, T. K., & Perera, A. H. (2017). Portraying Wildfires in Forest Landscapes as Discrete Complex Objects. In *Mapping Forest Landscape Patterns* (pp. 105-145). Springer, New York, NY.

Reyes-Gonzalez, A. (2017). Using Remote Sensing to Estimate Crop Water Use to Improve Irrigation Water Management.

Robson, P., & Roque, S. (2001). 'Here in the city, everything has to be paid for': locating the community in peri-urban Angola. *Review of African Political Economy*, *28*(90), 619-628.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93-104.

Rodrigues, Cristina U. (2009). Angolan cities: urban (re) segregation?. In *African cities* (pp. 37-54). Brill.

SASSCAL. (2016) [WebPage]
http://www.sasscal.org/wp-content/uploads/external-ssets/tasksheets/task_154_ao_bi.pdf. Accessed 05-05-18

Sebadduka, J. L. (2014). Mapping soil-landscape elements and the wetland in dambos and estimating CH4 and CO2 emissions from a dambo-terminated catena.

Kuching, S. (2007). The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis. *Journal of Computer Science*, *3*(6), 419-423.

Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, *70*, 78-87.

Shaw, H. A. (1947). The vegetation of Angola. *The Journal of Ecology*, 23-48.

Stehman, S. V. (2001). Statistical rigor and practical utility in thematic map accuracy assessment. *Photogrammetric Engineering and Remote Sensing, 67*(6), 727-734.

Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, *30*(20), 5243-5272.

Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote sensing of environment*, *64*(3), 331-344.

Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., ... & Woodcock, C. E. (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *European Communities, Luxembourg*, *51*(4).

Timberlake, J., & Chidumayo, E. (2011). Miombo ecoregion vision report. *Occasional Publications in Biodiversity*, *20*.

Tsendbazar, N. E., De Bruin, S., & Herold, M. (2015). Assessing global land cover reference datasets for different user communities. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 93-114.

Tuholske, C., Lopéz-Carr, D., & Roberts, D. (2017). Anthropogenic impacts on Roatán, Honduras: Years of land-cover and land-use change. *Plurimondi*, (16).

U.S. Geological Survey (2001). Landsat 7 Science Data User's, Project Science Office, NASA, Maryland

U.S. Geological Survey. (2016a) Landsat 8 (L8) Data Users Handbook**.** Sioux Falls: EROS, 106p. [WebPage] https://landsat.usgs.gov/landsat-8-l8-data-users-handbook. Acessed 15-06-18.

U.S. Geological Survey. (2016b). National Land Cover Database 2006 (NLCD2006) Product Legend. [WebPage] http://www.mrlc.gov/nlcd06_leg.phpNLCD 2006. Accessed 15-01-18.

Vanden Bossche, J. P., & Bernacsek, G. M. (1990). *Source book for the inland fishery resources of Africa* (Vol. 1). Food & Agriculture Org..

Vasconcelos, Francisco. B. (2015). *Desmatamento para produção de carvão e suas consequências. Estudo de caso: Comparação entre o Município do Lubango (Província da Huíla) e o Município da Bibala (Província do Namibe)* (Doctoral dissertation, Universidade José Eduardo dos Santos).

Vela, Hervé. (2015). *Cartografia de ocupação do solo derivada de imagens de satélite. Caso de Estudo: Município da Humpata. Angola* (Doctoral dissertation).

Vushoma, P. M. (2016). *Assessment of groundwater vulnerability in Norton Town, Zimbabwe* (Doctoral dissertation).

Wachilala, Piedade, Duarte, Isabel M., & Pinho, António B. (2016). Methods of earth building in the Huila Province, Angola.

Waser, L. T., Boesch, R., Wang, Z., & Ginzler, C. (2017). Towards Automated Forest Mapping. In *Mapping Forest Landscape Patterns* (pp. 263-304). Springer, New York, NY.

White, R. P., Murray, S., Rohweder, M., Prince, S. D., & Thompson, K. M. (2000). *Grassland ecosystems* (p. 81). Washington, DC, USA: World Resources Institute.

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., & Cunningham, S. J. (1999). Weka: Practical machine learning tools and techniques with Java implementations.

World Wildlife Fund, McGinley M (2008a) Angolan montane forest-grassland mosaic. In: Cleveland CJ (ed) Encyclopedia of Earth. Environmental Information Coalition, National Council for Science and the Environment, Washington. http://www.eoearth.org./article/Angolan_montane_forest-grassland_ mosaic. Accessed 15-06-17.

World Wildlife Fund, McGinley M (2008b) Angolan montane forest-grassland mosaic. In: Cleveland CJ (ed) Encyclopedia of Earth. Environmental Information Coalition, National Council for Science and the Environment, Washington. http://www.eoearth.org./article/Angolan_mopane_woodlands. Accessed 15-06-17.

Xian, G., Homer, C., Meyer, D., & Granneman, B. (2013). An approach for characterizing the distribution of shrubland ecosystem components as continuous fields as part of NLCD. *ISPRS journal of photogrammetry and remote sensing*, *86*, 136-149.

Yang, H., Li, S., Chen, J., Zhang, X., & Xu, S. (2017). The standardization and harmonization of land cover classification systems towards harmonized datasets: a review. *ISPRS International Journal of Geo-Information*, *6*(5), 154.

Zhao, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., ... & Radke, J. D. (2016). Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data. *Remote sensing of environment*, *183*, 170-185.

Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., ... & Auch, R. F. (2016). Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS Journal of Photogrammetry and Remote Sensing*, *122*, 206-221.

Annex 1 – Level II and Level I Classified Maps of Humpata Area