Enrivonmenta Project in Corioco, Boliva
A Case Study

Land Cover Classification by Remote Sensing

By Thomas Asbjørn Thomsen

Master Thesis M.Sc. Physical Geography at Aalborg University,

2014/2015

Aalborg University

School of Engineering and Science
Sofiendalsvej 9
9000 Aalborg

Title: Envionmental Project in Coroico, Bolivia

- A Case Study

Project period: Fall 2014 & Spring 2015

Deltagere:

Thomas Asbjørn Thomsen

Vejleder: Morten Lauge Pedersen

Oplægsantal: 3
Antal sider: 50
Antal sider i bilag: 7

The thesis will feature two main objectives, that will be referred to as the *Aims of the Thesis*. The Aims of the Thesis are based on conversation with EWB about the goals of their work in Coroico, combined with what is possible to do in the span of a year.
Of greatest importance to EWB was getting good maps of the Coroico Municipality in order to create a basis for their GIS-work in the region in future years.
The second goal of EWB is to point out certain areas that are contaminated by pesticides, in order to protect the locals from pesticide poisoning.
The mapping of contaminated sites will be carried out as a methodological project, where it will be explained how the mapping of pesticide contamination in the area could be approached, once the time is right for that.
The Aims of the Thesis can be seen as process, where it is necessary to complete the tasks chronologically. The Aims of the Thesis are:

1) Create a land use/land cover map of the Coroico Municipality.
2) Detect areas susceptible to pesticide contamination.

The first part of the thesis will focus on getting the GIS-data, and converting it in QGIS and ArcMap in order to create a useful map of Coroico. The creation of the map of Coroico will be beneficial in the second state of the thesis, when screening for potential contaminated sites. The screening will be done through Multiple-Criteria Decision Analysis (MCDA).

Preface

This thesis has been completed in the period from 1/9-2014 to 10/6-1015 and is worth 50 ECTS. This concludes my master of science in Physical Geography at the University of Aalborg.

I would like to thank my advisor Morten Lauge Pedersen, as well as Rikke Østergaard at Engineers Without Borders, and the numerous other volunteer at Engineers Without Borders.

A big thank also goes out to my three class mates during this master, Søren Land Nielsen, Henrik Rosenskjold and Ronni Fjordvald Søe. I can honestly say that I would have done this without your inputs and inspiration.

Thomas Asbjørn Thomsen
Aalborg 10/6-2015

Table of Contents

1 INTRODUCTION

Coroico is a municipality in the South American country of Bolivia. Currently the municipality is suffering from pesticide contamination of the drinking water as well as littering in the form of solid waste. The municipality and its' inhabitants, along with the local university (Unidad Académica Campesina – Carmen Pampa) and Engineers Without Borders (from hereon EWB), have decided to develop a local environmental action plan, to solve the two aforementioned environmental challenges. (Ingeniører uden Grænser, 2013)

EWB is a humanitarian organization largely comprised by volunteers. It is an international organization, but has a Danish division, named "Ingeniører uden Grænser". The Danish division has it's main seat in Copenhagen, but has branches in other cities, among those Aalborg.

On my 8th semester I was introduced to EWB, while looking into a study trip to Kyrgyzstan. The core values of the organization – voluntary involvement without subjugating to political or religious opinions (Ingeniører uden Grænser, 2013) – appealed to me, and went hand in hand with my initial wishes from when I began my geography studies. I therefore started looking into the possibility of doing my master thesis with the organization, and in the late summer of 2014 it became reality, through the goodwill of Rikke Østergaard, Project Leader of "Environmental Project in Coroico, Bolivia".

One of the big obstacles for EWB in Coroico is the lacking availability of useful digital maps, and I was assigned the task of helping with the creation of these through GIS (Geographical Information Systems). In order to minimize cost for it's partners in Coroico, EWB requested that the (GIS) application used in the project is cost free. As this is the case with QGIS, that program was chosen. For the project QGIS 2.6.0 – Brighton, which is the newest QGIS-software (as per November 24th, 2014), will be used. While QGIS includes most of the features present in most other GIS-programs, it have proved troublesome when asked to classify landscapes. Hence landscape classification will be created in ArcMap, while the rest of the maps will be created in QGIS. The maps will be composed of both vector and raster data.

The maps used in the thesis are mainly centered around the municipality of Coroico, but a few maps in the introduction of the thesis show the location of Bolivia in South America, the Department of La Paz, the Nor Yungas Providence and the Coroico Municipality within Bolivia.

2 AIMS OF THE THESIS

As opposed to most other thesis', this thesis will not have a specific *Problem Statement*. Instead, it will feature two main objectives, that will be referred to as the *Aims of the Thesis*. The Aims of the Thesis are based on conversation with EWB about the goals of their work in Coroico, combined with what is possible to do in the span of a year. Of greatest importance to EWB was getting good maps of the Coroico Municipality in order to create a basis for their GIS-work in the region in future years. The second goal of EWB is to point out certain areas that are contaminated by pesticides, in order to protect the locals from pesticide poisoning.

As time has progressed during the project period, it has become apparent that the work of EWB in Coroico will not be concluded by the time this thesis is due. The initial goal was to attain GPS-points of contaminated sites, and have those mapped. However getting these points have not been possible within the timeframe of the thesis. The mapping of contaminated sites will therefore be carried out as a methodological project, where I will explain how the mapping of pesticide contamination in the area could be approached, once the time is right for that.

The three Aims of the Thesis can be seen as process, where it is necessary to complete the tasks chronologically. The Aims of the Thesis are:

1) Create a land use/land cover map of the Coroico Municipality.
2) Detect areas susceptible to pesticide contamination.

The first part of the thesis will focus on getting the GIS-data, and converting it in QGIS and ArcMap in order to create a useful map of Coroico. The creation of the map of Coroico will be beneficial in the second state of the thesis, when screening for potential contaminated sites. The screening will be done through Multiple-Criteria Decision Analysis (MCDA).

3 DESCRIPTION OF STUDY AREA

This project is based on the city and municipality of Coroico in Bolivia. Bolivia is located in the central part of South America, West of Brazil (Figure 1). Bolivia is landlocked although only a couple of hundred kilometers of land in Peru and Chile separates it from the Pacific Ocean. As of 2013 Bolivia had a population of 10.671.200 (The World Bank, 2014) and covers an area of 1.090.249 km^2 and is split into nine departments, each with an administrative center (Figure 2). The Coroico Municipality is located in the Nor Yungas providence in the department of La Paz, where the country capital of the same name is also located. The municipality of Coroico covers 952 km^2 and covers approximately three fourths of the 1.298 km^2 of the Nor Yungas providence – the municipality of Coripata accounts for the remaining 346 km^2. As of 2010 there were 14.329 inhabitants in the Coroico Municipality (ESRI, u.d.).

Figure 1: Bolivia in South America (ESRI, u.d.).

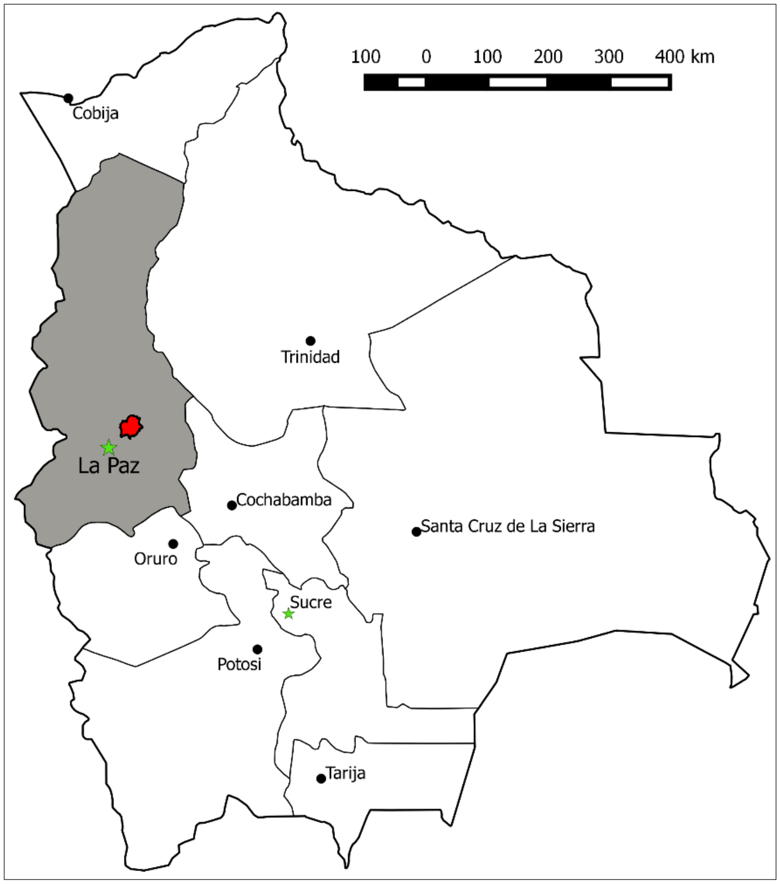

Figure 2: The nine departments of Bolivia, and their administrative centers. The municipality of Coroico is marked by a red color (ESRI, u.d.) (GeoBolivia, 2014).

The Coroico Municipality is located in the Andes Region of Bolivia. The city itself is located at an altitude of approximately 1700 m, and around the city, in the municipality, there are mountains as high as 4300 m, while the lowest river valleys are located at an altitude of 800 m (GeoBolivia, 2014). In Figure 3 the elevation of the municipality is displayed. With differences in elevation of such an extend, the slopes in the municipality are expected to be very steep, something that could potentially influence runoff flow of pesticides from fields to rivers and streams. The slopes of the municipality will be investigated further in the second Aim of the Thesis that is concerned with finding potential sites contaminated by pesticies.

Bolivia was part of the Inca Empire before it came under Spanish rule. During Spanish rule it was known as Upper Peru. In the start of the 19th century the inhabitants of Bolivia began wanting independence from Spain, this was in 1809. In 1825, on August 16th, after 16 years of war Bolivia was formed as a republic, named after Simón Bolivar (Arnade, 1957), (Morales, 2010).

Bolivia is a relatively poor country, ranked 117 out of 177 countries on the Human Development Index (HDI) by the United Nations, it is also the lowest ranking by any South American country (Morales, 2010). The rural areas away from the major cities are especially poor, and the difference in income between the highest earning 10%, and the lowest 10% in Bolivia, is the biggest of the countries from where data is available (Morales, 2010). Transport in the country is hampered by the fact that transport of goods by land is more expensive than by water. The transportation cost can therefore serve as a constraint, that leads to rural areas being cut off from major development (Morales, 2010).

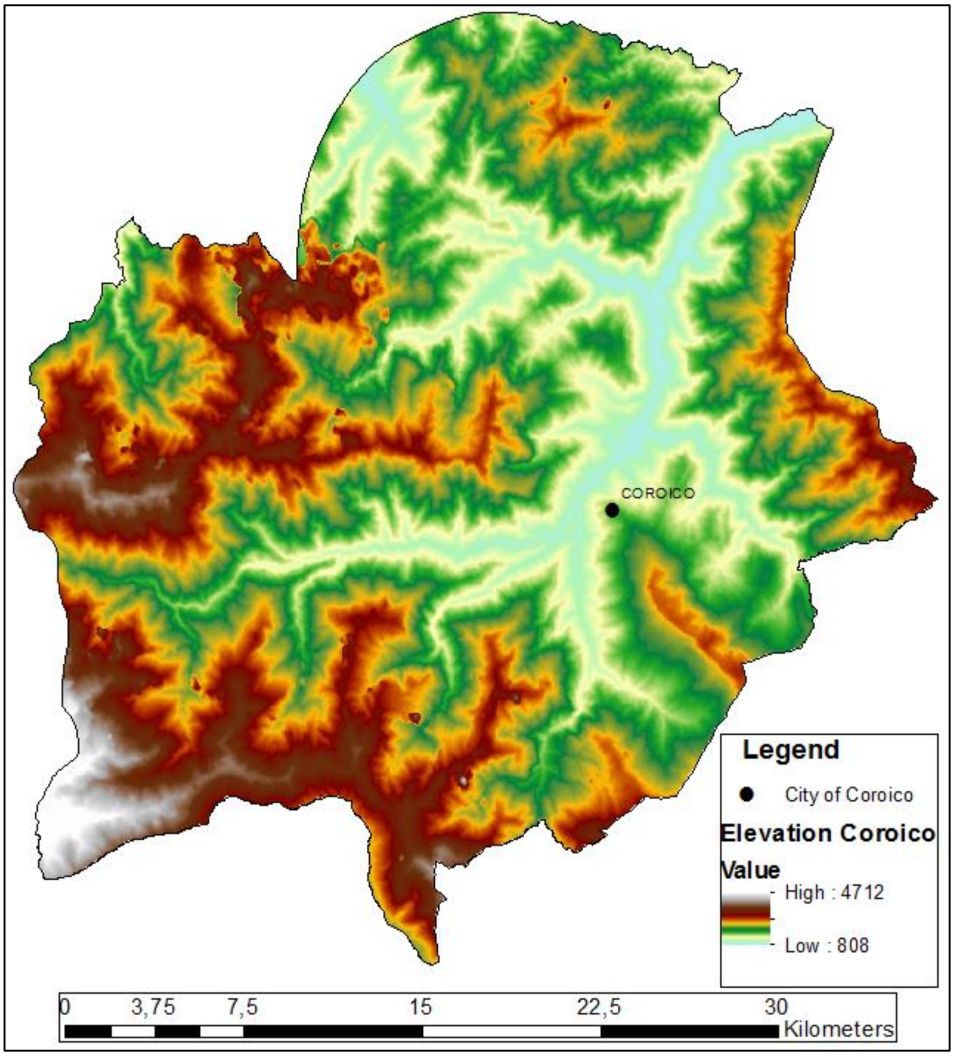

Figure 3: Altitude within the Coroico Municipality (GeoBolivia, 2014).

Much of the income in Coroico comes from agriculture, with an additional small income from service of weekend-tourist (Schouppe, 2006). Especially the production of coca-leafs is a big part of the agriculture. Chewing coca-leafs is legal in Bolivia, and comparable to smoking cigarettes in Europe, though probably more socially acceptable than cigarette smoking in Europe. The production of coca is an environmentally unfriendly process. In the Andes-region, coca is often produced on mountain slopes. The pesticide use in agricultural, and coca, production impact the environment, in an unsustainable matter (South, 1977).

In recent decades the expansion of agricultural lands in Bolivia have come at the expense of the forest, as these are cleared to facilitate the expansion (Kaimowitz, et al., 199) (Pacheco, 2006). In the long run clearing of forest and intense agricultural use can lead to permanent deforestation, something that should be a concern around Bolivia.

As briefly touched upon earlier, there are negative environmental impact from the use of pesticides. An uncontrolled use of pesticides only makes matters worse. It would appear, that the practice for applying pesticide for a long time has been, to use too much, rather than too little (Jørs, et al., 2014). Increasing the problem is the possibility of rainwater washing some of the pesticides out of the soil, so that they eventually might end up in downslope rivers; the same rivers that provide drinking water for the area.

While the coca production in the Andes-region is officially for chewing purposes, it is however a fact, that not all of the coca grown is used for chewing coca. A big portion of the coca is processed into cocaine (South, 1977). While the production of coca for chewing could be vastly improved, in terms of environment and sustainability, for now EWB do not want to partake in this, as it can be seen as an acceptance of the cocaine-production.

4 MAPPING COROICO – 1ST AIM OF THE THESIS

The first Aim of the Thesis is to meet the request of EWB, which was to create a detailed GIS-map of the Coroico Municipality. The first of two objective is to create a land use/land cover map, by landscape classification from images from the Landsat Satellite. Once this is completed, the second objective will be to add vector data from online sources in order to better visualize and present the area. The finished maps will be presented at the end of the chapter.

4.1. Land Use/Land Cover Classification

This first part of the chapter will be concerned with the Land Use and Land Cover Classification. The classification will be carried out in ArcMap via the Maximum Likelihood Tool. The basic principal of the tool is to group cells in a picture into predetermined groups, based on training samples that are determined by the ArcMap-user. The picture used in this case will be a satellite picture, and the product will be a raster map with a finite number of land uses or land covers. The final map will be analyzed and verified.

This progression of this chapter will be: A very brief introduction to raster and the Landsat Satellite. Hereafter the Maximum Likelihood Tool, as well as the Majority Filter Tool will be presented, both in theory and method. The last section of the chapter will be the analysis of the map, which will take point of departure in a verification of 300 points spread across the map.

4.1.1. Raster Theory

Raster-data come in a grid with square cells, with each cell containing a specific value, and has proven beneficial when representing data with various values. The use of raster in GIS began in the 1970's, as scientists had to develop a method capable of displaying the recordings from satellites orbiting Earth (Balstrøm, et al., 2010).

Raster-data varies in size and spatial resolution, from large-scale maps of the entire world in cells of 1x1 km representing land cover or altitude, to a mapping of a local forest with cell sizes of 1,6x1,6 m. A general rule of thumb is that the bigger the area, the bigger the cells. Dividing the entire world into cells of 1,6x1,6 m would, for instance, create an insurmountable amount of data. The raster cells that are used for land cover classification in this project, are 30x30 m. A raster-cell can only represent one value per layer, which can lead to mixed pixels (cells). This will be explained further in section 4.1.3., that is concerned with the Maximum Likelihood Tool for classification.

Every cell in a raster layer is represented by a pixel in a byte. 8 bits make up a byte, and each of those bits can have one of two values, 0 or 1. This creates $2^8 = 256$ possible distinct grayscale colors, and this is what makes up a black and white photo. When working with data displaying true colors, the three RGB-bands (red, green, blue) are used. As is the case with black and white photos, each of the bands represent 256 possible distinct colors within their own color. When the three bands are combined, it is possible to create 256^3 = 16.777.216 distinct colors. (Balstrøm, et al., 2010)

The value of a raster cell can be either a value representing an actual "real world" value, or a made up value. A "real world" value would be if one wanted to display the topology of an area. In that case one would assign the raster cells the specific elevation of that cell. If that was done for the entire world, the range would be roughly between 0 and 8.848, being sea level (although there are lower lying places) and Mount Everest. However mixed cells could change this, which, as previously stated, will be discussed in section 4.1.3., that is concerned with Maximum Likelihood Tool. Raster cells with made up values are seen in many instances. In this thesis they are encountered when looking at the land cover classification. As a raster cell cannot contain text or expressions, it is necessary to convert this into numbers, numbers that are then explained in an associated legend. Therefore forest (tree cover) could be assigned the value "6" while water got the value "21", which is the case in both instances in this thesis. All cells that in the map are of the value "6" is then given a green color indicating tree cover, while those with the value "21" are given a blue color since it is water. When looking at the entire map of the area the mosaic of the various colors and the related legend affords the viewer an opportunity to have an idea of what types of land cover are covering the area, albeit not showing the exact true colors of the area. (Balstrøm, et al., 2010)

4.1.2. Landsat Satellite Theory

To produce the map of the land cover for the Coroico Municipality, Landsat satellite data has been used. In 1972 the Landsat program started providing data of the Earth. Since then various satellites have been launched in order to upgrade or replace previous satellites. The upgrades since 1972 have amongst other things lead to finer resolutions and more bands, thus ensuring better and more precise data. The Landsat 8 satellite features 11 bands (U.S. Geological Survey, 2014). The swath of the satellite covers a width of 185 km (Garner, 2013), with a resolution of 30 m while it's repeat period is 16 days (U.S. Geological Survey, 2015).

On the webpage of the U.S. Geological Survey (USGS) data from Landsat satellites is available. Through the Global Visualization Viewer (glovis) it is possible to find any location of interest. The data is available in tiles of roughly 185x185 km. In the case of Coroico the entire municipality is located within one tile, thereby making the merging of tiles unnecessary. In *glovis* it is possible to find all the available data sets from this particular tile since 1984, and for each data set information about cloud cover (CC), date, and quality (Qlty) is avaiable. Once a data set that fits the needs and demands of the project is found, the dataset is selected and downloaded. It can then be opened in ArcMap, where it is possible to create a composite layer that makes it possible to mix and match three bands at a time, thereby visualizing the data in various ways.

For the Landsat 8 Satellite the 11 bands are (in numerical order):

1) Coastal Aerosol
2) Blue
3) Green
4) Red
5) Near Infrared (NIR)
6) SWIR 1 (Short Wavelength Infrared 1)
7) SWIR 2
8) Panchromatic
9) Cirrus
10) Thermal Infrared 1 (TIRS)
11) Thermal Infrared 2 (TIRS) (U.S. Geological Survey, 2014).

In ArcMap the sequence of the three channels in the composite layer is Red-Green-Blue (RGB). The band combination required to displays true colors (colors that the human eye would see, if observing the area from above) is therefore 4-3-2 as these are the band numbers of red, green, and blue, respectively (Figure 4).

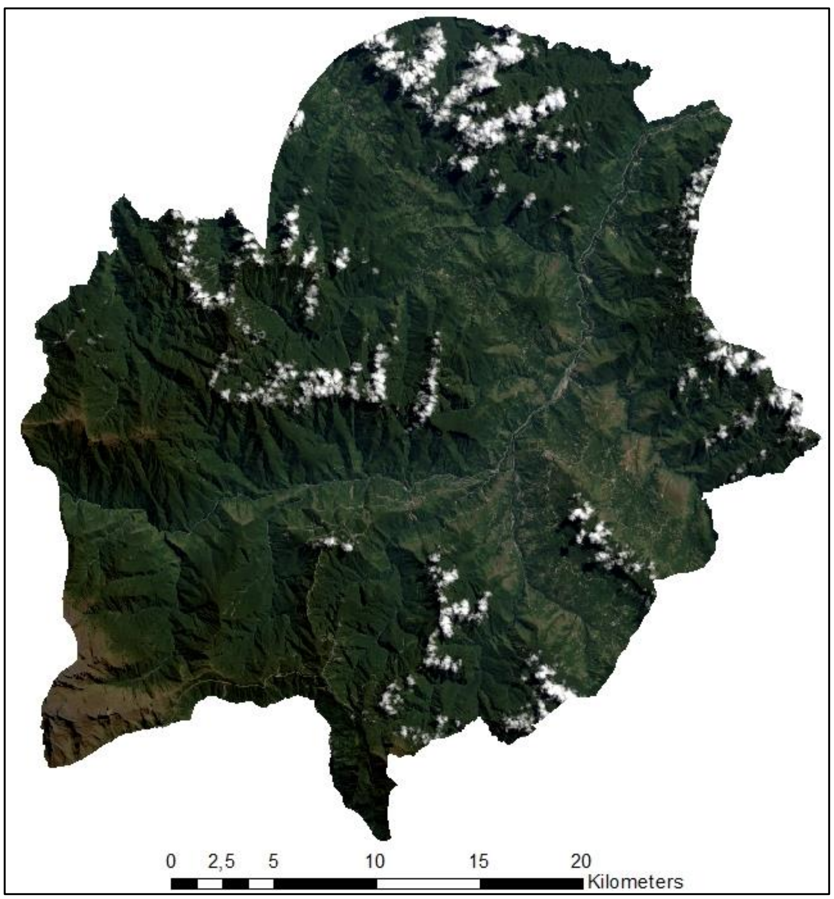

Figure 4: The Coroico Municipality shown in true colors, from Landsat Satellite data (U.S. Geological Survey, 2015).

If one wanted to show the area in the traditional color of infrared (CIR) (Figure 5) the combination would be 5-4-3 (U.S. Geological Survey, 2013). CIR is often used when the desire is to differentiate between vegetation and wetlands (Peters, 2015).

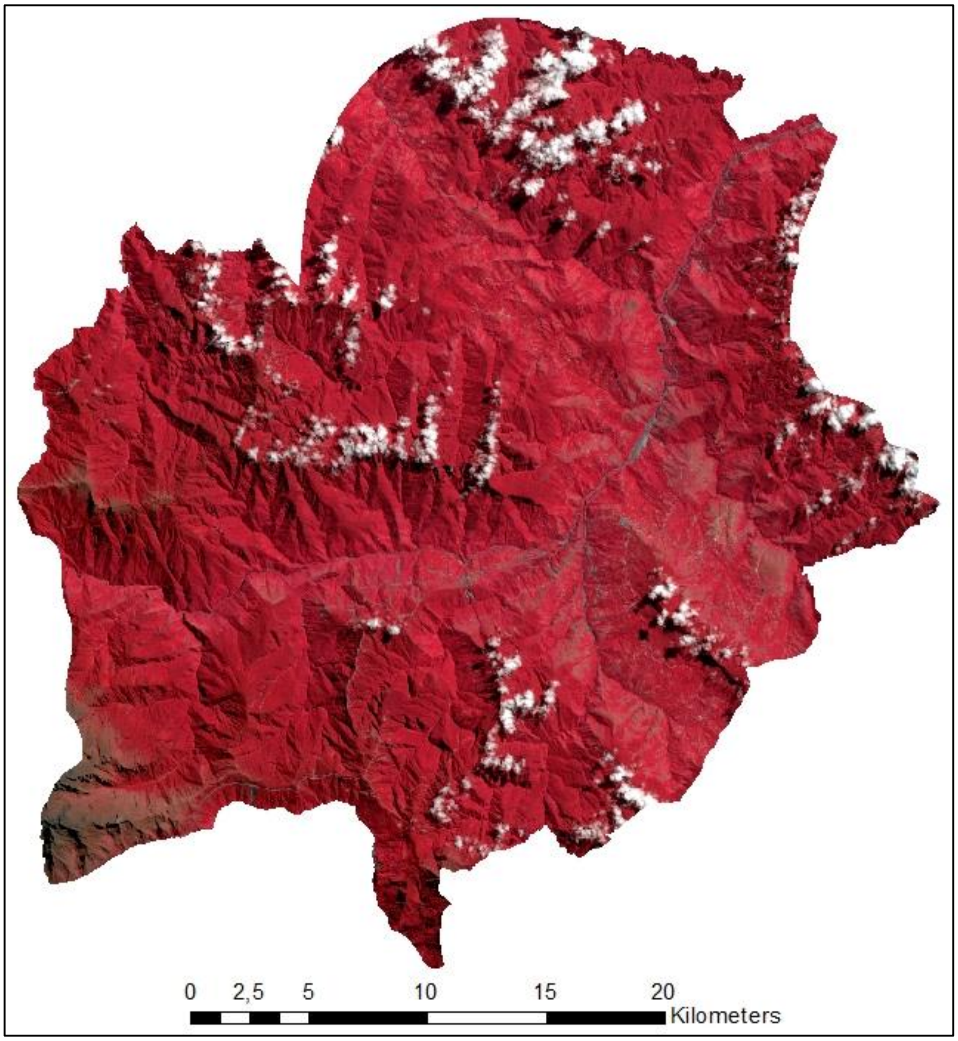

Figure 5: The Coroico Municipality shown in true colors, from Landsat Satellite data (U.S. Geological Survey, 2015).

Various other combinations have been used in order to single out cells that are of a certain landscape type, in order to create training samples for the Maximum Likelihood Tool to use, to classify the entire municipality.

The USGS satellite image used in this project is from April 19th 2013. This exact image was chosen based on the fact that the cloud cover was of minimal nature (8%) and that the image had a high quality of 9, on a scale from 1 to 10 (Appendix 1). A satellite image from October 2014 was also available, and while the cloud cover for the entire tile was actually lower (4%) than the one being used, the amount of clouds over the Coroico Municipality were considerably higher. Lastly the April-image displays the vegetation in the area, in a much clearer manner, than the October-image. While this would not make sense in Europe, or other places on the Northern Hemisphere, it is worth noting that on the Southern Hemisphere winter and summer is opposite of what we are used to. Therefore the image from April is comparable to a late summer/early fall image from the Northern Hemisphere, and that is why the amount of vegetation is visibly higher in the April-image, compared to that from October. Likewise the months with the highest precipitation in Bolivia are December, January, and February which further explain the high amount of vegetation in April (climate-data.org, u.d.).

4.1.3. Maximum Likelihood Theory and Method

To carry out a Land Cover Classification the Maximum Likelihood Tool is used in ArcMap. The purpose of the tool is to use Landsat satellite data to determine the land cover of each raster cell in a map. The reason for using the Maximum Likelihood Tool is that I have prior knowledge of it from my 8th semester. There I, along with three group members, wrote a project concerning different tools in ArcMap and their precision and usability when classifying landscapes and land cover types. What we found was that the Maximum Likelihood Tool was the best tool, both in terms of accuracy, but also when it came to optimizing time. The other methods used were the Iso Cluster Unsupervised Classification and a self-invented method that, just like the Supervised Maximum Likelihood Method, focused on training samples and classifying cells with the same characteristics. The Iso Cluster Tool proved to be very time consuming, while the self-invented Supervised Method only classified cells that were similar, and therefore one ended up with many unclassified cells. It might work if, in later stages one was concerned with explicitly only finding cells that met certain requirements, for instance if one could get a training sample of a couple of coca-fields, and then wanted to detect all coca-fields in the area.

In order to get ArcMap to classify raster cells, it is necessary to create training samples for the program to use. This is done by pointing out areas in the satellite picture, that one is sure are of a certain land cover type (ESRI, 2012). When creating training samples an ArcMap basemap was used as a reference, to help ensure that the training samples actually contained the land cover type that was intended. The reason for this is that while the satellite picture looks very detailed when viewing the entire municipality, the 30x30 meter resolution can be too coarse to reveal small open patches in a piece of forest. Therefore training samples were always crosschecked with the basemap. To ensure that the Maximum Likelihood Tool classifies a class correctly, it is useful to select training samples from all around the area, instead of just one spot. This ensures that, in the case of tree cover, that the training sample contains pixels of light green colors, as well as darker green, as is illustrated on the picture below (Figure 6). While not entirely fail proof, the two abovementioned steps do offer an extra security in terms of getting correct training samples.

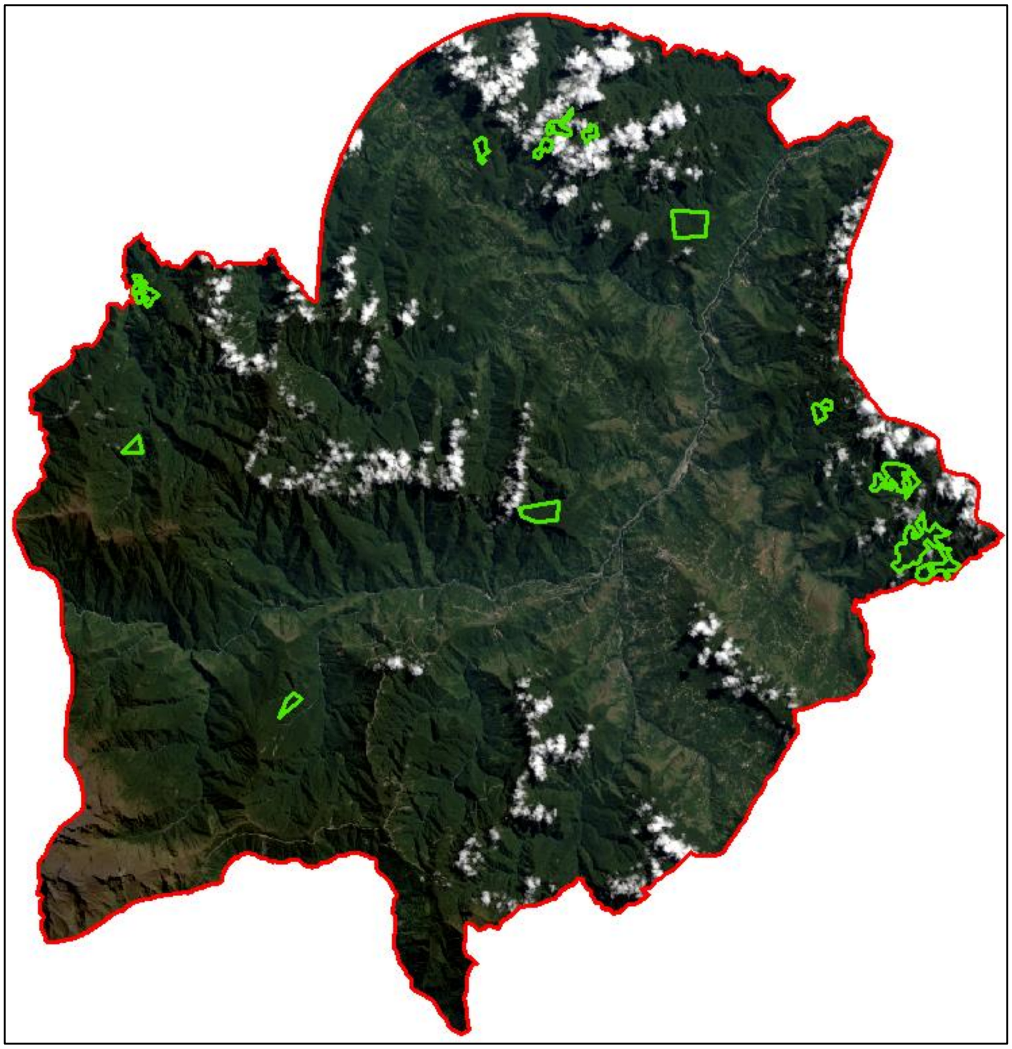

Figure 6: Training samples of tree cover in the municipality.

For this project specifically, the goal was to identify land cover features such as tree cover, shrub, fields, bare areas, and artificial land cover – mainly houses and roads. Also it was desired to determine areas with water. The desire to determine tree cover, shrub, fields and bare areas were, apart from them being dominant features in the landscape, the desire to later in the project to a screening for potential spots possibly contaminated by pesticides. The classes used in this project take point of departure in the classes of the Global Ecosystem Legend used by the United States Geological Survey, in their Global Land Cover Classification Data Base. In the Global Ecosystem Legend 25 out the 100 classes are some form of forest (U.S. Geological Survey, 2012). That type of precision and detail is not the focus of this project, and the various types of forest were therefore combined in one major class, named tree cover. To differentiate between trees and short green vegetation was another goal of the project. Again plenty of classes exist in the Global Ecosystem Legend, and these were combined in the class called Shrub. Fields were one of the initial interests of the project, as these were believed to be at great risk in terms of pesticide contamination, therefore fields were included. When doing a reconnaissance of the municipality it is evident that a noticeable part of the area is covered by bare mountain sides, and the class *Bare Areas* was therefore included. To account for clusters of houses and road the class *Artificial* was created. In the Global Ecosystems Legend it is termed *Urban*. The last class that was wished to classify was that of water. Pointing out the water of the municipality is paramount, as rivers and their aquatic organisms are at great risk when focusing on pesticide contamination. The final two classes, *clouds* and *shade* are created out of necessity. When observing the satellite images used, it is apparent that some mountain tops are covered in clouds. If specific classes had not been made for *clouds* and *shade*, these pixels would have been included in the previously mentioned classes, thus lowering the accuracy of these classes. As a result of this, the two classes were created.

The Maximum Likelihood Tool will only classify the classes that are already created as training samples. Therefore if one only creates two training samples, the tool can only classify cells as one of the two classes. The benefit of the tool is that one can decide exactly what classes should be shown in the map, however that can also serve as a confinement, as some classes might be left out entirely. Once the researcher has decided upon the classes to be used in the classification, the Maximum Likelihood Tool can be run. The tool classifies the pixels in the satellite image based on resemblance with the pixels of the training samples. All 11 bands in the image can be used to differentiate. By combining these 11 bands, the tool is able to make distinctions between cells that, for a human observing an image in RGB, might seem alike. The output will be a new raster layer, with the same cell sizes as those of the satellite image, with all cells classified as one of the eight predetermined classes.

When working with raster cells of 30x30 m, which is what is used in this project, one will a times encounter mixed cells. Mixed cells are cells that contain a mix of information, in this case land cover, and therefore cannot be defined indisputably. This could be the case in a cell where half of the cells is a lake, and the other half is forest. This will challenge the Maximum Likelihood Tool, as it cannot distinctively recognize water nor forest, and it might up classifying the cell as a third option, say fields or shrub. Likewise, if one was concerned with an area with a small stream or small lakes, those aquatic areas might not be portrayed in the map, as they would be dominated by the remaining land cover types in their respective cells, thus, for instance, classifying all of the cells as forest (Foody, 2002). The coarser the resolution of the satellite image, the more likely it is that mixed pixels will be present in a classification. When working with a resolution of 30x30 meters, mixed pixels will inevitably be a part of the classification (Jones & Vaughn, 2010). Another drawback of working with cells of 30x30 meters is the chance of human error, where some of the 300 control cells verified manually, are actually classified incorrectly by the researcher.

4.1.4. Majority Filter

Using the Majority Filter is one way to improve one's classification in ArcMap. What the tool does is to replace raster cells, if the majority of the surrounding cells are of another value. This can be exemplified by Figure 7. The figure illustrates how ArcMap converts cells based on the surrounding cell values. The method used in the example, and in the project is based on the *eight-neighbor-rule*.

This means that the conversion will be based on all eight neighbors, instead of only the four orthogonal neighbors, as is the case with the four-neighbor-rule. When determining the *Replacement Threshold* one has to determine whether to use *Majority* or *Half*. When using the *Majority Threshold*, which is the case in this project, the majority of the cells, five out of eight should be of the same value. If using the *Half Threshold* only half of the cells, as the name indicates, should be of the same value (ESRI, 2011).

$$\begin{matrix}1 & 1 & 1\\ 1 & 2 & 3\\ 1 & 3 & 3\end{matrix} = \begin{matrix}1 & 1 & 1\\ 1 & 1 & 3\\ 1 & 3 & 3\end{matrix}$$

Figure 7: First example of the Majority Filter.

The figure above (Figure 7) shows how the value of a cell (the middle cell) changes based on the *Majority Filter* using the *eight-neighbor-rule* and *Majority Threshold*. If the same *Majority Filter* was applied once more (Figure 8), the two cells (to the right in the middle row, and in the middle of the bottom row) with a value of 3 would also change to the value of 1, as three of their five neighbors now are of that value.

$$\begin{matrix}1 & 1 & 1\\ 1 & 1 & 3\\ 1 & 3 & 3\end{matrix} = \begin{matrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 3\end{matrix}$$

Figure 8: Second example of the Majority Filter.

After a third *Majority Filter* (Figure 9), the entire box would be filled with cells of the value 1.

$$\begin{matrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 3\end{matrix} = \begin{matrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{matrix}$$

Figure 9: Third example of the Majority Filter.

In the analysis of the classification, the effect of the Majority Filter in relation to this project will be discussed.

4.2. Classification Analysis

The classification of the Coroico Municipality via the Maximum Likelihood Classification Tool produced the following map, illustration the land use and land cover of the municipality (Figure 10). What is visually apparent is that the area is dominated by tree cover and shrub, and the cell counts support that observation as well. 671.693 of the 1.051.488 cells are determined to be tree cover, while 160.050 of them are shrub, which collectively is four fifths of the area. The third highest amount of cells is that of bare areas, which especially dominate the Southwestern and Western part of the map; 78.507 cells (or 7,47%) are bare areas. Accounting for just above 5% of the area each, clouds (61.423 cells (5,84%)) and artificial surfaces (58.754 cells (5,59%)) are the fourth and fifth most prevalent landscape types, albeit clouds cannot really be classified as a landscape type. The last three land cover types are shade (13.144 cells (1,25%)), water (4.668 cells (0,44%)), and fields (3.245 cells (0,31%)). Just as with clouds, shade cannot really be described as a land cover type, but it is necessary nonetheless, because some of the pixels in the area are dominated by shade from the clouds. Although only very few pixel indicating water are present in the area, they are still visible, and they follow the actual rivers of the area, which will be shown at the very end of this chapter.

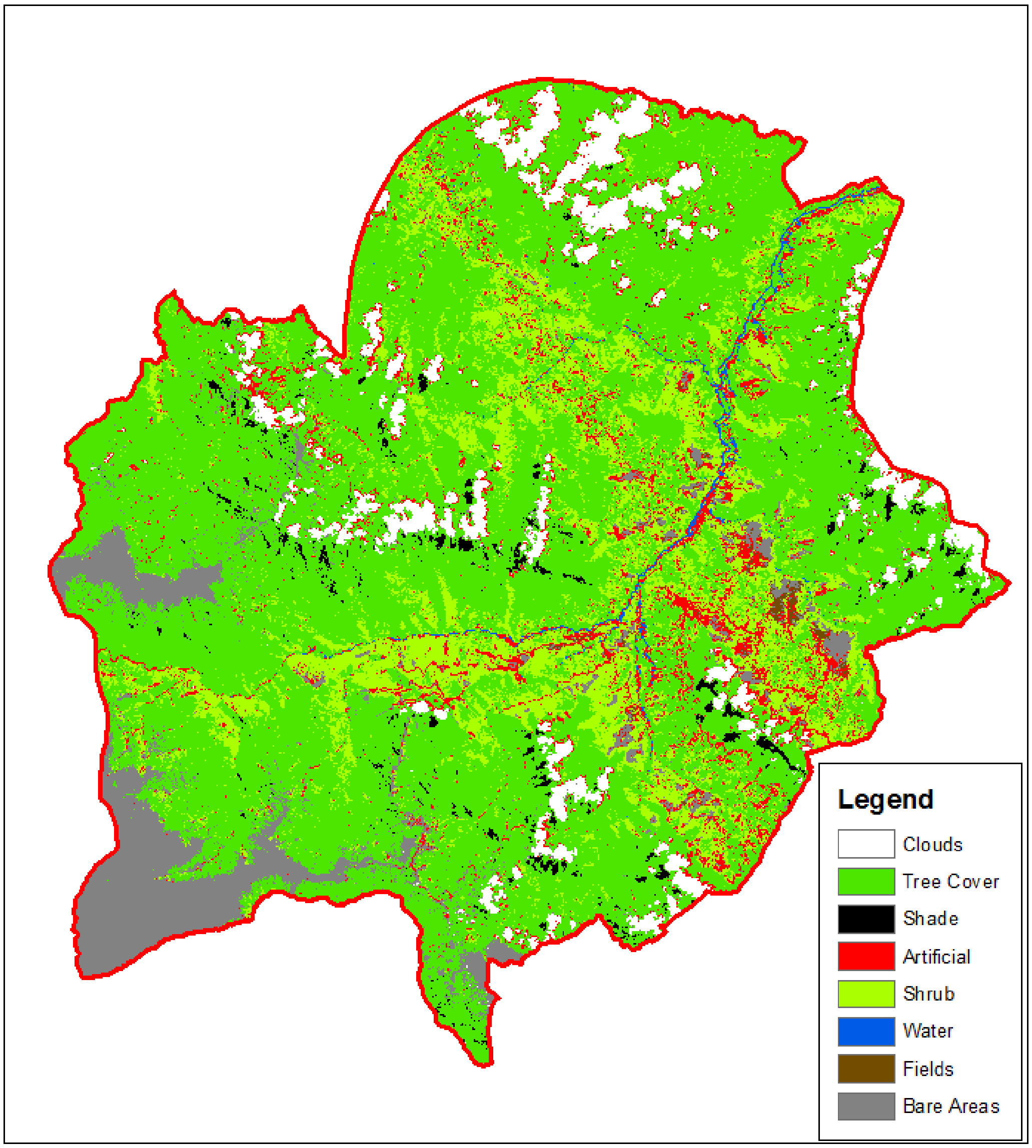

Figure 10: Maximum Likelihood Classification of the Coroico Municipality.

In order to determine whether or not the classification of the area is useful it is necessary to manually verify a number of point in the area, to detect whether or not what the pixels are classified as, is what they actually are. In order to determine how many points (n) are enough to verify a classification, it is necessary to determine the following: First it is necessary to decide on a expected accuracy (P) as a decimal number. Having knowledge from my previous experience with land cover classification, a somewhat conservative guess would be 0,75. Secondly an acceptable margin of error should be determined. For this project it is desired to predict the overall accuracy with 95% probability, therefore the acceptable margin of error (e) will be 0,05. Having decided upon these two numbers, it is now possible to use the following formula, from the book *Remote Sensing Digital Image Analysis* (Richards, 2012):

$$n = \frac{4P(1-P)}{e^2}$$

$$n = \frac{4 \times 0{,}75 \times (1 - 0{,}75)}{0{,}05^2}$$

$$n = 300$$

Based on the formula above, 300 points were randomly created, by using the data management tool *Create Random Points* in ArcGIS. By zooming in, and looking at both the satellite image and the basemap, the correct class was detected. This was then compared to what the Maximum Likelihood Tool had classified the cell as. To depict this, an Error Matrix was produced (Table 1) (Jones & Vaughn, 2010).

The Error Matrix presents detailed statistics on the accuracy of the classification. To the left (red background), and on the top (green background) of the matrix, the eight classes of the classification are aligned. The classes on the left are a review of what the Maximum Likelihood Tool classified the cells of the 300 points as, while the classes on the top are what the manual review, of the cells of the 300 points, turned out to be. To explain this further, the tree cover class will be used, as this class, by far, contains the most cells.

The blue diagonal cells represent the number of correct classifications. Of the 300 control cells, the number of cells classified as tree cover was 188, with 183 of these being correct. That means that the remaining five cells were classified as tree cover, but were in reality something else (one being clouds and four being shrub), which results in a *User's Accuracy* of 97%. The *User's Accuracy* is concerned with the percentage of classified points, in one class, that are correct (Jones & Vaughn, 2010).

On the other hand, 226 of the control cells were actually tree cover, albeit only 183 of them were being classified as such, giving a *Producer's Accuracy* of 81%. Especially the class of shrub created difficulties for the classification, with 27 cells that were actually tree cover, being classified as shrub. The remaining cells of actual tree cover were classified as: Artificial surfaces (9), bare areas (5), shade (1), and fields (1). None of the cells containing tree cover were classified as clouds or water. As opposed to the *User's Accuracy*, the parameter of the *Producer's Accuracy* is the percentage of control points of a certain class, that are actually classified as such (Jones & Vaughn, 2010).

The *Overall Accuracy* is an expression of the total amount of correctly classified points, divided by the total number of control cells. For the classification by the Maximum Likelihood Tool the *Overall Accuracy* is 0,763. Since the *Overall Accuracy* is a bit higher than the expected accuracy (P) the *Error* ± actually drops below the previously determined margin of error (e) at 0,05. The last barometer, included in this project, in terms of gauging the classification accuracy is the *Kappa Coefficient* (k). As opposed to the *Overall Accuracy*, the *Kappa Coefficient* also includes the *Commission* and *Omission Error*. The inclusion of these two parameters enables the *Kappa Coefficient* to conclude whether the success of the classification can be attributed to random chance, or if it is significantly better than chance. According to Jones & Vaughn:

> *"Values of kappa greater than about 0,75 indicate good to excellent classifier performance, while values less than 0,4 suggest rather poor performance"* (Jones & Vaughn, 2010). This is further backed up by John A. Richards, who presents the following index (Richards, 2012):

Kappa Coefficient	*Classification*
Below 0,4	Poor
0,41-0,6	Moderate
0,61-0,75	Good
0,76-0,8	Excellent
0,81 and above	Almost perfect

The *Kappa Coefficient* for the Maximum Likelihood classification in 0,53 thus indicating that the success of the classification can be regarded as moderate.

In an effort to optimize the results of the classification, the *Majority Filter Tool* was used.

Table 1: Error Matrix for Maximum Likelihood Classification.

Maximum Likelihood	Actual Land Cover: Clouds (1)	Tree Cover (6)	Shade (13)	Artificial (19)	Shrub (20)	Water (21)	Fields (22)	Bare Areas (93)	Total	User's accuracy	Commision error
Classification: (1) Clouds	15	0	0	0	0	0	0	0	15	1,00	0,00
(6) Tree Cover	1	183	0	0	4	0	0	0	188	0,97	0,03
(13) Shade	1	1	2	0	0	0	0	0	4	0,50	0,50
(19) Artificial	3	9	0	1	3	0	0	2	18	0,06	0,94
(20) Shrub	0	27	0	0	18	0	3	3	51	0,35	0,65
(21) Water	0	0	0	0	0	0	0	0	0	N/A	N/A
(22) Fields	0	1	0	1	1	0	0	0	3	0,00	1,00
(93) Bare Areas	0	5	1	1	4	0	0	10	21	0,48	0,52
Total	20	226	3	3	30	0	3	15	300		
Producer's accuracy	0,75	0,81	0,67	0,33	0,60	N/A	0,00	0,67	Kappa Coefficient:	0,530	
Omission error	0,25	0,19	0,33	0,67	0,40	N/A	1,00	0,33	Overall accuracy:	0,763	
									Error ± :	0,049	

While the change in classification after the first use of the *Majority Filter* might not be noticeable when observing the entire municipality of Coroico, the change is in *Overall Accuracy* improves 3 percentage points from 76,3% to 79,7%, and likewise the *Kappa Coefficient* improves from 0,53 to 0,578 (Table 2).

When applying the *Majority Filter* a second time, the *Kappa Coefficient* (up to 0,602) as well as the *Overall Accuracy* (up to 81%) once again increase, as can be seen in Table 3. Having proved that the use of the *Majority Filter* improves the classification, it is worth examining exactly where the classification improves.

For the class of clouds there is no change. All of the 15 cells classified as clouds are indeed clouds, however there are five other control points with an actual cloud cover, that are not classified as such, thus producing a *Producer's Accuracy* of 75%.

Although the number of cells classified as tree cover increase through the three classifications (188, 197, 199), the *User's Accuracy* stays put at 97%. The *Producer's Accuracy*, on the other hand, increases slightly through all three classifications. From 81, through 85, to 86%. This indicates that with every *Majority Filter* the classification succeeds in classifying a higher percentage of the actual tree cover, as tree cover.

Only three of the entire 300 control points are shade. One should therefore be wary of putting too much stock into the percentages hereof. In the initial classification four control points are actually classified as shade, thus the Maximum Likelihood Tool overestimates the amount of shade in the map. After the first Majority Filter one of the control cells incorrectly classified as shade, is classified as something else, thus improving the *User's Accuracy*.

The artificial surfaces are one of the major obstacles in the classification. As with shade, only three of the control points are actually artificial surfaces, which, in terms of *Producer's Accuracy* will result in very coarse jumps in percentages, which therefore is not to be relied on. The classification blatantly overestimates the artificial surfaces, as 18 cells are initially classified as such, with only one of the 18 actually being artificial surface, resulting in a *User's Accuracy* of 6%. By the conclusion of the second *Majority Filter*, the *User's Accuracy* is up to 14%, which still is far from impressive.

Shrub is the second most numerous of the classes. 10% of control points are actually shrub, but only 18 of these 30 points are correctly classified initially; this number then increases from 19 to 21, resulting in a final *Producer's Accuracy* of 70%. For shrub the big confusion seems to be with tree cover, when examining the *User's Accuracy*. More than half of the points classified as shrub are, in reality, not shrub, but for the most part tree cover instead, which is why the *User's Accuracy* of shrub is low (35, 40, and 45%); and while it does increase, it is obvious that shrub is one of the obstacles in terms of getting a solid classification.

Table 2: Error Matrix for Maximum Likelihood Classification after first Majority Filter

Maximum Likelihood with 1st Majority Filter	Actual Land Cover								Total	User's Accuracy	Commission Error
	Clouds (1)	Tree Cover (6)	Shade (13)	Artificial (19)	Shrub (20)	Water (21)	Fields (22)	Bare Areas (93)			
Classification (1) Clouds	15	0	0	0	0	0	0	0	15	1,00	0,00
(6) Tree Cover	1	192	0	0	4	0	0	0	197	0,97	0,03
(13) Shade	1	0	2	0	0	0	0	0	3	0,67	0,33
(19) Artificial	3	7	0	1	2	0	0	1	14	0,07	0,93
(20) Shrub	0	21	0	0	19	0	3	4	47	0,40	0,60
(21) Water	0	0	0	0	0	0	0	0	0	N/A	N/A
(22) Fields	0	1	0	1	1	0	0	0	3	0,00	1,00
(93) Bare Areas	0	5	1	1	4	0	0	10	21	0,48	0,52
Total	20	226	3	3	30	0	3	15	300		
Producer's Accuracy	0,75	0,85	0,67	0,33	0,63	N/A	0,00	0,67	Kappa Coefficient:	0,578	
Omission Error	0,25	0,15	0,33	0,67	0,37	N/A	1,00	0,33	Overall Accuracy:	0,797	
									Error ± :	0,046	

Table 3:Error Matrix for Maximum Likelihood Classification after second Majority Filter

Maximum Likelihood with 2nd Majority Filter	Actual Land Cover								Total	User's Accuracy	Commission Error
	Clouds (1)	Tree Cover (6)	Shade (13)	Artificial (19)	Shrub (20)	Water (21)	Fields (22)	Bare Areas (93)			
Classification (1) Clouds	15	0	0	0	0	0	0	0	15	1,00	0,00
(6) Tree Cover	1	194	0	0	4	0	0	0	199	0,97	0,03
(13) Shade	1	0	2	0	0	0	0	0	3	0,67	0,33
(19) Artificial	3	7	0	2	1	0	0	1	14	0,14	0,86
(20) Shrub	0	20	0	0	21	0	3	5	49	0,43	0,57
(21) Water	0	0	0	0	0	0	0	0	0	N/A	N/A
(22) Fields	0	1	0	0	0	0	0	0	1	0,00	1,00
(93) Bare Areas	0	4	1	1	4	0	0	9	19	0,47	0,53
Total	20	226	3	3	30	0	3	15	300		
Producer's Accuracy	0,75	0,86	0,67	0,67	0,70	N/A	0,00	0,60	Kappa Coefficient:	0,602	
Omission Error	0,25	0,14	0,33	0,33	0,30	N/A	1,00	0,40	Overall Accuracy:	0,810	
									Error ± :	0,045	

The sixth land cover class in the classification is water. As chance would have it, not a single one of the 300 control points are within cells either classified as water, or cells that actually are water, even though water is not the least commonly classified class, in the municipality (fields are). The absence of control points associated with water means that it is not possible to calculate the *User's Accuracy* nor the *Producer's Accuracy* for water.

While none of the 300 control points included water, three of the points in the classification are fields. However, not one of these three cells are actually fields, thereby producing a *User's Accuracy* of 0%. The *Producer's Accuracy* for fields is also 0%, as all three of the cells that manually were determined to contain fields, are classified as shrub. It can therefore be speculated that the classification is challenged when having to differentiate between shrub and fields.

The last of the eight classes is the class of bare areas. 15 of the control points are manually determined to be of this class. Only 10, 10, and 9 of these points are classified as bare areas, with the rest being confused with especially shrub, but also artificial surfaces. This results in a *Producer's Accuracies* of 67, 67, and 60%. The *User's Accuracy* for the class is even lower, at 48, 48, and 47%. This is due to the fact that almost half of the control points classified as bare areas, are indeed tree cover or shrub.

4.3. Final Classification

Having reviewed the accuracy of the three classifications, the one that will be used going forward is the classification after two majority filters. Having decided upon going forward with this classification, it is worth reviewing and analyzing the result, both numbers wise, and visually, with Figure 11 showing the final classification of the municipality.

By comparing the final classification (Figure 11) with the initial classification (Figure 10) the change should be noticeable. In many areas with scattering of cells in the initial classification, by the final classification these scattered cells are either combined into small clusters, or they are eradicated in favor of larger patches of another land cover type. The distribution of land use and land cover in municipality are as stated in Table 4 (the quantity of cells for each class, and each classification can be found in Appendix 2):

Table 4: Distribution of land use and land cover in the Coroico Municipality.

Land use / land cover	**Percentage**	**Cell Count**	**Area km^2**
Tree Cover	66,20	696.100	626,49
Shrub	14,65	154.060	138,65
Bare Areas	6,98	73.365	66,03
Clouds	5,88	61.853	55,67
Artificial Surfaces	4,65	48.902	44,01
Shade	0,98	10.252	9,23
Water	0,41	4.314	3,88
Fields	0,25	2.640	2,38

If calculating the total of square kilometers by addition, it will be apparent that the total area is 946,34 km^2, which is around 6 km^2 less than the 952 km^2 that were previously stated as the total area of the municipality. The explanation for the missing 6 km^2 is to be found at the borders of the municipality. The classification of the municipality is carried out in raster, while the correct area is based on vector data of municipal borders. When zooming in on the border of the map, it becomes apparent that some raster cells do not reach the municipal border, while others exceed it (Figure 12).

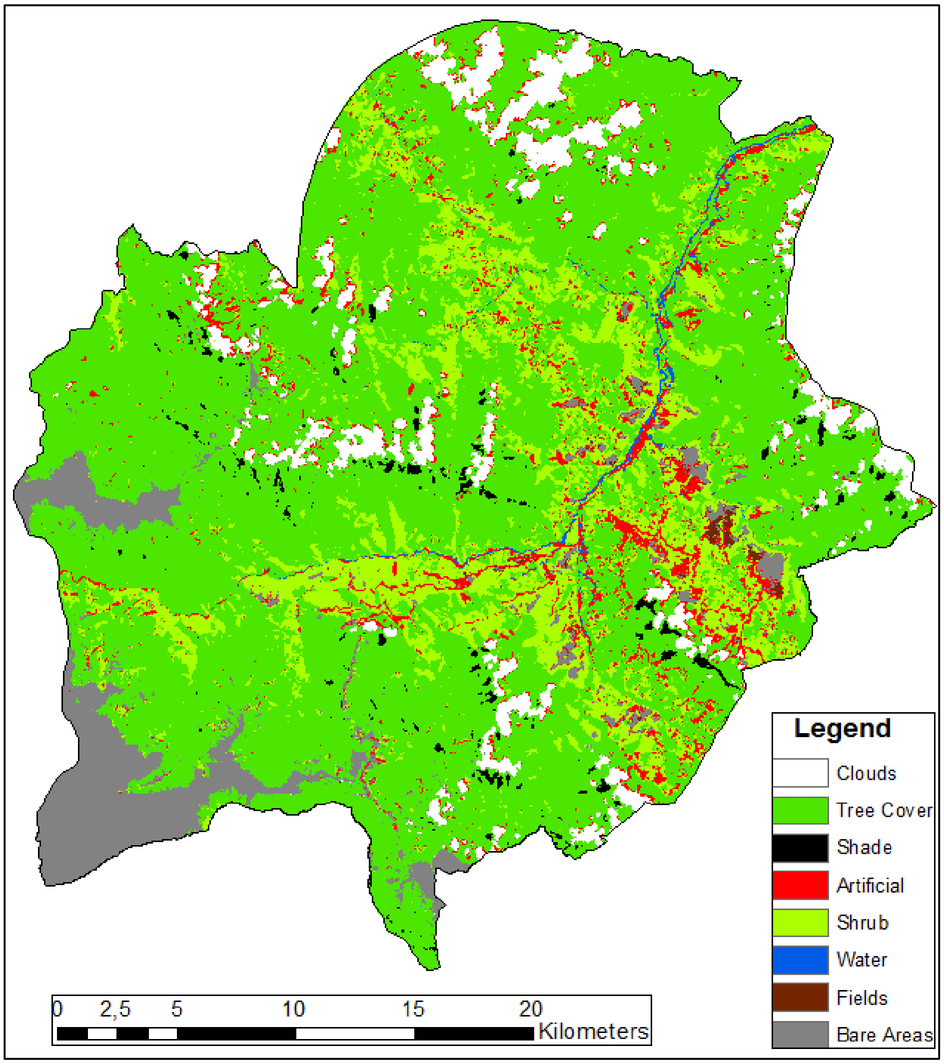

Figure 11: Final classification after two Majority Filters.

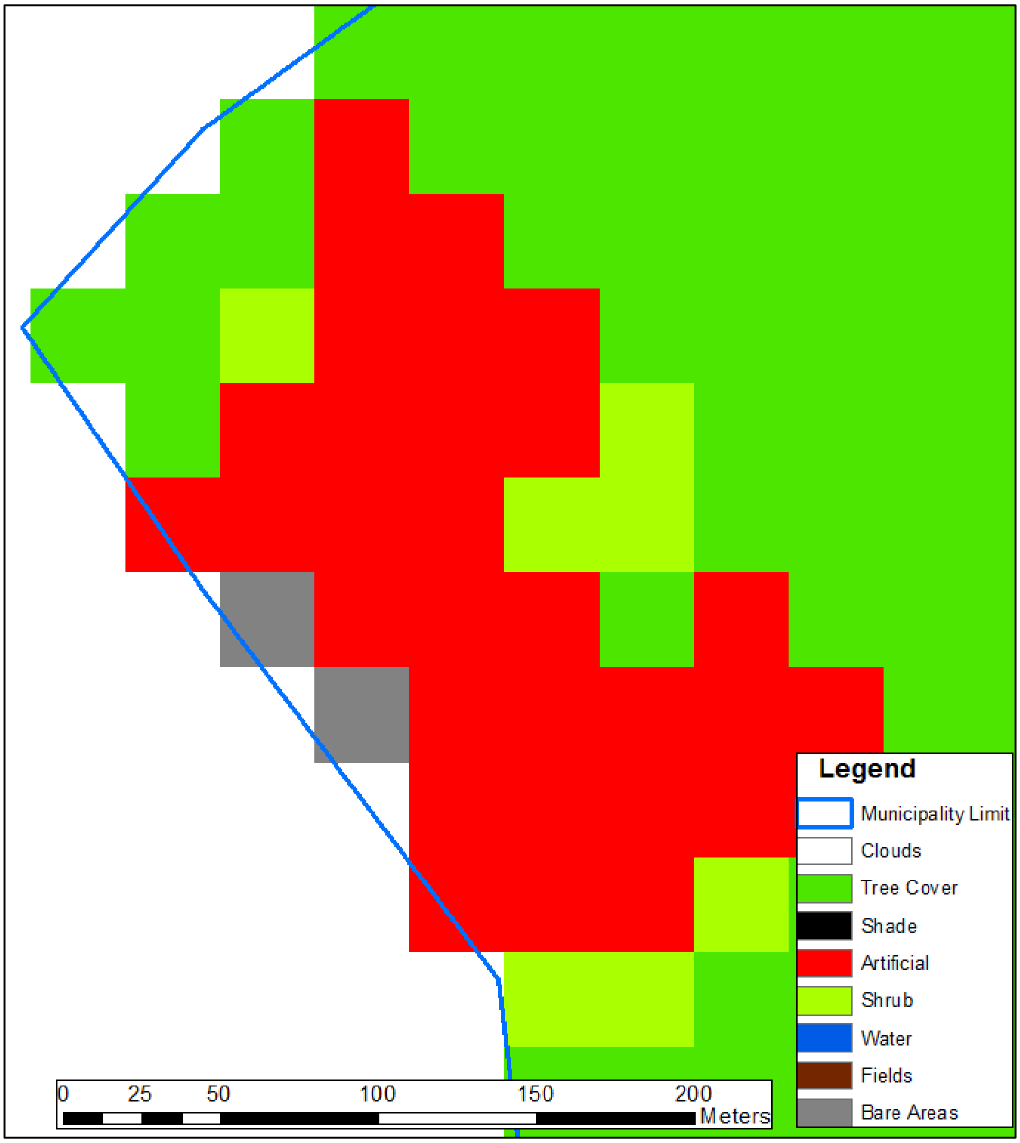

Figure 12: Example of raster cells not following the vector border of the Municipality.

4.4. Discussion of Land Cover Classification

In the previous section, the analysis, it was explained that there were no control points placed in cells either classified as water, or cells observed to be water. While it can might seem careless not to have any control points in cells associated with water, that is how the random points turned out. Also, my experience from a previous semester project leaves me confident that the Maximum Likelihood Tool is more than capable of classifying water at a satisfactory level. One reason for this confidence is that the spectral response of water is fundamentally different from that of any terrestrial land cover type, as it can be seen on the figure below (Figure 13), from the United States Geological Survey's webpage (U.S. Geological Survey, 2013).

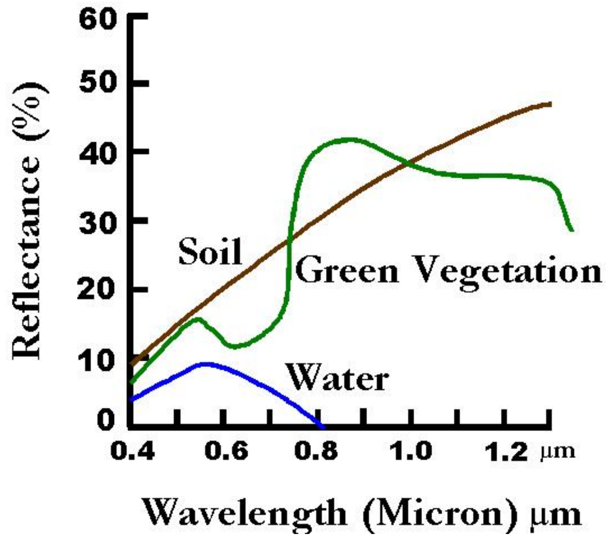

Figure 13:Spectral response of water, soli, and green vegetation (U.S. Geological Survey, 2013).

Another reason is that when combining band 5, 6, and 4 (in that order) in the composite layer in ArcGIS, singling out water is very straightforward (ESRI, 2013). Water will appear in shades of dark blue, while land appears in orange and green colors (Peters, 2015). This is depicted on Figure 14, illustrating that the classification does a fairly good job classifying water.

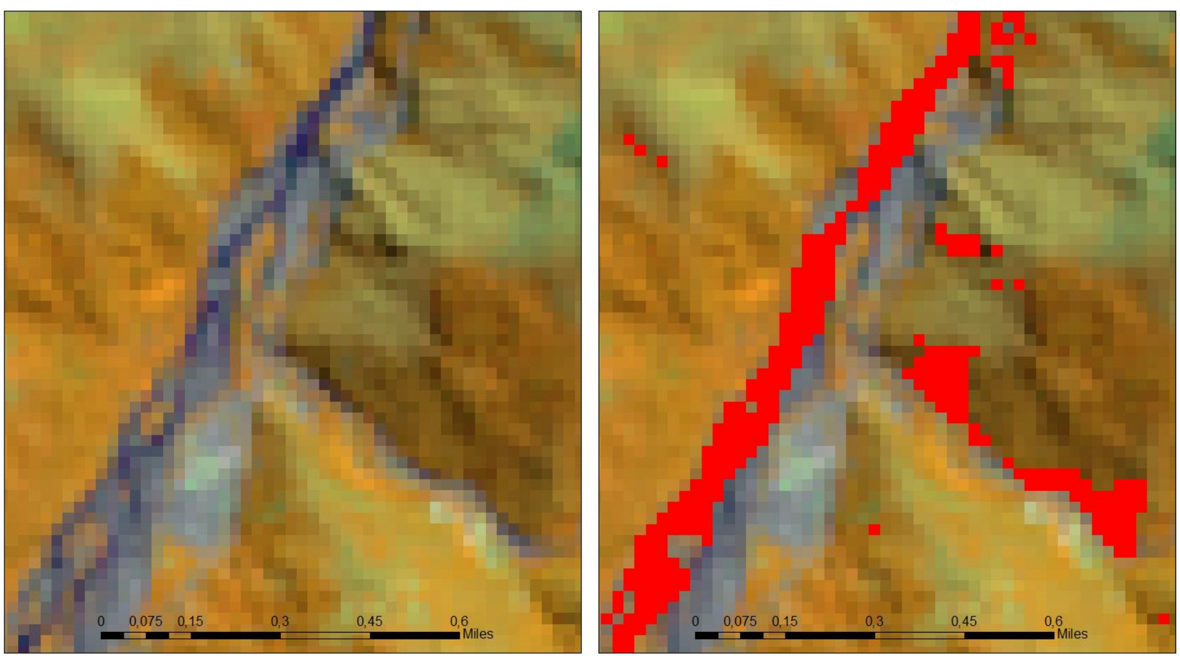

Figure 14: Picture to the left show an area with a river (blue color) using the 6-5-4-combination. On the picture to the right, cells classified as water are shown in red colors.

When using the *Majority Filter*, one has to be aware of the pros and cons of that tool. In the analysis it was described how the tool helped improve the *Overall Accuracy* and the *Kappa Coefficient*. In the review of the method, the mathematics behind the tool were explained. In Figure 15 is a visualization of how the tool smoothens a small part of the landscape in the Coroico Municipality. The image depicts the same three areas, but after the Maximum Likelihood Classification, the first Majority Filter, and after the second Majority Filter. For the viewer is should be obvious that the number of single, scattered pixel decreases through the three images, and that the images become more smooth. If paying attention to the red pixels, one will recognize that many of the single cells disappear, while the bigger clusters of artificial surfaces grow in size (number of pixels).

In the analysis it was noted that the Maximum Likelihood Tool, even when using Majority Filter, was challenged when having to differentiate between tree cover and shrub, as well as shrub and fields. Theoretically, different band combinations should be able to differentiate between the three different types of green vegetation and fields. These different band combinations were used when creating training samples, but despite that, it is still a big area of concern. When working on the MCDA in the second Aim of the Thesis, the three abovementioned classes, especially shrub and fields, will to a large extend be treated equally, in terms of potential pesticide contamination.

Getting an *Overall Accuracy* of 100% is probably close to impossible, when working with eight different classes. If one worked with only water and land, this might be possible, but still there, the cells on the border between the two, would most likely be a mix of the two, presenting the risk of incorrect classification because of mixed cells. Realistically, one will always encounter mixed cells.

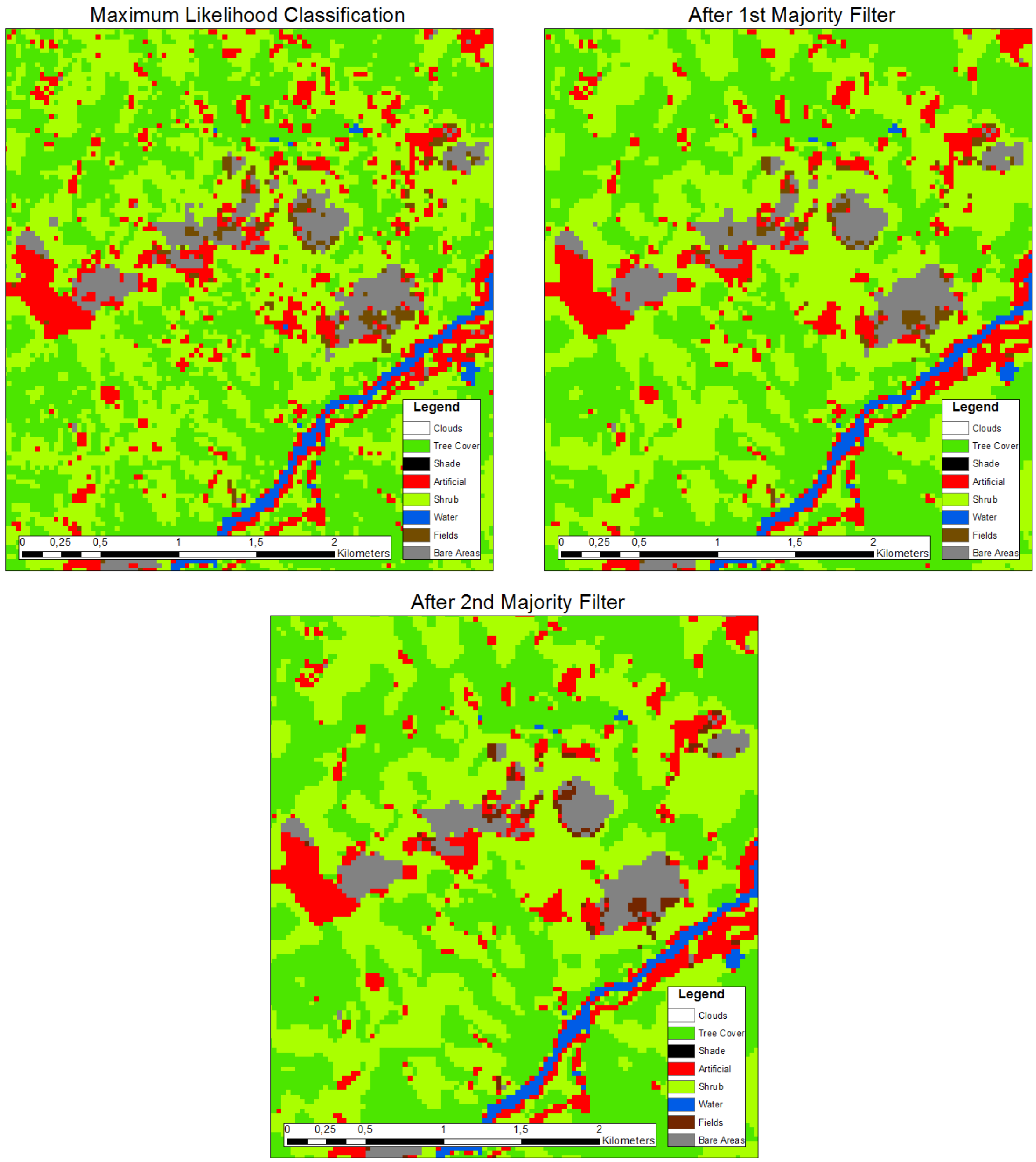

Figure 15: Visualization of the effect of the Majority Filter.

Going from the original Maximum Likelihood classification and through the two rounds of majority filter, the accuracy of the classification improved, both the *Overall Accuracy* and the *Kappa Coefficient*. However, it was observed that the accuracy of some classes suffered from this. This for instance was the case for bare areas, where the some of the originally correctly classified areas were being converted to shrub. Combined with the fact that the improvement from the first to the second majority filter was modest, it was decided not to run a third majority filter.

4.5. Vector Data from Online Resources

As the land cover classification of the municipality is concluded, the second part of the first Aim of the Thesis was to add vector data to the map, to place cities and villages, rivers and streams, and roads and paths in the area.

4.5.1. Vector Theory

When working with GIS-data there are two prevalent types of data used to visualize data. Those two are vector and raster (of which raster have already been described). Vectors are defined as being points, lines, and polygons. Vector datasets come with an attribute table, in which it is possible to obtain detailed information about the attributes of the feature in question (Balstrøm, et al., 2010). The information in the attribute table can be both numbers for instance inhabitants in a city displayed as a point, or the area of a municipality displayed as a polygon. Likewise the information can be names (cities, municipalities, or countries) or labels (land cover type, time zone, or climate classification) (Longley, et al., 2011). Which of the three types are used depends on what ones wants to visualize. In the case of the Coroico Municipality and Bolivia one could take point of departure in cities. When creating the overview-map showing all of Bolivia with it's departments and the cities that are their administrative centers, it makes sense to have the cities represented by points, and the departments as polygons divided by lines that represent the borders between the departments, as it is shown in Figure 2. A big river would likewise be shown as a line. However if we decided to zoom in on one of these cities and its' suburban areas it would not make sense to display the city as a point, as the rest of the canvas would then be left empty. In an example as that, it would be more beneficial to have the city, as well as the suburbs and a potential river defined by polygons, while features such as trees, telephone poles, bus stops, or certain amenities should be shown as points.

4.5.2. Vector Method

In this project, vector data is primarily used to visualize rivers, roads and cities, along with the municipal border. The vectors have been acquired from online sources. The acquisition of data and the sources will be explained in this section.

The data in order to represent Bolivia in South America is from ESRI and is available through direct download in ArcMap, via the "Add Data From ArcGIS online"-tool (ESRI, u.d.). When acquiring vector data used in the project, most of it, in it's original form, have covered an area larger than that of the particular interest, whether it was Bolivia or the Coroico Municipality. In both ArcMap and QGIS it is possible to clip the vectors to the extent of a selected area, by selecting the area in the attribute table, and then saving it as a new vector layer, thereby minimized the amount of unwanted data, and ensuring a more manageable attribute table. If the initial data set did not include borders to select the features from within, the geoprocessing tool "Clip" was used to clip the datasets to the templates of Bolivia and the Coroico Municipality.

GeoBolivia

GeoBoliva is a: "*project, which aims to provide institutions and users in general, geographical information of interest, regardless of the device from which it is accessed. Meaning have a relevant geographic information, harmonized and with quality to support social, economic, and environmental development."* (GeoBolivia, 2014)

> *"que pretende dotar a instituciones y usuarios en general, de información geográfica de interés, independientemente del dispositivo con el cuál se acceda; es decir, disponer de una información geográfica relevante, armonizada y de calidad para apoyar el desarrollo social, económico y ambiental del país."* (GeoBolivia, 2014)

It is a joint program between the government of Bolivia and the Swiss Agency for Development. As evident by the quote above, the main goal is to provide institutions as well as private users with cost free GIS-data, in order to facilitate social, economic, and environmental development. At GeoBolivia both raster and vector data of Bolivia is available. The vector data that have been used comprise of roads, rivers, locations, and municipalities.

OpenStreetMap

To support the vector data found at GeoBolivia, additional vector data from OSM (OpenStreetMap) was downloaded. OSM is an open source database, and is run much like Wikipedia and other open source information hubs. Independent users can map points, lines or polygons and then upload them to OSM, where anyone can access them cost free (OpenStreetMap, u.d.). Since most anyone can upload data to the database, the amount of data is widespread and numerous. When working with a rural and somewhat obscure region like the Coroico Municipality, the OSM database provides an excellent alternative to the more established information databases. However, when working with open source data, it is important to keep in mind that while cost free and easily accessible data is a positive, the drawback is that anyone can upload data, and in many instances, the data might not, be verified. Along with data maybe not being verified, the other obstacle is that much of the data is unimportant for the actual task, at hand. In the Coroico area this was exemplified by power towers in the point dataset. Power towers are not relevant for this project, however they do make up 47 of the 191 points in Coroico, which is also visualized on Figure 16. Therefore such points were deleted from the vector points dataset, in order to produce a "smoother" dataset, containing only relevant data.

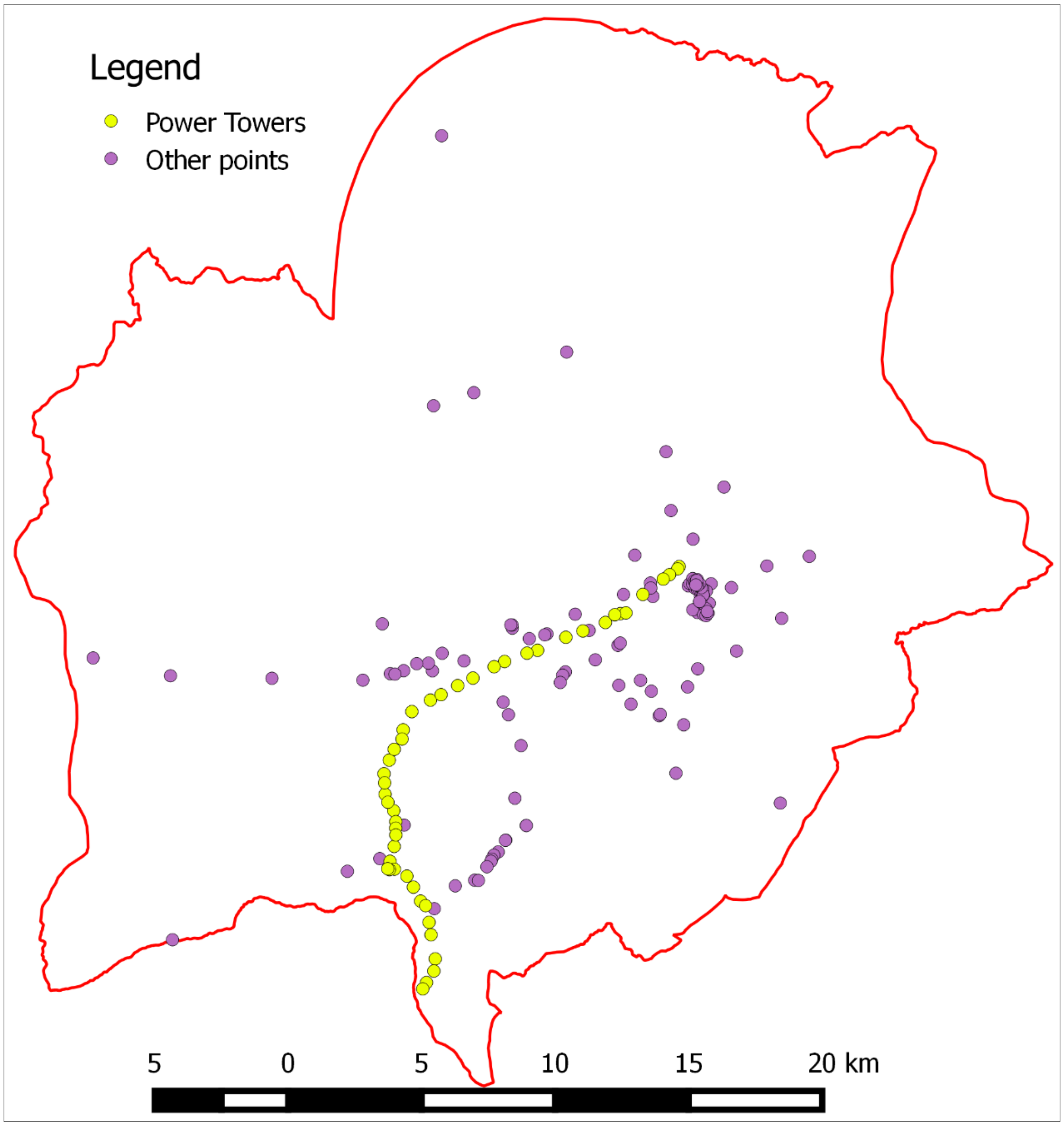

Figure 16: OSM points in the Coroico Municipality (OpenStreetMap, u.d.).

4.5.3. Vector Analysis

The layer showing rivers and streams is a merging between the two layers from GeoBolivia and OSM, respectively. While most of the waterways were portrayed in both layers, there were some instances were a waterway was present in one In the two maps below (Figure 17) and (Figure 18) waterways of the Coroico Municipality are shown on a background of the Landsat Satellite image and the digital elevation model. The digital elevation model is from GeoBolivia (GeoBolivia, 2014) and is a merging of two tiles, as the municipality spanned across more than just one tile. The digital elevation model is based on data from the ASTER (Advanced Spaceborne Thermal Emission and Reflection

Radiometer) that is mounted on the Earth Observing System (EOS) of NASA. The ASTER itself is a joint program between NASA, Japan's Ministry of Economy, Trade and Industry (METI), and Japan Space Systems (NASA, 2015).

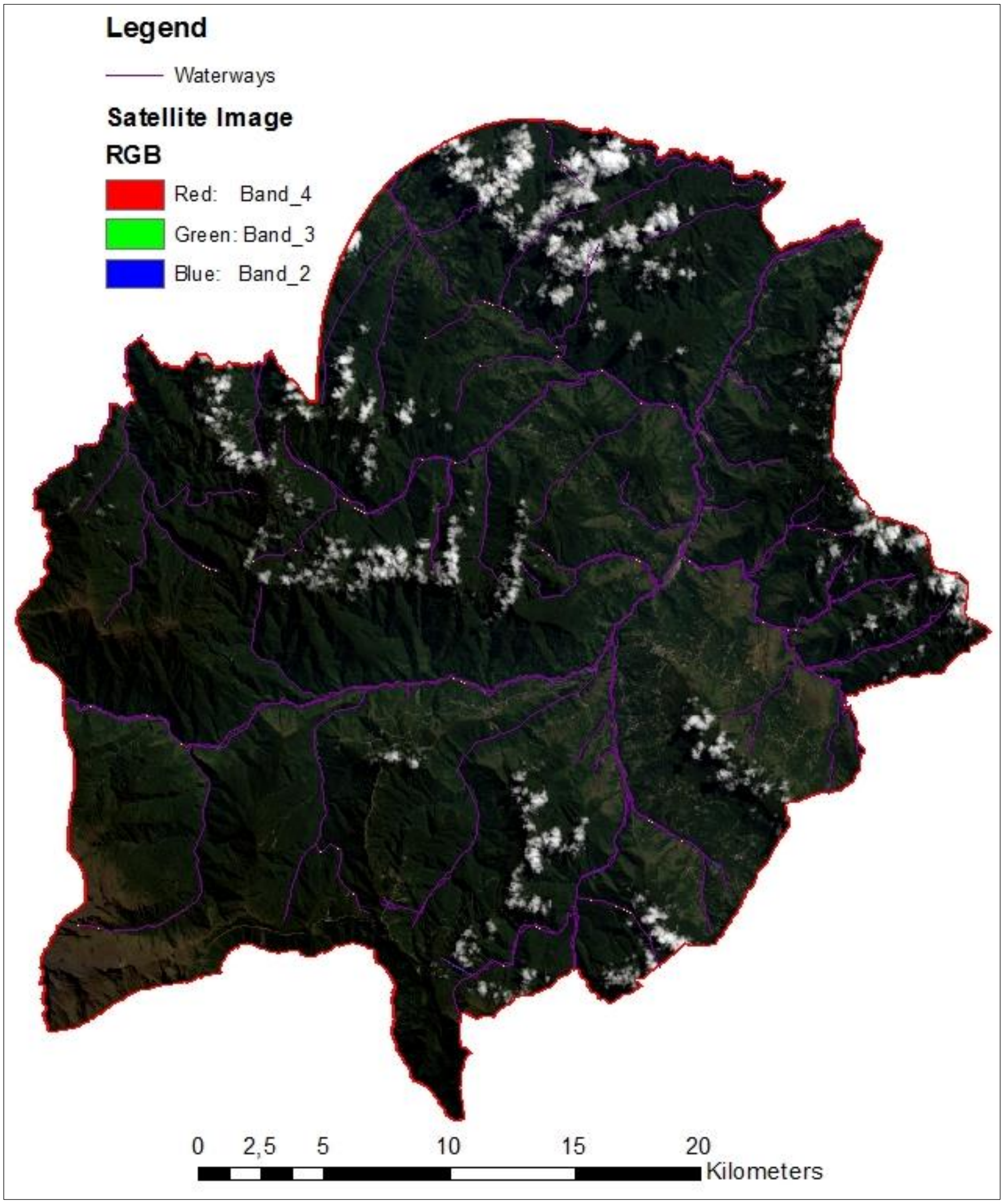

Figure 17: Waterways with satellite image as the background.

The placement of the waterways seem logic, when observing the two maps. On the first map (Figure 17) the true colors of the municipality are portrayed, and it is possible to get a visual representation of the mountain ridges in the municipality, and their mountain sides, down which rivers run towards the central river, running from the Southwest towards the Northeast, which, coincidentally, is the direction of the Amazonas. In the second map (Figure 18) the waterways are paired with the DEM.

Not surprisingly, the rivers run at the lowest elevations, again with smaller rivers and streams running down to the low elevations, from higher mountainous areas.

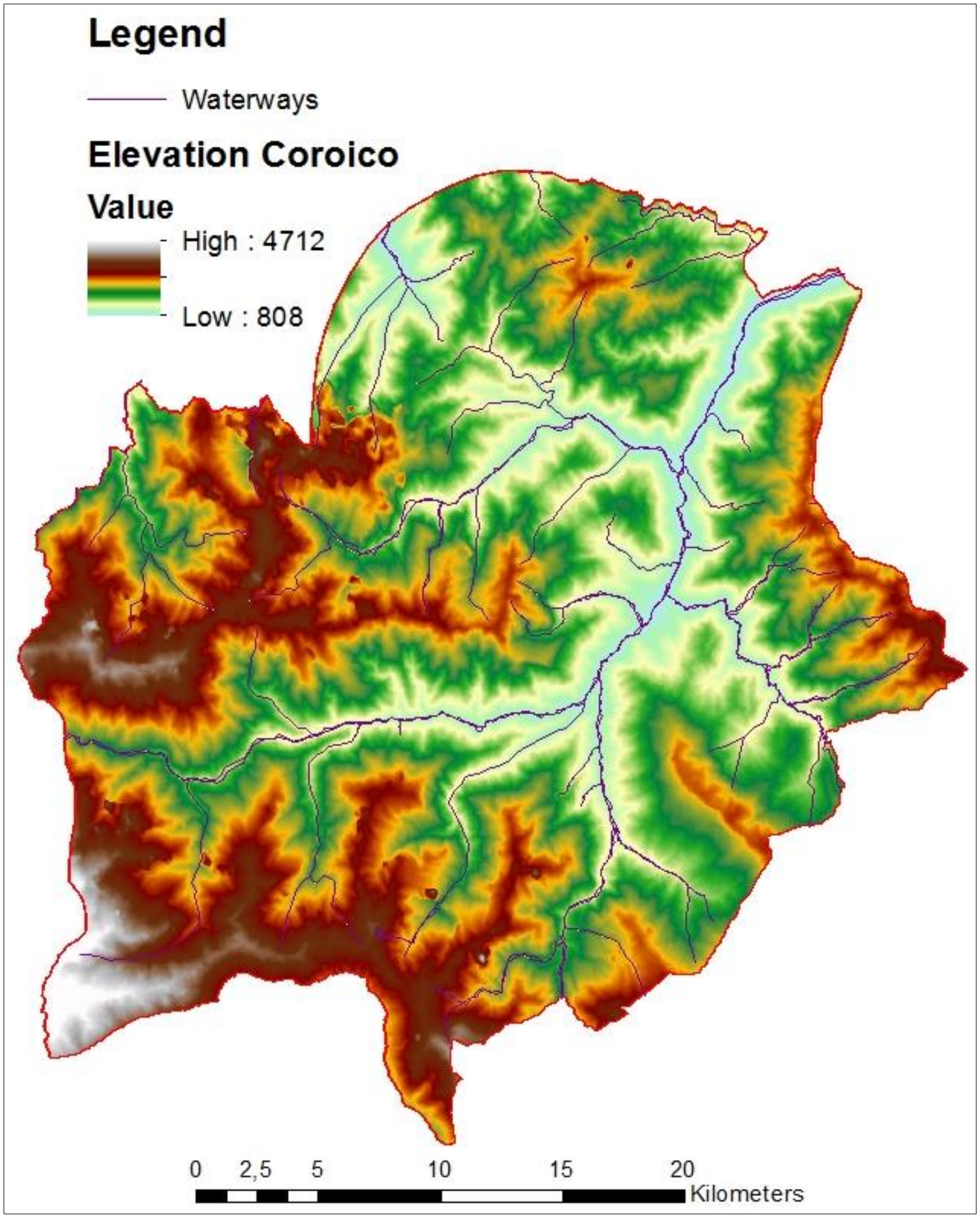

Figure 18: Waterways with DEM as background.

As with the waterway-layer, a layer showing roads and paths in the municipality was created, as a merging between the layers from GeoBolivia and OSM.

Many of the roads and paths are centered around the Southeastern part of the municipality which is also where the largest cities are located. Besides these roads, there are roads going North and West, towards other big cities. The Northwestern part of the municipality has very few roads, which can most likely be attributed to the high mountains there, some of them exceeding 4000 kilometers.

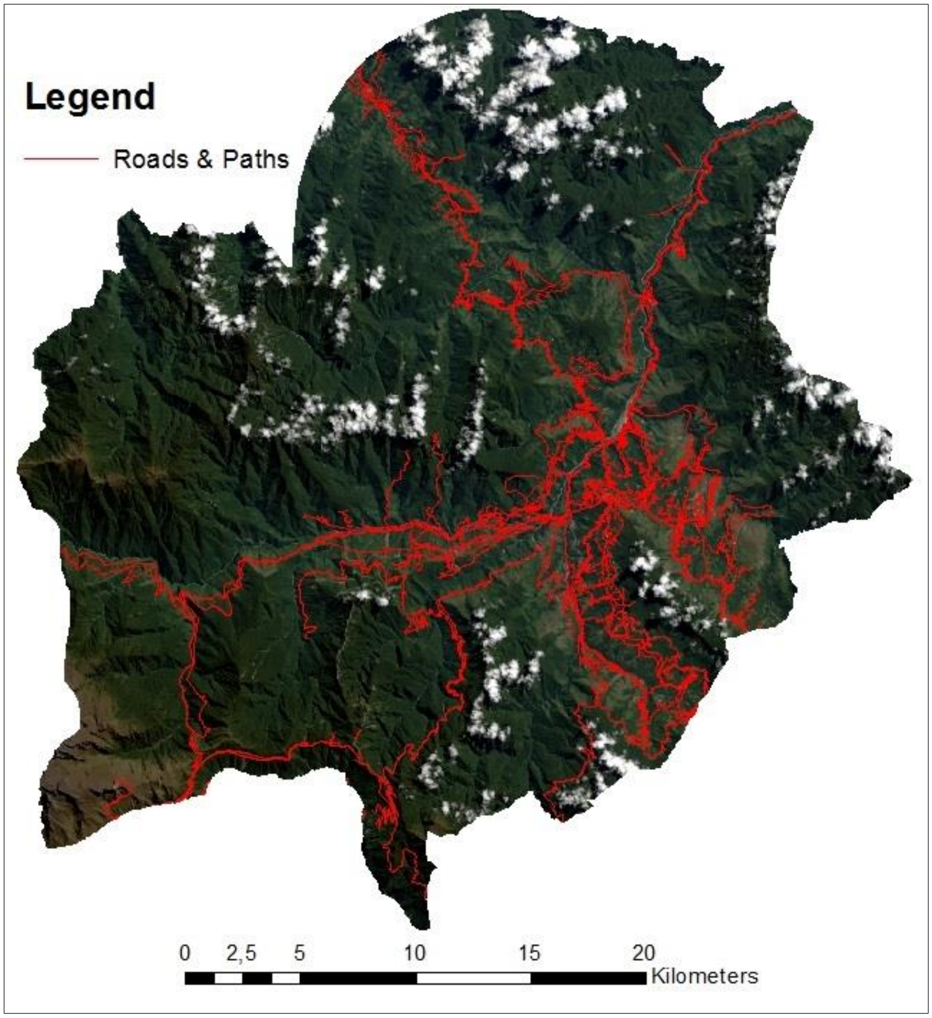

Figure 19: Roads and paths with satellite image as the background.

The last vector layer created is that of locations, villages, and towns (Figure 20). The method is exactly as with waterways and roads & paths. The reason for using a point vector layer showing locations, villages, and towns, is that it is highly likely that fields are placed within a certain distance from the nearest village.

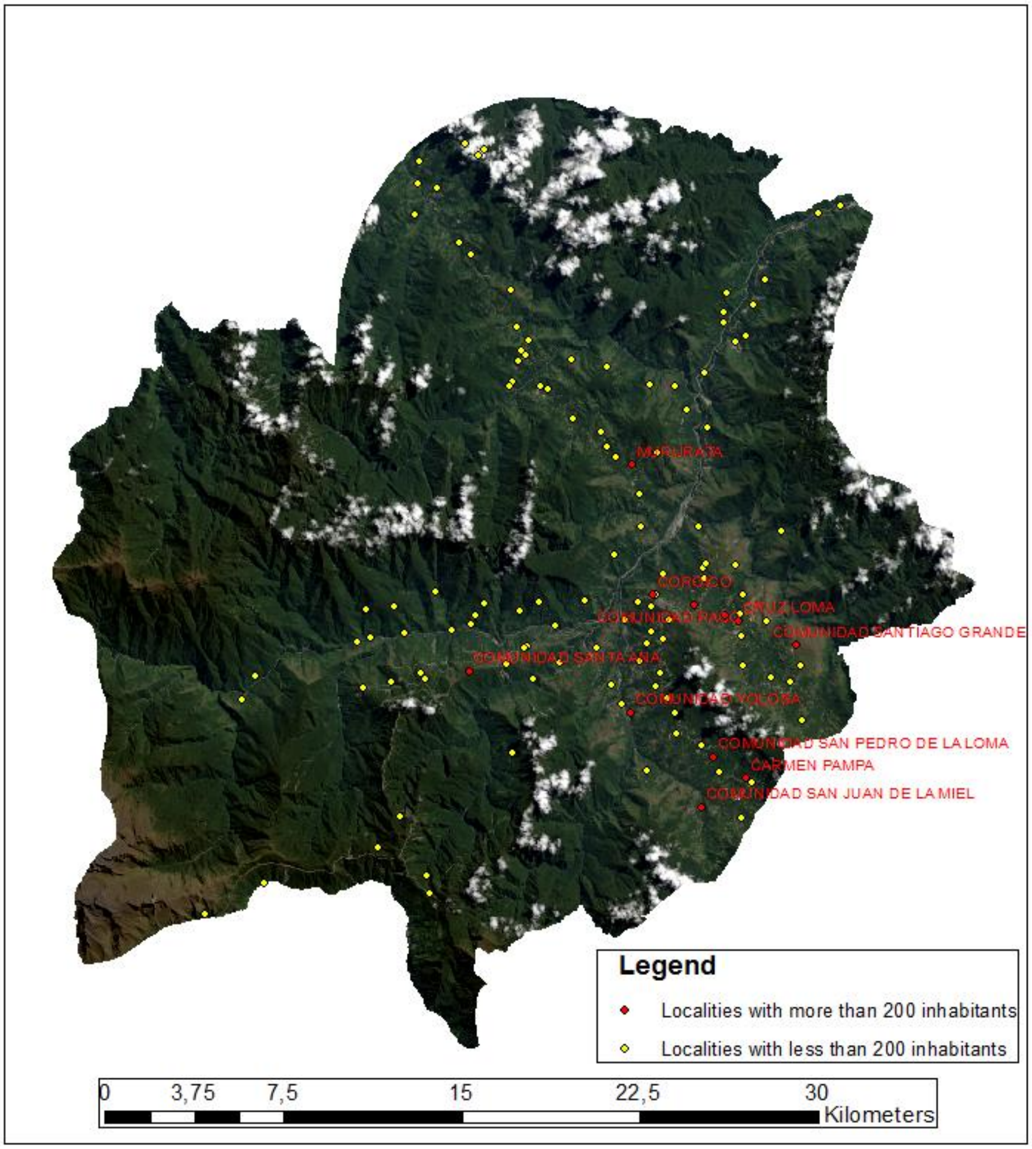

Figure 20: Localities with satellite image as the background.

In the municipality, there are 11 cities or communities with more than 200 inhabitants. They are primarily based in the Southeastern part of the municipality, fairly close to the municipal administrative center of Coroico, which is also the biggest city in the municipality, with 2197 inhabitants, which is about five times more than that of the second biggest city, Cruz Loma with 438 inhabitants.

4.6. Summarizing the 1st Aim of the Thesis

Both parts of the first Aim of the Thesis are hereby completed. The raster map in 30x30 meter cells of Coroico gives EWB, as well as local authorities, a map to use in future, in their work with mapping the area. The vector data from online sources afford an easy way to depict cities and villages, roads and paths, and rivers and streams. More so, the vector layers will be put to use in the MCDA in the following chapters.

This concludes the first Aim of the Thesis, and while both the raster land cover map and the vector layers are of satisfactory quality, a way to optimize those would have been through prior knowledge of the landscape, or in situ measurements. For EWB and the local authorities going forward, GPS-mapping of fields, shrub and tree cover would be of great value in order to improve the classification accuracy by way of better training samples.

5 CONTAMINATION SCREENING – 2ND AIM OF THE THESIS

In this, the second Aim of the Thesis, the focus will be on developing a screening tool that, via the maps created in GIS can be used to determine sites that are potentially contaminated by pesticides. The screening will be carried out by using Multiple-Criteria Decision Analysis (MCDA). The way to do this is by creating the maps as raster layers with the same cell size and extend, and then combine those layers. By standardizing the different layers and then multiplying them, it is possible to point out what sites have the highest risk of being contaminated.

5.1. MCDA Theory

As the name suggests, the MCDA is ideal at assisting in decision making, when having to account for multiple criteria. The examples are numerous, from locating new windmill farms, to finding the location of one's dream house. When using the MCDA it is possible to add a certain weighing to each layer, which allows one layer to be of higher importance than others (Balstrøm, et al., 2010). Another benefit of MCDA is the ability to use a GIS-program to ensure objective decisions in a geographic perspective. By setting up a set of rules before running the analysis, the researcher ensures that the impact of subjectivity is at a minimum (Longley, et al., 2011). In theory this should improve the validity of the analysis, since the researcher has the time to review the how the facts (input layers), should be handled and weighted in the analysis. A bad example of how to approach the MCDA would be to rush into the analysis, without carefully review the layers and their weighting, and then having to change those afterwards, in order to getting a new result, closer to what one had initially hoped to conclude.

5.2. MCDA Method

In this section the layers of the MCDA will be described, and how they came about.

5.2.1. Land Cover / Land Use

The cells most likely to be contaminated by pesticides are the ones classified as fields. They are therefore given the value 1, which is the highest, while 0 is the lowest. Shrub is believed to be second most likely and given the value 0,8, while tree cover is third at 0,6. Bare areas and artificial surfaces are set to 0,4 as there might be pesticide use there, as the classification had trouble with those areas. For the three classes water, cloud, and shade the value was set at 0 because it is not possible to say anything about land cover there, and because pesticides are not sprayed directly into water.

5.2.2. Elevation

In the article *"Coca in Bolivia*" it is described that optimal growing conditions for coca are between 1.500 and 1.600 feet (450-500 m) above sea level. As none of the areas in Coroico are that low, the elevation model is standardized with the lowest values getting 1, and the highest getting 0, with the rest of the values in between the two.

5.2.3. Slope

According to Robert South: *"The crop* (Coca) *is produced on steeply terraced mountain sides an on undulating terrain"* It can therefore be argued that areas with steep slopes are at bigger risk in terms of being contaminated by pesticides, in comparison to flat terrain. At the same token one has to consider that at one point, the slopes are simply too steep for man to navigate, and to grow crops. It was determined the optimal slope would be with a gradient between 20 and 45 and it is therefore given the value 1. A gradient below 20 was given the value of 0,75, while it was decided that at gradient above 45 and below 65 was harder to produce crops in than the gradient below 20. The gradient from 45 to 65 was therefore given the value 0,25. Gradients above 65 were deemed not possible to farm, and were therefore given the value 0.

5.3. Analysis and Discussion of screening

In the final screening of the three layers, the land cover was weighed at 0, while the elevation was weighed at 0,4. Time did not allow for inclusion of the slope in the MCDA.

What can be seen in the analysis (Figure 21) is that the areas most likely to be contaminated by pesticides are centered around the low areas of the municipality, which makes sense since the elevation is low here, Also the low lying areas of the municipality is where a lot of the fields and shrub land cover is located, which only heightens the chance of pesticides.

Another figure (Figure 22) is included to show the same map, but with buffers around the inhabited localities in the area. While it was not included in the MCDA, it does give a further indication of spots likely to be pesticide contaminated.

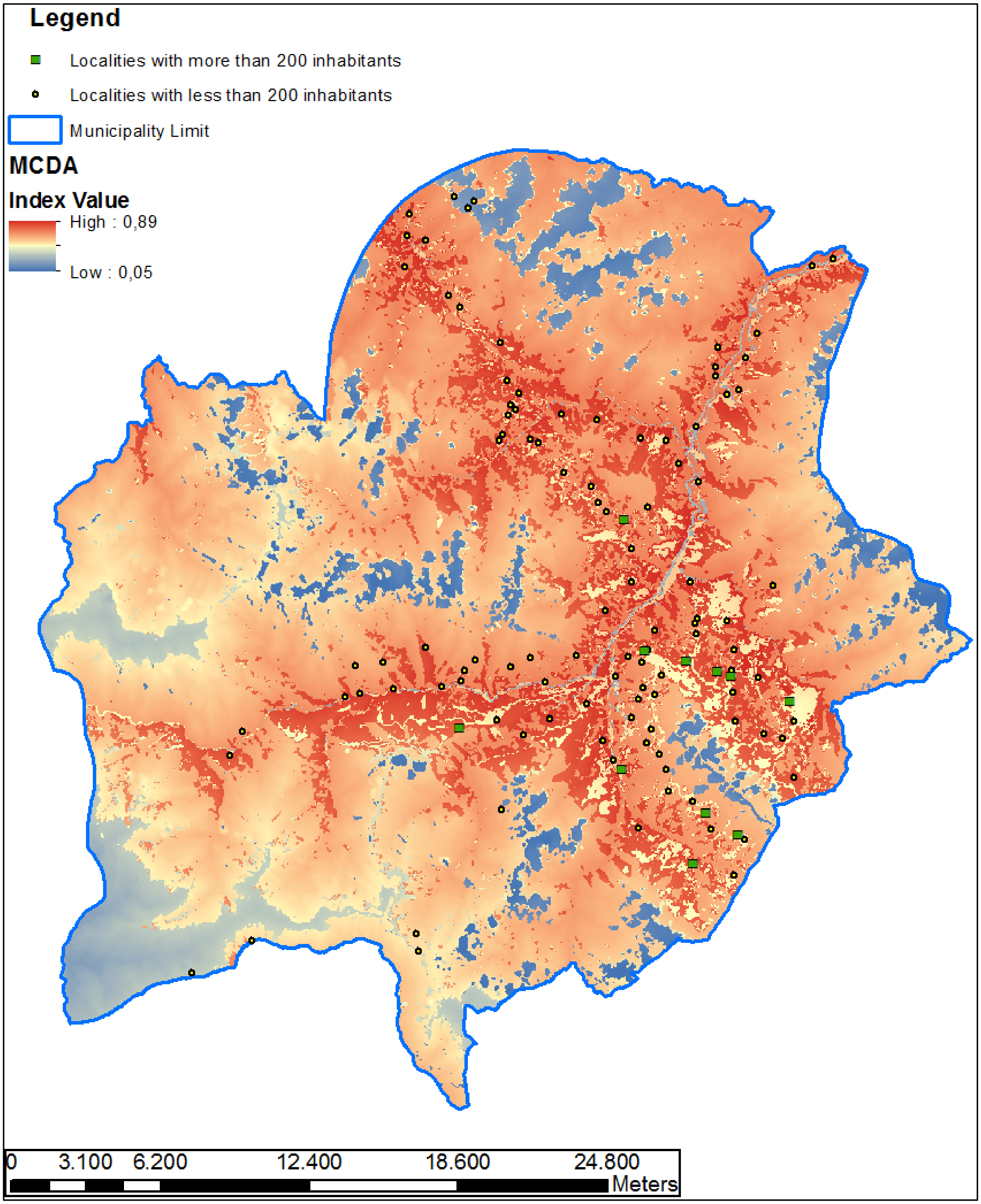

Figure 21: MCDA of Coroico.

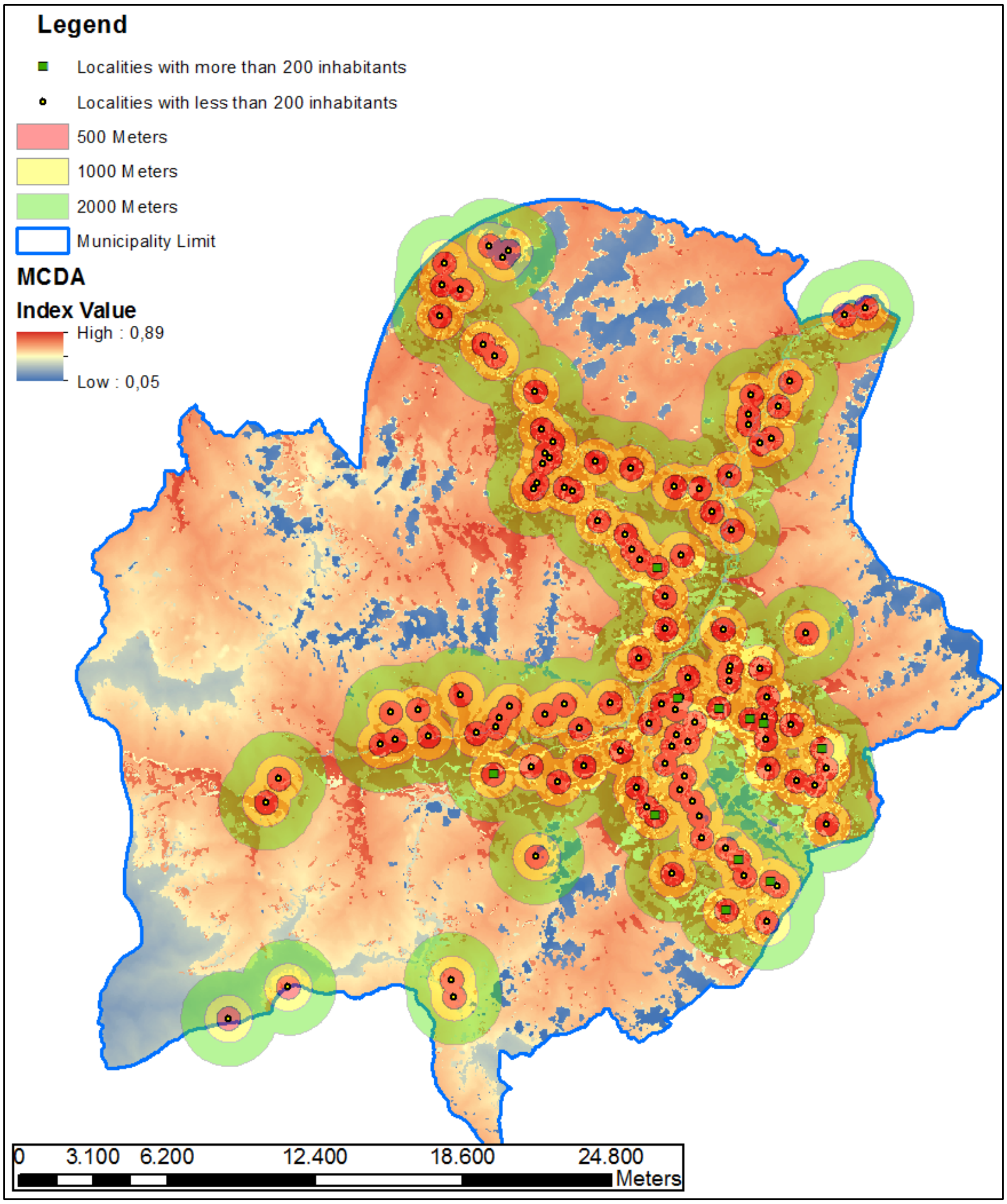

Figure 22: MCDA with an overlay of distances from localities.

.

6 CONCLUSION

The Land Use and Land Cover Classification produced a fairly accurate classification and the vectors of road, waterways and villages were successfully added, thus meeting the first Aim of the Thesis.

The second Aim of the Thesis was cut short due to time constraints, and leaves a lot of possibilities for Further Research.

7 Further Research

With additional time to continue the study it would be interesting to do a MCDA of the sites most likely to be contaminated by pesticides, in order to map the severity in terms of environmental impact and the impact on human settlements. This would take knowledge of flow direction and soil permeability, as well as knowledge of the various types of pesticides used in the area.

Also it would be hugely beneficial to create buffer zones around the road and paths, as well as the populated places in the area, in order to point out areas close to where people, which are going to be more likely to be where pesticide contaminated fields are located.

8 REFERENCES

Arnade, C. W., 1957. *The Emergence of the Republic of Bolivia.* Gainesville, Florida: University of Florida Press.

Balstrøm, T., Jacobi, O. & Bodum, L., 2010. Abstraktioner af den virkelige verden. In: *Bogen om GIS og geodata.* København NV: Forlaget GIS & Geodata, pp. 21-49.

Balstrøm, T., Jacobi, O. & Bodum, L., 2010. Analysemetoder. In: *Bogen om GIS og geodata.* København NV: Forlaget GIS & Geodata, pp. 171-229.

climate-data.org, n.d. *Climate: Coroico.* [Online]
Available at: http://en.climate-data.org/location/685849/
[Accessed 9. June 2015].

ESRI, 2011. *Smoothing zone edges with Boundary Clean and Majority Filter.* [Online]
Available at:
http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Smoothing_zone_edges_with_Boundary_Clean_and_Majority_Filter/009z0000003s000000/GUID-CBD68932-69B7-469B-9DCB-8A743B72B908/
[Accessed 5. June 2015].

ESRI, 2012. *Creating Training Samples.* [Online]
Available at: http://resources.arcgis.com/en/help/main/10.1/index.html#//00nv0000000n000000
[Accessed 3. June 2015].

ESRI, 2013. *Band Combinations for Landsat 8.* [Online]
Available at: http://blogs.esri.com/esri/arcgis/2013/07/24/band-combinations-for-landsat-8/
[Accessed 8. June 2015].

ESRI, n.d. *Data and Maps for ArcGIS.* [Online]
Available at: http://www.esri.com/data/data-maps/data-and-maps-dvd
[Accessed 25. November 2014].

Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment,* 1. April, pp. 185-201.

Garner, R., 2013. *Landsat 8 Instruments.* [Online]
Available at: https://www.nasa.gov/content/landsat-8-instruments/#.VW2Rpc_tlBc
[Accessed 2. June 2015].

GeoBolivia, 2014. *Institución.* [Online]
Available at: http://geo.gob.bo/#institution
[Accessed 26 March 2015].

GeoBolivia, 2014. *Servicios Y Ayuda.* [Online]
Available at: http://geo.gob.bo/#services
[Accessed 8. June 2015].

Ingeniører uden Grænser, 2013. *Miljøprojekt i Bolivia.* [Online]
Available at: http://iug.dk/projekter/projekter/projekter/miljoprojekt-i-bolivia#.VGHiVvmG9-4
[Accessed 11. November 2014].

Ingeniører uden Grænser, 2013. *Om IUG.* [Online]
Available at: http://iug.dk/content/om-iug/om-iug/om-iug#.VGHi7PmG9-4
[Accessed 11. November 2014].

Jones, H. G. & Vaughn, R. A., 2010. Sampling, error, and scaling. In: *Remote Sensing of Vegetation.* Dundee, Scotland: Oxford University Press, pp. 249-270.

Jones, H. G. & Vaughn, R. A., 2010. Use of spectral information for sensing vegetation properties and for image classification. In: *Remote Sensing of Vegetation.* Dundee, Scotland: Oxford University Press, pp. 163-195.

Jørs, E. et al., 2014. Do Bolivian small holder farmers improve and retain knowledge to reduce occupational pesticide poisonings after training on Integrated Pest Management. *Environmental Health,* 13(75), pp. 2-9.

Kaimowitz, D., Thiele, G. & Pacheco, P., 199. The Effects of Structural Adjustment on Deforestation and Forest Degradation in Lowland Bolivia. *World Development,* 27(3), pp. 505-520.

Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W., 2011. Representing Geography. In: R. Flahive, ed. *Geographic Information Systems & Science.* London: John Wiley & Sons, Inc, pp. 75-98.

Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W., 2011. Spatial Modeling with GIS. In: R. Flahive, ed. *Geographic Information Systems & Science.* London: John Wiley & Sons, Inc, pp. 403-424.

Morales, W. Q., 2010. *A Brief History of Bolivia.* 2. ed. New York: Lexington Associates.

NASA, 2015. *ASTER.* [Online]
Available at: https://asterweb.jpl.nasa.gov/
[Accessed 8. June 2015].

OpenStreetMap, n.d. *About OpenStreetMap.* [Online]
Available at: https://www.openstreetmap.org/about
[Accessed 10. June 2015].

Pacheco, P., 2006. Agricultural expansion and deforestation in lowland Bolivia: the import substitution versus the structural adjustment model. *Land Use Policy,* Volume 23, pp. 205-225.

Peters, J., 2015. *The Many Band Combinations of Landsat 8.* [Online]
Available at:
http://www.exelisvis.com/Home/NewsUpdates/TabId/170/ArtMID/735/ArticleID/14305/The-Many-Band-Combinations-of-Landsat-8.aspx
[Accessed 8. June 2015].

Richards, J. A., 2012. Image Classification in Practice. In: *Remote Sensing Digital Image Analysis.* 5. ed. Canberra: Springer, pp. 381-435.

Schouppe, J., 2006. *Coroico.* [Online]
Available at: http://www.coroico.info/
[Accessed 9. June 2015].

South, R. B., 1977. Coca in Bolivia. *Geographical Review,* 67(1.), pp. 22-33.

The World Bank, 2014. *World Development Indicators.* [Online]
Available at: http://databank.worldbank.org/data/views/reports/tableview.aspx
[Accessed 23. September 2014].

U.S. Geological Survey, 2012. *Global Land Cover Characteristics Data Base Version 2.0.* [Online]
Available at: http://edc2.usgs.gov/glcc/globdoc2_0.php#app1
[Accessed 5. June 2015].

U.S. Geological Survey, 2013. *Landsat 8 band combinations.* [Online]
Available at: http://landsat.usgs.gov/L8_band_combos.php
[Accessed 2. June 2015].

U.S. Geological Survey, 2013. *Remote Sensing Applications.* [Online]
Available at: http://landsat.usgs.gov/remote_sensing_applications.php
[Accessed 8. June 2015].

U.S. Geological Survey, 2014. *Landsat Missions - Band designations for the Landsat satellites.* [Online]
Available at: http://landsat.usgs.gov/band_designations_landsat_satellites.php
[Accessed 2. June 2015].

U.S. Geological Survey, 2015. *EarthExplorer.* [Online]
Available at: http://earthexplorer.usgs.gov/
[Accessed 9. June 2015].

U.S. Geological Survey, 2015. *Landsat Missions - Landsat 8.* [Online]
Available at: http://landsat.usgs.gov/landsat8.php
[Accessed 2. June 2015].

9 APPENDIX 1

```
GROUP = L1_METADATA_FILE
  GROUP = METADATA_FILE_INFO
    ORIGIN = "Image courtesy of the U.S. Geological Survey"
    REQUEST_ID = "0101411083948_00009"
    LANDSAT_SCENE_ID = "LC80010712013109LGN01"
    FILE_DATE = 2014-11-08T19:47:32Z
    STATION_ID = "LGN"
    PROCESSING_SOFTWARE_VERSION = "LPGS_2.4.0"
  END_GROUP = METADATA_FILE_INFO
  GROUP = PRODUCT_METADATA
    DATA_TYPE = "L1T"
    ELEVATION_SOURCE = "GLS2000"
    OUTPUT_FORMAT = "GEOTIFF"
    SPACECRAFT_ID = "LANDSAT_8"
    SENSOR_ID = "OLI_TIRS"
    WRS_PATH = 1
    WRS_ROW = 71
    NADIR_OFFNADIR = "NADIR"
    TARGET_WRS_PATH = 1
    TARGET_WRS_ROW = 71
    DATE_ACQUIRED = 2013-04-19
    SCENE_CENTER_TIME = 14:36:52.0206102Z
    CORNER_UL_LAT_PRODUCT = -14.90844
    CORNER_UL_LON_PRODUCT = -69.06880
    CORNER_UR_LAT_PRODUCT = -14.89937
    CORNER_UR_LON_PRODUCT = -66.96054
    CORNER_LL_LAT_PRODUCT = -16.90457
    CORNER_LL_LON_PRODUCT = -69.06949
    CORNER_LR_LAT_PRODUCT = -16.89422
    CORNER_LR_LON_PRODUCT = -66.94033
    CORNER_UL_PROJECTION_X_PRODUCT = 492600.000
    CORNER_UL_PROJECTION_Y_PRODUCT = -1648200.000
    CORNER_UR_PROJECTION_X_PRODUCT = 719400.000
    CORNER_UR_PROJECTION_Y_PRODUCT = -1648200.000
    CORNER_LL_PROJECTION_X_PRODUCT = 492600.000
    CORNER_LL_PROJECTION_Y_PRODUCT = -1869000.000
    CORNER_LR_PROJECTION_X_PRODUCT = 719400.000
    CORNER_LR_PROJECTION_Y_PRODUCT = -1869000.000
    PANCHROMATIC_LINES = 14721
    PANCHROMATIC_SAMPLES = 15121
```

```
  REFLECTIVE_LINES = 7361
  REFLECTIVE_SAMPLES = 7561
  THERMAL_LINES = 7361
  THERMAL_SAMPLES = 7561
  FILE_NAME_BAND_1 = "LC80010712013109LGN01_B1.TIF"
  FILE_NAME_BAND_2 = "LC80010712013109LGN01_B2.TIF"
  FILE_NAME_BAND_3 = "LC80010712013109LGN01_B3.TIF"
  FILE_NAME_BAND_4 = "LC80010712013109LGN01_B4.TIF"
  FILE_NAME_BAND_5 = "LC80010712013109LGN01_B5.TIF"
  FILE_NAME_BAND_6 = "LC80010712013109LGN01_B6.TIF"
  FILE_NAME_BAND_7 = "LC80010712013109LGN01_B7.TIF"
 FILE_NAME_BAND_8 = "LC80010712013109LGN01_B8.TIF"
  FILE_NAME_BAND_9 = "LC80010712013109LGN01_B9.TIF"
  FILE_NAME_BAND_10 = "LC80010712013109LGN01_B10.TIF"
  FILE_NAME_BAND_11 = "LC80010712013109LGN01_B11.TIF"
  FILE_NAME_BAND_QUALITY = "LC80010712013109LGN01_BQA.TIF"
  METADATA_FILE_NAME = "LC80010712013109LGN01_MTL.txt"
  BPF_NAME_OLI = "LO8BPF20130419141201_20130419145808.01"

  BPF_NAME_TIRS = "LT8BPF20130419140808_20130419144739.01"
  CPF_NAME = "L8CPF20130401_20130627.08"
  RLUT_FILE_NAME = "L8RLUT20130211_20431231v09.h5"
END_GROUP = PRODUCT_METADATA
GROUP = IMAGE_ATTRIBUTES
  CLOUD_COVER = 7.79
  IMAGE_QUALITY_OLI = 9
  IMAGE_QUALITY_TIRS = 9
  ROLL_ANGLE = -0.001
  SUN_AZIMUTH = 47.86199035
  SUN_ELEVATION = 50.79894912
  EARTH_SUN_DISTANCE = 1.0045092
  GROUND_CONTROL_POINTS_VERSION = 2
  GROUND_CONTROL_POINTS_MODEL = 250
  GEOMETRIC_RMSE_MODEL = 8.345
  GEOMETRIC_RMSE_MODEL_Y = 5.284
  GEOMETRIC_RMSE_MODEL_X = 6.458
  GROUND_CONTROL_POINTS_VERIFY = 73
  GEOMETRIC_RMSE_VERIFY = 4.131
END_GROUP = IMAGE_ATTRIBUTES
GROUP = MIN_MAX_RADIANCE
  RADIANCE_MAXIMUM_BAND_1 = 753.25427
  RADIANCE_MINIMUM_BAND_1 = -62.20398
  RADIANCE_MAXIMUM_BAND_2 = 771.34155
  RADIANCE_MINIMUM_BAND_2 = -63.69764
  RADIANCE_MAXIMUM_BAND_3 = 710.78461
  RADIANCE_MINIMUM_BAND_3 = -58.69683
```

```
    RADIANCE_MAXIMUM_BAND_4 = 599.37366
    RADIANCE_MINIMUM_BAND_4 = -49.49647
    RADIANCE_MAXIMUM_BAND_5 = 366.78677
    RADIANCE_MINIMUM_BAND_5 = -30.28937
    RADIANCE_MAXIMUM_BAND_6 = 91.21654
    RADIANCE_MINIMUM_BAND_6 = -7.53269
    RADIANCE_MAXIMUM_BAND_7 = 30.74486
    RADIANCE_MINIMUM_BAND_7 = -2.53892
    RADIANCE_MAXIMUM_BAND_8 = 678.32599
    RADIANCE_MINIMUM_BAND_8 = -56.01638
    RADIANCE_MAXIMUM_BAND_9 = 143.34859
    RADIANCE_MINIMUM_BAND_9 = -11.83777
    RADIANCE_MAXIMUM_BAND_10 = 22.00180
    RADIANCE_MINIMUM_BAND_10 = 0.10033
    RADIANCE_MAXIMUM_BAND_11 = 22.00180
    RADIANCE_MINIMUM_BAND_11 = 0.10033
  END_GROUP = MIN_MAX_RADIANCE
  GROUP = MIN_MAX_REFLECTANCE
    REFLECTANCE_MAXIMUM_BAND_1 = 1.210700
    REFLECTANCE_MINIMUM_BAND_1 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_2 = 1.210700
    REFLECTANCE_MINIMUM_BAND_2 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_3 = 1.210700
    REFLECTANCE_MINIMUM_BAND_3 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_4 = 1.210700
    REFLECTANCE_MINIMUM_BAND_4 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_5 = 1.210700
    REFLECTANCE_MINIMUM_BAND_5 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_6 = 1.210700
    REFLECTANCE_MINIMUM_BAND_6 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_7 = 1.210700
    REFLECTANCE_MINIMUM_BAND_7 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_8 = 1.210700
    REFLECTANCE_MINIMUM_BAND_8 = -0.099980
    REFLECTANCE_MAXIMUM_BAND_9 = 1.210700
    REFLECTANCE_MINIMUM_BAND_9 = -0.099980
  END_GROUP = MIN_MAX_REFLECTANCE
  GROUP = MIN_MAX_PIXEL_VALUE
    QUANTIZE_CAL_MAX_BAND_1 = 65535
    QUANTIZE_CAL_MIN_BAND_1 = 1
    QUANTIZE_CAL_MAX_BAND_2 = 65535
    QUANTIZE_CAL_MIN_BAND_2 = 1
    QUANTIZE_CAL_MAX_BAND_3 = 65535
    QUANTIZE_CAL_MIN_BAND_3 = 1
    QUANTIZE_CAL_MAX_BAND_4 = 65535
    QUANTIZE_CAL_MIN_BAND_4 = 1
```

```
    QUANTIZE_CAL_MAX_BAND_5 = 65535
    QUANTIZE_CAL_MIN_BAND_5 = 1
    QUANTIZE_CAL_MAX_BAND_6 = 65535
    QUANTIZE_CAL_MIN_BAND_6 = 1
    QUANTIZE_CAL_MAX_BAND_7 = 65535
    QUANTIZE_CAL_MIN_BAND_7 = 1
    QUANTIZE_CAL_MAX_BAND_8 = 65535
    QUANTIZE_CAL_MIN_BAND_8 = 1
    QUANTIZE_CAL_MAX_BAND_9 = 65535
    QUANTIZE_CAL_MIN_BAND_9 = 1
    QUANTIZE_CAL_MAX_BAND_10 = 65535
    QUANTIZE_CAL_MIN_BAND_10 = 1
    QUANTIZE_CAL_MAX_BAND_11 = 65535
    QUANTIZE_CAL_MIN_BAND_11 = 1
  END_GROUP = MIN_MAX_PIXEL_VALUE
  GROUP = RADIOMETRIC_RESCALING
    RADIANCE_MULT_BAND_1 = 1.2443E-02
    RADIANCE_MULT_BAND_2 = 1.2742E-02
    RADIANCE_MULT_BAND_3 = 1.1742E-02
    RADIANCE_MULT_BAND_4 = 9.9013E-03
    RADIANCE_MULT_BAND_5 = 6.0591E-03
    RADIANCE_MULT_BAND_6 = 1.5068E-03
    RADIANCE_MULT_BAND_7 = 5.0789E-04
    RADIANCE_MULT_BAND_8 = 1.1206E-02
    RADIANCE_MULT_BAND_9 = 2.3680E-03
    RADIANCE_MULT_BAND_10 = 3.3420E-04
    RADIANCE_MULT_BAND_11 = 3.3420E-04
    RADIANCE_ADD_BAND_1 = -62.21642
    RADIANCE_ADD_BAND_2 = -63.71038
    RADIANCE_ADD_BAND_3 = -58.70857
    RADIANCE_ADD_BAND_4 = -49.50637
    RADIANCE_ADD_BAND_5 = -30.29543
    RADIANCE_ADD_BAND_6 = -7.53420
    RADIANCE_ADD_BAND_7 = -2.53943
    RADIANCE_ADD_BAND_8 = -56.02759
    RADIANCE_ADD_BAND_9 = -11.84014
    RADIANCE_ADD_BAND_10 = 0.10000
    RADIANCE_ADD_BAND_11 = 0.10000
    REFLECTANCE_MULT_BAND_1 = 2.0000E-05
    REFLECTANCE_MULT_BAND_2 = 2.0000E-05
    REFLECTANCE_MULT_BAND_3 = 2.0000E-05
    REFLECTANCE_MULT_BAND_4 = 2.0000E-05

    REFLECTANCE_MULT_BAND_5 = 2.0000E-05
    REFLECTANCE_MULT_BAND_6 = 2.0000E-05
    REFLECTANCE_MULT_BAND_7 = 2.0000E-05
```

```
    REFLECTANCE_MULT_BAND_8 = 2.0000E-05
    REFLECTANCE_MULT_BAND_9 = 2.0000E-05
    REFLECTANCE_ADD_BAND_1 = -0.100000
    REFLECTANCE_ADD_BAND_2 = -0.100000
    REFLECTANCE_ADD_BAND_3 = -0.100000
    REFLECTANCE_ADD_BAND_4 = -0.100000
    REFLECTANCE_ADD_BAND_5 = -0.100000
    REFLECTANCE_ADD_BAND_6 = -0.100000
    REFLECTANCE_ADD_BAND_7 = -0.100000
    REFLECTANCE_ADD_BAND_8 = -0.100000
    REFLECTANCE_ADD_BAND_9 = -0.100000
  END_GROUP = RADIOMETRIC_RESCALING
  GROUP = TIRS_THERMAL_CONSTANTS
    K1_CONSTANT_BAND_10 = 774.89
    K1_CONSTANT_BAND_11 = 480.89
    K2_CONSTANT_BAND_10 = 1321.08
    K2_CONSTANT_BAND_11 = 1201.14
  END_GROUP = TIRS_THERMAL_CONSTANTS
  GROUP = PROJECTION_PARAMETERS
    MAP_PROJECTION = "UTM"
    DATUM = "WGS84"
    ELLIPSOID = "WGS84"
    UTM_ZONE = 19
    GRID_CELL_SIZE_PANCHROMATIC = 15.00
    GRID_CELL_SIZE_REFLECTIVE = 30.00
    GRID_CELL_SIZE_THERMAL = 30.00
    ORIENTATION = "NORTH_UP"
    RESAMPLING_OPTION = "CUBIC_CONVOLUTION"
  END_GROUP = PROJECTION_PARAMETERS
END_GROUP = L1_METADATA_FILE
END
```

10 Appendix 2

Maximum Likelihood Classification							
#	Class	Cell count	Percentage	Area km^2	1 cell:	30x30	m
1	Clouds	61423	5.84	55.28	1 cell:	900	m^2
6	Tree Cover	671693	63.88	604.52	1 cell:	0.0009	km^2
13	Shade	13148	1.25	11.83			
19	Artificial	58754	5.59	52.88			
20	Shrub	160050	15.22	144.05			
21	Water	4668	0.44	4.20			
22	Fields	3245	0.31	2.92			
93	Bare Areas	78507	7.47	70.66			
		1051488	100.00	946.34			

After first Majority Filter				
#	Class	Cell count	Percentage	Area km^2
1	Clouds	61746	5.87	55.57
6	Tree Cover	689024	65.53	620.12
13	Shade	11224	1.07	10.10
19	Artificial	52008	4.95	46.81
20	Shrub	155595	14.80	140.04
21	Water	4419	0.42	3.98
22	Fields	2778	0.26	2.50
93	Bare Areas	74689	7.10	67.22
		1051483	100.00	946.33

After second Majority Filter				
#	Class	Cell count	Percentage	Area km^2
1	Clouds	61853	5.88	55.67
6	Tree Cover	696100	66.20	626.49
13	Shade	10252	0.98	9.23
19	Artificial	48902	4.65	44.01
20	Shrub	154060	14.65	138.65
21	Water	4314	0.41	3.88
22	Fields	2640	0.25	2.38
93	Bare Areas	73365	6.98	66.03
		1051486	100.00	946.34