Centre for Geo-Information Science and Remote Sensing

Thesis Report GIRS-2016-29

MONITORING 40 YEARS OF LAND USE CHANGE IN THE MAU FOREST COMPLEX, KENYA

A LAND USE CHANGE DRIVER ANALYSIS

Rebecca Swart

WAGENINGEN UNIVERSITY
WAGENINGEN UR

MONITORING 40 YEARS OF LAND USE CHANGE IN THE MAU FOREST COMPLEX, KENYA

A LAND USE CHANGE DRIVER ANALYSIS

Rebecca Elisabeth Swart
890630819130

Supervisors

Prof. Dr. M. Herold[1]
Dr. B. Mora[1]
Dr. Mariana C. Rufino[1,2]

[1]Laboratory of Geo-Information Science and Remote Sensing, Wageningen, The Netherlands
[2]Center for International Forestry Research, Nairobi, Kenya

A thesis submitted in partial fulfilment of the degree of Master of Science at Wageningen University and Research, The Netherlands.

August, 2016
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis report: GIRS-2016-29
Wageningen University and Research
Laboratory of Geo-Information Science and Remote Sensing

ACKNOWLEDGEMENTS

In the first place I would like to thank my supervisors Brice Mora and Mariana Rufino for their feedback and ideas. In particular for their support and patience in a difficult period. Brice for reviewing every idea I had and giving me a lot of advice. Mariana for giving me context and providing me with information about the Mau forest, as well making me aware of the consequences of the land use changes for society and environment in this region. Then I would like to thank Sytze de Bruin (WUR) for his statistical advice: explaining and clarifying the statistical framework was very helpful. Furthermore, I want to thank Dennis Ojwang (CIFOR) for conducting the fieldwork and his effort in clarifying the classification and segmentation procedures. And then of course my mom, for our discussions about this research and reviewing my thesis. Last but not least, I want to thank all the Geo-Information Science students in the thesis room. It was nice that we could all help, support and listen to each other, and in particular the illegal coffee card made thesis writing a lot more bearable.

ABSTRACT

Land use systems are a fundamental part of the Earth's surface. Change of land use has significant impacts on climate, biodiversity, hydrological cycles, biogeochemical processes and human society. The Mau forest complex in Kenya is a montane forest ecosystem where significant land use changes have occurred: over the last few decades a quarter of the forest cover has been lost. The complex is one of the largest closed-canopy montane forests in Eastern Africa and is part of the global biodiversity hotspots; ecosystems that are exceptional rich in biodiversity and where abrupt and significant environmental degradation takes place. In addition, the forest fulfils a crucial socio-economic function because several millions of Kenyans rely on the products and services it provides. Currently there is no comprehensive overview of the driving factors of land use change in the Mau forest complex. Knowledge of these drivers is important for policy making and modelling future processes. To address this gap, this study analysed the land use changes and its drivers in the Mau forest complex in the period 1973-2013. Remote sensing and GIS techniques combined with multiple logistic regression modelling were used to identify the dynamics and drivers of land use change. This study shows that the main land use changes in the Mau forest complex in the period 1973-2013 were loss of forest and rangeland, while smallholder agriculture extended. More precisely, the largest land use change in the study area was a conversion from forest to smallholder agriculture. Hence, smallholder agriculture can be considered the most important proximate driver of deforestation in the Mau forest complex in every time period analysed. Based on the accuracy assessment and land use change dynamics analysis two land use change models were fitted: forest conversion and smallholder agricultural expansion. The regression analysis showed that biophysical and socio-economic factors were significant driving forces in both models. Drivers such as aspect East, curvature, the topographical wetness index, population density, distance to towns and distance to roads increased the odds of forest conversion, and in particular the distance variables became more important in more recently periods (1994-2003 and 2003-2013). In the agricultural expansion model, biophysical factors had mainly an influence in the second period (1984-1994), while the socio-economic underlying drivers were more important in the third and fourth periods (1994-2003 and 2003-2013). In the overall period 1973-2013, both the forest conversion and agricultural expansion model showed that a growth in population density increased the chance of land use change. In conclusion, this research demonstrates that land use change and its drivers show different spatial-temporal trends. The models revealed an increasing importance of socio-economic variables which means there is a need for better understanding the socio-economic aspects behind land use change. Therefore, future research and policies should be time and space specific and focus more on socio-economic drivers of land use change.

Keywords: Remote sensing, GIS, Land use change, Deforestation, Proximate drivers, Underlying drivers, Multiple logistic regression, Spatial-temporal analysis

TABLE OF CONTENTS

1. INTRODUCTION

1.1 BACKGROUND

Land use systems are a fundamental part of the Earth's surface and change of land use has significant impacts on climate, biodiversity, hydrological cycles, biogeochemical processes and human society (Baldyga et al., 2007; Lambin et al., 2001; Were et al., 2013). In fact, land use change, and in particular the process of deforestation, is one of the largest anthropogenic contributors to carbon dioxide emissions (Le Quéré et al., 2009). Given that carbon dioxide is the primary driver of global warming and climate change, understanding the causes of land use change (e.g. deforestation) is essential for mitigation policies such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation in developing countries) and modelling future scenarios (Hansen et al., 2009; DeFries et al., 2010; Were et al., 2014: Kissinger et al., 2011).

Kenya's Mau forest complex is an area where significant land use changes occurred: over the last few decades a quarter of the forest cover has been removed (Baxter, 2014; KWS, 2009; UNEP, 2011). The forest complex is one of the largest closed-canopy montane forests in Eastern Africa and is part of the global biodiversity hotspots; ecosystems that are exceptional rich in biodiversity and where abrupt and significant environmental degradation takes place (Mittermeier et al., 2011; Myers et al., 2000; Sloan et al., 2014; UNEP, 2006; Were et al., 2013). The Mau forest complex has a valuable function as carbon sink and is an essential catchment area as it forms the water catchment of several of Kenya's main rivers and feeds major lakes such as Lake Victoria, Lake Baringo and Lake Nakuru (Le Quéré et al., 2009; Olang and Kundu, 2011; Were et al., 2013). In addition, the complex fulfils a crucial socio-economic function because several millions of Kenyans rely on the products and services it provides (Sena, 2006; UNEP, 2012). The urban centres depend on its water supply and the area is important for economic sectors such as tourism and agriculture, as well for the tea industry and energy sector (Baxter, 2014; KWS, 2009; Olang and Kundu, 2011; UNEP, 2006; UNEP, 2012). Not only the human population needs the forest for its services, the water is essential for many domestic and wild animals that have their habitat there (Sena, 2006). Thus, the Mau forest complex plays an essential role for environment and society, and destruction of the area has substantial consequences.

Several studies analysed the land use changes that occurred in different areas in and around the Mau forest, and primarily observed a process of deforestation driven by agricultural expansion (Baldyga et al., 2008; Kiage et al., 2006; KWS, 2009; KFS, 2013; Mutoko et al., 2014; Olang and Kundu, 2011; Were et al., 2014). According to these studies, the land use changes can be explained by climatic variations (Kiage et al., 2006), an increased demand for land because of a growing human population (Olang and Kundu, 2011; KFS, 2013), topographical and soil-related factors (Were et al., 2014), and institutional arrangements and poor governance (KWS, 2009). It can be argued that in the Mau forest complex, various proximate and underlying drivers – which can be biophysical and socio-economic – and their interactions, determine land use changes (Geist and Lambin, 2002; Jaimes et al., 2010; Kissinger et al., 2011; Were et al., 2014).

Remote sensing and geographic information systems (GIS) techniques combined with statistical analysis methods are effective tools to identify, analyse and understand land use change dynamics (DeFries et al., 2010; Kiage et al., 2006; Long et al., 2007; Serneels and Lambin, 2001; Verburg et al., 2004; Were et al., 2014). Currently there is no comprehensive overview of the drivers of land use change in the Mau forest complex, and since knowledge of these drivers is important for policy making and modelling future processes, this research focuses on monitoring land use changes and analysing its drivers.

1.2 RESEARCH OBJECTIVE AND QUESTIONS

The main objective of this study is **to analyse land use change and the drivers of these changes in the Mau forest complex in the period 1973 - 2013**. Analysing historical data - in this case a 40-year period - helps understanding land use change patterns. In this way trends can be detected which contributes to future land use change strategies (Kissinger et al., 2011). Furthermore, focusing on areas where key drivers are most prevalent, risks are high and possibilities for successful intervention seem greatest, is essential for effective national policy making (Herold and Skutch, 2011).

In order to achieve the aim stated above, the following research questions were formulated:

1. What **land use changes** can be identified and what are the **proximate drivers** of land use change in the Mau forest complex?
2. What are the **underlying drivers** of land use change in the Mau forest complex?
3. What are the main **spatial-temporal trends** of land use change in the Mau forest complex, and how can they inform on future land use change trends?

1.3 METHODS

The figure 1.1 shows the overview of the methodology applied in this research (chapter 3 shows the methods in more detail). The first research question was studied by validating five remote sensing derived land use maps of a previous research and characterising the land use dynamics in the Mau forest complex. Secondly, the possible underlying drivers were identified by use of a literature study and data was collected to quantify these drivers. To finally answer the second question, and the third research question as well, a statistical analysis was performed: in particular, multiple logistic regression models were computed. This helped characterising the underlying drivers and the main spatial-temporal trends in the study area.

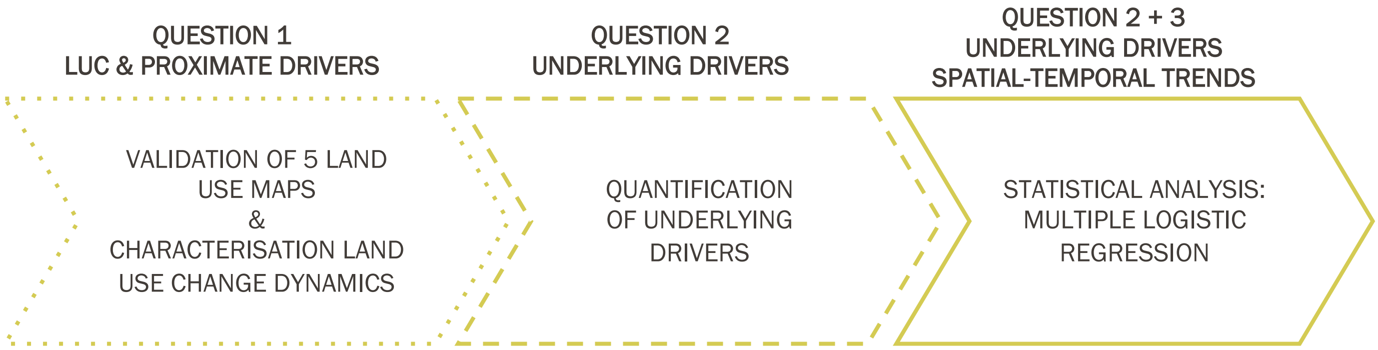


Figure 1.1: The methodology

1.4 STRUCTURE

This report starts with a literature review where proximate and underlying drivers and the modelling of land use change will be discussed. The next chapter elaborates on the methods used in this study: the methods used for validation of the land use maps, the characterisation of the land use change dynamics in the Mau forest, the quantification of land use change drivers, and the regression modelling framework will be explained. In chapter 4, the results of the validation, the characterisation of the land use dynamics and the statistical analysis are described. The results are reviewed and discussed in chapter 5. The report concludes in chapter 6 by answering the research questions and give recommendations for further research.

2. LITERATURE REVIEW

This chapter gives an overview of the literature that underpins the research. First the drivers of land use change will are described, and next the modelling of land use changes.

Land cover and *land use* are not the same concepts. According to Lambin et al. (2001:262) land cover points to 'the biophysical attributes of the Earth's surface', whereas land use is 'the human purpose or intent applied to these attributes' – the way land cover is used. In this report the applied terminology is *land use* because the classes that are used in the classification maps (see table 3.3) are land use categories according to the widely applied definitions of the IPCC (2000) and FAO (2005). Other studies use the concepts interchangeably as well, primarily because land use is often very similar to land cover (see for instance Kiage et al., 2006). Thus, in this research the drivers of *land use* change are studied without explicitly making a difference in terminology.

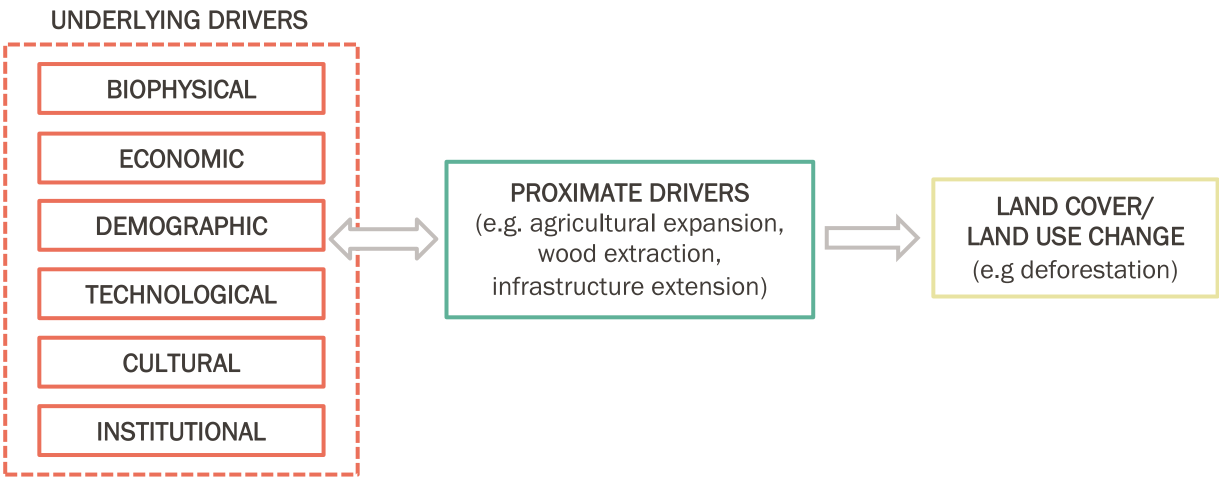


Figure 2.1: Framework drivers of land use change (based on Geist and Lambin, 2002; Kissinger et al., 2011)

2.1 THE DRIVERS OF LAND USE CHANGE

Land use change can be caused by various factors and considerable research has been conducted to identify the drivers of land use changes: from deforestation in tropical regions (DeFries et al., 2010; Geist and Lambin, 2002; Houghton, 2012) to urban processes of land use change (Braimoh and Onishi, 2006; Lambin et al., 2001; Seto & Kaufman, 2003), agricultural changes and land use changes in mountainous ecosystems (Alexander et al., 2015; Mottet et al., 2006; Serra et al., 2008). Geist and Lambin (2002) developed a theoretical framework that identified and analysed the drivers of tropical deforestation, which is widely adopted in other studies, policy programmes such as REDD+, and used to identify causes for different kinds of land use changes (Mutoko et al., 2014; Jaimes et al., 2010; Kissinger et al., 2011).

In Geist and Lambin's framework, the drivers are broadly divided into first, the proximate or direct drivers, and second, the underlying or indirect drivers of land use change (see figure 2.1). The complex interactions between the proximate and underlying causes describe land use change processes, or as Geist and Lambin (2002:149) argue in the context of deforestation: "there is no single or key variable ... that unilaterally impacts forest cover change; synergies between proximate causes and underlying driving forces best explain tropical forest cover losses". Thus, the combined effects of the proximate and underlying drivers explain land cover and land use changes. Drivers of land use change differ in time and space (Hosonuma et al., 2012).

2.1.1 Proximate drivers

The first group, the proximate or direct drivers, are the human activities and actions that have a direct effect on land cover or land use (Geist and Lambin, 2002; Jaimes et al. 2010; Kissinger et al., 2011). In the context of deforestation, multiple studies argue that the most important proximate driver is agriculture: the development of subsistence or commercial agriculture can cause forest loss (Alexander et al., 2015; Hosonuma et al., 2012; Kissinger et al., 2011). Other direct factors that explain deforestation can be the extraction of wood (for either commercial use or fuelwood for domestic use) or for instance the development of infrastructure (Geist and Lambin, 2002; Hosonuma et al., 2012; Kissinger et al., 2011). Another important proximate driver of deforestation – and land use change in general - is urban expansion (Long et al., 2007; Lambin et al. 2001).

2.1.2 Underlying drivers

Second, the underlying drivers, are 'fundamental (social) processes that underpin the proximate causes and either operate at the local level or have an indirect impact from national or global level' (Geist and Lambin, 2002: 143) or as Kissinger et al. (2011:5) state: 'they are complex interactions of social, economic, political, cultural and technological processes that affect the proximate drivers'. Several studies mention an extra group of environmental or biophysical drivers as explaining forces for land use changes (Aguiar et al., 2007; Jaimes et al., 2010; Were et al., 2014). Thus, broadly these underlying drivers can be divided into environmental or biophysical, economic, demographic, technological, cultural, and institutional drivers (see figure 2.1). They can be considered as interconnected concepts, all linked to each other, operating on multiple scales. Lambin et al. (2001) stress the importance of these scales: according to them, global forces are becoming the key factors of land use change. Therefore, it is essential to consider drivers of land use change from a local, regional, national or global level.

Environmental or biophysical

Climatic, geological, geomorphological and soil related factors are often documented as significant drivers for land use change (Aguiar et al., 2007; Mottet et al., 2005; Serra et al., 2008; Verburg et al., 2002; Were et al., 2014). First, climatic variability can explain changes in forest or agricultural land use: temperature, precipitation and solar radiation can have substantial influence (Serra et al., 2008). Next, geological features as rock types or sediments can be of importance for land use change (Verburg et al., 2002). Geomorphological factors, for instance elevation, slope and aspect, can have an effect on land use, and are in particular of importance in mountainous areas (Huang et al., 2007; Mottet et al., 2005). For instance, slopes give an indication about soil erosion and land degradation which is related to the way land is used (Kiage et al., 2006). Furthermore, the quality of the soil is of importance for land use change (Aguiar et al., 2007; Were et al., 2008).

Economic

Several studies argue that economic drivers are essential to consider when explaining land use changes (DeFries et al., 2010; Geist and Lambin, 2002; Kissinger et al., 2011). In fact, Geist and Lambin (2002) show that, globally, economic factors are the primary drivers for tropical deforestation. Market growth, rising income of population, commercialization or for instance change in poverty rates can all have an influence on the conversion of land use (Aguiar et al., 2007; Geist and Lambin, 2002). Other research shows that accessibility to markets is often identified as one of the economic determinants (Aguiar et al., 2007; Braimoh and Onishi, 2006; Serneels and Lambin, 2001). Moreover, a change in agricultural production can be an economic underlying driver. DeFries et al. (2010) identify - in their study - agricultural trade as one of the main determinants of deforestation. According to them, a rising demand for agricultural products in urban and international centres causes forest loss.

Demographic

Demographic drivers, and in particular population growth and population density, are extensively discussed in literature as an important driver for land use change (Alexander et al., 2015; Kiage et al., 2006; Kissinger et al., 2011). Alexander et al. (2015) emphasize that population growth is the largest driver for agricultural expansion on a global scale. Rising income that changes diet, the production of animal products and use of agriculture for bioenergy are consequences of population growth, and this in turn has an effect on agricultural land use (Alexander et al., 2015). In particular, in the humid tropics, the growth of urban population places pressure on rural landscapes for commercial agriculture (DeFries et al., 2010). Mutoko et al. (2014) argue that population growth does increase the demand for food, however their study shows that growth of population does not per se lead to agricultural intensification – other factors were more important in their study such as economic and technological changes.

Technological

Technological development can have an influence on land use change, and in particular agro-technological changes can be of importance (Geist and Lambin, 2002; Mutoko et al., 2014). Change of soil conservation measures, fertilizers and for instance improvement of crop varieties can advance farming practices, which can have an effect on land use (Mutoko et al., 2014). Furthermore, Geist and Lambin (2002) argue that technological changes in the wood sector can be a driver of tropical deforestation.

Cultural

Cultural factors are often underlying economic and institutional drivers: public attitudes, values and beliefs towards environment can be important (Geist and Lambin, 2002). For instance, society can be unaware of environmental issues (and the need of forest protection) as governments and policies might be focused on economic development or modernisation. Understanding values and beliefs of communities is essential: for example, indifference about nature and environment, in particular towards others and future generations, can contribute to land use changes (Geist and Lambin, 2002). In addition, when considering cultural factors, individual and household behaviour is important too.

Policy or institutional

The institutional and/or policy framework implemented is of high influence on how land is used - a change of regulations can have enormous effects on land use. Multiple studies identify this driver as very important one in the context of deforestation (Geist and Lambin, 2002) or urban expansion (Seto and Kaufman, 2003). Kissinger et al. (2010) mention the weak governance in some countries, and in Kenya, specifically towards the forest sector, that drives land use changes. As a matter of fact, in the Mau forest, the effects of forest conversion can be related to a change in regulations and weak policies in Kenya (Were et al., 2013).

2.2 MODELLING LAND USE CHANGE

A way to understand land use change dynamics and its drivers is by modelling the land use changes (Huang et al., 2007; Jaimes et al., 2010; Lambin and Geist, 2006; Serneels and Lambin, 2001; Serra et al., 2008; Seto and Kaufmann, 2003; Veldkamp and Lambin, 2001; Verburg et al. 2002; Were et al., 2014). Studies that model these changes can be divided in two broad categories: non-spatial and spatial. The first category models the magnitude and rate of land use change, without considering spatial variation. The second, on the other hand, focuses on land use changes at a specific spatial level (for instance administrative units) and detects spatial variation in land use change and the biophysical and socio-economic context (Seto and Kaufmann, 2003; Huang et al. 2007).

Spatial land use change models are thus of great significance for understanding land use change processes Knowledge of the drivers in time and space is needed to do this: the identification of proximate drivers is necessary for spatial prediction of changes, whereas insight in the underlying drivers is essential for the prediction of future drivers of land use change (Serneels and Lambin, 2001). Specifically, detecting historical trends of drivers helps constructing future scenarios because it broadens knowledge about past and recent drivers (Veldkamp and Lambin, 2001; Kissinger et al., 2011). For instance, information about the development of an underlying driver such as population growth is useful for predicting future land use change (e.g. continuous population growth keeps on affecting land use) (Kissinger et al., 2011).

In particular, empirical-statistical models of land use change are used to explain the driving forces of land use changes (Aguiar et al., 2007; DeFries et al., 2010; Millington et al., 2007; Serra et al., 2008; Serneels and Lambin, 2001; Verburg et al., 2004; Were et al., 2014). These models show the relationship between the observed land use changes and the drivers: they examine the statistical significance of the drivers and help predicting land use changes (Millington et al., 2007; Yin et al., 2014). More specifically, regression methods are used for this: from linear regression, to geographically weighted regression, regression tree methods and multiple logistic regression (Aguiar et al., 2007; DeFries et al., 2010; Jaimes et al., 2010; Verburg et al., 2004).

Especially the last method, *multiple logistic regression,* has proved to be suitable for spatial land use change modelling, and has been widely applied in land use change research (Braimoh and Onishi, 2006; Millington et al., 2007; Serneels and Lambin, 2001; Serra et al., 2008; Verburg et al., 2004; Were et al., 2014; Yin et al., 2014; Zar, 2015). In these models land use changes are often represented as discrete events (e.g. forest changes to agriculture). Logistic regression is used to measure the probability of the presence of a particular land use change process – given a selection of drivers (Yin et al., 2014). A benefit of logistic regression modelling is that the method can be used for prediction of future land use changes (Millington et al., 2007; Were et al., 2014).

3. METHODS

Figure 3.1 shows the methodology of the research (see also figure 1.1 in the introduction). The validation of the land use maps and the characterisation of the land use dynamics were of importance to answer research question 1 (the yellow arrows in the figure 3.1) and helped identifying the dependent variables for regression modelling. The quantification of the underlying drivers (part of research question 2, the red arrow) resulted in finding the independent variables. Last, the dependent and independent variables were statistically analysed by using multiple logistic regression models (green arrow). This resulted in the identification of the underlying drivers and the main spatial-temporal trends of land use change, which helped answering both research question 2 and research question 3.

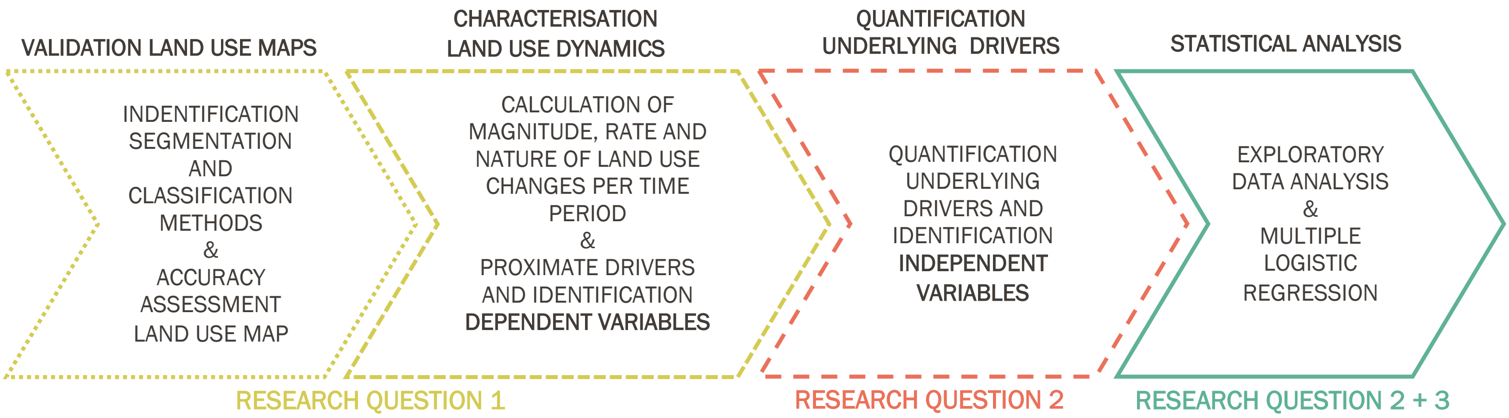


Figure 3.1: Framework methodology

In this chapter the study area is described first. Next, the data that was used is discussed: the remotely sensed data for the land use maps, the fieldwork data for validation, and the datasets that were used for regression modelling (paragraph 3.2). Furthermore, the methods that were applied to validate the land use maps and to characterise the dynamics of change are discussed. In addition, the methodology to conduct the dependent and independent variables for statistical modelling is explained. The last paragraph will focus on the logistic regression modelling framework.

3.1 STUDY AREA

The study area, the Mau forest complex, is located in the Rift Valley province in Kenya, between latitudes 0°91' N - 1°49' S and longitudes 34°9' - 36° 6' E (see figure 3.2). It covers a total area of about 24,000 km^2 and 13 counties: Baringo, Bomet, Elgeyo Marakwet, Kericho, Kiambu, Kisumu, Nakuru, Nandi, Narok, Nyamira, Nyandarua and Uasin Gishu (see figure F1 in Appendix F). The most dominant land use in the study area is smallholder agriculture (50.7%), followed by rangeland (23.7%) and forest (17.7%) (see figure 4.1). Furthermore, several rivers run through the study area and a few large lakes are located in the county Nakuru: Lake Naivasha and Lake Nakuru. Important urban centres are Nakuru and Kericho. The altitude ranges between 1000 to 3200 meters above sea level, with the most highly elevated areas located in the middle part of the study area: the northern part of the county Narok and western areas of Nakuru (NASA, 2011).

The area falls into different climate zones: equatorial tropical rainforest climates with high monthly rainfalls and tropical savannah climates with dry seasons. In the study area, the rainfall pattern is bimodal and the long rainy season is from March to May, and short rainy season October to December (Mutoko et al., 2014). Depending on the exact location in the area, dry season generally runs from January to March, and May to September. Thus, there are some differences in the amount of rainfall across the study area: the annual monthly rainfall is higher in the western counties of the study area (western parts of Kisimu, Kericho and Bomet), whereas the north and south eastern counties prove to be dryer (Baringo, Elgeyo

Marakwat, Nakuru, Narok) (Worldbank, 2016). Annual monthly temperatures vary spatially as well: the highly elevated parts show low annual temperatures (minimums of 10.6 °C), while the most northern parts in Elgeyo Marakwet and Baringo have high annual temperatures (maximum temperatures of 24.6 °C) (Worldbank, 2016).

The estimated total human population in the study area is around 4.8 million, with Nakuru as largest county in terms of inhabitants (KNBS, 2015). The Mau forest complex is not only important for the livelihoods of the people in the study area, it support livelihoods of people in the Rift Valley province and western parts of Kenya as well (KWS, 2009). As mentioned, agriculture is the most dominant land use, and is foremost important for food security. In addition, economic reasons motivate agriculture: it functions as a source of income. The farmers mostly export their product to other places within Kenya (Atele et al., 2012). Internationally, the study area is of importance for the tea industry and tourism (KWS, 2009).

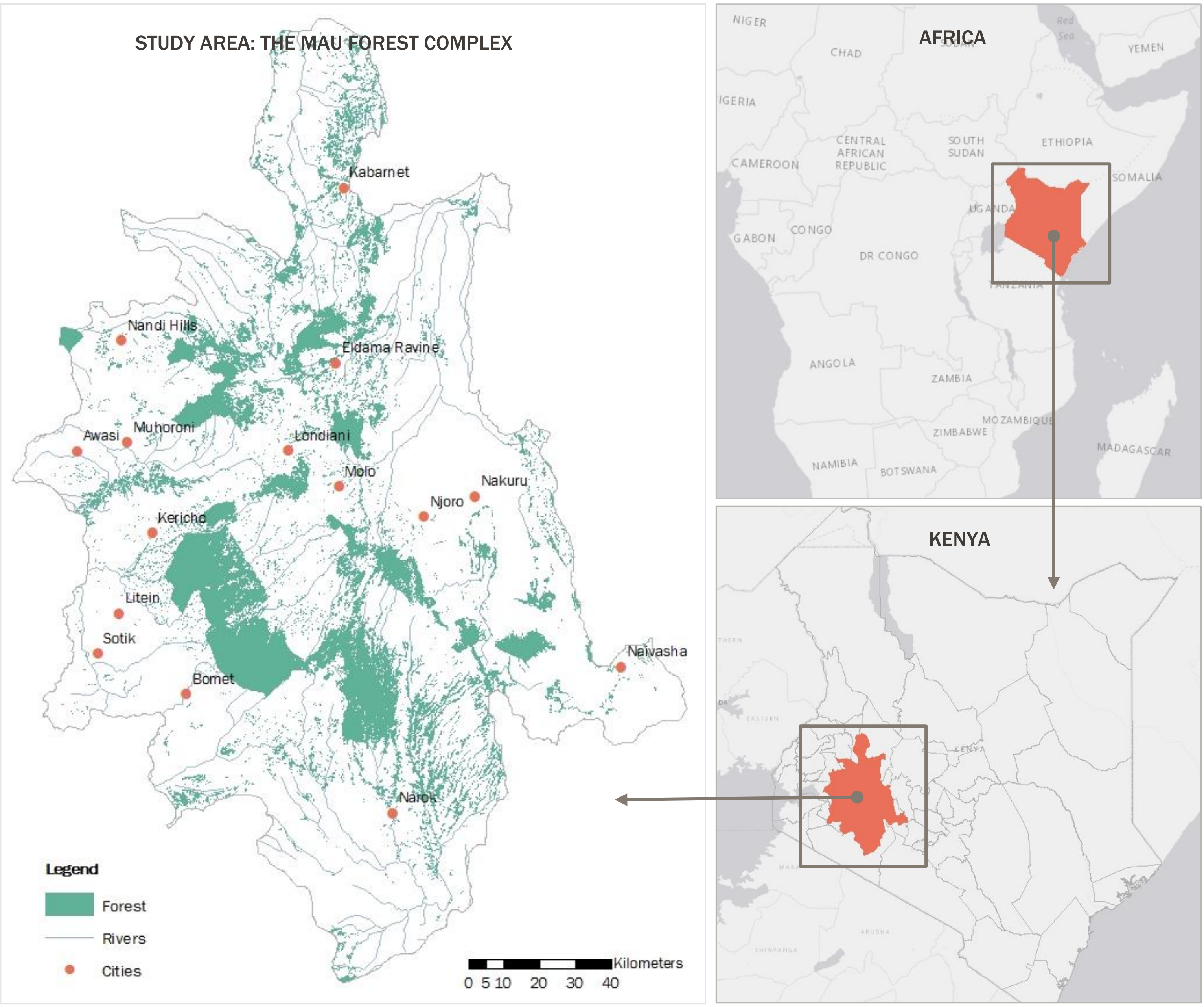


Figure 3.2: The study area (data from ILRI, 2000; KNBS, 2015; WRI, 2016)

3.2 DATA

The data used in this study were Landsat satellite images (see table 3.1), which were used to produce land use maps (Vita et al., 2014), fieldwork data for validation of these maps, and multiple datasets for the logistic regression modelling, which are summarized in table 3.2.

3.2.1 Remotely sensed data

Table 3.1 shows the Landsat data Vita et al. (2014) used for the segmentation and classification of five land use maps. The maps for the years 1973, 1984, 1994, 2003 and 2013 were created by use of different Landsat sensors (obtained via USGS): Multispectral Scanner System (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper (ETM+) and Operational Land Imager (OLI). As in tropical areas it is cloudy throughout the year, selecting images without cloud cover can be difficult. The images shown in table 3.1 were chosen by Vita et al. (2014) based on temporal resolution, in every time period an image, and data availability, thus only cloud-free images.

Table 3.1: Implemented images for classification

LANDSAT SENSOR	SCENE ID (PATH-ROW)	DATE	SPATIAL RESOLUTION
MSS	181-060	31 Jan 1973	60m
MSS	181-061	31 Jan 1973	60m
MSS	182-060	01 Feb 1973	60m
TM	169-060	01 Jul 1984	30m
TM	169-061	01 Jul 1984	30m
TM	169-060	17 Oct 1994	30m
TM	169-061	04 Dec 1994	30m
ETM	169-060	04 Feb 2003	30m
ETM	169-061	04 Feb 2003	30m
OLI	169-060	30 May 2013	30m
OLI	169-061	17 Jul 2013	30m

3.2.2 Fieldwork data

Fieldwork was conducted in February 2016 to validate the extracted land use maps of 2013. Appendix A shows a map of the fieldwork plan. 180 sampling points were collected by use of Garmin 30 handheld GPS device. Validation points were collected in particular areas (see the boxes figure A1 in Appendix A). Reasons to focus on specific areas were the large nature of the study area, the poor road network, and the limited available time for validation in the field. The areas were appointed based on a validation check of the land use maps by use of Google Earth high resolution imagery (see paragraph 3.3.2 methods accuracy assessment).

3.2.3 Logistic regression modelling datasets

The data that was used for the logistic regression analysis is summarized in table 3.2. The table shows the dependent variables, the independent variables, the data types, the units of the data, and the data sources. The binary dependent variables were derived from the land use classification maps created by Vita et al., (2014). Rainfall and temperature data was derived from WorldClim, slope, aspect, curvature and the topographical wetness index (TWI) were derived from the ASTER GDEM elevation data, and soil data obtained via World Soil Information. Population data was derived from several census reports from Kenya Central Bureau of Statistics. To calculate the other socio-economic variables, the accessibility variables, data of roads was derived via the Centre for International Forestry Research, data of rivers via the World Resource Institute and data of towns via the International Livestock Research Institute and Central Bureau of Statistics. Paragraph 3.4 and 3.5 explain the exact methods that were used to compute the dependent and the independent variables.

Table 3.2: The variables used for logistic regression (R stats package version 3.4.0)

VARIABLES	TYPE	UNIT	SOURCE
DEPENDENT VARIABLES			
Forest conversion, 1973-1984	Binary	0 - 1	Land use classification maps of Landsat imagery (Vita, 2014)
Forest conversion, 1984-1994	Binary	0 - 1	
Forest conversion, 1994-2003	Binary	0 - 1	
Forest conversion, 2003-2013	Binary	0 - 1	
Forest conversion, 1973-2013	Binary	0 - 1	
Smallholder agriculture expansion, 1973-1984	Binary	0 - 1	
Smallholder agriculture expansion, 1984-1994	Binary	0 - 1	
Smallholder agriculture expansion, 1994-2003	Binary	0 - 1	
Smallholder agriculture expansion, 2003-2013	Binary	0 - 1	
Smallholder agriculture expansion, 1973-2013	Binary	0 - 1	
INDEPENDENT VARIABLES			
Biophysical			
Rainfall	Continuous	mm	worldclim.org
Temperature	Continuous	ºC * 10	worldclim.org
Elevation	Continuous	m	ASTER GDEM[1]
Slope	Continuous	º	DEM extract
Aspect	Continuous	º	DEM extract
Curvature	Continuous	1/100 m	DEM extract
TWI	Continuous	-	DEM extract
Soil pH	Continuous	water * 10	ISRIC[2]
Soil CEC	Continuous	cmol/kg	ISRIC
Socio-economic			
Population density, 1979	Continuous	pers/km^2	KCBS[3]
Population density, 1989	Continuous	pers/km^2	KCBS
Population density, 1999	Continuous	pers/km^2	KCBS
Population density, 2009	Continuous	pers/km^2	KCBS
Population density, change 1979-2009	Continuous	%	KCBS
Distance to roads	Continuous	km	CIFOR[4]
Distance to rivers	Continuous	km	WRI[5]
Distance to towns	Continuous	km	KCBS, ILRI[6]

[1]DEM via https://asterweb.jpl.nasa.gov/gdem.asp
[2]ISRIC World Soil Information datasets via www.isric.org
[3]Kenya Central Bureau of Statistics: population census reports of 1979, 1989, 1999 and 2009
[4]Centre For International Forestry Research road network dataset of Kenya
[5]World Resources Institute river network dataset
[6]International Livestock Research Institute towns dataset

The remote sensing software that was used to create land use classification maps was eCognition Developer 8.9. The GIS software used for computation of variables was ArcMap 10.1 and QGIS 2.14. For pre-processing and statistical analysis R version 3.3.0 was used, supplemented with statistical software IBM SPSS Statistics 22.

3.3 VALIDATION LAND USE MAPS

This paragraph describes the methodology of the validation of the land use maps. First the exact methods used for segmentation and classification were identified. Secondly an accuracy assessment was applied.

3.3.1 Image segmentation and classification

Figure **3.3** shows the methodology Vita et al. (2014) applied for image segmentation and classification. The input for object based segmentation in eCognition was a Landsat 2003 image with the following parameters: band 2, 3 and 4 received a weight of 1, whereas the Normalised Differenced Vegetation Index received a weight of 2 (see box 1 and 2 in the figure **3.3)**. The scale parameter was 9 pixels (30x30m), shape value 0.1 and compactness 0.5. The output was a dataset consisting of 121.000 land use polygons. Next, the dataset was intersected with the Africover classification data: a land cover classification scheme created in 2004 by the FAO. This resulted in a land use map where each of the 121.000 polygons were coded with one of the 59 Africover classes (box 5 in the figure).

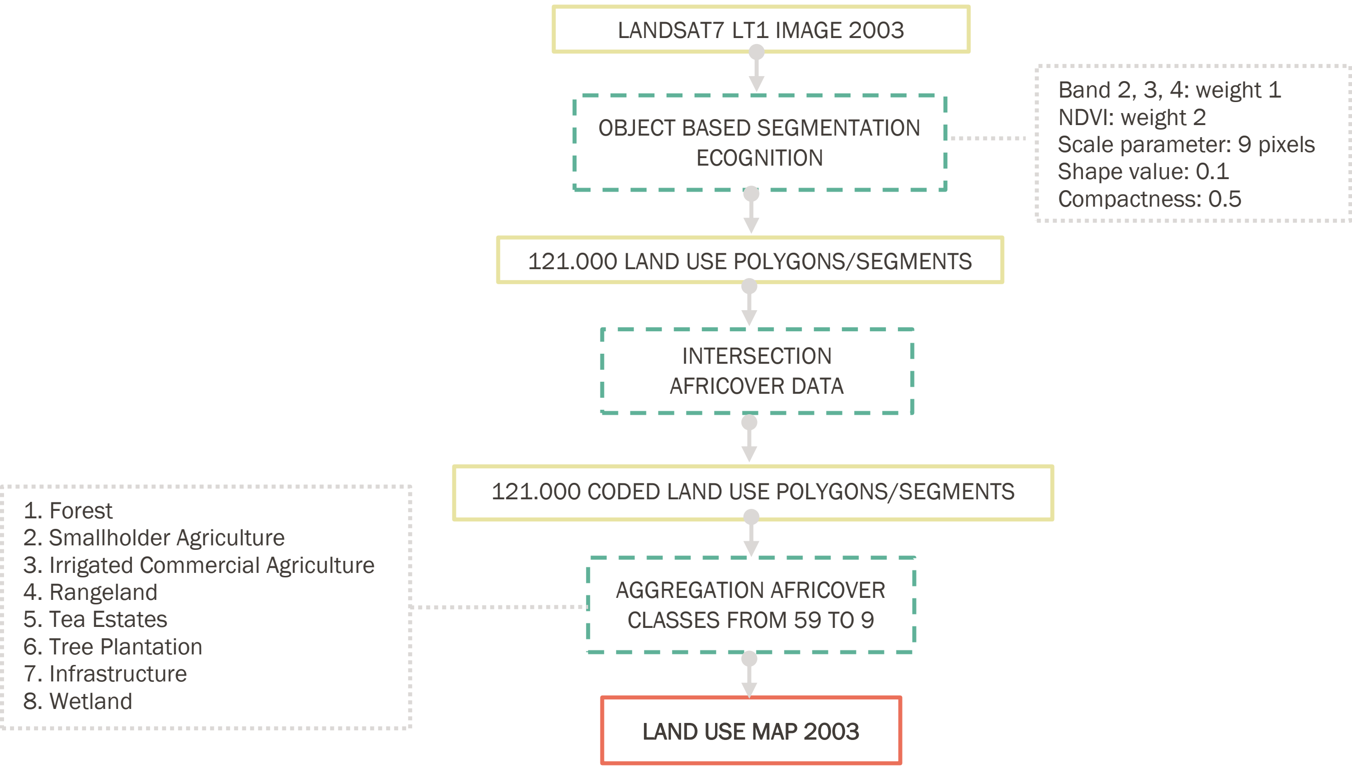


Figure 3.3: Image segmentation and classification methods Vita et al. (2014)

Thus, a classification scheme was designed by Vita et al. (2014) based on the Africover land cover classification. Table 3.3 shows the Africover land use classes and the aggregation of the final 9 land use classes (see table B1 in Appendix B for the full scheme of the 59 classes). The result of the aggregation of the land use classes was a 2003 land use map of the study area consisting of 10910 land use polygons. Finally this map was used to produce the land use maps for 1973, 1984, 1994 and 2013.

Table 3.3: Land use classes

CLASS	DEFINITION	LAND USE DESCRIPTION	AFRICOVER CLASS
F	Forest	Tree height > 2 m, crown coverage > 30%, includes bamboo forest	Closed trees with shrubs, closed trees bamboo
SA	Smallholder Agriculture	Includes small scale agriculture, horticulture, rain fed maize , agriculture frontier, small (<2ha) tea orchards intercropped with food	Rain fed shrub crops small fields, rain fed herbaceous crop large to medium fields, large fields, tree crop small fields,

		crops, small tree/shrub orchards intercropped with herbaceous crops	irrigated herbaceous crop small fields sugarcane, large fields coffee
IC	Irrigated Commercial Agriculture	High input crops, irrigated, use of mechanization, includes paddy rice	Irrigated crops large to medium field, includes sugarcane and rice, irrigated crop sugarcane large fields
TE	Tea Estates	Large tea estates (>2 ha), high use of fertilizers	Tea, rain fed shrub crops
TP	Tree Plantations	Timber plantation of eucalyptus, pines, cypress	Forest plantation
R	Rangeland	Includes grassland and shrub land, trees may be present but < 30% of crown coverage, may be suitable for pasture	Closed to very open shrubs, closed to very open herbaceous, sparse to very open trees, sparse shrubs
I	Infrastructure	Built up area or excavation site	Rural, quarry, airport, urban areas
WL	Wetland	Temporary or permanently flooded rangeland	Shrubs/open herbaceous on temporarily and permanently flooded land
WB	Waterbodies	Lake (natural or artificial) or river	Natural lakes, artificial lakes, lake shore

3.3.2 Accuracy assessment

The previously produced land use maps were not thoroughly validated and therefore another accuracy assessment procedure was conducted on the 2013 land use map. Vita et al. (2014) created a 5 by 5 kilometres grid consisting of 961 sampling locations covering the study area. This grid was validated by using high resolution imagery and supplemented with fieldwork sampling points (see Appendix **C** for the final validation grid). The accuracies of the other four land use maps could not be assessed because there was no ground validation data available, nor where there historical maps available for the years 1973, 1984, 1994 and 2003.

First, the 961 sampling points were imported into Google Earth, and land use was evaluated per sampling point. Depending on the data availability, imagery from 2009 until 2016 (Spot 6, 7) was used to identify the land use classes of 2013. Based on this assessment some land use classes indicated a high uncertainty (mainly due to limited image availability and/or low resolution imagery). Next, a fieldwork campaign was organised to check the land use classes with high uncertainty: rangeland and tree plantations. As the study area is large and the road network poor (dirt roads), two areas were appointed for data collection: in the west and south of the Mau forest complex (see Appendix A for fieldwork map). During the fieldwork conducted in January 2016, 180 points were sampled which in the end were overlaid with the validation grid (Appendix C).

Finally an accuracy assessment was performed by using this validation grid: the observed land use classes were compared to the classified land use classes. A confusion matrix was computed, and the overall accuracy (measure of agreement or accuracy that considers the diagonal in the matrix), the producer's- and user's accuracies of the land use classes were calculated (Foody, 2002). The producer's accuracy is the omission error: it gives the probability that a reference pixel is correctly classified. The user's accuracy represents the commission error: the probability that a pixel classified on the map actually is that land use on the ground (Story and Congalton, 1986). In addition, Kappa coefficient was computed, which is another measure of accuracy, and considers the off-diagonal elements in the confusion matrix (Congalton, 1991). The following formula was used to calculate Kappa:

$$K = \frac{(\text{overall classification accuracy} - \text{expected classification accuracy})}{(1 - \text{expected classification accuracy})} \quad (1)$$

3.4 CHARACTERISATION LAND USE CHANGE DYNAMICS

The five land use maps were used to make a characterisation of the magnitude, rate and nature of land use change in four time periods: 1973-1984, 1984-1994, 1994-2003 and 2003-2013. For each land use class in each time period the area was calculated, and by overlaying the land use segments (polygons), the change of those areas in kilometres (magnitude of change) and percentages (the rate of change) were calculated. In addition, the nature of the land use changes were computed via change matrices and land use change maps were created to show the spatial patterns. The characterisation of the land use dynamics resulted in the identification of the main conversions and the proximate drivers of land use change for every time period – which was of importance for selecting the dependent variables for regression modelling (see next paragraph).

3.5 DEPENDENT VARIABLES

The land use maps that were produced by Vita et al. (2014) were used to identify the land use conversions and proximate drivers in the study area, which resulted in the characterisation of the independent variables for logistic regression modelling (see figure 3.1 for the methodology and see paragraph 4.2 for results of the land use change dynamic analysis). The foremost land use change that was identified in the Mau forest complex in the period 1973-2013 was a conversion from forest to smallholder agriculture. Next to that, forest conversion and agricultural expansion were identified as main changes in the study area. Another important land use change that was identified was rangeland conversion (mainly rangeland to agriculture see paragraph 4.2.2). However, the accuracy assessment revealed a misclassification in the rangeland class. Based on these primary land use changes and the results of the accuracy assessment, two main land use changes were selected for the regression modelling: *forest conversion* and *smallholder agriculture expansion*.

The first model included the change from forest land use to non-forest land use, and the second model the conversion of any land use to smallholder agricultural land use. These variables were transformed into binary variables: coded with 0 if the area was constant in time, coded with 1 if land use changed (see figures 3.4a and 3.4b). Thus, forest that remained forest was coded 0, while land use that changed from forest to another land use was coded 1. In addition, smallholder agriculture that stayed smallholder agriculture was coded 0, whereas any land use type that changed to smallholder agriculture was coded 1. This resulted in a total of 10 dependent variables: for each time period and the overall period a land use change variable was created (see table 3.2).

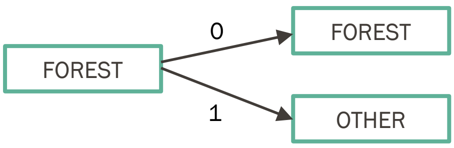


Figure **3.4a:** Forest conversion variable

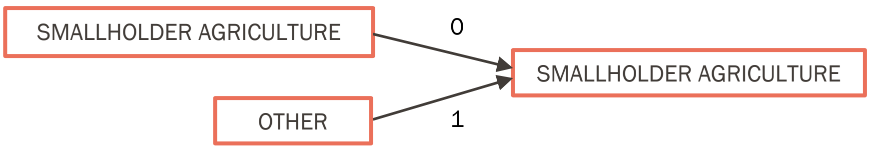


Figure **3.4b:** Smallholder agriculture expansion variable

3.6 INDEPENDENT VARIABLES

The factors that could have an effect on land use changes and on the proximate drivers – the independent variables in the regression analysis – were identified by use of a literature study (see 2.1.2 for literature framework). Verburg et al. (2002) emphasize the importance of selecting only variables for which a theoretical relationship between land use and drivers is known. Absence of causality is often a problem in land use driver studies and in this way the occurrence of spurious correlations can be prevented. Studies showed that biophysical, economic, demographic, technological, cultural and institutional drivers could be underlying drivers of land use changes (see the table D1 in Appendix D with all possible underlying drivers that were identified by use of literature).

For the statistical analysis the drivers were divided into *biophysical* and *socio-economic* variables. The latter covers both economic and demographic drivers. The technological, cultural and institutional drivers could not be quantified due to a lack of available or useful data. As a matter of fact, cultural and institutional drivers are difficult to quantify and easier to assess qualitatively by use of, for instance, stakeholder interviews or household surveys (Mutoko et al., 2014). Unfortunately this was not possible within the time frame and therefore literature was used to obtain extra information about these drivers (e.g. changes in policies). Table 3.4 summarizes the fourteen independent variables that were finally used in the analysis (see also table 3.2) and gives a short description of them. The remainder of this paragraph will explain how each variable was computed.

Table 3.4: The independent variables with description

VARIABLE		DESCRIPTION	RESOLUTION
Biophysical			
1	Rainfall	Average annual rainfall in Kenya in the period 1950-2000	925x925m
2	Temperature	Average annual temperature in Kenya in the period 1950-2000	925x925m
3	DEM	Digital Elevation Model	30x30m
4	Slope	DEM derivative	30x30m
5	Aspect north (cosine)	DEM derivative, slope direction south-north	30x30m
6	Aspect east (sine)	DEM derivative, slope direction west-east	30x30m
7	Curvature	DEM derivative, indication about flow and drainage	30x30m
8	TWI	Topographical Wetness Index extracted from DEM (contributing area and the slope)	30x30m
9	Soil pH	Soil pH mean 0 - 200 cm depth	250x250m
10	CEC	Cation Exchange Capacity mean 0 - 200 cm depth	250x250m
Socio-economic			
11	Population Density	Population density per county for the years 1979, 1989, 1999 and 2009, and population density change between 1979-2009	30x30m
12	Distance Roads	Euclidean distance to road network in Kenya, use of excellent, good and fair quality roads: gives an indication about the availability of the area	30x30m
13	Distance Towns	Euclidean distance to towns in Kenya with more than 10.000 inhabitants: gives an indication about the availability of markets	30x30m
14	Distance Rivers	Euclidean distance from grid cell to nearest river in Kenya: gives an indication about the availability of irrigation	30x30m

3.6.1 Biophysical variables

Different climate, surface and soil variables were selected for the regression analysis: rainfall, temperature, elevation, slope, aspect , curvature, topographical wetness index (TWI), soil pH and soil cation exchange capacity (CEC). The selection of these specific variables mainly depended on data availability.

Rainfall

The dataset of rainfall in Kenya was obtained from WorldClim. Two tiles that covered the study area, with a resolution of 30 arc seconds (925x925m at the equator) were downloaded. They showed the average monthly rainfall in the period 1950-2000 in mm. To use the data for the analysis, the average annual rainfall (the average of 12 months) in this time period was calculated. Finally the data was clipped to the extent of the study area. Rainfall values range between 51.5 and 148.6 mm in the study area (see the figure E1 in Appendix E), with the Western counties such as Kericho and Kisimu, showing higher average annual rainfall than the other counties in the study area. No other rainfall datasets were available for more recent time periods nor were there higher resolution datasets available.

Temperature

The temperature data in the study area was obtained via WorldClim as well. Two tiles with a resolution of 30 arc seconds (925x925m at the equator) were downloaded, and showed the average monthly temperature in the period 1950-2000 in Celcius *10. For this variable, the average annual temperature was calculated and the data was clipped to the correct extent. The values in the study area vary between 10.5 and 24.7 Celsius (see figure E2 for a temperature map in Appendix E). The Northern, lower elevated, counties show higher temperatures, whereas the highly elevated parts of the study area show lower temperatures. For temperature there were no datasets available with a higher spatial or temporal resolution.

Digital Elevation Model (DEM)

The DEM was downloaded from ASTER GDEM website and is a 30x30m resolution. The dataset shows the elevation in the study area. Some pre-processing was performed to improve the quality of the data, which is of importance when using the DEM for geomorphological analysis (Reuter et al., 2009). First, the outliers were removed from the elevation data: average elevation in a 3x3 window was calculated and the difference between these values and the elevation at the core cell was derived. A spatial threshold was selected to identify the pixels that were spikes or pits: if the difference was larger than two times the standard deviation in a 5x5 window, a pixel was selected as outlier (see Reuter et al., 2009: 100).

In this research the elevation data was used for hydrological purposes – the movement of water is of importance for land use change - and therefore the next step was removing supposedly erroneous sinks in the elevation data (Reuter et al., 2009). First, the maximum depth of the sinks was calculated: 112 meters. Following, this value was used as the z-limit for filling in sinks of the DEM. The final DEM used in the analysis was a filled elevation model without outliers. The map in Appendix E shows the elevation data (figure E3). The lowest areas are around 959 meters, whereas the highest are 3107 meters. The middle part of the study areas shows the highest elevation.

Slope

The slope in degrees was derived from the elevation dataset and it is the first derivative of the DEM. The slope gradient is the maximum rate of change in elevation and gives an indication about the rate of movement downslope (the slope rate) (Buckley, 2010; Olaya, 2009). The slope map in Appendix E shows the slope values in the study area (figure E4). The mean slope value in the study area is around 9 degrees, whereas values range between 0 and 69 degrees.

Aspect north and east

The aspect was also extracted from the DEM. The variable indicates the compass direction of the slope, and thus the flow-line direction: water follows the aspect direction (Olaya, 2009). In addition, aspect can give an indication about the sunny side of a cell, which may be of importance for land use change. The output values range from -1, which is flat, to 360 degrees, which indicates a northern slope direction. To better interpret the aspect variable in the regression analysis, the aspect was converted to cosine and sine values (Jenness, 2007; Olaya, 2009; Rutherford, 2008). In this way a distinction between north-south and east-west was made. For the calculation, the aspect was first converted to radians, and then the cosine and sine were calculated. The cosine from the aspect ranges from -1, which is south, to 1, which is north. The sine, on the other hand, ranges from -1, which is west, to 1, which points to an eastern slope direction. This resulted in two independent variables: aspect north (the cosine of the aspect) and aspect east (the sine of the aspect) (see Appendix E figure E5 and E6 for maps of the aspect).

Curvature

The curvature gives an indication about flow and drainage across a surface and was calculated by use of the DEM. Curvature is called the second derivative of the elevation: it is the slope of the slope and gives information about the convexity and concavity of the surface (Buckley, 2010; Olaya, 2009). Values generally range between -1 and 1, and the units are 1/100 of the z-unit (meters in this study). A positive curvature value indicates that the surface is upwardly convex (the surface curves outward), whereas negative points to upwardly concave (the surface curves inward). A value of zero represents flat areas. The figure 3.5 shows how curvature can influence flow across a surface: it can have effects on convergence, divergence, deceleration and acceleration of flow, and thus on erosion and deposition (Buckley, 2010; Olaya, 2009). The figure E7 in Appendix E shows the curvature values around Lake Naivasha in the study area: ridges (positive values – darker colour) and valleys (negative values – lighter colour) are clearly visible.

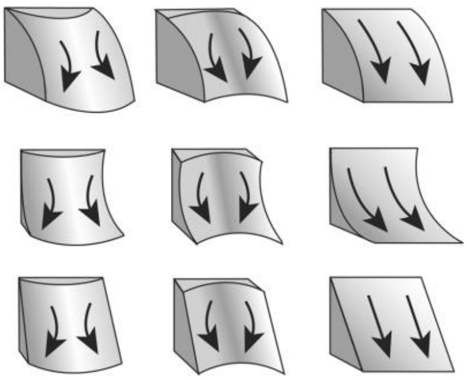


Figure 3.5: Surface curvature (Buckley, 2010)

Topographic Wetness Index (TWI)

The TWI describes the wetness conditions in the study area: the tendency of a cell to accumulate water (Gruber and Peckham, 2009). To calculate the TWI, the following formula was used:

$$\mathrm{TWI} = \ln(\alpha/\tan\beta) \qquad (2)$$

where α is the local upslope *contributing area* and $\tan\beta$ the local *slope* angle (Sørensen et al., 2005). The contributing area is the basin area, the upslope area or the flow accumulation: the area where water aggregates (Gruber and Peckham, 2009).

To compute the TWI for the study area, first the elevation data was used to derive a flow direction raster, of which in turn a flow accumulation raster could be obtained. The flow accumulation raster in m^2 gives the local upslope contributing area. Next, the slope in degrees was converted into radians (degree multiplied by π divided by 180°). Then the flow accumulation and the slope raster were used to conduct the TWI in the study area. The final TWI dataset is unit-less, and gives a relative measure: high values represent wet areas, or drainage depressions, whereas low values indicate dry areas such as ridges or crests (see figure E8 in Appendix E for the TWI values around Lake Naivasha in the study area).

Soil pH

The soil pH gives an indication about the acidity or alkalinity in the soil: a value below 7.0 is acid, whereas above 7.0 the soil is alkaline. The data of the pH values in the soil were obtained via ISRIC world soil information and the highest spatial resolution available was 250x250m. To use the dataset in the regression analysis, the average pH value was calculated for all soil depths (0-200 cm). The map shows that soil appears to be more acid than alkaline in the study area, and values range between 40 and 89 (see figure E9 in Appendix E).

Cation Exchange Capacity (CEC)

Soil quality is of importance for land use change. Soils hold on to nutrient reserves which are supplied to plants. The cation exchange capacity, or CEC, gives an indication about the fertility in the soil. It is the relative capacity of the soil to store one particular group of plants nutrients, the cations (Mengel, 1993).

Examples of nutrients are calcium, magnesium, aluminium, iron, zinc and copper. The more clay and organic matter in the soil, the higher the CEC, the higher the water holding capacity (CUCE, 2007).

The CEC was obtained via ISRIC world soil information and the average value for all soil depths was calculated (0-200 cm). The variable is expressed in centimoles of charge per kilogram of exchanger (cmol(+)/kg). In the study area the values range between 4 and 52, with the counties on the Eastern side of the study area showing higher values than the counties on the Western side (see figure E10 in Appendix E).

3.6.2 Socio-economic variables

Socio-economic drivers of land use change were difficult to quantify due to the limited data available. In the end four predictor variables were conducted: population density, distance to roads, distance to towns and distance to rivers.

Population density

To get an indication about population growth and thereby population pressure, population density in the area was computed. Verburg et al. (2002) emphasize the importance of calculating the population density over larger areas: population can have effects on land use, not only locally, but in particular over certain distances. Statistical reports from the Kenya Central Bureau of Statistics were used to calculate the population density in the Mau forest complex (KNBS, 1979; 1989a; 1989b; 1999; 2009b). In Kenya population is measured once in a decade (in the years 1979, 1989, 1999 and 2009 respectively) and therefore data of four moments in time were used matching each time period of the land use maps (see table 3.5).

Table 3.5: KCBS population census reports

	LUC TIME PERIOD	STATISTICAL DATA REPORT
1	1973 - 1984	1979
2	1984 - 1994	1989
3	1994 – 2003	1999
4	2003 - 2013	2009

The statistical data was only available on a county level. Currently, the Mau forest complex overlaps 12 counties in Kenya (see figure F1 in Appendix F). However, over the past 40 years the administrative boundaries of the districts and/or counties in Kenya changed due to different policies and regulations. Older maps and reports helped identifying the original districts, the change of the districts, and the present counties (Hassan, 2013; Kariuku, 1989; Muandu, 1999; Statoids, 2013; Trapman, 1974). Table F1 in Appendix F shows an overview of the development of the districts and counties. The population density numbers per district and county that were extracted from the four population census can be found in Appendix F as well.

Finally, for the regression analysis, the population density per county per time period was used. For the models of the overall period 1973-2013 a population density growth variable was calculated: the percentage change between 1979 and 2009. The maps of population density in 1979, 1989, 1999 and 2009 in Appendix E reveal that population density in the study area increased over the years (figure E11, E12, E13, E14, and E15). In particular in the counties Kisumu, Kericho and Bomet population density increased. Figure E15 shows that the highest growth in population density occurred in the southern part of the study area (between 250 and 300%).

Distance to roads

The distance to a road gives an indication about the accessibility of an area. The data of the road network of Kenya was provided by the Centre for International Forestry Research. The dataset showed all the roads in the country classified from excellent to very poor roads. The roads that had a poor to very poor surface condition (local bad accessible sandy roads) were excluded from this dataset. Next, the road network was used to calculate the Euclidean distance from a raster cell to the nearest road. It should be noted that when calculating the distance variables, the area outside the study area was considered as well (data of the whole country was used to calculate the Euclidean distances inside the study area). The resolution used was a cell size of 30x30m (based on the DEM, which is the smallest spatial unit in the analysis) (see Appendix E figure E16 for a map of the distance to roads). The map shows that the southern part of the study area is not easy accessible.

Distance to towns

The availability of markets in an area can influence how people use land and therefore the distance to towns (the centre for markets) was calculated. A spatial dataset of towns from the International Livestock Research Institute (ILRI, 2000) and a list of cities from the Population Census Report 2009 (KNBS, 2009a) were combined to create a dataset of towns with more than 10000 inhabitants. This threshold was used because there was almost no data available on population numbers of smaller towns. The created dataset of towns was used to calculate the Euclidean distance in a 30x30m resolution, and clipped according the extent of the study area. Figure E17 in Appendix E shows the towns in the area and the distances from each grid cell to the nearest town.

Distance to rivers

The distance to rivers can give an indication about the availability of irrigation, which can have an effect on land use. The river dataset was obtained via the World Resources Institute (WRI, 2016) and shows the main permanent and non-permanent rivers in Kenya. Next, the distance from each grid cell to the nearest river was calculated which resulted in an Euclidian distance map in a 30x30m resolution. Figure E18 in Appendix E reveals this distance to the rivers.

3.7 LOGISTIC REGRESSION MODELLING

3.7.1 Pre-processing dataset

The logistic regression was performed on the land use segment level due to the nature of the dependent variables: land use segments (polygons). In similar studies the regression analysis was conducted on the pixel level because of the pixel based classification of the land use maps (see for instance Serneels and Lambin. 2001, Verburg et al, 2004, Yin et al., 2014). Because this study used object-based classification land use maps, the regression analysis was performed on this level: the land use objects.

First, the land use maps were overlaid with the maps of the independent variables, which were of raster format. Next, the mean values of the independent variables per land use segment were calculated. Although a large part of the data was not normally distributed, and the median might have been a suitable measurement as well, descriptive statistics showed that the mean value represented the central tendency of the independent variables the best (see Appendix G for descriptive statistics and histograms). By use of zonal statistics the mean values of the independent variables were assigned to each land use segment. This resulted in a geodatabase with the mean values of every independent variable attached to the specific land use polygons. Finally eight datasets were produced: the two types of land use conversions (forest conversion and smallholder agricultural expansion) coupled to the independent variables for four time periods (1973-1984, 1984-1994, 1994-2003, 2003-2013).

3.7.2 Exploratory data analysis

Multi-collinearity, or the collinearity, between independent variables can influence the results of the regression analysis. When independent variables are highly correlated it is hard to separate out the predictive values of the variables (Ott and Longnecker, 2010). In addition, collinearity implies that regression coefficients will change as variables are added to or deleted from a regression model - the accuracy of the slope estimates decreases. For these reasons, multi-collinearity was tested with Spearman's rank correlation coefficient and the Variance Inflation Factor (VIF). The first was chosen because descriptive statistics showed that the data was partly non-parametric (see figures Appendix G for data distribution of the variables). Correlation above 0.8 needed extra attention. Secondly, the VIF measures how much the variance of a coefficient increases due to collinearity. A value of 1 points to no collinearity, whereas 10 or higher indicates that variables are highly correlated (Ott and Longnecker, 2010: 689). Therefore, independent variables that revealed VIF values higher than 10 were excluded from the models (see also Rossiter and Loza, 2012).

3.7.3 Statistical framework

In this study a multiple logistic regression model was used to analyse the drivers of land use change in the Mau forest complex. Because the dependent or response variables were dichotomous - either land use changed or did not change, and the independent or explanatory variables were continuous, a multiple logistic regression model was used. In this way, the probability of the presence of a particular land use change process (the dependent variables), given a set of underlying drivers (the independent variables), could be computed.

The dependent or response variable y was thus a binary variable. In case of forest conversion, the independent variable y = 1 if forest land use changed, and y = 0 if land use remained forest. If the response variable is binary, the distribution of y reduces to a single value, the probability p = Pr(y = 1) (Ott and Longnecker, 2010). To relate p (which ranges from zero to one) to a linear combination of the independent variables (which vary between -infinity to +infinity), the probabilities need to be transformed into an odds ratio (the ratio of the probability that an event happens to the probability that it does not happen) (Ott and Longnecker, 2010;). Values of the odds ratio vary between zero and infinity. However, mathematically this is still problematic, and a solution is to calculate the logarithm of the odds ratio (the log odds). In this way values will vary between -infinity to +infinity, when probabilities vary between zero and one (Rossiter and Loza, 2012).

Simple logistic regression model

Thus, in the logistic regression model the natural logarithm of the odds ratio is related to the independent variables by a linear model. When there is one independent variable, $p(x)$ is the probability that y equals 1 when the independent variable equals *x*. The simple logistic regression model is given by:

$$\ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x \qquad (3)$$

where the inverse transform *p(x)* is:

$$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1}} \qquad (4)$$

For instance, $p(x)$ would be the probability of forest conversion to occur in an area exposed to *x* units of annual rainfall. The values β_0 and β_1 are estimated from the observed data by use of maximum likelihood estimation. Parameter β_0 is the intercept or constant, which gives the estimation of the probability of the event associated with y = 1 when the explanatory variable x = 0. The slope parameter β_1 measures the degree of association between the probability of the event occurring and the value of the independent variable *x*.

Multiple logistic regression model

Including multiple independent variables, then x_1, x_2,... x_k are the k predictor variables of the binary response variable y, the multiple logistic regression model is:

$$\ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \qquad (5)$$

The probability values can be expressed in terms of the independent variables by:

$$p(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\dots+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\dots+\beta_k x_k}} \tag{6}$$

where *p(x)* is the probability that the dependent variable *y* = 1 (e.g. forest conversion), parameter β_0 is the intercept or constant (the log-likelihood of change when predictor variables are zero), $\beta_1\beta_2 \dots \beta_k$ are the estimated coefficients of the predictor variables with values $x_1 x_2 \dots x_k$ (thus β_1 is the slope of the log-likelihood of change for predictor variable with value x_1). The ratio $\frac{p(x)}{1-p(x)}$ is the odds and $\ln\left(\frac{p(x)}{1-p(x)}\right)$ is the log-odds of *p(x)*.

Thus, the probability of forest conversion can be given by predictor variables such as rainfall, temperature, distance to roads and distance to towns with values x_1, x_2, x_3 and x_4 for a specific area.

Odds ratio

An important interpretation of the logistic regression is the odds ratio, which gives an indication about how much more likely (or unlikely) it is for the outcome to be present for a set of values of independent variables (Serneels and Lambin, 2001). If the predictor variable x_i increases with one unit, the odds of the event are multiplied by e^{β_i}, when the other predictor variables are held constant (Ott and Longnecker, 2010). If the odds ratio is lower than 1, an increase in the predictor or independent variables by one unit means a decrease in the odds of the event. On the other hand, if the odds ratio is higher than 1, an increase in the predictor by one unit indicates an increase in the odds of the event.

Model fitting

A few methods can be applied to fit the models: stepwise models, 'best subset' models or predefined conceptual models. The latter is used by as Serneels and Lambin (2001) and was applied in this research as well. The conceptual framework (see chapter 2 and figure 2.1) indicated multiple independent variables (see paragraph 3.6) and these were all tested in the models. The other options for model fitting, such as using maximum likelihood and set-wise forward and backward elimination were also tested, however this did not improve the explanatory power of the models, and therefore full models were analysed to identify which variables were of significant influence to land use change.

Evaluating models: the pseudo R-Square

In regression modelling the R-Square is used as measure of predictive power: it gives an indication about the model performance or how well you can predict the dependent variables based on the independent variables (Allison, 2013). However, in logistic regression modelling is it not possible to compute an exact R-Square, and therefore the pseudo R-Square was used (it is called 'pseudo' because the measure looks similar as values range from 0 to 1). According to Serneels and Lambin (2001) the pseudo R-Square should be interpreted differently than in standard regression analysis: the values are generally lower. In this study the pseudo R-Square was tested with Cox & Snell R^2 and Nagelkerke R^2. The ratio in these measures gives an indication about to what degree the parameters of the model improve from the null model (predicting the dependent variable without independent variables) to the fitted model (predicting dependent variable with independent variables) (IDRE, 2011). A small ratio indicates a high improvement and thus a high R-Squared (IDRE, 2011). The first measure, Cox & Snell R^2, is given by:

$$R^2{}_{C\&S} = 1 - \left(\frac{L_0}{L_m}\right)^{2/N} \tag{7}$$

where L_0 is the value of the likelihood of the intercept model (the null model without independent variables), L_m is the likelihood of the full model (the fitted model that is estimated) (Allison, 2013), and *N* the sample size. Cox and Snell R^2 has an upper bound that is not 1: if the model predicts perfectly, its value is lower than 1. To be precisely, the maximum possible value is $1-(L_0)^{2/N}$. A solution for this is

Nagelkerke R^2, which modifies Cox & Snell R^2 by dividing it by its maximum value (IDRE, 2011), so that values can range from 0 to 1, and is given by:

$$R^2{}_N = \frac{1-\left(\frac{L_0}{L_m}\right)^{2/N}}{1-(L_0)^{2/N}} \tag{8}$$

For every model both measures were computed.

Evaluating models: ROC curve and AUC

In addition to the pseudo R-Square, the goodness of fit was assessed with the receiving operating characteristic (ROC) curve (Rossiter and Loza, 2012; Serra et al. 2008; Verburg et al., 2002; Were et al., 2014). The ROC is a plot of the sensitivity against the specificity. The sensitivity is the proportion of true positives, in this case the polygons that actually changed land use. The specificity is the proportion of false positives, which are the true negatives, in this case the correctly predicted no change polygons. The curve is plotted at a series of thresholds: 0.5 was selected as threshold, which means change is equally likely or not (Rossiter and Loza, 2012).

The ROC graph shows on the y-axis the true positive rate – the sensitivity – and on the x-axis the false positive rate – the specificity. The model is accurate if the curve is close to the left top border: it predicts most true positives with a few false positives. If the curve comes close to the diagonal, the model is less accurate. The diagonal is the random case: if the models would predict at random, the chances would be equally like to be true or false positive (Rossiter and Loza, 2012).

More specifically, the area under the curve (AUC) was used to evaluate the success of the logistic regression models. Basically it measures the success of the model in correctly classifying land use polygons that did and did not actually change. A value of 0.5 indicates a random model, whereas 1.0 indicates a model without error (Serra et al., 2008; Verburg et al., 2002). According to Rossiter and Loza (2012), 0.6 can be considered sufficient, 0.7 good, 0.8 very good and 0.9 an excellent discriminatory power.

Spatial autocorrelation: Moran's I

One assumption in multiple logistic regression analysis is that the data has to be statistically independent, and in particular for spatial land use analysis this is of importance (Millington et al., 2005; Overmars et al., 2003). However, spatial data tend to be dependent, as Waldo Tobler stated in what he called the first law of geography: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970: 236). To overcome this problem and minimize spatial autocorrelation of the regression results, a random sample can be taken (Verburg et al., 2002). In this study all data was used, and for that reason the spatial autocorrelation of the regression results (the residuals) was tested with Global Moran's I. The measure was calculated in ArcGIS. The values range between -1 and +1, with negative values indicating a dispersed pattern and positive values a clustered pattern of the data. Values around 0 indicate that the data is random: no clustering (ESRI, 2016). Figure 3.6 shows what the data looks like when dispersed, random or clustered.

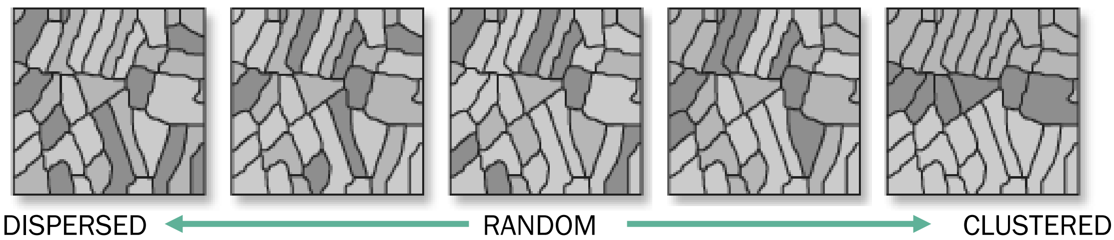


Figure 3.6: Data clustering (adjusted from ESRI, 2016)

4. RESULTS

In this chapter first the outcome of the accuracy assessment and the analysis of the land use change dynamics are discussed. Then the results of the exploratory data analysis and the land use change models are described: the forest conversion models and the agricultural expansion models per time period and the overall period.

4.1 ACCURACY ASSESSMENT

The accuracy assessment showed an overall accuracy of the classified 2013 land use map of 74.2% with a Kappa coefficient of 0.62 (see table 4.1). These results indicate that there is a moderate agreement between the classified land use map and the reference data.

Forest, smallholder agriculture and tea estates prove to be classified quite accurate: producer and user accuracies are above 70% (see table 4.1). Irrigated commercial agriculture on the other hand, shows very low accuracies: only 20% was correctly classified and 60% of the irrigated commercial agriculture segments was actually irrigated commercial agriculture. The table shows that there is a lot of confusion with smallholder agriculture. This low accuracy is not that problematic because irrigated commercial agriculture covers only 2.9% of the study area (see figure 4.1). In addition, the tree plantations class had low accuracies as well: only half of the polygons was correctly classified. Furthermore, rangeland shows a high producer accuracy, but a very low user accuracy: even though 85.8% of rangeland was correctly classified, only 55.2% of the rangeland polygons were actually rangeland - which suggests a substantial misclassification in the rangeland land use class. The confusion matrix shows that rangeland is mainly mixed with smallholder agriculture.

The results of infrastructure, wetland and waterbodies were not considered because not enough sampling locations were assessed, and therefore results are not statistically valid (see Congalton and Green, 2008).

Table **4.1:** Confusion matrix land use map 2013

CONFUSION MATRIX MAU FOREST 2013												
	GROUND TRUTH											
CLASSIFIED	F	SA	IC	TE	TP	R	I	WL	WB	**TOTAL**	**PRODUCER'S**	**USER'S**
FOREST	**133**	5	0	0	0	27	0	1	0	**166**	80.1	80.1
SMALLHOLDER AGRICULTURE	16	**346**	4	0	8	116	3	0	1	**494**	70.0	89.2
IRRIGATED COMMERCIAL AGRICULTURE	0	18	**6**	0	0	5	1	0	0	**30**	20.0	60.0
TEA ESTATES	1	0	0	**13**	0	0	0	0	0	**14**	92.9	92.9
TREE PLANTATION	1	3	0	1	**10**	3	0	0	0	**18**	55.6	52.6
RANGELAND	14	16	0	0	1	**187**	0	0	0	**218**	85.8	55.2
INFRASTRUCTURE	0	0	0	0	0	0	**4**	0	0	**4**	100.0	50.0
WETLAND	1	0	0	0	0	1	0	**1**	0	**3**	33.3	50.0
WATERBODIES	0	0	0	0	0	0	0	0	**10**	**10**	100.0	90.9
TOTAL	**166**	**388**	**10**	**14**	**19**	**339**	**8**	**2**	**11**	**957**		
OVERALL ACCURACY	**74.2%**											
KAPPA	**0.62**											

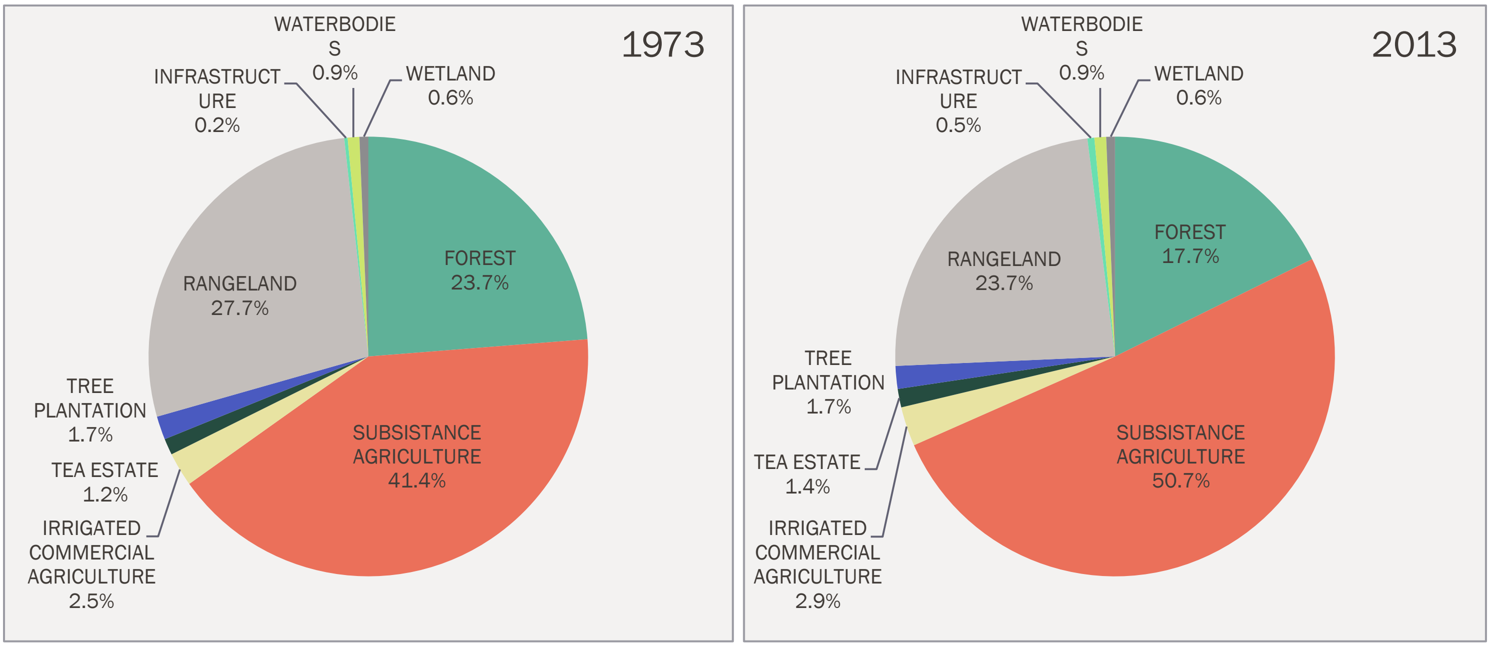


Figure 4.1: The land use types in the study area (in % of the total area)

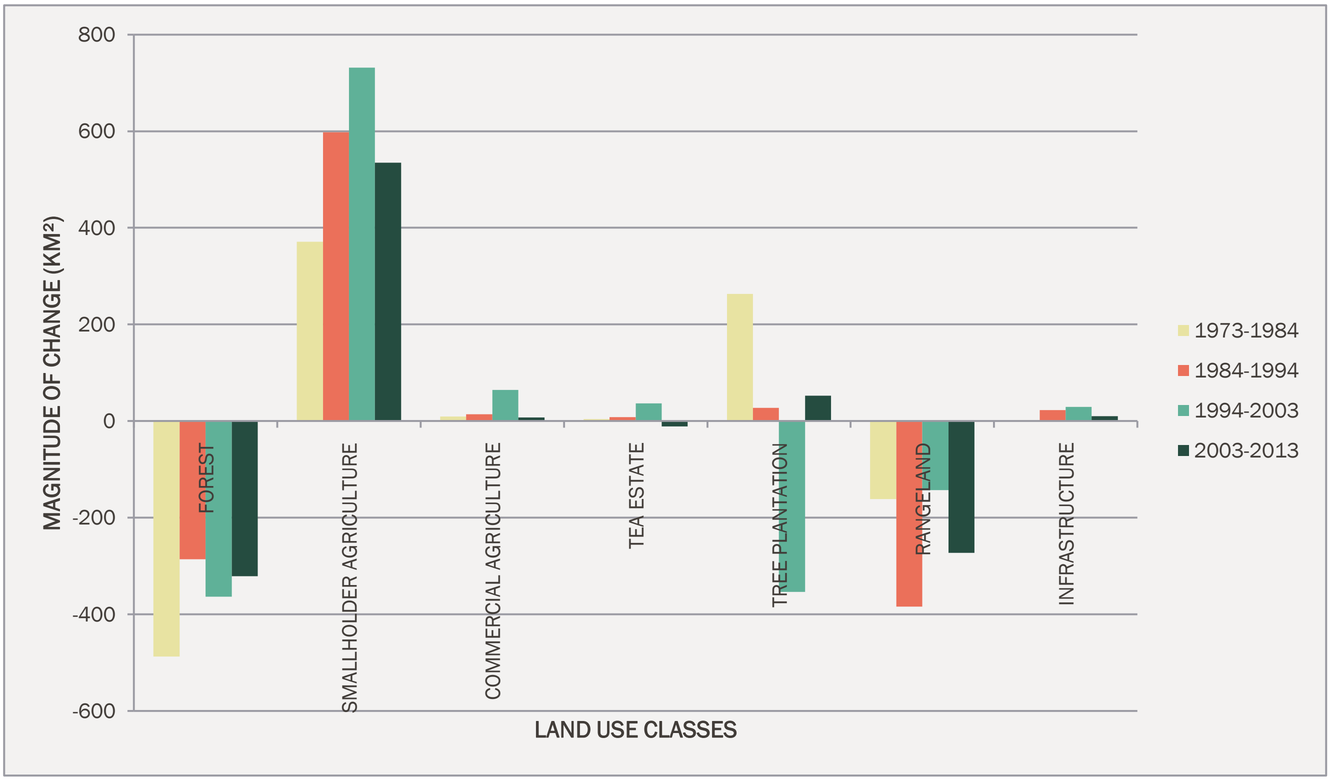


Figure 4.2: The change in km^2 per land use class

4.2 LAND USE DYNAMICS

For each land use class the results of the land use dynamics analysis are described: the magnitude of change, the rate of change and the nature of the changes.

4.2.1 Magnitude and rate of change

In 2013, the dominant land use types in the study area are smallholder agriculture (50.7%), followed by rangeland (23.7%) and forest (17.7%). The first increased in area in the period 1973-2013, whereas the latter two decreased (figure 4.1 and table 4.3). In fact, figure 4.2 shows that forest, smallholder agriculture and rangeland have high magnitudes of change.

Thus, a forest decline in the period 1973-2013 can be detected. To be precise, 1458.5 km^2 of forest was lost, which is a decline of 25.5% in a period of 40 years (see table 4.4 and 4.5). In the first time period (1973-1984) the rate of change was the highest (table 4.6). However, the tables show an on-going decrease of forest area: in the period 1973-2013 an annual decrease of 0.6% can be noticed. Smallholder agriculture on the other hand, shows an annual growth rate of 0.6%. In a 40-year period, smallholder agriculture increased by 2235.3 km^2, which is an increase of 22.4%. The highest increase can be detected in the period 1994-2003: 731.3 km^2 (table 4.4). In addition, irrigated commercial agriculture increased with 15.9% (95.7 km^2). In particular in the period 1994-2003 the increase was high. Next, an overall increase of tea estates of 13% can be noticed, mainly in the period 1994-2003 with an annual growth rate of 1.4%. The tree plantations class shows an interesting change of its area: in the first period a high increase (63.3%), the second period a smaller increase (4.1%), while the third period shows a very large decrease (50.1%), and the last period an increase again (14.9%) (see also table 4.5). Over the 40 years tree plantation decreased with 10.9 km^2 (2.6%) – which is little. Moreover, rangeland shows a decline of 14.4% and decreased mostly in the period from 1984-1994 (with 380.4 km^2). Infrastructure is the land use class with the highest growth: an increase of 107.2%. It expanded from 59.4 km^2 to 123 km^2 (which is, as figure 4.1 showed, a very small part of the study area, namely 0.5%). Infrastructure expansion occurred mostly between 1984 and 2003: an annual growth rate between 3.7% and 3.9% can be detected (see table 4.6). Last, waterbodies and wetland prove to be steady areas over time, the small changes for these classes can probably be related to classification errors. The decrease of wetlands by 1%, may be explained by drying off as well, however this needs further investigation.

It can be concluded that forest and smallholder agriculture showed the highest magnitudes of change: the first a decrease of 25.5%, the second an increase of 22.4%. In addition, rangeland lost much of its area as well: a decrease of 14.4%. The change in the tree plantation class in the period 1973-1984 and the period 1994-2003 were quite abrupt. Infrastructure showed an annual growth rate of 2.7% - by far the highest of the land use classes, and irrigated commercial agriculture and tea estates show a notable increase in period 3. The small variations of waterbodies and wetland can be attributed to classification errors.

Table 4.3: The area of the land use classes in km^2

AREA IN KM^2	1973	1984	1994	2003	2013
FOREST	5724.0	5236.8	4950.5	4586.9	4265.4
SMALLHOLDER AGRICULTURE	9981.2	10352.2	10950.2	11681.5	12216.4
IRRIGATED COMMERCIAL AGRICULTURE	602.0	611.4	625.5	690.2	697.7
TEA ESTATE	289.5	293.4	301.5	338.2	326.9
TREE PLANTATION	415.6	678.7	706.2	352.4	404.7
RANGELAND	6666.1	6504.5	6120.4	5977.5	5705.0
INFRASTRUCTURE	59.4	60.8	83.3	112.6	123.0
WATERBODIES	207.6	208.0	208.0	207.6	207.6
WETLAND	156.4	156.0	156.0	154.9	154.9

Table 4.4: The change of the area of the land use classes in km²

AREA CHANGE IN KM²					
	1973-1984	**1984-1994**	**1994-2003**	**2003-2013**	**TOTAL PERIOD**
FOREST	-487.2	-286.3	-363.6	-321.4	-1458.5
SMALLHOLDER AGRICULTURE	370.9	598.1	731.3	534.9	2235.3
IRRIGATED COMMERCIAL AGRICULTURE	9.4	14.1	64.6	7.5	95.7
TEA ESTATE	3.9	8.1	**36.7**	-11.4	37.4
TREE PLANTATION	263.1	27.6	-353.9	52.4	-10.9
RANGELAND	-161.6	-384.1	-142.9	-272.5	-961.1
INFRASTRUCTURE	1.42	22.6	29.2	10.4	63.7
WATERBODIES	0.43	0	-0.4	0	0
WETLAND	-0.42	0	-1.1	0	-1.5

Table 4.5: The change of the area of the land use classes in %

AREA CHANGE IN %					
	1973-1984	**1984-1994**	**1994-2003**	**2003-2013**	**TOTAL PERIOD**
FOREST	-8.5	-5.5	-7.3	-7.0	**-25.5**
SMALLHOLDER AGRICULTURE	3.7	5.8	6.7	4.6	**22.4**
IRRIGATED COMMERCIAL AGRICULTURE	1.6	2.3	**10.3**	1.1	15.9
TEA ESTATE	1.4	2.8	**12.2**	-3.4	12.9
TREE PLANTATION	**63.3**	4.1	**-50.1**	14.9	-2.6
RANGELAND	-2.4	-5.9	-2.3	-4.6	**-14.4**
INFRASTRUCTURE	2.4	**37.1**	**35.1**	9.3	**107.2**
WATERBODIES	0.2	0.0	-0.2	0.0	0.0
WETLAND	-0.3	0.0	-0.7	0.0	-1.0

Table 4.6: The annual rate of change of the land use classes in %

ANNUAL RATE OF CHANGE IN %					
	1973-1984	**1984-1994**	**1994-2003**	**2003-2013**	**TOTAL PERIOD**
FOREST	**-0.9**	**-0.5**	**-0.8**	**-0.7**	**-0.6**
SMALLHOLDER AGRICULTURE	0.3	0.6	0.7	0.5	0.6
IRRIGATED COMMERCIAL AGRICULTURE	0.1	0.2	1.1	0.1	0.4
TEA ESTATE	0.1	0.3	1.4	-0.3	0.3
TREE PLANTATION	**5.8**	0.4	**-5.6**	1.5	-0.1
RANGELAND	-0.2	-0.6	-0.3	-0.5	-0.4
INFRASTRUCTURE	0.2	**3.7**	**3.9**	0.9	2.7
WATERBODIES	0.0	0.0	0.0	0.0	0.0
WETLAND	0.0	0.0	-0.1	0.0	0.0

4.2.2 Nature of changes

Figure 4.3 shows the nature of the main land use changes in the period 1973-2013 based on the change matrices (see table 4.7 and in Appendix H table H1, H2, H3 and H4 for change matrices). The foremost land use change in a 40-year time period was forest to smallholder agriculture: a conversion of 1343.7 km^2 (see table 4.7). Furthermore, forest lost to rangeland, tree plantations and tea estates. The second large conversion in the study area was from rangeland to smallholder agricultural land use: 988.9 km^2. Moreover, rangeland lost to irrigated commercial agriculture, tree plantations and forest. The last important change was smallholder agriculture gaining area from tree plantations (mainly in period 3): 149 km^2. Thus, smallholder agriculture gained at the expense of forest (1343.7 km^2), rangeland (988.9 km^2) and tree plantations (149.0 km^2). Based on these key changes, three processes could be identified: forest conversion (figure 4.3a), rangeland conversion (figure 4.3b) and smallholder agricultural expansion (figure 4.3c).

Then, two other changes are noteworthy. First, irrigated commercial agriculture gained area from smallholder agriculture, mainly in the third period (see table H3 in Appendix H). Second, the dynamics and rate of change analysis showed that infrastructure had the highest annual growth rate, which is, according to table 4.7, primarily at the expense of smallholder agriculture.

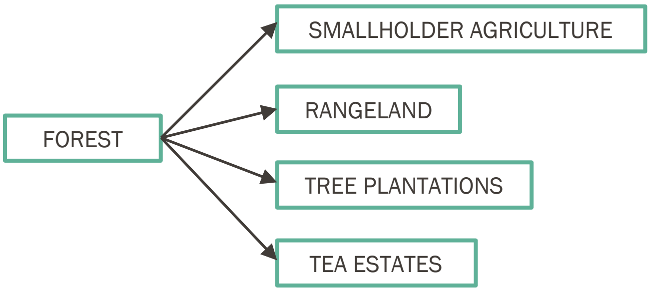


Figure 4.3a: Forest conversion

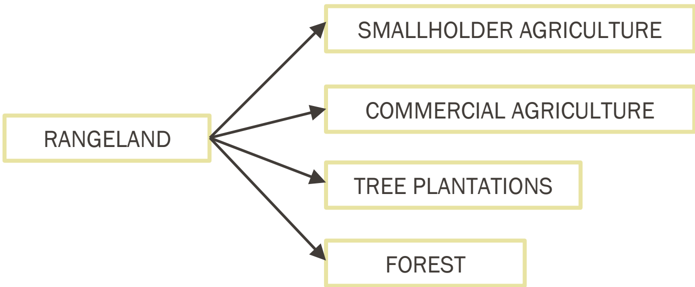


Figure 4.3b: Rangeland conversion

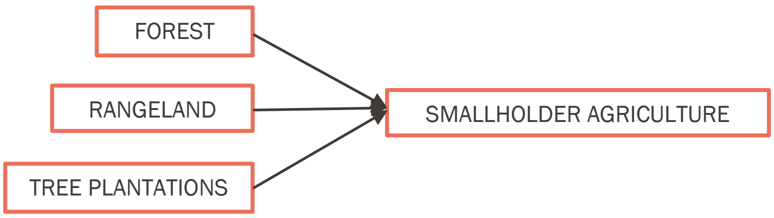


Figure 4.3c: Smallholder agriculture expansion

Table 4.7: Change matrix indicating the nature of land use change period 1973-2013

1973 - 2013 (IN KM^2)												
		1973										
		F	**SA**	**CA**	**TE**	**TP**	**R**	**I**	**WB**	**WL**	**TOTAL**	**GAIN**
2013	**Forest**	4165.4	72.0	0	0	3.7	24.4	0	0.0	0	4265.4	100.1
	Smallholder Agriculture	**1343.7**	9734.1	0	0.8	**149.0**	**988.9**	0	0.0	0	12216.4	**2482.3**
	Irrigated Commercial Agriculture	2.3	49.8	602.0	0	0	42.1	0	1.5	0	697.7	95.7
	Tea Estates	**47.6**	6.3	0	268.5	4.5	0	0	0.0	0	326.9	58.3
	Tree Plantations	**75.3**	44.4	0	20.1	240.1	24.9	0	0.0	0	404.7	164.7
	Rangeland	**89.1**	20.3	0	0	15.1	5580.5	0	0.0	0	5705.0	124.5
	Infrastructure	0.7	**54.3**	0	0	3.2	5.4	59.3	0.0	0	123.0	63.7
	Waterbodies	0	0	0	0	0	0	0	154.9	0	154.9	0.0
	Wetland	0	0	0	0	0	0	0	0.0	207.6	207.6	0.0
TOTAL		5724.0	9981.2	602.0	289.5	415.6	6666.1	59.3	156.4	207.6	24101.7	
LOSS		**1558.6**	247.0	0	21.0	175.5	**1085.6**	0	1.5	0		

4.2.3 Spatial pattern changes

The land use maps of the study area in 1973 and 2013 are shown in figure 4.4 (see figures in Appendix I for land use maps of the years 1984, 1994 and 2003). When comparing the land use change maps, growth in infrastructure, urban in particular, can be noted. The top arrow shows an extensive increase of urban area around Lake Nakuru. A second spatially notable change is a decrease of forest and tree plantations in favour of smallholder agriculture in the central part of the study area: in Nakuru county (the second arrow in figure 4.4). In addition, the Southern part of the study area, Narok county, shows clearly visible land use changes: an expansion of smallholder agriculture at the expense of mainly forest and rangeland.

Furthermore, the land use change maps for every time period (see Appendix J) revealed that land use changes are quite spread out over the study area, with the most Northern and Southern rangeland parts most steady over time. The maps show that the main changes occurred in and around the forest blocks – the central part of the study area, overlapping multiple counties – which can be observed in figure 4.4 as well. However, when focusing on the most important land use changes (forest conversion, rangeland conversion, smallholder agriculture expansion), more variation can be detected. In period 1 the western parts of Nakuru and Baringo show much deforestation. The northern parts of Narok and western parts of Nakuru experience most forest conversions and smallholder agricultural expansion during every time period (see the maps in Appendix K).

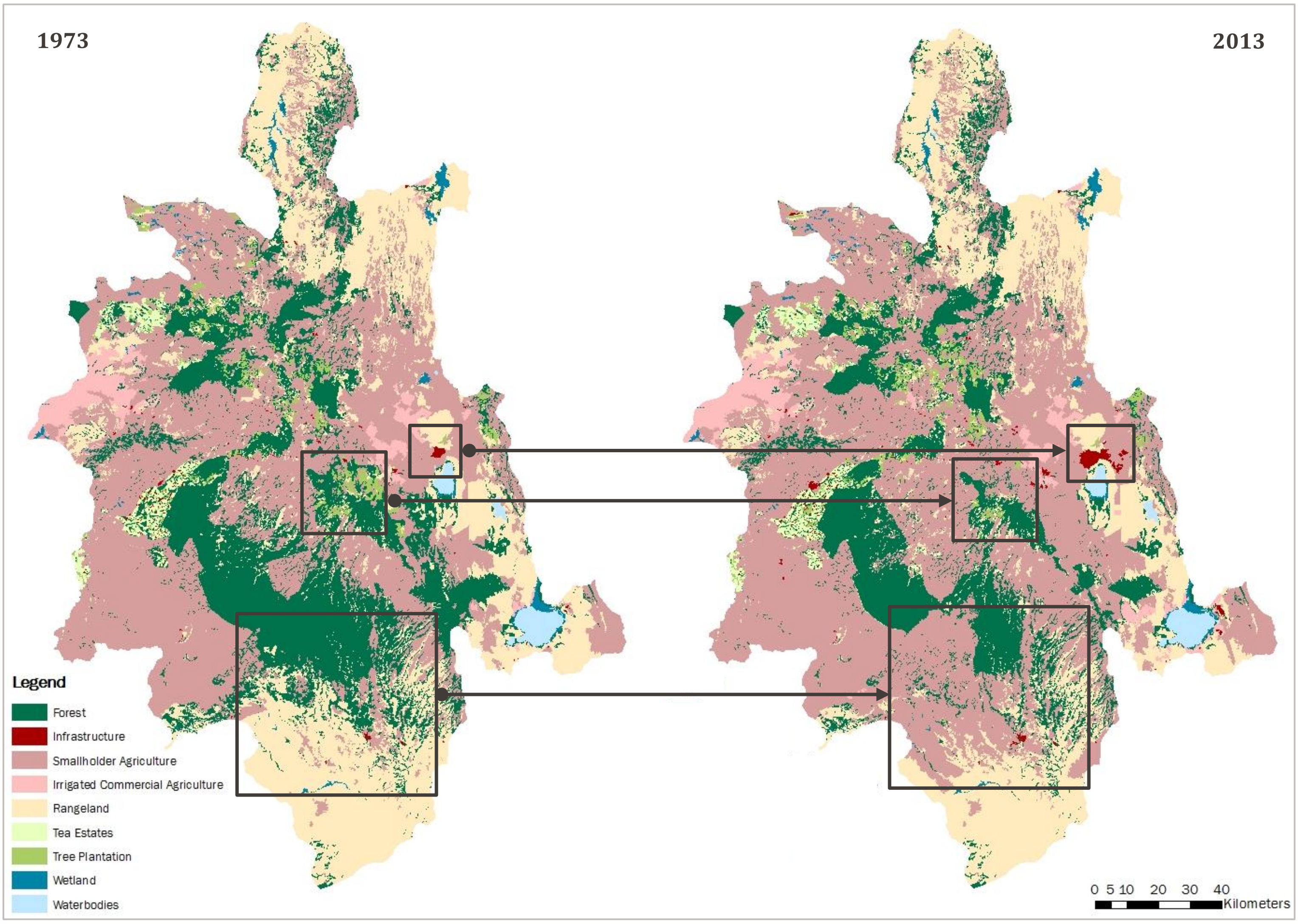


Figure 4.4: Land use changes 1973-2013

4.3 EXPLORATORY DATA ANALYSIS

The accuracy assessment previously described and the land use dynamics analysis, were essential steps before regression modelling. Based on these results the dependent variables could be identified (see paragraph 3.5). Besides the dependent variables, multiple independent variables were selected (see paragraph 3.6), and this section describes the exploratory data analysis: the collinearity between variables was analysed (see 3.7.2 for methods).

In general, Spearman's rank correlation coefficient (table 4.8) shows low correlations between the independent variables. However, high correlations between the population density variables and between temperature and elevation (-0.97) were detected. The population density variables were never used together as input. The temperature and elevation variables were reconsidered in every model, and depending on the VIF value, either elevation or temperature was excluded. In addition, Spearman's rho indicates that there is a stronger correlation between rainfall and soil pH (-0.81), TWI and slope (-0.79), and rainfall and population density (0.75-0.76). As a consequence, in every statistical model the correlations were reassessed.

Besides Spearman's rho, the multi-collinearity was tested with the VIF. As mentioned in the methodology, a threshold of 10 was applied**.** Thus, the variables that were strongly correlated according to Spearman, and showed VIF values that exceeded 10, were excluded from the models. For the both models this resulted in discarding either temperature or elevation (see the regression models tables 4.9, 4.10, 4.11, 4.12 and 4.13 for the variables that were included and the VIF values of the variables**)**. The other independent variables all showed VIF values below 10, with most of them around 1, which points to no correlation.

Table 4.8 Spearman's rho

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Rainfall	1.00																
2	Temperature	-0.10	1.00															
3	Elevation	0.24	**-0.97**	1.00														
4	Slope	0.24	-0.18	0.24	1.00													
5	Aspect North	0.06	0.06	-0.03	0.06	1.00												
6	Aspect East	-0.13	-0.03	0.03	-0.06	-0.05	1.00											
7	Curvature	-0.01	-0.03	0.05	0.07	-0.01	0.02	1.00										
8	TWI	-0.21	0.20	-0.28	**-0.79**	-0.01	0.02	-0.46	1.00									
9	Soil pH	**-0.81**	0.35	-0.47	-0.33	0.02	0.05	-0.01	0.32	1.00								
10	Soil CEC	-0.36	-0.36	0.30	0.06	-0.04	0.08	0.02	-0.06	0.33	1.00							
11	Distance Towns	-0.27	-0.18	0.17	0.07	-0.05	0.05	0.01	-0.07	0.14	0.26	1.00						
12	Distance Roads	-0.17	-0.09	0.05	0.02	-0.03	0.04	0.02	-0.03	0.05	0.04	0.30	1.00					
13	Distance Rivers	-0.11	-0.12	0.10	0.02	0.03	-0.03	-0.01	0.00	0.06	0.03	-0.18	-0.06	1.00				
14	Population Density '79	**0.76**	0.08	0.03	0.07	0.08	-0.11	-0.04	-0.05	-0.56	-0.30	-0.39	-0.16	0.00	1.00			
15	Population Density '89	**0.76**	0.08	0.03	0.07	0.08	-0.11	-0.04	-0.05	-0.55	-0.30	-0.39	-0.16	0.00	**1.00**	1.00		
16	Population Density '99	**0.76**	0.08	0.03	0.07	0.08	-0.11	-0.04	-0.05	-0.55	-0.30	-0.39	-0.16	0.00	**1.00**	**1.00**	1.00	
17	Population Density '09	**0.75**	0.07	0.04	0.06	0.08	-0.11	-0.04	-0.05	-0.55	-0.29	-0.38	-0.17	0.00	**1.00**	**1.00**	**1.00**	1.00

4.4 TIME PERIOD 1: 1973-1984

The results of the forest conversion and agricultural expansion models for the first period are shown in table **4.9**. The regression coefficients (β), the odds ratio and the VIF values of the independent variables in two land use change models are shown. In case $\beta > 0$, the probability that the events occurs (e.g. forest conversion or agricultural expansion) increases, when the value of the independent variable increases. When $\beta < 0$, the probability that the event occurs decreases, when the value of the independent increases. In particular the odds ratio of each independent variable is of importance when interpreting the models. It gives an indication about the effect of the variable and tells something about how much more likely or unlikely it is for the outcome to be present given the independent variables. It can be understood in a similar way as β: if the odds ratio is higher than 1, one unit increase in the explanatory variable, increases the odds of forest conversion (the darker colours in the table). However, if the odds ratio is lower than 1, one unit increase of the independent variable, decreases the odds of the event occurring (lighter colours in the table). For example, table 4.9 shows that if predictor variable rainfall increases with one unit (which would be 1 mm in this case), the odds of the event are multiplied by 0.97 (which points to a decrease in the odds), when the other k – 1 predictor variables are held constant (see also paragraph 3.7.3 for methods of the statistical framework). Furthermore, the number of polygons of land use change included are mentioned (N), the pseudo R-Square measures Nagelkerke R^2 and Cox and Snell R^2, the AUC goodness of fit measure, and Moran's I for spatial autocorrelation. The remainder of this section describes the results of the forest model first, followed by the agricultural expansion model.

Table 4.9: Forest conversion and agricultural expansion models for period 1

PERIOD 1	**FOREST CONVERSION MODEL**			**AGRICULTURAL EXPANSION MODEL**		
	β	**Odds ratio**	**VIF**	**β**	**Odds ratio**	**VIF**
Rainfall	-0.030***	0.97	3.45	-0.005	1.00	4.05
Temperature[1]	-	-	-	-	-	-
Elevation[1]	0.002***	1.00	2.14	0.002***	1.00	2.54
Slope	-0.164***	0.85	2.04	-0.011	0.99	2.88
Aspect North	-0.143	0.87	1.04	-0.001	1.00	1.07
Aspect East	0.846***	**2.33**	1.05	-0.328	0.72	1.08
Curvature	0.155	1.17	1.48	0.110	1.12	1.61
TWI	-0.867**	0.42	4.10	0.042	1.04	3.17
Soil pH	-0.122***	0.89	3.56	-0.032	0.97	4.06
CEC	-0.074***	0.93	2.06	-0.090***	0.91	1.43
PopDens_79	-0.006***	0.99	2.29	-0.005	1.00	2.45
Distance_Towns	-0.047***	0.95	1.47	0.035***	**1.04**	1.32
Distance_Roads	0.058**	**1.06**	1.12	-0.021	0.98	1.15
Distance_Rivers	0.023	1.02	1.19	0.012	1.01	1.21
Intercept	14.030***			-1.928		
N	4721			1789		
Nagelkerke R^2	**0.20**			**0.19**		
Cox & Snell R^2	**0.11**			**0.10**		
AUC	0.766			0.769		
Moran's I	0.168			0.276		

[1]Either temperature or elevation was excluded due to multi-collinearity
*p <0.05, **p<0.01,*** p<0.001

4.4.1 Forest conversion model

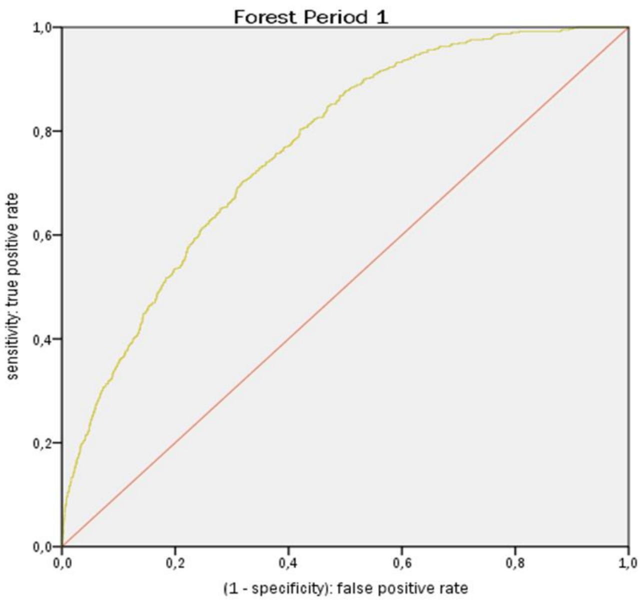
The table shows that the effects of rainfall, elevation, slope, aspect east, TWI, soil pH, soil CEC, population density, distance to towns and distance to roads are significant. Aspect north, curvature and distance to rivers are not. According to table 4.9 an increase of one unit in aspect east and distance to roads, increases the odds of a forest conversion: the odds are multiplied by 2.33 and 1.06. Furthermore, an increase of rainfall, slope, TWI, soil pH, CEC, population density and distance to towns, decreases the odds of forest conversion occurring in the first period (odds are multiplied with 0.97, 0.85, 0.42, 0.89, 0.93, 0.99 and 0.95 respectively). The odds ratio of 1.00 for elevation indicates that an increase of one unit (meters) has no effect on the likelihood of forest conversions.

The pseudo R-Square (measured with Nagelkerke R^2 and Cox & Snell R^2) indicates that the model explained between 11% and 20% of the variability in forest conversions (see table 4.9). In addition, the discriminating power of the model is good: the AUC gives a value of 0.766. Figure 4.5 shows the receiving operating characteristic (ROC) curve, which indicates that the model is quite accurate: the curve reaches towards the left top and is not close to the diagonal (which represents the random case). Thus, the power of the model to correctly classify the areas that actually changed and that did not change is satisfactory. Furthermore, Moran's I indicates that the spatial autocorrelation is weak, but positive (0.168).

4.4.2 Agricultural expansion model

Only three variables were statistically significant in the agricultural expansion model: elevation, soil CEC and distance to towns. For every one unit of increase of distance to towns, the odds of agricultural expansion multiplied by 1.04. In contrast, the soil CEC variable, shows that one unit of increase results in a decrease of the likelihood of occurrence of agricultural expansion. And for this model as well, elevation neither increases of decreases the chance of agricultural expansion to occur.

The agricultural expansion model shows a good discriminatory power (AUC = 0.769). The ROC curve indicates that the model is accurate as well: the curve is close to the left-hand top border (see figure 4.5). The pseudo R-Square shows that between 10% and 19% of variation in the agricultural expansion model has been explained. Moran's I shows a quite weak but positive spatial autocorrelation (0.276) : the residuals tend to be more clustered than dispersed.



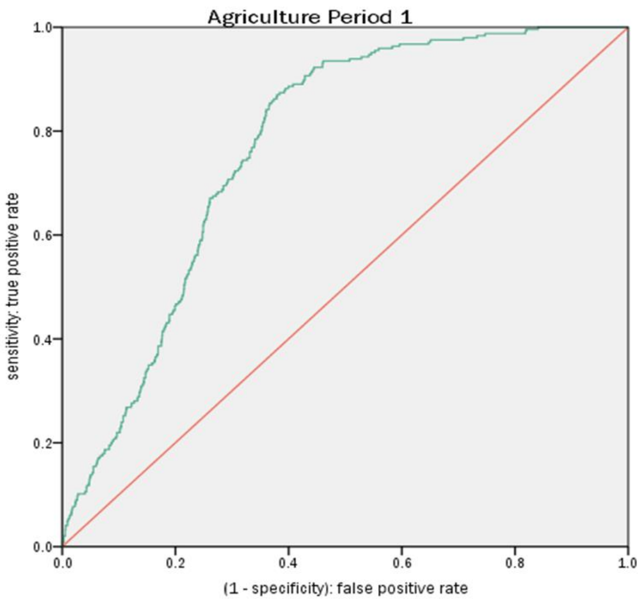


Figure 4.5: ROC curve showing the performance of the forest conversion model (left, yellow curve) and the agricultural expansion model (right, green curve) in period 1

4.5 TIME PERIOD 2: 1984-1994

Table 4.10: Forest conversion and agricultural expansion model period 2

PERIOD 2	FOREST CONVERSION MODEL			AGRICULTURAL EXPANSION MODEL		
	β	Odds ratio	VIF	β	Odds ratio	VIF
Rainfall	0.014*	**1.01**	3.33	0.001	1.00	4.30
Temperature[1]	-0.028***	0.97	2.50	-	-	-
Elevation[1]	-	-	-	0.003***	1.00	2.76
Slope	0.010	1.01	2.28	0.063*	**1.07**	3.35
Aspect North	-0.678***	0.51	1.04	-0.140	0.87	1.08
Aspect East	0.616***	**1.85**	1.06	-0.479**	0.62	1.05
Curvature	0.932	2.54	1.45	1.645*	**5.18**	1.74
TWI	0.754*	**2.13**	3.89	1.062**	**2.89**	3.93
Soil pH	0.029	1.03	4.13	0.095***	**1.10**	4.93
CEC	-0.067**	0.93	1.95	-0.130***	0.88	1.34
PopDens_89	-0.006***	0.99	2.26	-0.001	1.00	2.25
Distance_Towns	0.019*	**1.02**	1.48	0.007	1.01	1.33
Distance_Roads	-0.037	0.96	1.14	-0.086*	0.92	1.14
Distance_Rivers	-0.115***	0.89	1.17	-0.021	0.98	1.21
Intercept	-3.892			-18.658***		
N	4128			2058		
Nagelkerke R^2	**0.10**			**0.19**		
Cox & Snell R^2	**0.04**			**0.11**		
AUC	0.725			0.756		
Moran's I	0.114			0.225		

[1]Either temperature or elevation is excluded due to multi-collinearity
*p <0.05, **p<0.01,*** p<0.001

4.5.1 Forest conversion model

In the second time period rainfall, temperature, aspect east, TWI, soil CEC, population density, distance to towns and distance to rivers prove to be statistically significant. One unit increase of rainfall, aspect east, TWI and distance to towns indicate an increase in the likelihood of forest conversion (see table 4.10). The odds of forest conversion are multiplied by 1.01, 1.85, 2.13 and 1.02 respectively. An increase in one unit of temperature, aspect north, soil CEC, population density, and distance to rivers significantly decrease the likelihood of forest conversion: the odds are multiplied by 0.97, 0.51, 0.93, 0.93, 0.99 and 0.89 respectively.

Nagelkerke R^2 and Cox & Snell R^2 indicate that the model only explains between 4% and 10% of the variability in the occurrence of forest conversion – which is very weak. The ROC curve and the AUC (0.725) show that the model's discriminatory power is acceptable. Last, Moran's I shows a weak and positive autocorrelation of the residuals: 0.114.

4.5.2 Agricultural expansion model

The biophysical variables slope, curvature, TWI and soil pH have a significant effect on the likelihood of agricultural expansion to occur. The odds ratios are 1.07, 5.18, 2.89 and 1.10 respectively. On the other hand, an increase in one unit of aspect east, soil CEC and distance to roads decreases the odds of agricultural expansion by 0.62, 0.88 and 0.92. Elevation is of importance in this model, although it neither increases or decreases the odds of agricultural expansion. Rainfall, aspect north, population density, distance to roads and distance to rivers show to have no significant effect in the second period.

The pseudo R-Square indicates that between 11% and 19% of the variance has been explained. The ROC curve shows an acceptable performance of the model (see figure 4.6) and the AUC is 0.756 which indicates that the power to differentiate the actual agricultural expansions was good. Moran's I indicates a weak positive spatial autocorrelation: 0.225.

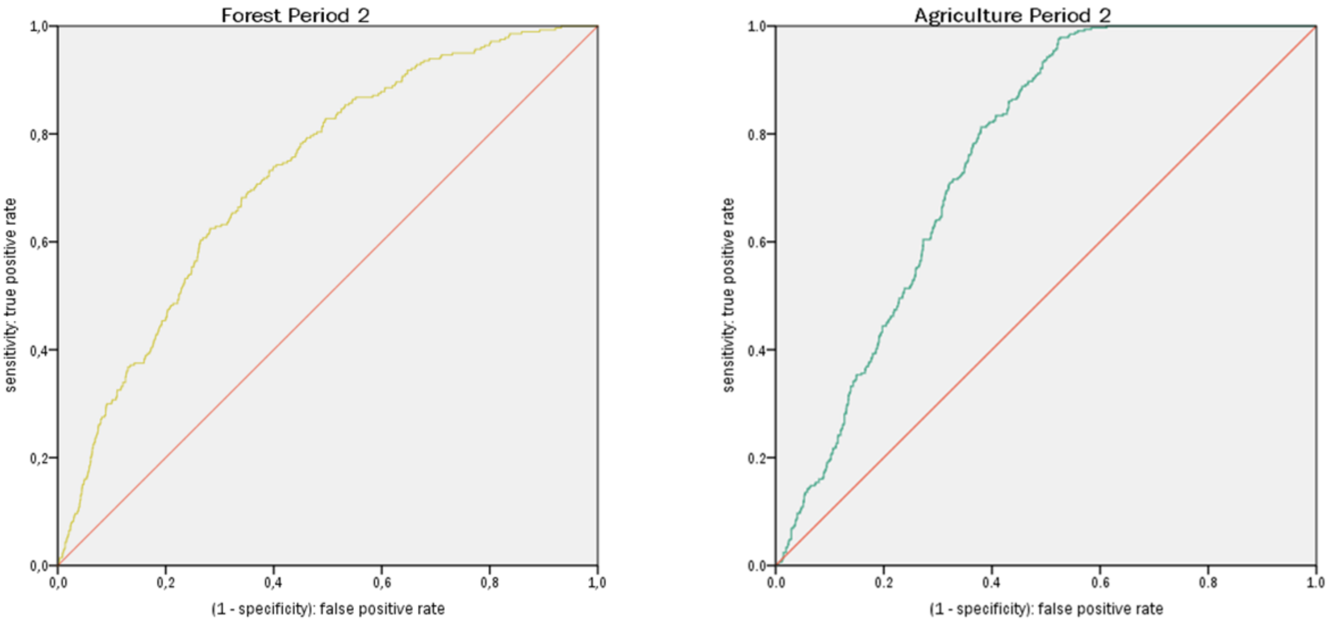


Figure 4.6: ROC curve showing the performance of the forest conversion model (left, yellow curve) and the agricultural expansion model (right, green curve) in period 2

4.6 TIME PERIOD 3: 1994-2003

Table 4.11: Forest conversion and agricultural expansion model period 3

PERIOD 3	FOREST CONVERSION MODEL			AGRICULTURAL EXPANSION MODEL		
	β	Odds ratio	VIF	β	Odds ratio	VIF
Rainfall	0.003	1.00	3.26	0.008	1.01	3.05
Temperature[1]	-0.025***	0.98	2.48	-	-	-
Elevation[1]	-	-	-	0.003***	1.00	2.11
Slope	0.002	1.00	2.27	-0.030	0.97	1.64
Aspect North	-0.429***	0.65	1.04	0.437***	**1.55**	1.04
Aspect East	0.571***	**1.77**	1.06	0.437	0.91	1.04
Curvature	1.066*	**2.90**	1.44	-0.097	0.39	1.27
TWI	1.031***	**2.81**	3.90	-0.935	1.19	2.89
Soil pH	-0.014	0.99	4.16	0.176	0.99	3.33
CEC	-0.037*	0.96	1.96	-0.006***	0.93	1.83
PopDens_99	-0.001	1.00	2.18	-0.067*	1.00	2.06
Distance_Towns	0.017**	**1.02**	1.47	-0.037***	0.96	1.34
Distance_Roads	0.054**	**1.06**	1.16	0.048*	**1.05**	1.10
Distance_Rivers	-0.059**	0.94	1.17	0.028	1.03	1.21
Intercept	-2.936			-6.331**		
N	3874			3075		
Nagelkerke R^2	**0.12**			**0.32**		
Cox & Snell R^2	**0.07**			**0.23**		
AUC	0.707			0.788		
Moran's I	0.159			0.157		

[1]Either temperature or elevation is excluded in the models due to multi-collinearity
*p <0.05, **p<0.01,*** p<0.001

4.6.1 Forest conversion model

Table 4.11 shows that in the third period an increase of one unit of aspect east, curvature, TWI, and distance to towns and roads, indicate an increase in the odds of forest to change. The odds ratio are 1.77, 2.90, 2.81, 1.02 and 1.06 respectively. Temperature, aspect north, soil CEC and distance to rivers indicate a decrease in the likelihood of forest conversion. The distance variables are becoming more important compared to the first two periods. Moreover, the variables rainfall, slope, soil pH and population density were not statistically significant in this time period.

The pseudo R-Square shows an explanatory power of the model between 7% and 12%. The ROC curve indicates a moderate performance, just as the AUC shows: 0.707 (see figure 4.7). Moran's I indicates a weak and positive spatial autocorrelation of the residuals (0.157).

4.6.2 Agricultural expansion model

Three biophysical variables prove to be statistically significant in the agricultural expansion model in the third period: elevation, aspect north and soil CEC. In addition, the socio-economic variables population density, distance to towns and distance to roads are significant as well.

One unit increase of aspect north and distance to rivers increases the odds of agricultural expansion. The odds are multiplied with 1.55 and 1.05. In contrast, an increase of soil CEC and distance to towns decreases the odds of agricultural expansion. Elevation and population density are of importance for agriculture expansion, however the odds ratio of 1.00 indicates that there is almost no effect on the increase of decrease of the likelihood of agricultural expansion.

The agricultural expansion model in the third period shows the highest pseudo R-Square of all agricultural models: between 23% and 32% of the variance is explained. The ROC curve reaches the top-left border which indicates that the model uses little false positives for prediction (see figure 4.7). The AUC clarifies this as well: 0.788 is a good score. The spatial autocorrelation is positive and weak (Moran's I = 0.157).

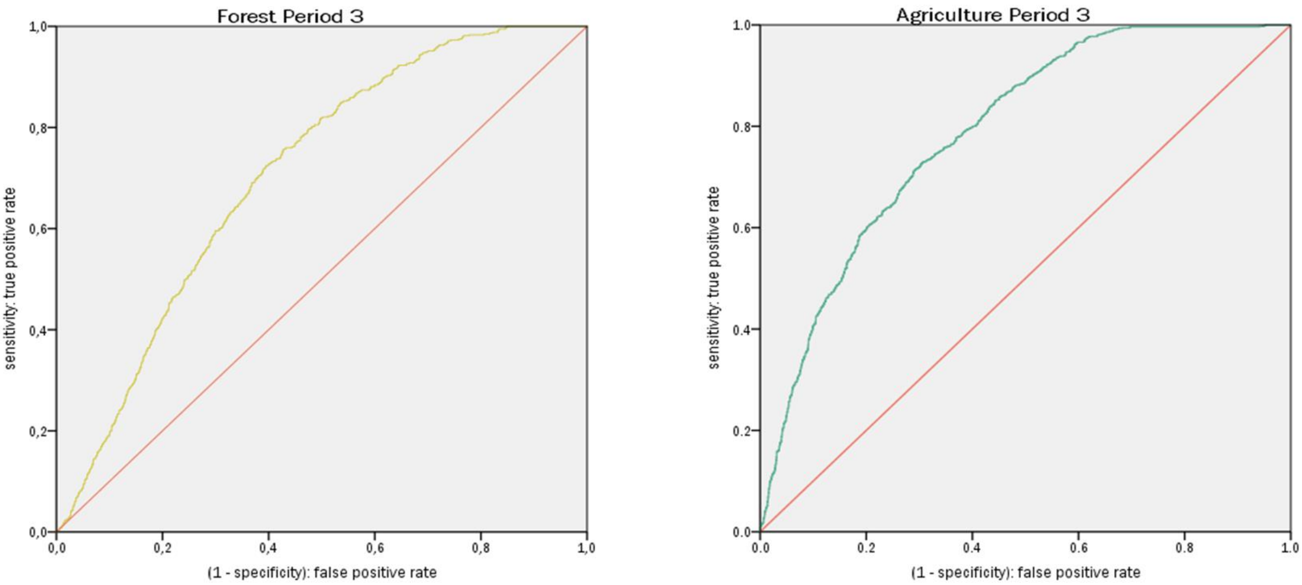


Figure 4.7: ROC curve showing the performance of the forest conversion model (left, yellow curve) and the agricultural expansion model (right, green curve) in period 3

4.7 TIME PERIOD 4: 2003-2013

Table 4.12: Forest conversion and agricultural expansion model period 4

PERIOD 4	FOREST CONVERSION MODEL			AGRICULTURAL EXPANSION MODEL		
	β	Odds ratio	VIF	β	Odds ratio	VIF
Rainfall	-0.024**	0.98	3.16	-0.003	1.00	3.36
Temperature[1]	-	-	-	-	-	-
Elevation[1]	0.001***	1.00	2.05	0.002***	1.00	2.18
Slope	-0.110***	0.90	2.09	0.002	1.00	1.81
Aspect North	0.257	1.29	1.04	0.123	1.13	1.03
Aspect East	0.508***	**1.66**	1.05	-0.549***	0.58	1.04
Curvature	-0.129	0.88	1.48	0.670	1.95	1.34
TWI	-0.863*	0.42	4.48	-0.689***	0.50	3.70
Soil pH	-0.146***	0.86	3.75	0.030*	**1.03**	3.87
CEC	0.070**	**1.07**	2.04	0.002	1.00	1.68
PopDens_09	-0.007***	0.99	2.11	-0.008***	0.99	2.16
Distance_Towns	0.040***	**1.04**	1.47	-0.033***	0.97	1.36
Distance_Roads	0.084***	**1.09**	1.17	0.137***	**1.15**	1.10
Distance_Rivers	-0.082**	0.92	1.19	-0.052*	0.95	1.19
Intercept	11.087***			-2.281**		
N	3321			3390		
Nagelkerke R^2	**0.28**			**0.16**		
Cox & Snell R^2	**0.13**			**0.09**		
AUC	0.833			0.738		
Moran's I	0.218			0.225		

[1]Either temperature or elevation is excluded in the models due to multi-collinearity
*p <0.05, **p<0.01,*** p<0.001

4.7.1 Forest conversion model

In the last period two biophysical variables significantly increase the odds of forest conversion to occur: every one unit increase of aspect east and soil CEC multiplies the odds by 1.66 and 1.07 (see table 4.12). The distance variables are again of importance in the forest model: the probability of forest conversion increases by 4% (odds ratio of 1.04) and 9% (odds ratio of 1.09) for every meter increase in distance to towns and roads. On the other hand, an increase in one unit of the variables rainfall, slope, TWI, soil pH, population density and distance to rivers significantly decrease the odds of forest conversion to occur. And once more, elevation does not increase or decrease the odds of conversion. Aspect north and curvature prove to be not statistically significant.

The forest conversion model in the fourth period shows the highest pseudo R-Square of all the forest conversion models: between 13% and 28% of the variance is explained by this model. The ROC curve and the AUC (0.833) indicate that the model is accurate. The curve is close to the top left border which means that it predicts true positives with a few false positives (see figure 4.8). Moran's I is weak and positive (0.218), which means that the residuals tend to be more clustered than random.

4.7.2 Agricultural expansion model

The variables elevation, aspect east, TWI, soil pH and the socio-economic variables prove to be significant, whereas rainfall, slope, aspect north, curvature and soil CEC are not. In the last time period only soil pH and distance to roads increase the odds of agricultural expansion to occur: the odds are multiplied with 1.03 and 1.15. Noteworthy is the change of the latter variable: compared to the third period, distance to roads shows a higher odds ratio in the fourth period. For every 1 meter increase in distance to roads to probability of agricultural expansion increases by 5% in the third period (odds ratio of 1.05) and 15% in the fourth period (odds ratio of 1.15).

Furthermore, one unit increase of aspect east, TWI, population density, distance to towns and distance to rivers, decreases the likelihood of agricultural expansion. The odds are multiplied with 0.58, 0.50, 0.99, 0.97 and 0.95 respectively. Elevation shows an odds ratio of 1.00 and thus no effect on expansion of agriculture.

The explanatory power of the model is weak: between 9% and 16% of the variability of agriculture expansion is explained by the model. The ROC curve and the AUC indicate a satisfactory model: AUC = 0.738. The spatial autocorrelation is positive and quite weak: Moran's I = 0.225.

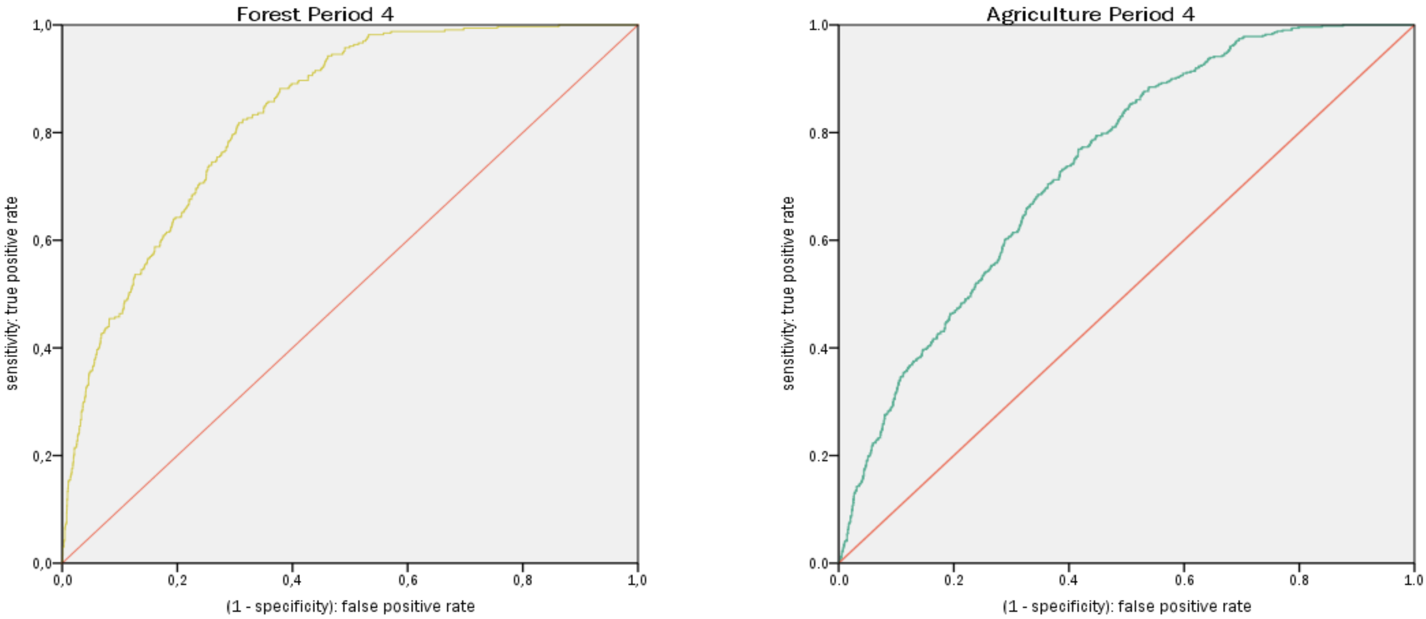


Figure 4.8: ROC curve showing the performance of the forest conversion model (left, yellow curve) and the agricultural expansion model (right, green curve) in period 4

4.8 OVERALL TIME PERIOD: 1973-2013

Table 4.13 shows the results of the forest conversion and agricultural expansion model of the overall period: 1973-2013. These last two models used the exact same variables as the other models except for the population density variable: the change of population density between the first and last period in percentages was introduced in this model.

Table 4.13: Forest conversion and agricultural expansion model period 1973 - 2013

ALL PERIOD	FOREST CONVERSION MODEL			AGRICULTURAL EXPANSION MODEL		
	β	Odds ratio	VIF	β	Odds ratio	VIF
Rainfall	0.004	1.00	3.33	0.024***	**1.02**	2.33
Temperature[1]	-	-	-	-	-	-
Elevation[1]	0.001***	1.00	2.55	0.002***	1.00	1.92
Slope	-0.055***	0.95	3.77	-0.008	0.99	2.78
Aspect North	-0.255**	0.77	1.06	0.325**	**1.38**	1.04
Aspect East	0.870***	**2.39**	1.06	-0.443***	0.64	1.04
Curvature	0.804*	**2.23**	1.97	-0.276	0.76	1.68
TWI	0.121	1.13	4.48	-0.048	0.95	3.59
Soil pH[2]	-0.099***	0.91	3.33	0.000	1.00	1.00
CEC[2]	-0.004	1.00	1.86	-	-	-
PopDens_%Change	0.013***	**1.01**	2.35	0.013***	**1.01**	2.14
Distance_Towns	-0.005	1.00	1.53	-0.009	0.99	1.28
Distance_Roads	0.059***	**1.06**	1.12	-0.009	0.99	1.07
Distance_Rivers	-0.014	0.99	1.17	0.028	1.03	1.21
Intercept	-0.294			-9.129***		
N	4721			3377		
Nagelkerke R^2	**0.27**			**0.36**		
Cox & Snell R^2	**0.20**			**0.27**		
AUC	0.764			0.775		
Moran's I	0.164			0.170		

[1]In both models temperature is excluded due to multi-collinearity with elevation

[2]In the agricultural expansion model CEC is excluded due to multi-collinearity with soil pH

*p <0.05, **p<0.01,*** p<0.001

4.8.1 Forest conversion model

In the forest conversion model of the overall period, the variables elevation, slope, aspect North, aspect East, curvature, soil pH, population density change and distance to roads are significant, while rainfall, TWI, CEC, distance to towns and distance to rivers have no significant influence on deforestation in the study area (see table 4.13). An increase of one unit aspect east, curvature, population density growth and distance to roads multiplies the odds of forest conversion by 2.39, 2.23, 1.01 and 1.06 respectively. On the other hand, an increase is one unit slope, aspect north and soil pH, decreases the chance that deforestation occurs. The odds ratio of 1.00 for elevation indicates that an increase of one unit (meters) has almost no effect on the likelihood of forest conversions.

Nagelkerke R^2 and Cox & Snell R^2 indicate that the model explained between 20% and 27% of the variability in the presence of forest conversions (see table 4.13). In addition, the discriminating power of the model is good: the AUC gives a value of 0.764. Figure 4.9 shows the ROC curve, which indicates that the model is quite accurate: the curve reaches towards the left top and is not close to the diagonal (which represents the random case). Thus, the power of the model to correctly classify the areas that actually changed and that did not change is satisfactory. Furthermore, Moran's I indicates that the spatial autocorrelation is weak, but positive (0.164).

4.8.2 Agricultural expansion model

The agricultural expansion model on the other hand, shows only five significant variables. An increase in rainfall, aspect North and population density, multiplies the odds of agricultural expansion by 1.02, 1.38 and 1.01. An increase is aspect East decreases the chance of change by 0.64. And again, elevation has almost no effect on the land use change.

The results of the pseudo R-Squares are the highest of all the agricultural expansion models: between 27% and 36% of the variance is explained by this model (which is still weak). The AUC of 0.775 indicates that the model's discriminatory power is good, which is also visible in figure 4.9: the ROC curve is not close to the diagonal. Moran's I gives a value of 0.170: the residuals tend to be more clustered than dispersed, however the spatial autocorrelation is weak.

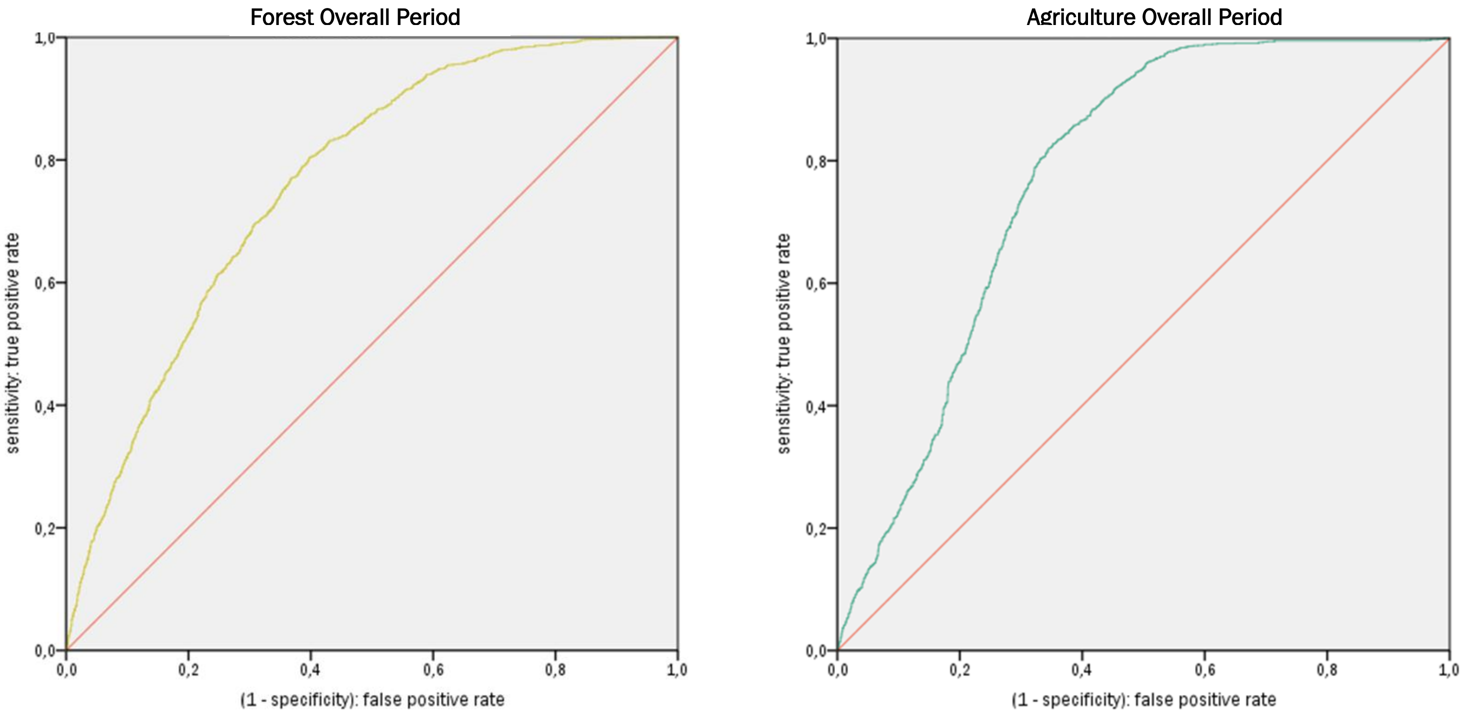


Figure 4.9: ROC curve showing the performance of the forest conversion model (left, yellow curve) and the agricultural expansion model (right, green curve) in the overall period 1973-2013

4.9 SUMMARY OF THE MODELLING RESULTS

Table 4.14, figure 4.10 and figure 4.11 show an overview of the underlying drivers that significantly contributed to land use change in the Mau forest complex. Depending on the model and the time period, some drivers increased the odds of a land use change to occur (+ in the table), while other drivers decreased the odds of a land use change to occur (- in the table). Elevation was statistically significant in both models, however the effect of the variable was little and the likelihood of land use change neither increased or decreased much (+/- in table).

Based on the results it can be concluded that for the forest conversion model, both biophysical and socio-economic factors were of importance. Rainfall, aspect East (in all periods), curvature, TWI, soil CEC, population density, distance to towns and distance to roads increased the odds of forest conversion and in particular the distance variables became more important in more recent time periods. On the other hand, the agricultural expansion model revealed that fewer variables were significantly contributing to agricultural land use change: the climatic factor temperature was not significant in any time period. The other biophysical drivers in this model mainly have an influence in the second time period, whereas the socio-economic factors are more important in the third and fourth time period. In particular a larger distance to roads results in an increase of the likelihood of agricultural expansion between 1994 and 2013.

Both models show the highest explanatory power in the overall model: the period 1973-2013 (see table 4.14). However, for agriculture only a few variables were significant: rainfall, elevation (almost no effect), aspect and population density. Noteworthy is that in these overall models another population density variable was introduced, namely the change of population density in percentages between the first and the last period. The results of both models show that this variable has a significant effect in the land use changes in the Mau forest in the period 1973-2013: the higher the growth in population density, the more likely forest conversion or agricultural expansion occurs.

Table 4.14: Overview model performance and odds

	FOREST CONVERSION MODEL					**AGRICULTURAL EXPANSION MODEL**				
	73-84	**84-94**	**94-03**	**03-13**	**ALL**	**73-84**	**84-94**	**94-03**	**03-13**	**ALL**
Rainfall	-	+		-						+
Temperature		-	-							
Elevation	+/-			+/-	+/-	+/-	+/-	+/-	+/-	+/-
Slope	-			-	-		+			
Aspect North		-	-		-			+		+
Aspect East	+	+	+	+	+		-		-	-
Curvature			+		+		+			
TWI	-	+	+	-			+		-	
Soil pH	-			-	-		+		+	
Soil CEC	-	-	-	+		-	-	-		
Population Density	-	-		-	+			+/-	-	+
Distance_Towns	-	+	+	+		+		-	-	
Distance_Roads	+		+	+	+		-	+	+	
Distance_Rivers		-	-	-					-	
Pseudo R^2	0.11–0.20	0.04–0.10	0.07-0.12	0.13– 0.28	**0.20-0.27**	0.10-0.19	0.11-0.19	0.23–0.32	0.09–0.16	**0.27-0.36**

+ = increase in the odds, - = decrease in the odds, +/- = negligible effect

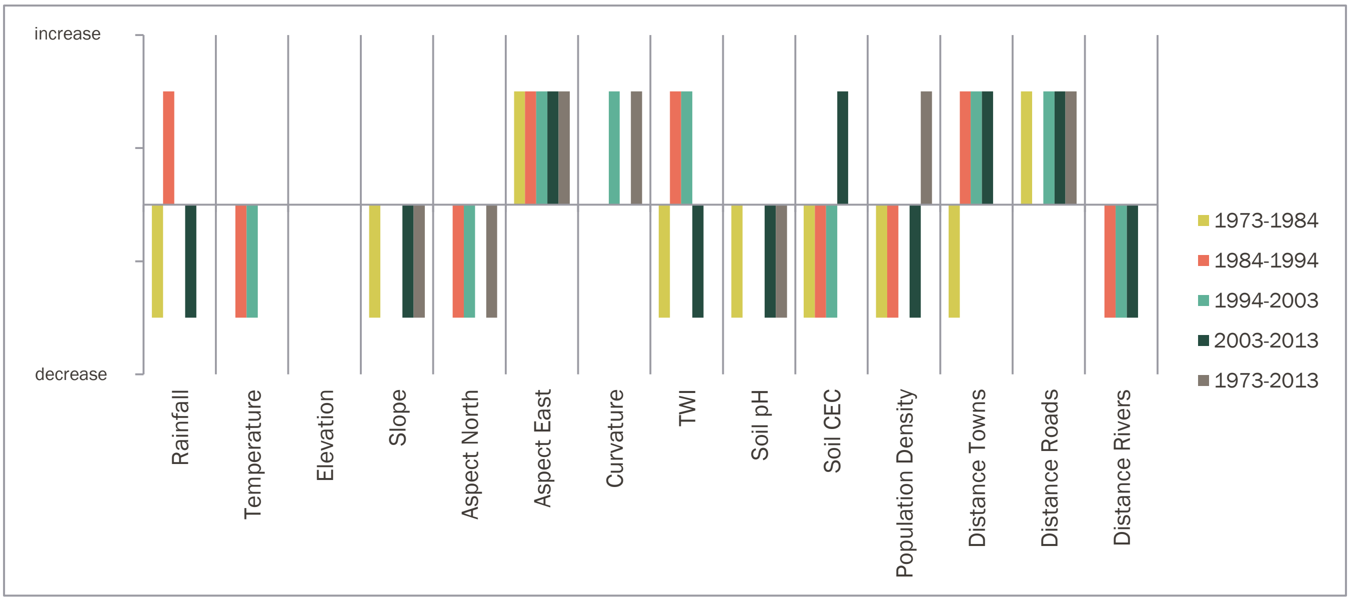


Figure 4.10: Results odds ratio (increase/decrease) per time period forest conversion model

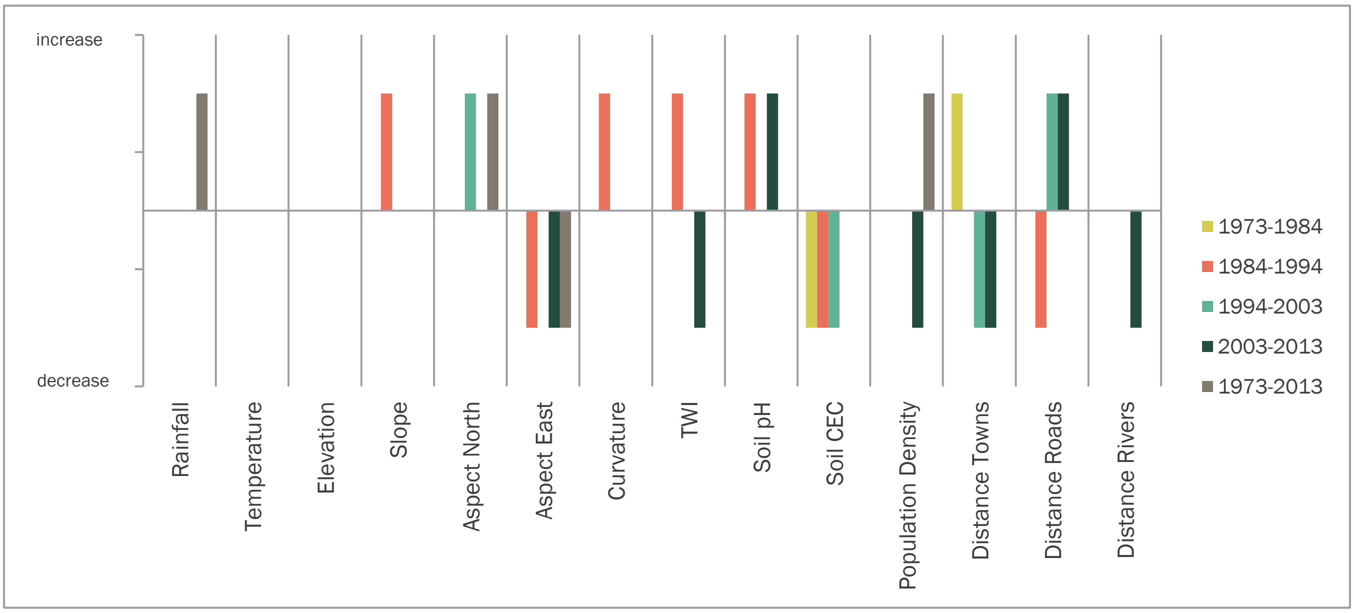


Figure 4.11: Results odds ratio (increase/decrease) per time period agricultural expansion model

5. DISCUSSION

This chapter first discusses the uncertainty in the data, then the land use changes and the proximate drivers, and finally the land use change models and the underlying drivers.

5.1 UNCERTAINTY IN THE DATA

The results of the accuracy assessment, land use dynamics analysis and the regression modelling demonstrated that there was uncertainty in the data that was used. First, the accuracy assessment showed poor results, namely an overall accuracy of 74.2%, which is below the 85% threshold generally applied (see Foody, 2012). Not to mention the low producer's and user's accuracies of some land use classes, which should in general not be lower than 70% (some classes were below 50%) (Foody, 2012). Low accuracies can usually be explained by errors in the classification, and in this study some concerns regarding the land use classification maps can be identified. The resolution of the Landsat data is coarse: 60 by 60 meters for the 1970s and 30 by 30 meters for the other time periods, which indicates that misclassification already occurs due to the large pixel size: some land use classes that are smaller than the pixels size (such as rivers and roads) are missed in these coarse resolutions. Nevertheless, for the 40-year period of this research, this was the most accurate data available. In addition, the dates of the images that were selected for classification are disputable because seasonality was not considered, which might have resulted in misclassifications. The weather in the Mau forest complex follows a bimodal rainfall pattern and experiences rain and dry seasons. Some images were selected in the dry season, while others were selected in the rain season, which makes distinguishing between different land uses types difficult.

Besides the aforementioned classification issues, the low accuracy can be written to the fact that the assessment itself was repeated and adjusted from a previous research. Although a systematic sampling procedure was applied (the samples were selected at equal intervals of 5 km), the exact methods were not transparent and the sampling size of the land use classes was not considered. When doing an accuracy assessment the sampling scheme should precisely be examined: random, systematic, stratified random or cluster sampling (Foody, 2012). Next to that, it is essential to select and calculate the sample size per land use class. For instance, this accuracy assessment checked for three land use classes (waterbodies, wetlands and infrastructure) less than 20 points, which makes the results not statistically valid (Congalton and Green, 2008). In this study we tried to improve the classification by incorporating other data and context: a fieldwork was organised to decrease the errors in the reference points. However, a pitfall here is that the reference data was collected in 2016 and the remote sensing data used was from 2013. Ground data should be collected close in time to the remote sensing data because landscapes can change fast (Congalton and Green, 2008).

The last source of uncertainty in the data could lie in the independent variables. For the climatic variables rainfall and temperature, values over a 50 year period in a coarse resolution were used: pixels of 925 by 925 meters. Because of the average values of the climatic data, the influence of variations could not be tested, which is a drawback as climatic variations might explain land use change (Kiage et al., 2007). A coarse resolution was the case for the soil variables as well: a spatial resolution of 250 meters. The elevation data and its derivatives have the highest spatial resolution: 30 by 30 meters, which makes it the most accurate data. In addition, the demographic variable population density showed a lot of uncertainty. In Kenya the administrative boundaries of the districts and counties changed over the past 40 years which made it difficult to trace back the exact population numbers per county. Moreover, the final population density numbers have a low spatial resolution: the data is on a county scale. The datasets of the road network, the rivers and towns, which were used to calculate Euclidean distances, are not perfectly reliable because they were created based on different data sources. Thus, all biophysical data, except the geomorphological data (elevation and its extracts), and all socio-economic data indicate a high degree of uncertainty. Nonetheless, the data collected, used and created, was the best available data.

5.2 LAND USE CHANGES AND THE PROXIMATE DRIVERS

The study revealed that three main land use change processes could be identified in the Mau forest complex in the period 1973-2013: forest and rangeland showed high losses, while smallholder agriculture increased in extend (see figure 4.3, table 4.5 and table 4.7). These results are consistent with a study of Olang and Kundu (2011) in the same area, and similar to studies in the Eastern Mau forest reserve and the Lake Nakuru drainage basin (Were et al., 2013), the Vihiga district which is close to the study area (Mutoko et al. 2014) and the River Njoro watershed which falls within the study area (Baldyga et al., 2007). The most important land use change in the study area was a conversion from forest to smallholder agriculture. Hence, smallholder agriculture can be considered the most important proximate driver of deforestation in the Mau forest complex in every time period. Moreover, larger scale agriculture is a proximate driver of deforestation: forest lost area to tree plantations in the period 1973-1984 and to tea estates in the first three periods (1973 till 2003) (Appendix H). These findings are in agreement with several studies that identified agriculture as proximate driver of deforestation in African countries (Hosonuma et al., 2012; Kissinger et al. 2012).

Because of the outcome of the accuracy assessment (and the poor quality of the classification), the results of the land use dynamics analysis had to be interpreted with caution. Even though land use classes as rangeland, irrigated commercial agriculture and infrastructure showed notable changes, due to the low accuracies, we decided not to interpret the results nor use them in the land use change models. In the end only the most accurate land use classes and the related changes were used: forest conversion and agricultural expansion (see paragraph 3.5). It was expected that in this way the models would show the most reliable results and would give the best indication about the main underlying drivers of land use change.

5.3 THE UNDERLYING DRIVERS

The driving forces of the change processes were identified by use of logistic regressions models and the results demonstrated that different biophysical and socio-economic factors were of importance for forest conversion and smallholder agricultural expansion. In general the model performances were weak and revealed low pseudo R-Squares. Nagelkerke R^2 and Cox and Snell R^2 indicated that the weakest models were able to explain between 4% and 10% of the variability, whereas the best performing models – the overall models - explained between 27% and 36% of the variability in the presence of a land use conversion (table 4.14). In the first place, these low scores can be explained by the poor data quality of both the dependent and independent variables (the badly classified land use maps and coarse resolution data of the drivers - as discussed in paragraph 5.1). In addition to that, the models take local conditions into account, while drivers are able to operate over larger distances. For example, population can have effects over larger distances. In fact, Verburg et al. (2002) argue that considering multi-scale characteristics might improve model performance. We tried to overcome this problem, by calculating a population density over a larger area, nevertheless in the end the variable showed no significant contribution to most of the models due to a resolution that was too coarse.

Another reason for the low explanatory power of the models is that many other drivers are able to explain the forest conversions or agricultural expansions as well. These drivers could not be incorporated in the models because of the absence of spatial data and the difficulties of quantifying some of the variables. For example, factors that are missing in the models are: poverty, technological changes, environmental governance and policies, international drivers and behaviour of people (see Appendix D for possible underlying drivers). Literature showed that economic, social, cultural and political underlying drivers influence each other, they interact and do not operate independently (see Geist and Lambin (2002); Kissinger et al. (2010)). It is not possible to capture all these interactions in models, the models give an indication of the drivers that can contribute to land use change, however they do not give a cause-effect relationship. A growth in population does not by itself explain a land use change, other drivers should always be considered, and to incorporate all possible other drivers is nearly impossible.

Especially the changes of policy and governance is an important driver for the Mau forest because most loss of forest can be written to unplanned settlements and illegal extraction of resources (logging and charcoal burning) (UNEP, 2011). In particular in the 1990s and beginning of 2000s a lot of forest area was lost due to these weak environmental policies (Klopp and Sang, 2011; KWS, 2009). Being aware of these drivers is essential and therefore in future research these drivers should be addressed as well: not in a quantitative, but in a qualitative way (see recommendations paragraph 6.2).

5.3.1 Forest conversion

For the forest conversion model several remarks can be made with regard to the underlying drivers. First, the goodness of fit of the forest model improves over time: the model better explains the forest conversion in the last decade than in the periods before. In particular all socio-economic variables are of significance in the period 2003-2013, even though some of the effects are small. An important explanation is that humans, and human induced actions, appear to have more effect on land use changes now than in the past. Another reason is that in the last period the data of the independent variables used in the models is more accurate and up to date. Most datasets of the independent variables correspond with more recent periods, whereas little data was available from earlier time periods (e.g. no road network data in the 1970s).

Second, the results showed that in every period, the more eastern-facing the aspect is, the higher chance of forest conversion: the east-facing slopes appear to have a higher chance of deforestation. This cannot be explained by the sun side as at the equator the sun goes straight overhead and shines almost equally at all slopes. However, it might be explained by the fact that in Kenya the East and South-Eastern slopes are wetter due to moist winds from the Indian Ocean, and therefore may be more suitable for agriculture.

Then, the forest conversion model demonstrated an important influence of the topographical wetness index. In the second and third period (1984-1994 and 1994-2003) an increase in TWI demonstrated an increase in the odds of forest conversion, whereas for the first and last period the inverse was true. The land use change analysis showed that smallholder agriculture increased the most in the second and third periods (see figure 4.2). It appears that forest conversions, and mainly the change from forest to smallholder agriculture, are closely related to a high TWI, or a high degree of wetness.

A last notable result of the forest conversion model is that in the period 1994-2003 and 2003-2013 the accessibility factors distance to towns and roads become more important: the further from roads and towns, the more likely that deforestation occurs, and the closer to roads and towns, the less likely deforestation occurs. Many factors could possibly explain this as socio-economic variables are closely related to each other. Population growth may be related to this: in the Mau forest complex population increased enormously in the last few decades (see Appendix F). The results of the overall model revealed this as well: the higher the growth of population density, the more likely deforestation occurs. When more people need forest products and smallholder agricultural products for their livelihood, or need to be better connected to the larger urban areas, this may lead to deforestation.

5.3.2 Smallholder agricultural expansion

Overall, in the agricultural expansion model, many variables are not significantly contributing to the land use changes (see table 4.14). In the second period, mainly biophysical drivers explain expansion, whereas in the other periods socio-economic factors are important. More exactly, the biophysical factors that were of influence in the second time period were the drainage factors. Curvature and the wetness index have a large effect on smallholder agricultural expansion. The more convex (curve outward) the surface and the wetter the area, the higher the chance of agricultural expansion in the period 1984-1994.

Another interesting outcome is that the models showed that distance to roads becomes more important in the period 1994-2003 and 2003-2013 which is consistent with a study of Were et al. (2014) in the Eastern Mau forest reserve and Lake Nakuru drainage basin. The influence of this driver can probably be attributed to the changes in Narok county. The analysis of the spatial pattern of the land use changes, showed that in particular Narok county experienced a high degree of agricultural expansion at the expense of forest and rangeland (Southern part study area see figure 4.4), and that the county was badly accessible in terms of road infrastructure (see figure E11 in Appendix E). An outcome that may be related to population growth as well: the model of the overall period revealed that population density is one of the main drivers of agricultural expansion, and in particular Narok county experienced the highest growth in population density in the period 1973-2013 (an increase between 250% and 300%, see figure E15 in Appendix E).

Furthermore, a similar result to the forest conversion model is that the socio-economic factors become more important over time: in the period 2003-2013 all socio-economic factors are of significant influence on agricultural expansion. Another resemblance is that in the agricultural expansion model of the overall period a significant influence of population density was detected, whereas the models of the intermediate periods did not reveal that trend. The effect of this variable may be stronger in the overall period because land use changes in the period 1973-2013 are larger.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

The main land use changes in the Mau forest complex in the period 1973-2013 were a loss of forest and rangeland, while smallholder agriculture expanded. To be more precise, the foremost land use change in the study area was a conversion from forest to smallholder agriculture. Hence, smallholder agriculture can be considered the most important proximate driver of deforestation in the Mau forest complex in every time period analysed. The conversions occurred around the borders of the forest blocks, with the northern parts of county Narok experiencing most changes.

Based on the accuracy assessment and land use change dynamics analysis two land use change models were identified: forest conversion and smallholder agricultural expansion. The regression analysis showed that biophysical and socio-economic factors were significant driving forces in both models. Underlying drivers such as aspect East, curvature, the topographical wetness index, population density change, distance to towns and distance to roads increased the odds of forest conversion, and in particular the distance variables became more important in more recent time periods (1994-2003 and 2003-2013). In the agricultural expansion model, biophysical factors mainly had an influence in the second time period (1984-1994), while the socio-economic underlying drivers were more important in the third and fourth time period (1994-2003 and 2003-2013). Especially in county Narok a larger distance to roads resulted in an increase of the likelihood of agricultural expansion. In the overall period 1973-2013, both the forest conversion and agricultural expansion model showed that a growth in population density increased the chance of land use change.

6.2 RECOMMENDATIONS

Several recommendations can be formulated that would result in better outcomes or improve future research. First, to reduce uncertainty in data, in particular with regard to the classified land use data, a pixel based classification method, and thereby pixel based land use maps, would be more suitable in a land use change driver analysis. In this study a pixel level analysis was not possible due to the nature of the land use maps: object based classification maps. Because of the use of land use segments (polygons) some of the data of the independent variables with higher resolutions was lost: values of central tendency had to be calculated per land use polygon. Almost all the independent variables were of raster format and therefore pixel based land use maps would have reduced the loss of data of the independent variables. In addition, if a pixel based classification was used, a sample could have been taken for the regression modelling, which would have reduced the spatial autocorrelation of the data as well.

Actually, a recommendation for this particular research would be to start over: produce new land use maps. Carefully consider the remote sensing images that will be used: examine seasonality, spatial and temporal resolutions. For better data quality, and for better driver analysis results, possibly adjust the time period of the research: study a 20-year or 30-year period instead of a 40-year period. This study showed that in particular in earlier periods there are difficulties to obtain high quality data, and therefore studying a more recent and smaller time period results in more reliable independent variables. Moreover, consider the conditions of this specific area and the possible difficulties for classifying (e.g. the mixing of rangeland and smallholder agricultural land use classes). Last, use a pixel based classification method and create new land use maps, apply a statistically sound accuracy assessment where both sampling scheme and size are well-thought-out, and organise a fieldwork close in time to the reference points.

In addition, incorporating other variables in the models would improve results. To do this more time is necessary, because more and better resolution spatial data is needed. Also incorporating variables that consist of multi-scale characteristics is important. Include drivers that operate over larger distances, on higher scales: not only local, but consider regional, national and international forces that may influence land use change. For example, economic demands from outside the country, such as the export of tea products (because the tea industry is an important economic sector in the study area), or on a national level the gross domestic product of the population. Yet important is that this data should be of a high spatial resolution for the results to be significant.

Moreover, another regression method could be applied. For example, multinomial regression modelling considers all land use changes. It uses a categorical dependent variable instead of a binary dependent variable. This research showed poorly classified land use classes - only the forest and smallholder agriculture class could be used for analysis - and therefore this option could not be tested. However future research might benefit from these methods (see Lin et al., 2014).

Finally, a qualitative assessment is highly recommended: visit the area, conduct a fieldwork and interviews, and try to obtain expert knowledge. It is essential to have knowledge of the context of an area because in this way it is easier to identify land use change processes and helps to select and decide what factors might contribute to land use change. A qualitative assessment may help identifying other drivers, and the relevance of the independent variables can be checked based on interviews and field observations. In this way cultural and political drivers can be assessed as well. Also the evaluation and interpretation of the outcome of a research is challenging without visiting an area.

In conclusion, this research shows the importance of assessing data quality in land use change driver analysis and demonstrates that land use change and its drivers show different spatial-temporal trends (see figure 4.1 and 4.2 for land use changes, figure 4.4 for spatial patterns of the changes, and figures 4.10 and 4.11 for the drivers per time period). The work provides an overview of spatial-temporal knowledge of the drivers of land use change in the Mau forest complex, which is essential for national policy making and for mitigation policy programs such as REDD+. In addition, the models that were developed function as a good base for other land use change driver analysis. The study reveals that in a 40-year period the socio-economic drivers increase in importance, which means there is a need for better understanding the socio-economic aspects behind land use change. Therefore, future research and policy should be time and space specific and focus more on the socio-economic drivers of land use change.

REFERENCES

Aguiar, A. P. D., Câmara, G., & Escada, M. I. S. (2007). Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. Ecological Modelling, 209(2), 169-188.

Allison, P. (2013). What's the Best R-Squared for Logistic Regression?. Statistical Horizons. Accessed via: http://statisticalhorizons.com/r2logistic

Alexander, P., Rounsevell, M. D., Dislich, C., Dodson, J. R., Engström, K., & Moran, D. (2015). Drivers for global agricultural land use change: the nexus of diet, population, yield and bioenergy. Global Environmental Change, 35, 138-147.

Atela, J. O., Denich, M., Kaguamba, R., & Kibwage, J. (2012). Agricultural land allocation in small farms around Maasai Mau forest, Kenya and the implications on carbon stocks. Journal of Ecology and the Natural Environment, 4(4), 98-108.

Baldyga, T. J., Miller, S. N., Driese, K. L., & Gichaba, C. M. (2008). Assessing land cover change in Kenya's Mau Forest region using remotely sensed data. *African Journal of Ecology*, *46*(1), 46-54.

Baxter, J. (2014). Researchers study future of Kenya's 'last big mountain forest'. Accessed via: http://blog.cifor.org/25664/kenya-mau-forest-water-study-climate-change?fnl=en

Braimoh, A. K., & Onishi, T. (2007). Spatial determinants of urban land use change in Lagos, Nigeria. Land Use Policy, 24(2), 502-515.

Buckley, A. (2010). ESRI ArcGIS Blog. Understanding curvature rasters. Accessed via: https://blogs.esri.com/esri/arcgis/2010/10/27/understanding-curvature-rasters/

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

Congalton, R. G., & Green, K. (2008). Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

DeFries, R. S., Rudel, T., Uriarte, M., & Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*, *3*(3), 178-181.

ESRI (2016). ArcGIS Pro Tool Reference, Spatial Autocorrelation (Global Moran's I). Accessed via: http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/spatial-autocorrelation.htm

FAO (2005). Food and Agricultural Organisation of the United Nations. World Programme for the Census of Agriculture 2010. A system of integrated agricultural censuses and surveys, Volume 1. Rome, Italy.

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

Geist, H. J., & Lambin, E. F. (2002). Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *BioScience*, *52*(2), 143-150.

Gruber, S., & Peckham, S. (2009). Land-surface parameters and objects in hydrology. In: Hengl, T. & Reuter, H.I. (eds.) 2009. Geomorphometry. Concepts, Software, Applications. Developments in Soil Science, Volume 33. Elsevier: 171-194

Hansen, M. C., Stehman, S. V., & Potapov, P. V. (2010). Quantification of global gross forest cover loss. Proceedings of the National Academy of Sciences, 107(19), 8650-8655.

Hassan, M. (2013) District Creation in Kenya Under President Moi. Department of Political Science, University of Michigan, USA.

Herold, M., & Skutsch, M. (2011). Monitoring, reporting and verification for national REDD+ programmes: two proposals. Environmental Research Letters, 6(1), 014002.

Hosonuma, N., Herold, M., De Sy, V., De Fries, R. S., Brockhaus, M., Verchot, L., Angelsen, A. & Romijn, E. (2012). An assessment of deforestation and forest degradation drivers in developing countries. Environmental Research Letters, 7(4), 044009.

Houghton, R. A. (2012). Carbon emissions and the drivers of deforestation and forest degradation in the tropics. Current Opinion in Environmental Sustainability, 4(6), 597-603.

Huang, Q. H., Cai, Y. L., & Peng, J. (2007). Modeling the spatial pattern of farmland using GIS and multiple logistic regression: a case study of Maotiao River Basin, Guizhou Province, China. Environmental Modeling & Assessment, 12(1), 55-61.

IPCC (2000). Intergovernmental Panel on Climate Change. Special Report on Land-Use, Land-Use Change, and Forestry. Cambridge University Press.

Institute for Digital Research and Education (IDRE) (2011). What are pseudo R-Squareds? UCLA: Statistical Consulting Group. accessed via: http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm/

International Livestock Research Institute (ILRI) (2000): GIS database. ILRI towns dataset: 1620 towns digitized by International Livestock Research Institute, ILRI, GIS group from Kenya topographic sheets scale 1:250,000 for Northern Kenya and 1:50,000 for rest of Kenya from Survey of Kenya. Created in August 2000.

Jaimes, N. B. P., Sendra, J. B., Delgado, M. G., & Plata, R. F. (2010). Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. Applied Geography, 30(4), 576-591.

Jenness, J. (2007): Some Thoughts on Analyzing Topographic Habitat Characteristics. USA. Accessed via: http://www.jennessent.com/downloads/topographic_analysis_online.pdf

Kariuku, D.P. (1989). Theileriosis in Kenya via FAO Corporate Document Repository. National Veterinary Research Centre, Kikuyu, Kenya.

Kenya Forest Service (KFS) (2013). Report Ministry of Forest and Wildlife Kenya,. Analysis of drivers and underlying causes of forest cover change in the various forest types in Kenya. July 2013: Nairobi, Kenya.

Kenya National Bureau of Statistics (KNBS) (1979), 1979 Population Census - Volume II - Analytical Report, 1979

Kenya National Bureau of Statistics (KNBS) (1989a), Kenya Population Census, 1989 - Volume I, Central Bureau Of Statistics, March 1994

Kenya National Bureau of Statistics (KNBS) (1989b), Kenya Population Census 1989 - Analytical Report Volume III - Population Dynamics Of Kenya (Central Bureau Of Statistics, April 1996)

Kenya National Bureau of Statistics (KNBS) (1999), 1999 Population And Housing Census - Volume I - Population Distribution By Administrative Units, January 2001

Kenya National Bureau of Statistics (KNBS) (2009a). The 2009 Kenya Population and Housing Census – Volume IA – Population Distribution by Administrative Units, August 2010.

Kenya National Bureau of Statistics (KNBS) (2009b), The 2009 Kenya Population And Housing Census - Volume IB - Population Distribution By Political Units, August 2010

Kenya National Bureau of Statistics (KNBS) (2015). Statistical Abstract 2015. Nairobi.

Kenya Wildlife Service (KWS) (2009). Report of the Prime Minister's Task Force on the Conservation of The Mau Forests Complex. March 2009: Nairobi, Kenya.

Kiage, L. M., Liu, K. B., Walker, N. D., Lam, N., & Huh, O. K. (2007). Recent land-cover/use change associated with land degradation in the Lake Baringo catchment, Kenya, East Africa: evidence from Landsat TM and ETM+. International Journal of Remote Sensing, 28(19), 4285-4309.

Kissinger, G., Herold, M., & De Sy, V. (2012). Drivers of deforestation and forest degradation: a synthesis report for REDD+ policymakers. Lexeme Consulting, Vancouver Canada, August 2012.

Klopp, J. M., & Sang, J. K. (2011). Maps, power, and the destruction of the Mau Forest in Kenya. Georgetown Journal of International Affairs, 125-134.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., ... & Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change*, *11*(4), 261-269.

Lambin, E.F. & Geist, H.J. (2006). Land-Use and Land-Cover Change. Local Processes and Global Impacts. Berlin: Springer.

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., ... & Woodward, F. I. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, *2*(12), 831-836.

Lin, Y., Deng, X., Li, X., & Ma, E. (2014). Comparison of multinomial logistic regression and logistic regression: which is more efficient in allocating land use?. Frontiers of Earth Science, 8(4), 512-523.

Long, H., Tang, G., Li, X., & Heilig, G. K. (2007). Socio-economic driving forces of land-use change in Kunshan, the Yangtze River Delta economic area of China. Journal of Environmental Management, 83(3), 51-364.

Mengel, D.B., 1993. Fundamentals of soil cation exchange capacity (CEC). Agronomy guide AY-238. Cooperative Extension Service, Purdue University, West Lafayette, USA.

Millington, J. D., Perry, G. L., & Romero-Calcerrada, R. (2007). Regression techniques for examining land use/cover change: a case study of a Mediterranean landscape. Ecosystems, 10(4), 562-578.

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011). Global biodiversity conservation: the critical role of hotspots. In Biodiversity hotspots (pp. 3-22). Springer Berlin Heidelberg.

Mottet, A., Ladet, S., Coqué, N., & Gibon, A. (2006). Agricultural land-use change and its drivers in mountain landscapes: a case study in the Pyrenees. *Agriculture, ecosystems & environment*, *114*(2), 296-310.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403(6772), 853-858.

Muandu, P. M. (1999) Traditional Food Plants of Kenya. National Museum of Kenya, Nairobi, Kenya.

Mutoko, M. C., Hein, L., & Bartholomeus, H. (2014). Integrated analysis of land use changes and their impacts on agrarian livelihoods in the western highlands of Kenya. *Agricultural Systems*, *128*, 1-12.

NASA (2011). Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Map. Accessed via: https://asterweb.jpl.nasa.gov/gdem.asp

Olang, L.O. & Kundu, P.M. (2011). Land Degradation of the Mau Forest Complex in Eastern Africa: A Review for Management and Restoration Planning, Environmental Monitoring, InTech, Accessed via: http://www.intechopen.com/books/environmental-monitoring/land-degradation-of-the-mau-forest-complex-ineastern-africa-a-review-for-management-and-restoration

Olaya, V. (2009). Basic Land-Surface Parameters. In: Hengl, T. & Reuter, H.I. (eds.) 2009. Geomorphometry. Concepts, Software, Applications. Developments in Soil Science, Volume 33. Elsevier

Overmars, K. P., De Koning, G. H. J., & Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. Ecological Modelling, 164(2), 257-270.

Ott, R. L., & Longnecker, M. T. (2010). An Introduction to Statistical Methods and Data Analysis, Sixth Edition. Texas A&M University. Brooks/Cole, Cengage Learning.

Reuter, H. I., Hengl, T., Gessler, P., & Soille, P. (2009). Preparation of DEMs for geomorphometric analysis. In: Hengl, T. & Reuter, H.I. (eds.) 2009. Geomorphometry. Concepts, Software, Applications. Developments in Soil Science, Volume 33. Elsevier: 87-120.

Rossiter, D. G., & Loza, A. (2012). Analyzing land cover change with logistic regression in R. University of Twente, Faculty of Geo-Information Science & Earth Observation (ITC), Enschede, The Netherlands.

Rutherford, G. N., Bebi, P., Edwards, P. J., & Zimmermann, N. E. (2008). Assessing land-use statistics to model land cover change in a mountainous landscape in the European Alps. Ecological Modelling, 212(3), 460-471.

Sena, K. (2006). Mau forest: Killing the goose but still wanting the golden eggs. Indigenous affairs, (4), 30-35.

Serneels, S., & Lambin, E. F. (2001). Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. Agriculture, Ecosystems & Environment, 85(1), 65-81.

Serra, P., Pons, X., & Saurí, D. (2008). Land-cover and land-use change in a Mediterranean landscape: a spatial analysis of driving forces integrating biophysical and human factors. Applied Geography, 28(3), 189-209.

Seto, K. C., & Kaufmann, R. K. (2003). Modeling the drivers of urban land use change in the Pearl River Delta, China: integrating remote sensing with socioeconomic data. Land Economics, 79(1), 106-121.

Sloan, S., Jenkins, C. N., Joppa, L. N., Gaveau, D. L., & Laurance, W. F. (2014). Remaining natural vegetation in the global biodiversity hotspots. Biological Conservation, 177, 12-24.

Sørensen, R., Zinko, U., & Seibert, J. (2005). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences Discussions*, *2*(4), 1807-1834.

Statoids (2013). Administrative Divisions of Countries. Districts of Kenya. Accessed via: http://www.statoids.com/yke.html

Story, M., & Congalton, R. G. (1986). Accuracy assessment - A user's perspective. Photogrammetric Engineering and remote sensing, 52(3), 397-399.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*, 234-240.

Trapman, C. (1974). Change in Administrative Structures. A Case Study of Kenyan Agricultural Development. Overseas Development Institute Ltd., London, UK.

United Nations Environment Programme (UNEP) (2006). Eastern and South West Mau Forest Reserves: Assessment and way forward.

United Nations Environment Programme (UNEP) (2011). Mau Ecosystem Rehabilitation Gains Momentum with Donor Support and UN Backing. November 2011. Nairobi.

United Nations Environment Programme (UNEP) (2012) .The role and Contribution of Montane Forests and Related Ecosystem Services to the Kenyan Economy. February 2012.

Verburg, P. H., Van Eck, J. R., de Nijs, T. C., Dijst, M. J., & Schot, P. (2004). Determinants of land- use change patterns in the Netherlands. Environment and Planning B, 31(1), 125-150.

Vita, L. (2014). Monitoring 40 Years of Land Use Change in The Mau Forest Complex, Kenya. Technical Report. Wageningen University and Research, The Netherlands.

Were, K. O., Dick, Ø. B., & Singh, B. R. (2013). Remotely sensing the spatial and temporal land cover changes in Eastern Mau forest reserve and Lake Nakuru drainage basin, Kenya. *Applied Geography*, *41*, 75-86.

Were, K., Dick, Ø. B., & Singh, B. R. (2014). Exploring the geophysical and socio-economic determinants of land cover changes in Eastern Mau forest reserve and Lake Nakuru drainage basin, Kenya. *GeoJournal*, *79*(6), 775-790.

Worldbank (2016). Climate Change Knowledge Portal. Average Monthly Temperature and Rainfall for Kenya from 1990-2012. Accessed via: http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisRegion=Africa&ThisCCode=KEN#

World Resources Institute (WRI) (2016). River network data. Accessed via: http://www.wri.org/publication/natures-benefits-kenya

Zar, J.H. (2010). Biostatistical Analysis (5th ed.). New Jersey: Prentice Hall Pearson.

*Figure cover page Mau forest (Kurosoi, South-West Mau): Rufino, M. (2014). CIFOR PowerPoint Presentation. Resilience, adaptation and mitigation in Kenya. Accessed via: http://blog.cifor.org/25664/kenya-mau-forest-water-studyclimate-change?fnl=en

APPENDICES

APPENDIX A – FIELDWORK PLAN

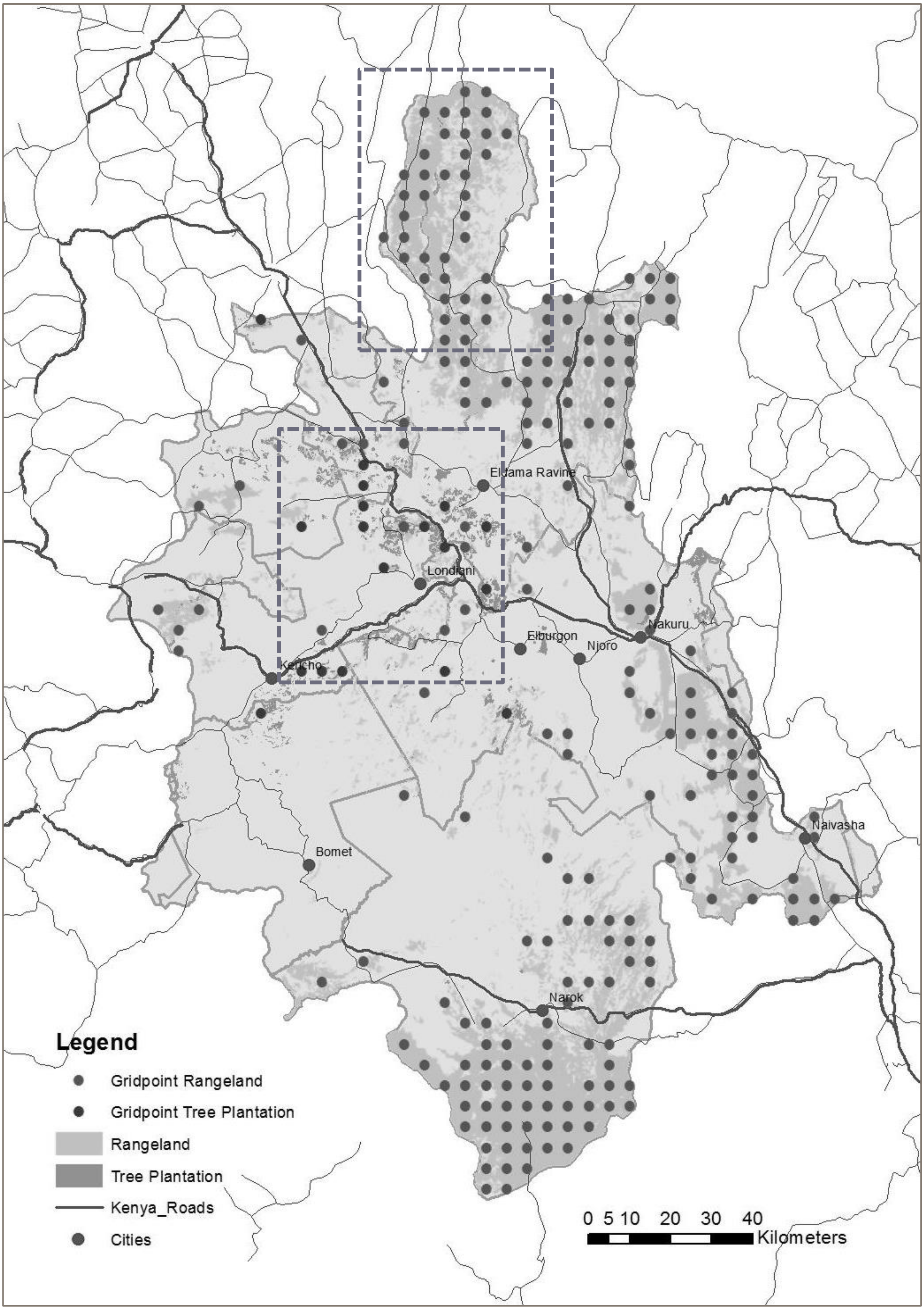


Figure A1: Map fieldwork plan

APPENDIX B – CLASSIFICATION SCHEME AFRICOVER

Table B1: Classification scheme Africover land use classes

AFRICOVER	AFRICOVER CLASS DESCRIPTION	LU CLASS	DESCRIPTION
2TC-B	Closed Trees - Bamboo	F	Forest
2TC8	Closed trees with shrubs	F	Forest
5UR	Rural settlements	I	Infrastructure
5Q	Quarry	I	Infrastructure
5A	Airport	I	Infrastructure
5U	Urban areas (general)	I	Infrastructure
HD57	Irrigated Herbaceous Crop, Large to Medium Fields	IC	Commercial Agriculture
HM57	Irrigated Herbaceous Crop, Medium Fields	IC	Commercial Agriculture
HD57-s	Irrigated Herbaceous Crop, Large to Medium Fields - Sugarcane	IC	Commercial Agriculture
HM57-s	Irrigated Herbaceous Crop, Medium Fields - Sugarcane	IC	Commercial Agriculture
2WP6	Open general woody with herbaceous	R	Rangeland
2TV28	Very open trees (broadleaved deciduous) with closed to open shrubs	R	Rangeland
2TP8	Open general trees with shrubs, woodland with shrubs	R	Rangeland
2WC7	Closed woody with sparse trees	R	Rangeland
2TO28	Open trees (broadleaved deciduous) with closed to open shrubs	R	Rangeland
2TO268	Open trees (broadleaved deciduous) with closed to open herbaceous and sparse shrubs	R	Rangeland
2TC8	Closed trees with shrubs	F	Forest
2TV268	Very open trees (broadleaved deciduous) with closed to open herbaceous and sparse shrubs	R	Rangeland
GRZ-r	Cultivated Aquatic or Regularly Flooded Areas, Cereals, Rice - Small, Medium, Large Fields	IC	Commercial Agriculture
2SP6	Open general shrubs with closed to open herbaceous	R	Rangeland
2SOJ67	Open shrubs with closed to open herbaceous and sparse trees	R	Rangeland
2SVJ67	Very open shrubs with closed to open herbaceous and sparse trees	R	Rangeland
2SV6	Very open shrubs with closed to open herbaceous	R	Rangeland
2SCJ7	Closed shrubs with sparse trees	R	Rangeland
2SCJ	Closed shrubs	R	Rangeland
2H(CP)	Closed to very open herbaceous	R	Rangeland
2H(CP)8	Closed to very open herbaceous with sparse shrubs	R	Rangeland
2H(CP)78	Closed to very open herbaceous with sparse trees and shrubs	R	Rangeland
SR47V-t	Rainfed Shrub Crop, Small Fields - Tea	SA	Smallholder Agriculture
SR47V	Rainfed Shrub Crop, Small Fields	SA	Smallholder Agriculture
HM4	Rainfed Herbaceous Crop, Medium Fields	SA	Smallholder Agriculture
HD4-w	Rainfed Herbaceous Crop, Large to Medium Fields - Wheat	SA	Smallholder Agriculture
HD4	Rainfed Herbaceous Crop, Large to Medium Fields	SA	Smallholder Agriculture
HL4-w	Rainfed Herbaceous Crop, Large Fields - Wheat	SA	Smallholder Agriculture

HL4	Rainfed Herbaceous Crop, Large Fields	SA	Smallholder Agriculture
TR47V	Rainfed Tree Crop - Small Fields	SA	Smallholder Agriculture
HR4	Rainfed Herbaceous Crop, Small Fields	SA	Smallholder Agriculture
HR57-s	Irrigated Herbaceous Crop, Small Fields - Sugarcane	SA	Smallholder Agriculture
SL47V-c	Rainfed Shrub Crop, Large Fields - Coffee	SA	Smallholder Agriculture
SL47V-t	Rainfed Shrub Crop, Large Fields - Tea	TE	Tea Estate
TL47PL	Forest Plantation - Large Fields	TP	Tree Plantation
8WP	Natural Lakes	WB	Water
7WP	Artificial Lakes or Reservoirs	WB	Water
8WN2	Lake shore	WB	Water
4H(CP)F8	Closed to very open herbaceous with sparse shrubs on temporarily flooded land - fresh water	WL	Wetland
4H(CP)FF	Closed Herbaceous (on permanently flooded land - Fresh Water)	WL	Wetland
4HCF	Closed herbaceous on temporarily flooded land - fresh water	WL	Wetland
4TPF6	Open general trees with closed to open herbaceous on temporarily flooded land - fresh water	WL	Wetland
4WPF6	Open general woody with closed to open herbaceous on temporarily flooded land - fresh water	WL	Wetland

APPENDIX C – VALIDATION GRID

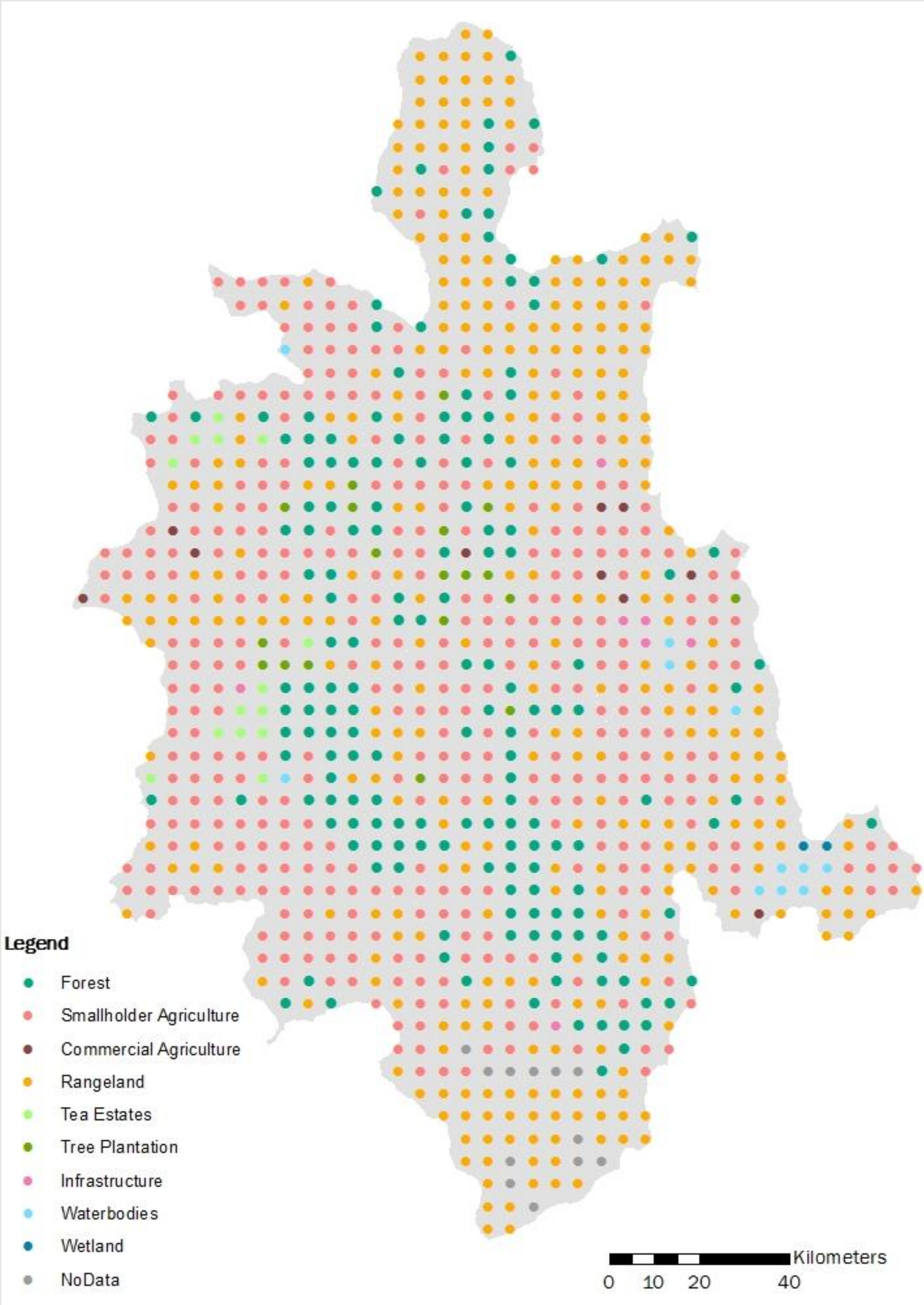


Figure C1: Validation grid Mau Forest Complex

APPENDIX D – UNDERLYING DRIVERS

Table D1: Drivers of land use change identified in literature*

DRIVER GROUP	INDEPENDENT VARIABLE
ENVIRONMENTAL/BIOPHYSICAL	elevation
	slope
	aspect
	curvature
	soil type
	soil quality
	soil pH
	soil CEC
	average humidity
	rainfall
	temperature
	geological (rock type, sediment type)
	erosion (degree of erosion)
	topographical wetness index
ECONOMIC	GDP per capita
	annual income/GDP growth
	market growth
	poverty
	accessibility of the area/to markets
	distance to roads
	distance to urban centres
	distance to rivers/permanent water
	travel time centre/harbour
DEMOGRAPHIC	population growth/density
	urban population growth
	rural population growth
	migration
TECHNOLOGICAL	net agricultural trade per capita
	% of agricultural production exported
	production animal products
	bioenergy
	agro-technological change
	technological level of farmers
CULTURAL	public attitudes
	values and beliefs
INSTITUTIONAL	percentage of protected areas
	change in land tenure
	reform measures/policy change

* Aguiar et al. (2007), Alexander et al. (2015), Braimoh and Onishi (2007), DeFries et al. (2010), Geist and Lambin (2002), Huang et al. (2007); Jaimes et al. (2010), Kiage et al. (2007), Kissinger et al. (2012), Long et al. (2007), Mottet et al. (2006), Mutoku et al (2014), Serra et al. (2008), Serneels and Lambin (2001), Seto and Kaufmann (2003), Verburg et al. (2002), Were et al. (2014)

APPENDIX E – MAPS INDEPENDENT VARIABLES

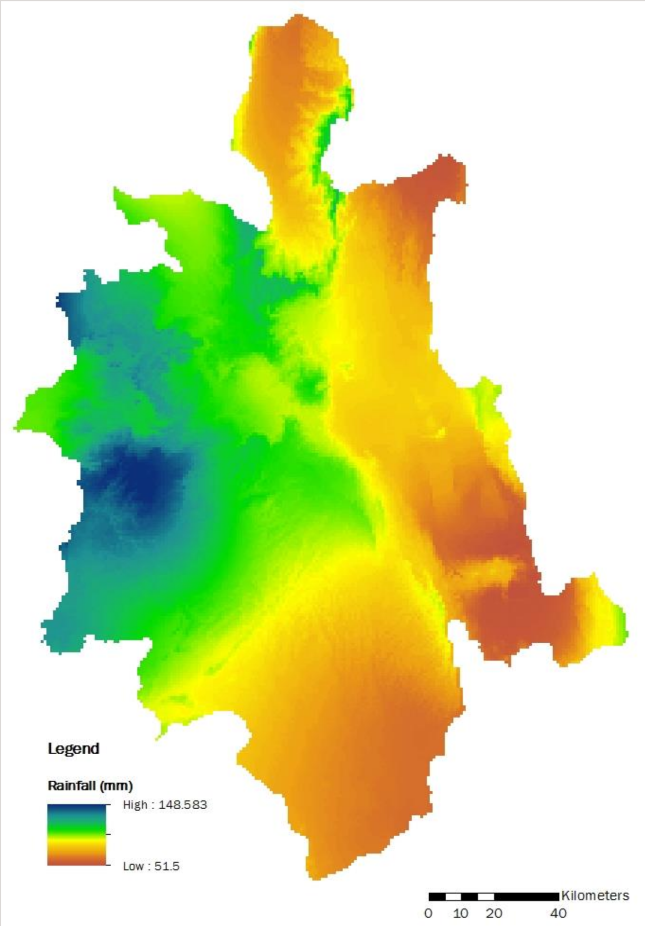


Figure E1: Rainfall

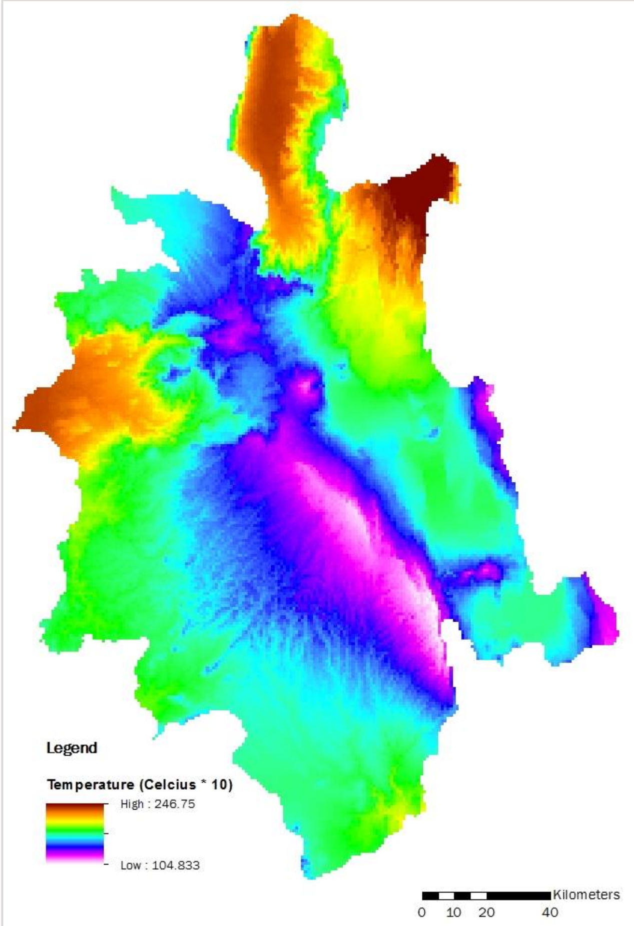


Figure E2: Temperature

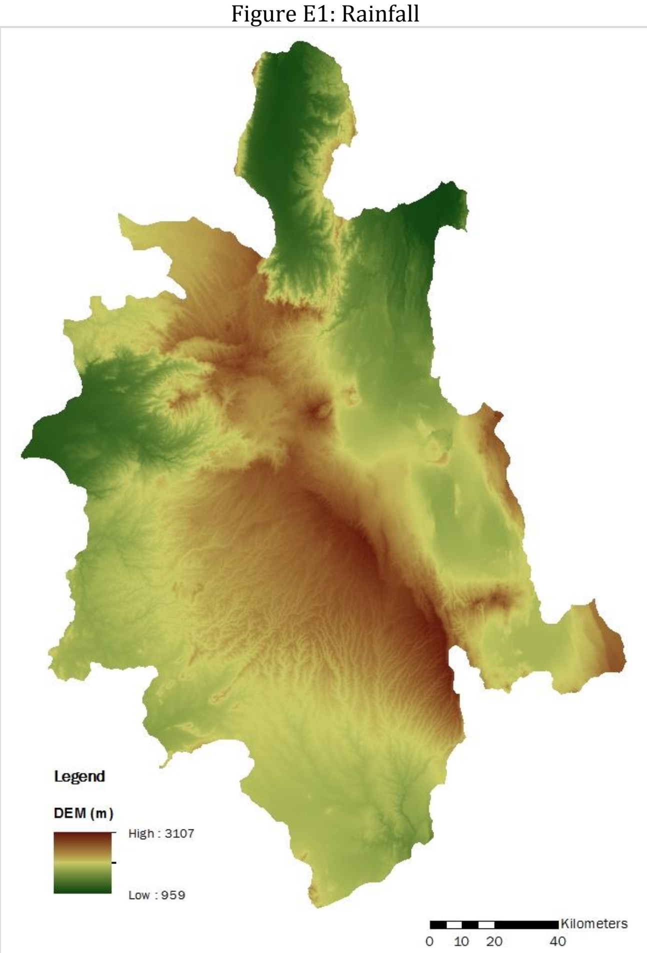


Figure E3: DEM

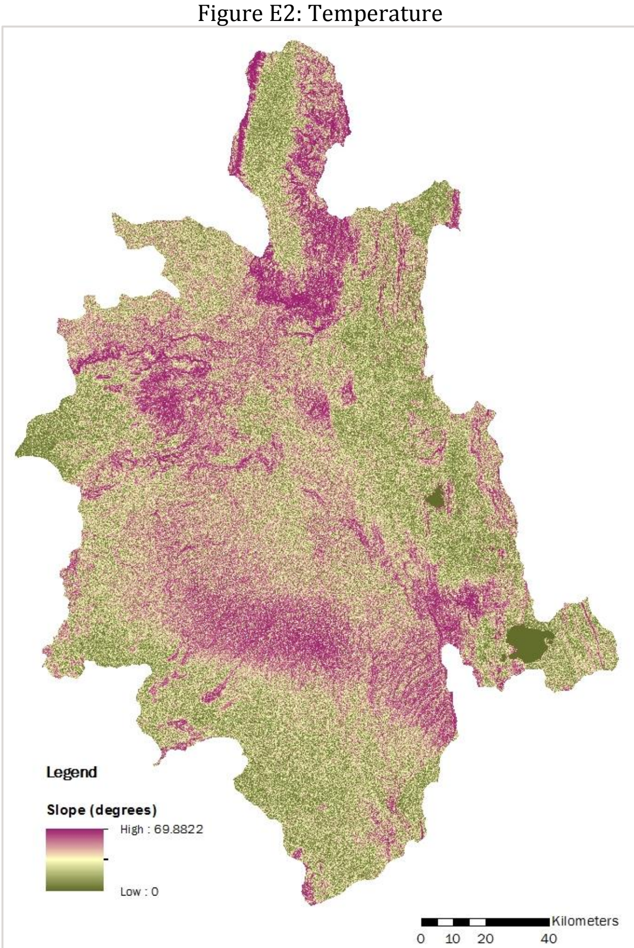


Figure E4: Slope

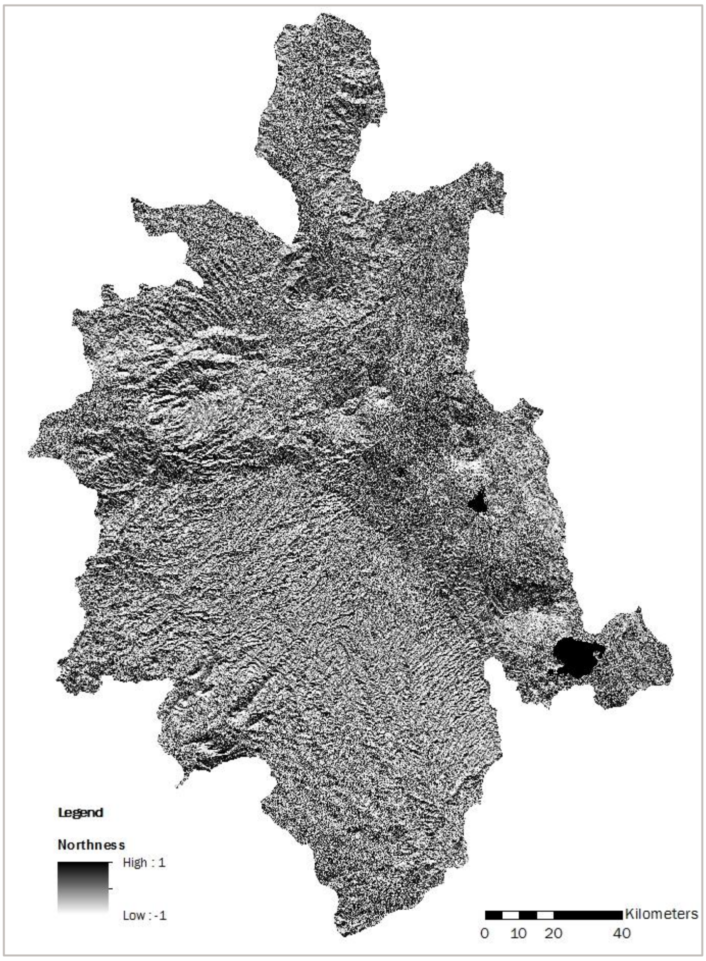


Figure E5: Aspect north

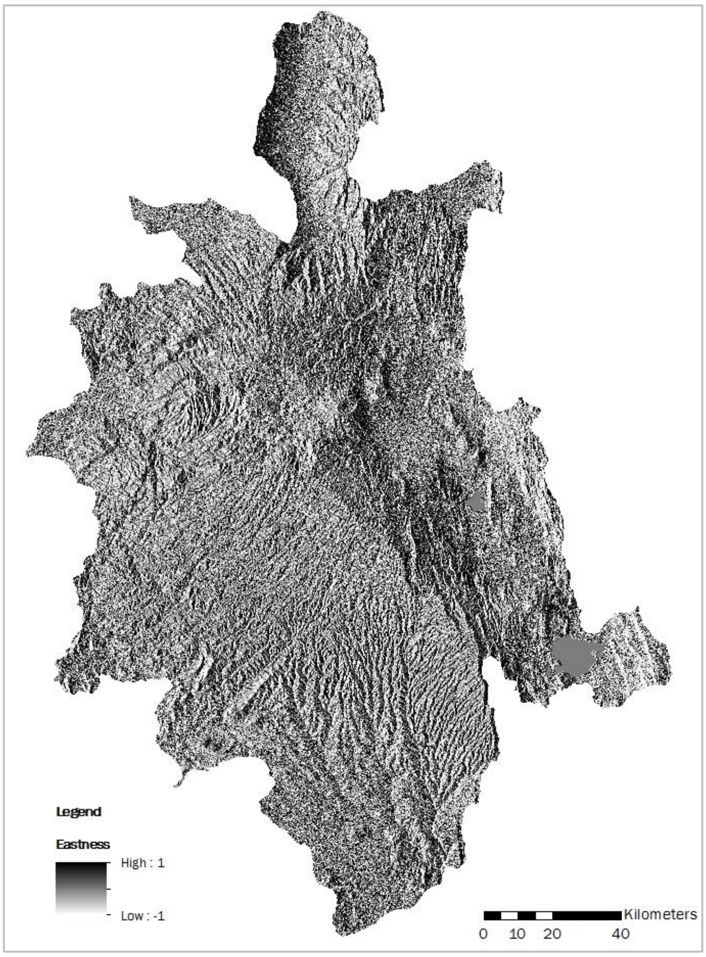


Figure E6: Aspect east

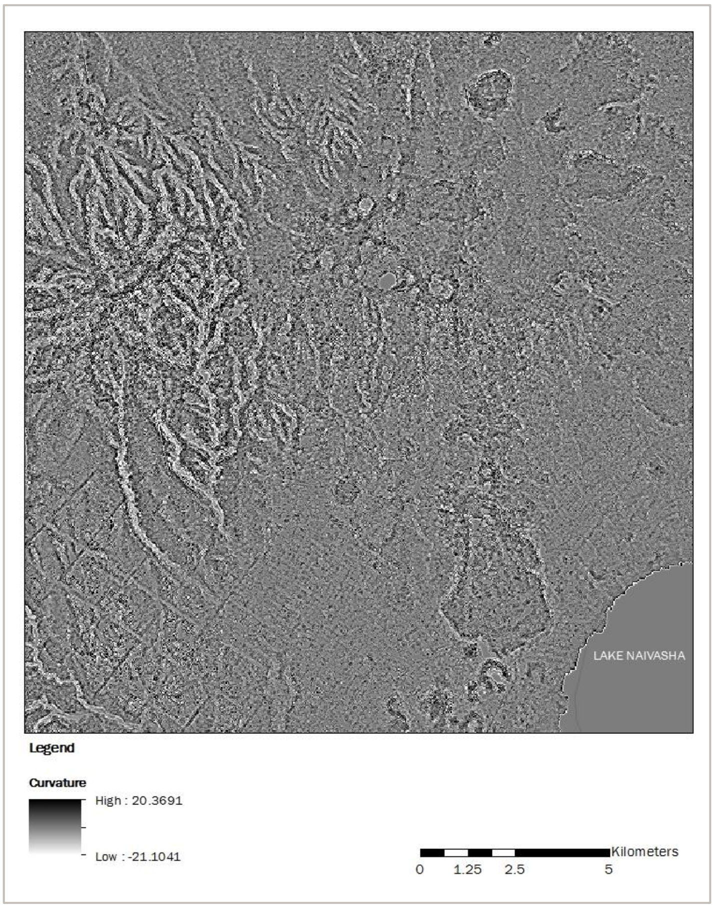


Figure E7: Curvature

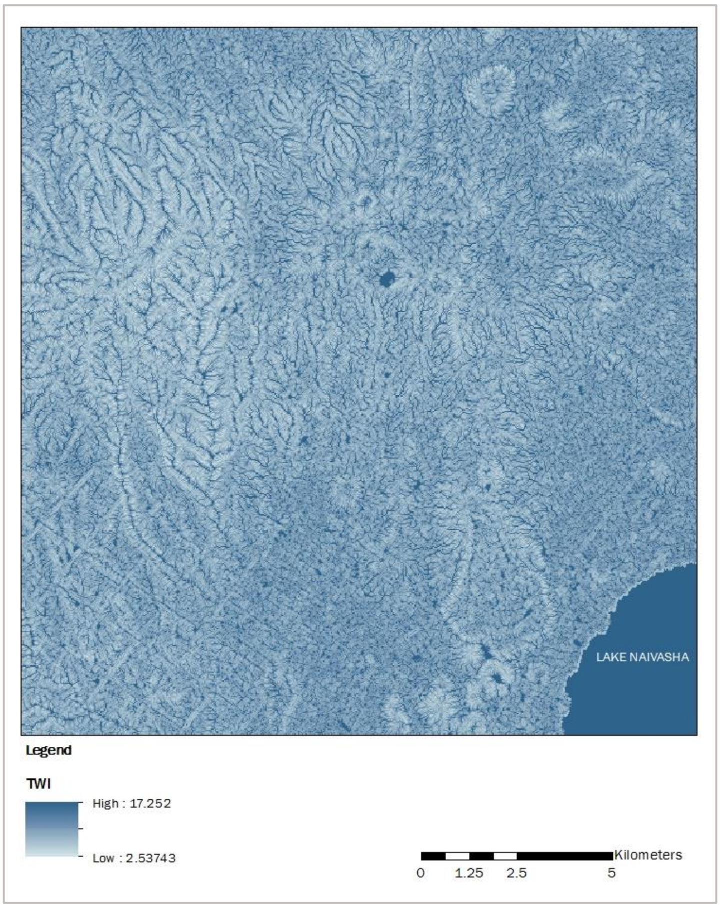


Figure E8: TWI

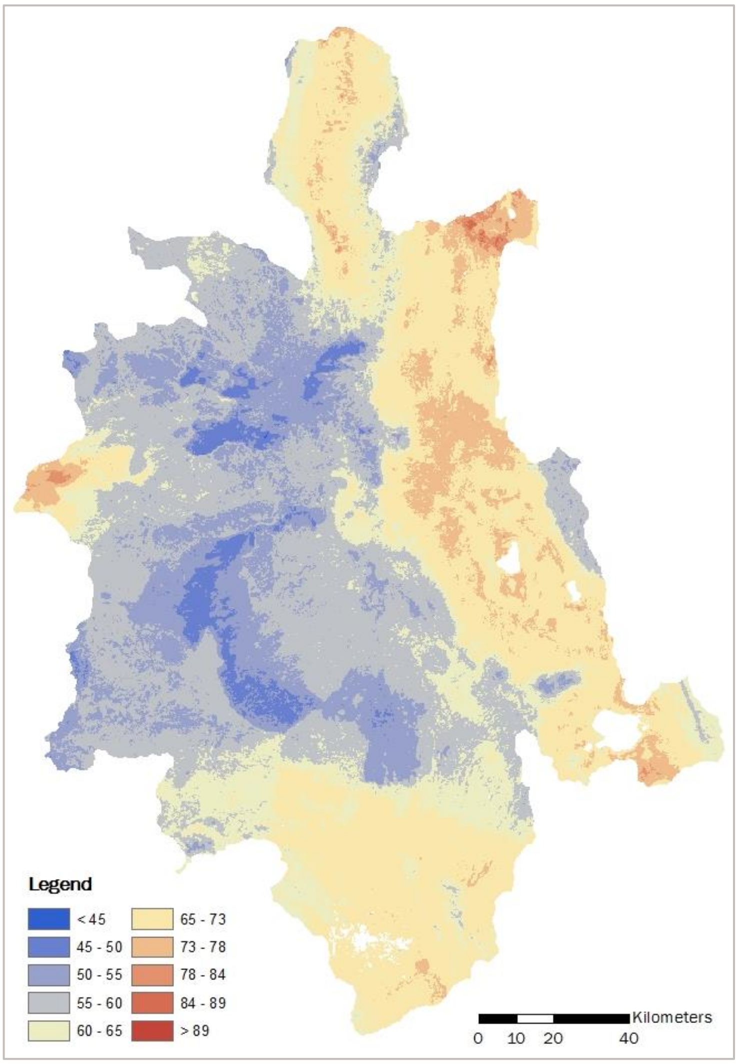


Figure E9: Soil pH

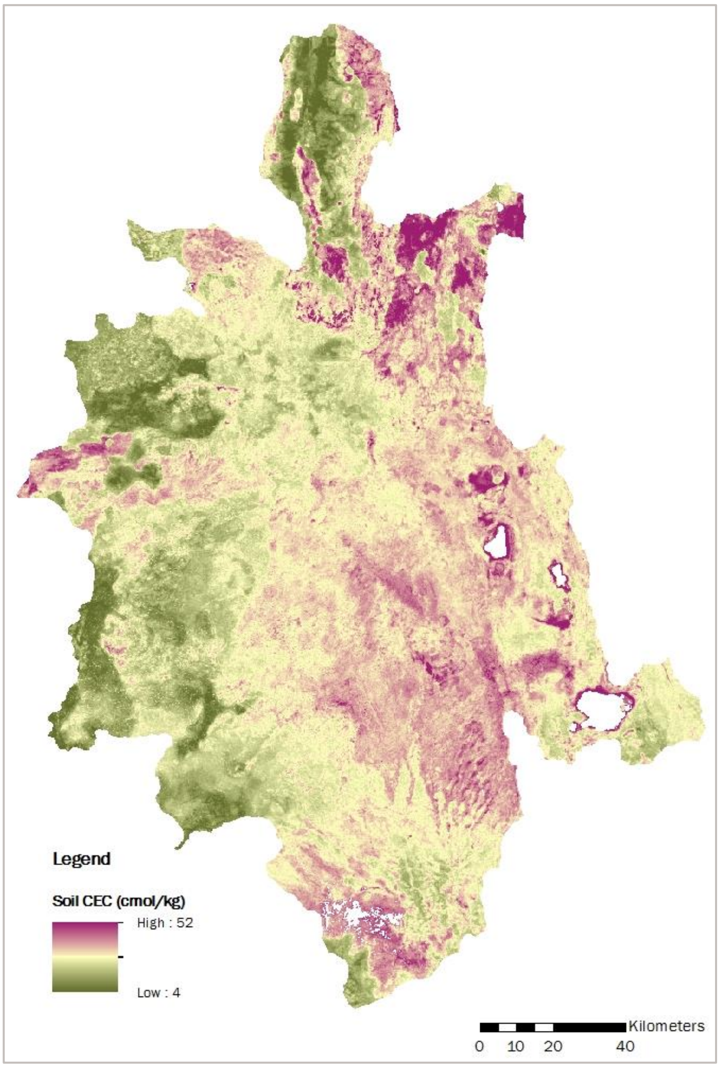


Figure E10: Soil CEC

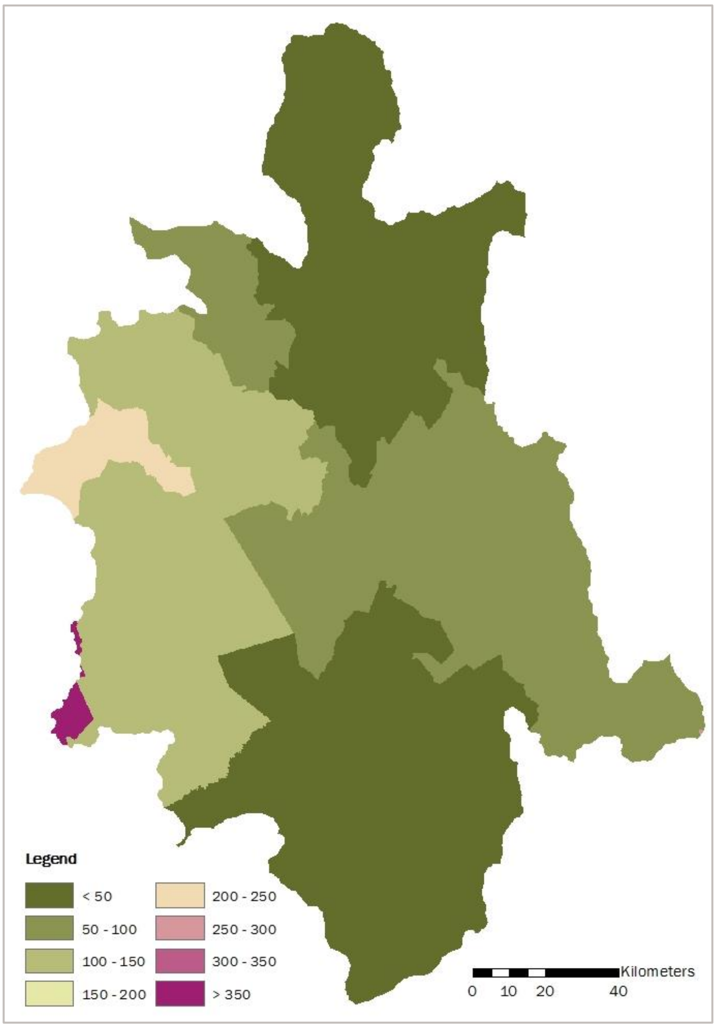


Figure E11: Population density 1979

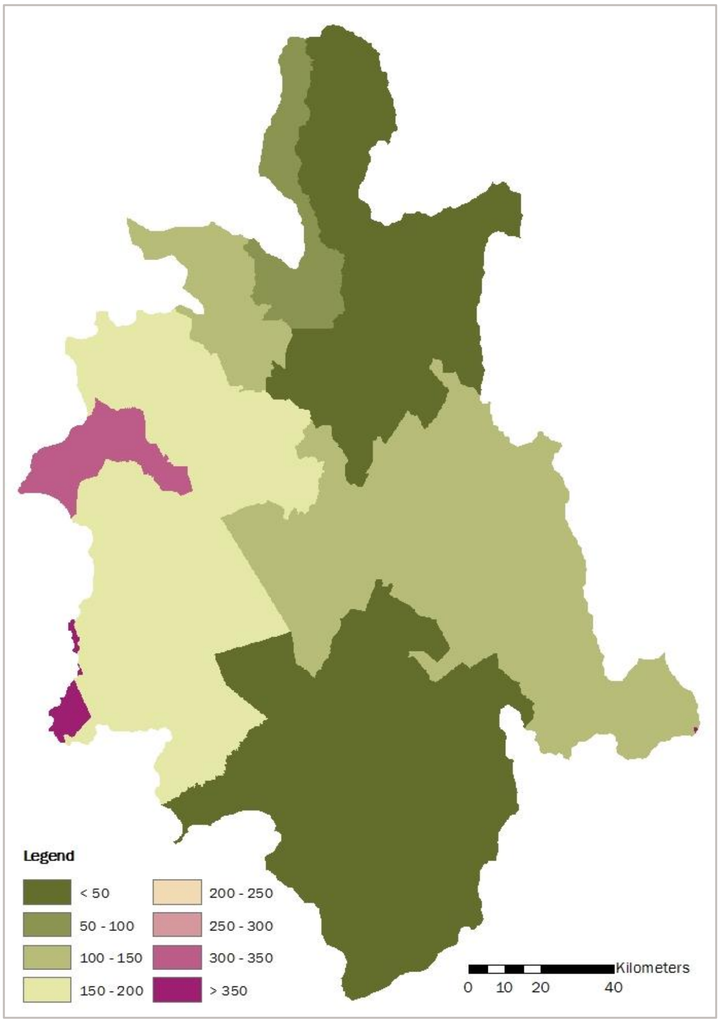


Figure E12: Population density 1989

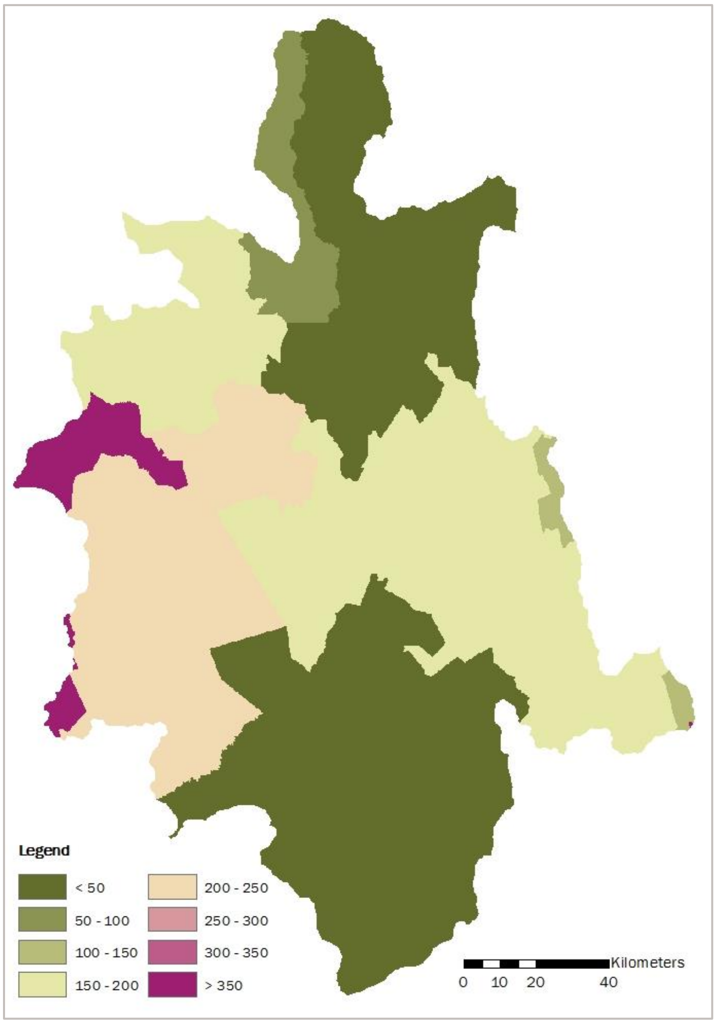


Figure E13: Population density 1999

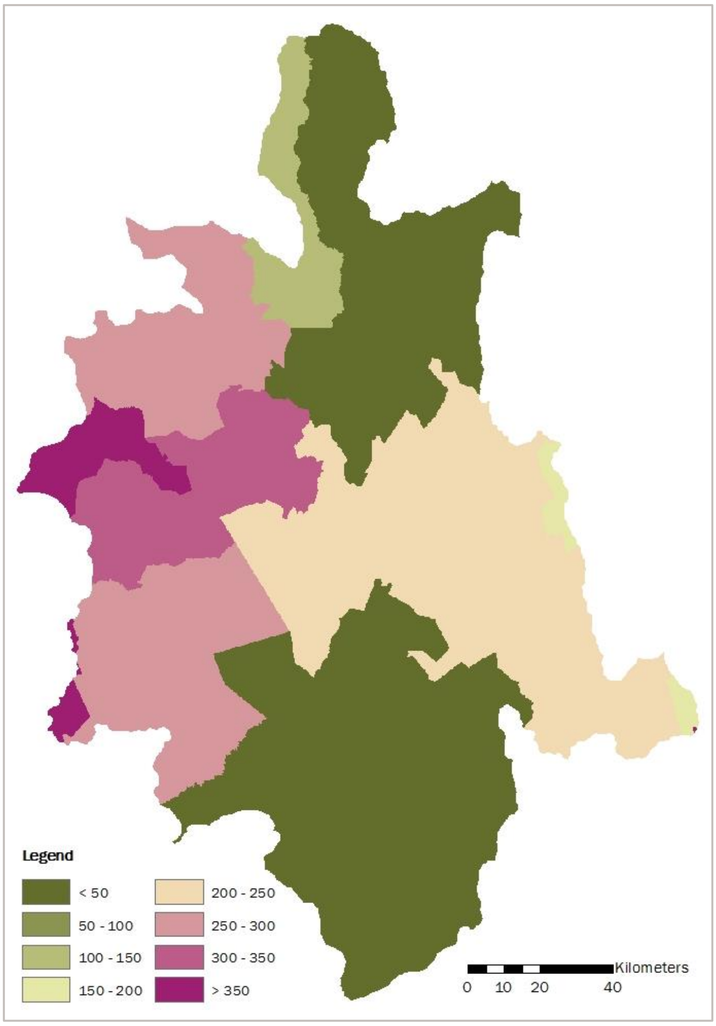


Figure E14: Population density 2009

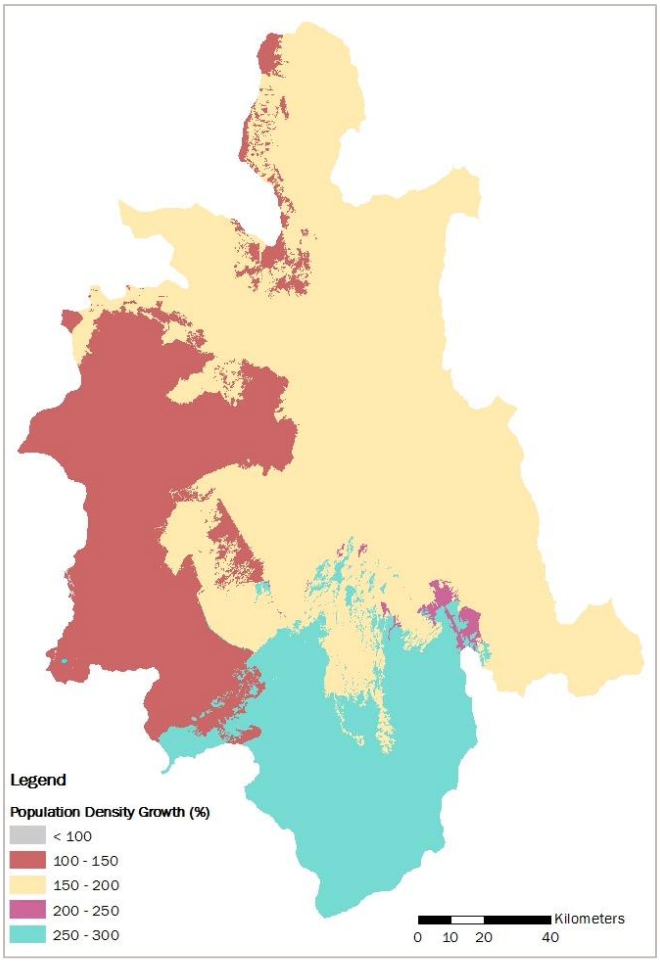


Figure E15: Population density growth 1973 - 2013

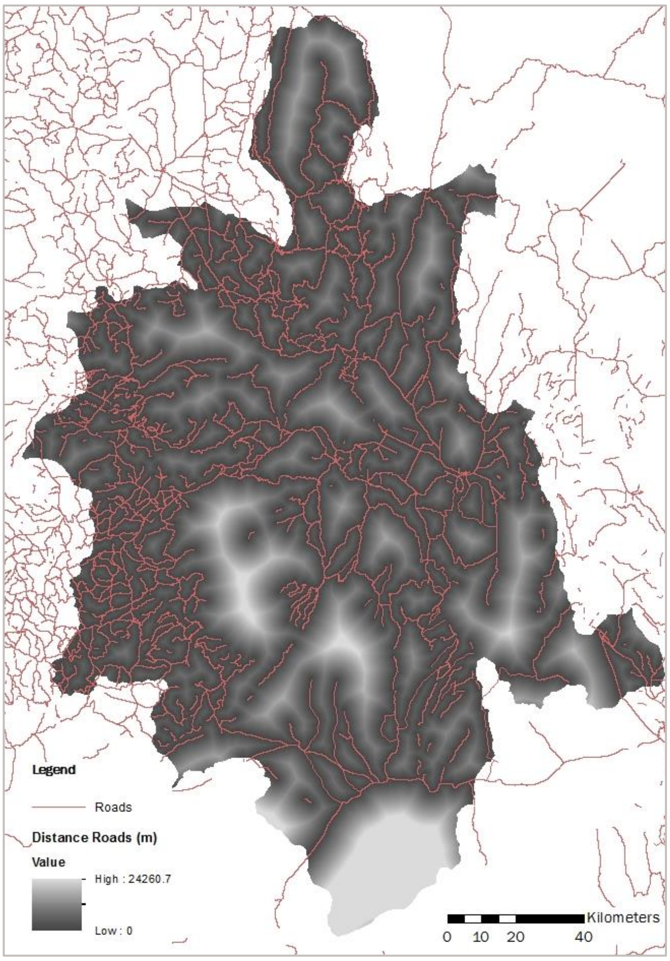


Figure E16: Distance to roads

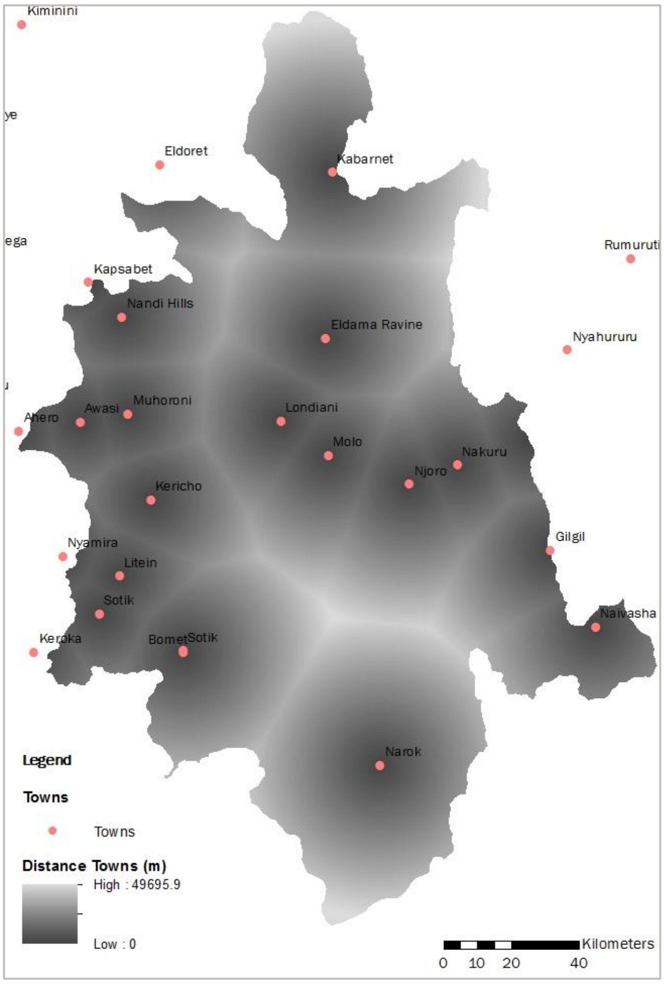


Figure E17: Distance to towns

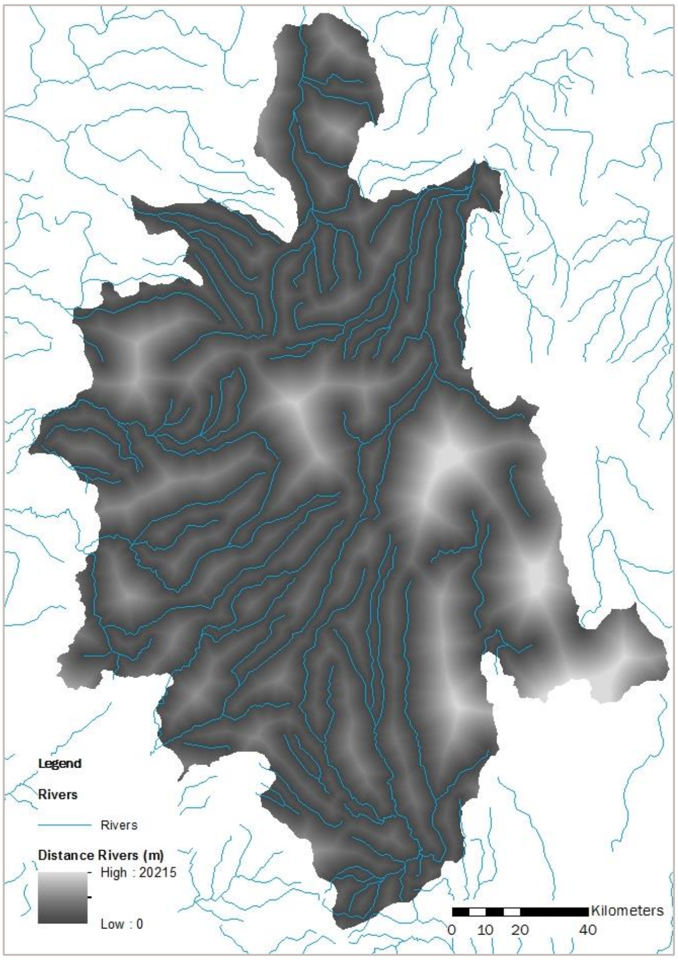


Figure E18: Distance to rivers

APPENDIX F – POPULATION NUMBERS

Figure F1: The counties the study area overlaps

Table F1: The development of the counties in the study area 1973-2013

PROVINCE	DISTRICTS/COUNTIES		
	ORIGINAL	**1999**	**2009**
	Baringo	Baringo	Baringo
		Koibatek (1995)	
	Elgeyo-Marakwet	Keiyo	Elgeyo-Marakwat
		Marakwet (1994)	
RIFT VALLEY	Kericho	Kericho	Kericho

		Bomet (1992)	Bomet
		Buret (1998)	
	Nakuru	Nakuru	Nakuru
	Nandi	Nandi	Nandi
	Narok	Narok	Narok
		Trans-Mara	
	Uasin Gishu	Uasin Gishu	Uasin Gishi
	Kisumu	Kisumu	Kisumu
		Nyando (1998)	
NYANZA	Kisii	Kisii Central	Kisii
		Gucha (1995)	Nyamira
		Nyamira (1992)	
	Nyandurua	Nyandurua	Nyandurua
CENTRAL	Kiambu	Kiambu	Kiambu
		Thika	

Table F2: The area in km² per district/county

PROVINCE	DISTRICTS/COUNTIES			AREA IN KM²			
	ORIGINAL	1999	2009	1969/79	1989	1999	2009
	Baringo	Baringo	Baringo	9,885	10,954	8,646	11,015.30
		Koibatek (1995)				2,306	
	Elgeyo-Marakwet	Keiyo	Elgeyo-Marakwat	2,279	3,049	1,439	3,029.80
		Marakwet (1994)				1,588	
	Kericho	Kericho	Kericho	3931	4,940	2,111	2,479.00
		Bomet (1992)	Bomet			1,882	2,471.30
RIFT VALLEY		Buret (1998)				955	
	Nakuru	Nakuru	Nakuru	5769	7,190	7,242	7,495.10
	Nandi	Nandi	Nandi	2745	2,784	2,899	2,884.20
	Narok	Narok	Narok	16115	18,002	15,098	17,933.10
		Trans-Mara				2,846	
	Uasin Gishu	Uasin Gishu	Uasin Gishi	3378	3,218	3,328	3,345.20
	Kisumu	Kisumu	Kisumu	2093	2,077	919	2,085.90
		Nyando (1998)				1,168	
NYANZA	Kisii	Kisii Central	Kisii	2,196	2,198	649	1,317.50
		Gucha (1995)	Nyamira			661	899.3
		Nyamira (1992)				896	
	Nyandurua	Nyandurua	Nyandurua	3,528	3,373	3,304	3,245.30
CENTRAL	Kiambu	Kiambu	Kiambu	2,448	2,587	1,324	2,543.40
		Thika				1,960	

Table F3: Population numbers per district/county

PROVINCE	DISTRICTS/COUNTIES			POPULATION				
	ORIGINAL	1999	2009	1969	1979	1989	1999	2009
RIFT VALLEY	Baringo	Baringo	Baringo	161,741	203,792	286,490	264,978	555,561
		Koibatek (1995)					138,163	

	Elgeyo-Marakwet	Keiyo	Elgeyo-Marakwat	159,265	148,868	216,487	143,865	369,998
		Marakwet (1994)					140,629	
	Kericho	Kericho	Kericho	479,135	633,348	900,934	468,493	758,339
		Bomet (1992)	Bomet			724,186	382,794	724,186
		Buret (1998)					316,882	
	Nakuru	Nakuru	Nakuru	290,853	522,709	849,096	1,187,039	1,603,325
	Nandi	Nandi	Nandi	209,068	299,319	433,613	578,751	752,965
	Narok	Narok	Narok	125,219	210,306	398,272	365,750	850,920
		Trans-Mara					170,591	
	Uasin Gishu	Uasin Gishu	Uasin Gishi	191,036	300,766	445,530	622,705	894,179
NYANZA	Kisumu	Kisumu	Kisumu	400,643	482,327	664,086	504,359	968,909
		Nyando (1998)					299,930	
	Kisii	Kisii Central	Kisii	675,041	869,512	1,137,054	491,786	1,152,282
		Gucha (1995)	Nyamira				460,939	598,252
		Nyamira (1992)					498,102	
CENTRAL	Nyandurua	Nyandurua	Nyandurua	176,928	233,302	345,420	479,902	596,268
	Kiambu	Kiambu	Kiambu	475,576	686,290	909,762	744,010	1,623,282
		Thika					645,713	

Table F4: Population density per district/county

PROVINCE	DISTRICTS/COUNTIES			POPULATION DENSITY						
	ORIGINAL	1999	2009	1969	1979	1989	1999	1999*	2009	2009*
RIFT VALLEY	Baringo	Baringo	Baringo	15	19	32	31	36.8	50	50
		Koibatek (1995)					60			
	Elgeyo-Marakwet	Keiyo	Elgeyo-Marakwat	52	49	71	100	94.0	122	122
		Marakwet (1994)					89			
	Kericho	Kericho	Kericho	97	128	182	222	236.1	306	306
		Bomet (1992)	Bomet				203		293	293
		Buret (1998)					332			
	Nakuru	Nakuru	Nakuru	40	73	118	164	163.9	214	214
	Nandi	Nandi	Nandi	75	108	156	200	199.6	261	261
	Narok	Narok	Narok	7	12	22	24	29.9	47	47
		Trans-Mara					60			
	Uasin Gishu	Uasin Gishu	Uasin Gishi	59	93	138	187	187.1	267	267
NYANZA	Kisumu	Kisumu	Kisumu	193	232	320	549	385.4	465	465
		Nyando (1998)					257			
	Kisii	Kisii Central	Kisii	307	396	517	758	657.7	875	790
		Gucha (1995)	Nyamira				698		665	
		Nyamira (1992)					556			
CENTRAL	Nyandurua	Nyandurua	Nyandurua	52	69	102	145	145.2	184	184
	Kiambu	Kiambu	Kiambu	184	265	352	562	423.2	638	638
		Thika					329			

*the average population density of the total counties in the original districts

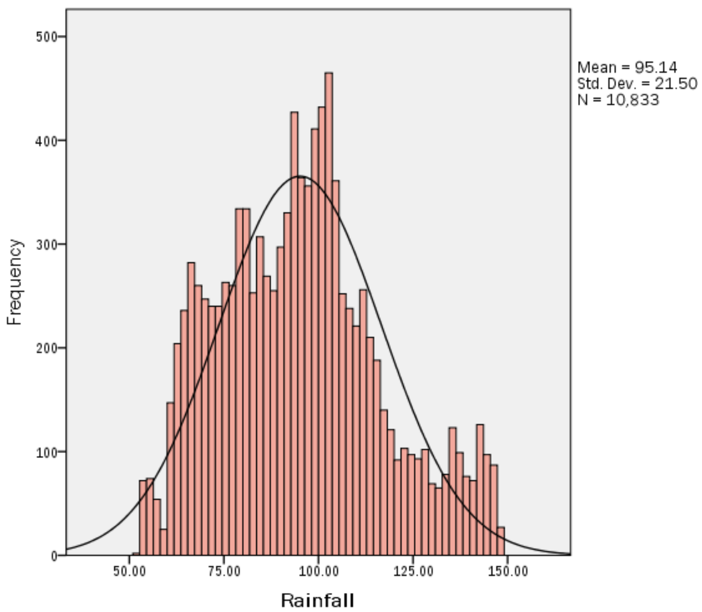
APPENDIX G – DESCRIPTIVE STATISTICS

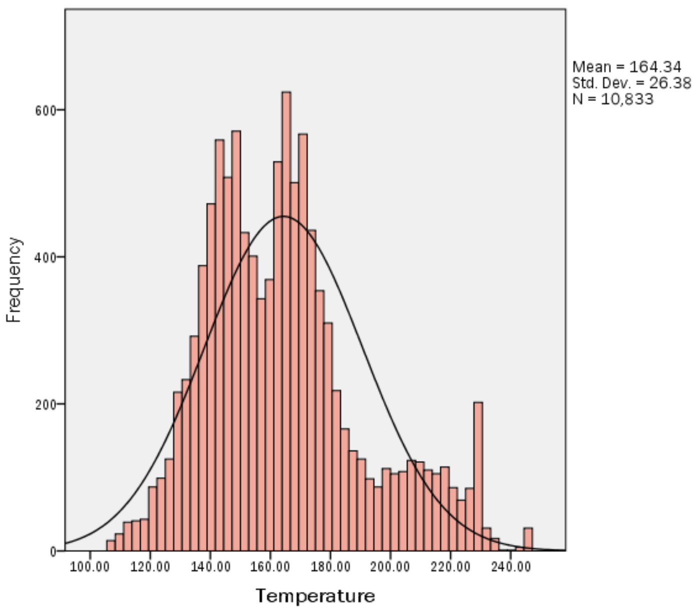
Table G1: Descriptive statistics independent variables

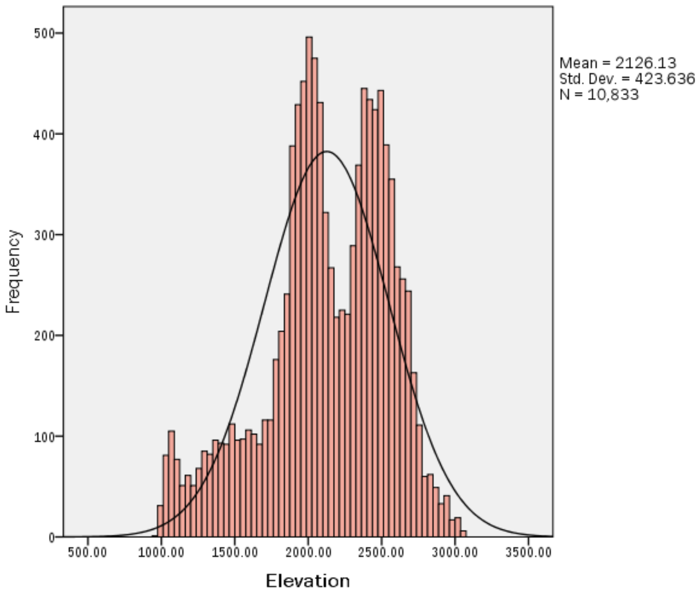
	min	max	mean	standard deviation
Rainfall	51.92	148.20	95.14	21.50
Temperature	106.61	246.45	164.34	26.38
Elevation	974.02	3068.73	2126.13	423.64
Slope	0.33	48.81	9.05	4.87
Aspect north	-0.99	0.98	0.01	0.36
Aspect east	-1.00	1.00	-0.05	0.41
Curvature	-1.36	2.80	0.00	0.13
TWI	4.35	10.97	6.36	0.42
Soil pH	46.96	138.89	60.36	6.53
Soil CEC	9.31	68.39	25.86	4.67
Distance to Towns	0.34	49.09	19.27	10.30
Distance to Roads	0.00	19.89	2.14	2.12
Distance to Rivers	0.03	15.94	3.04	2.72
Population Density '79	12.0	396.0	62.9	50.3
Population Density '89	22.0	517.0	95.0	69.6
Population Density '99	30.0	658.0	125.2	90.5
Population Density '09	47.0	790.0	164.8	113.1

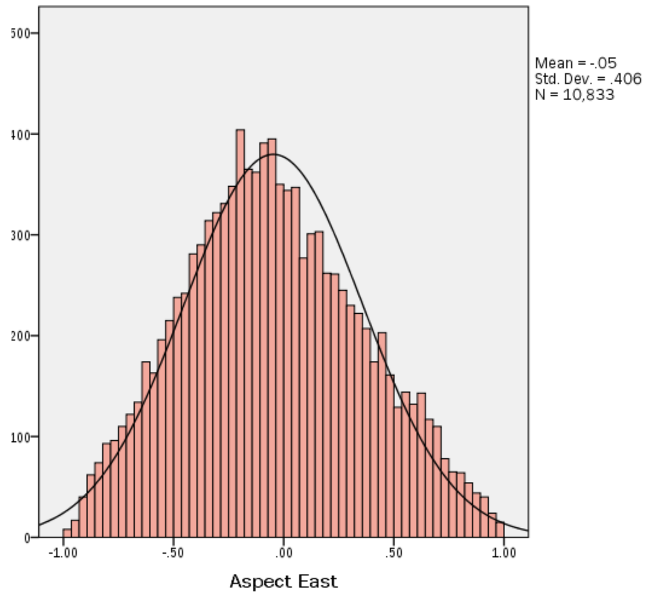
Data distribution and histograms

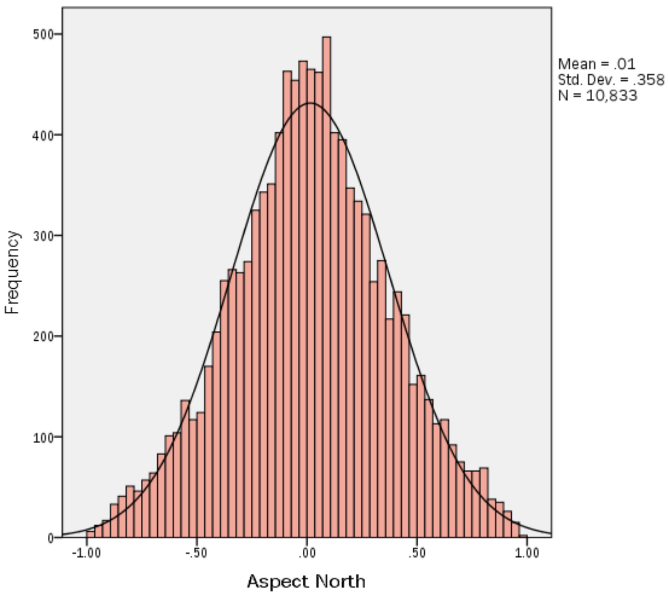
Geophysical variables

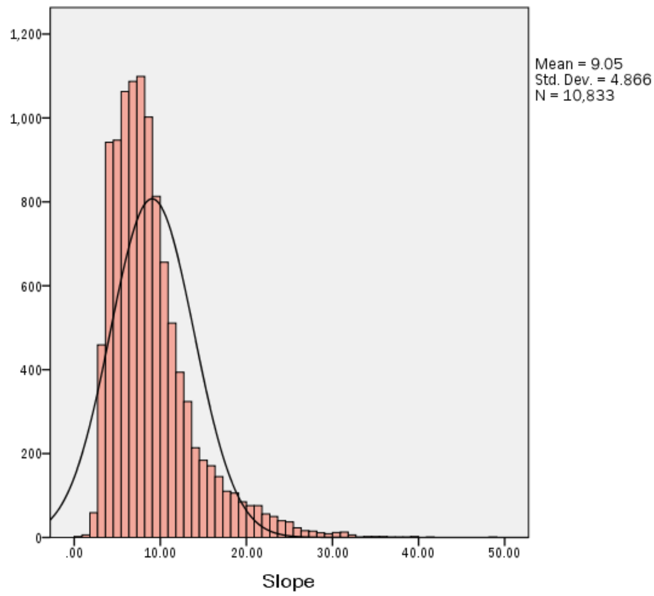


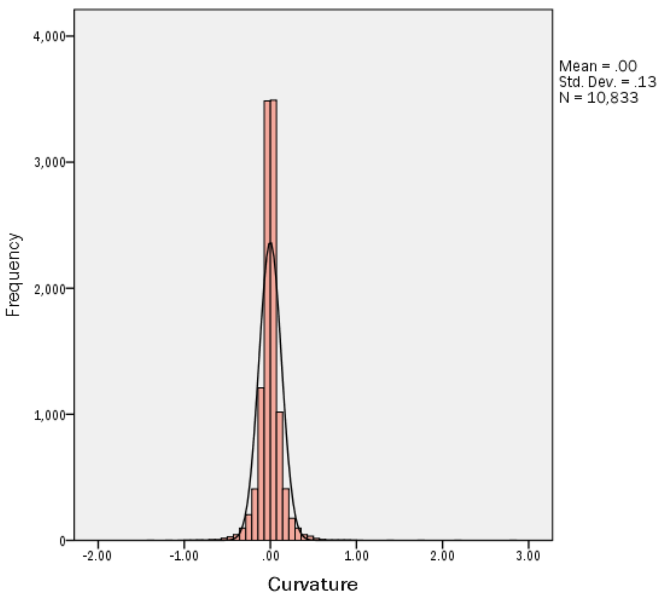


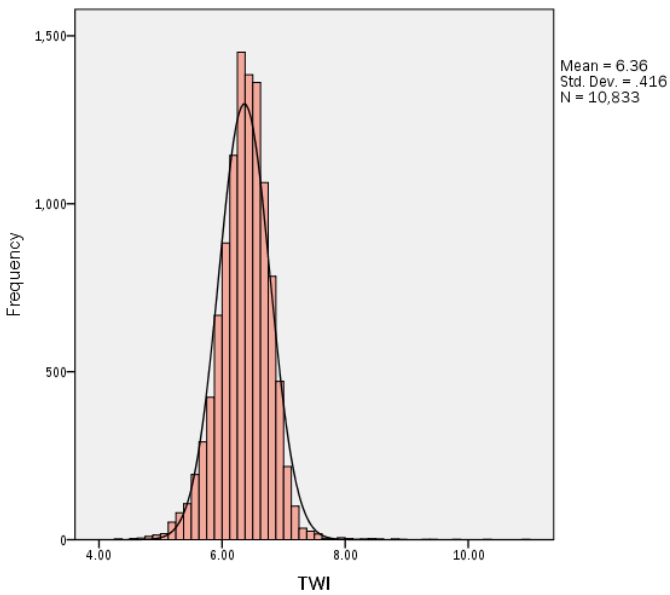

Mean = 2126.13
Std. Dev. = 423.636
N = 10,833
Frequency
Elevation

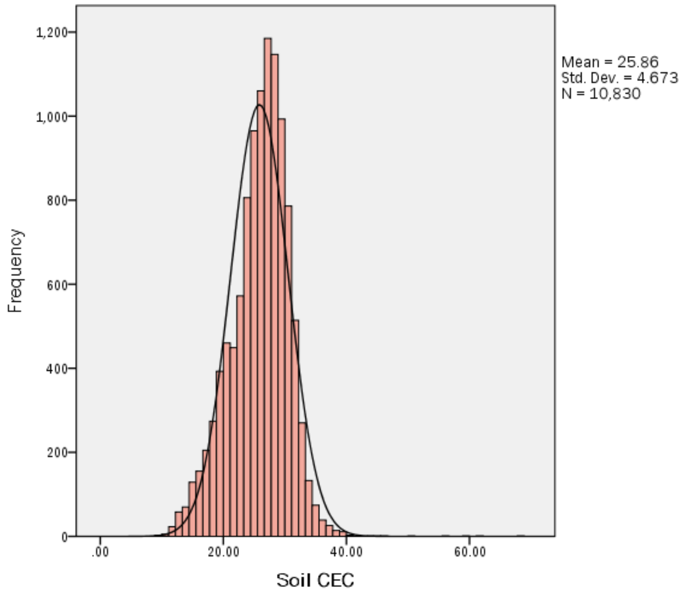

Mean = -.05
Std. Dev. = .406
N = 10,833
Aspect East

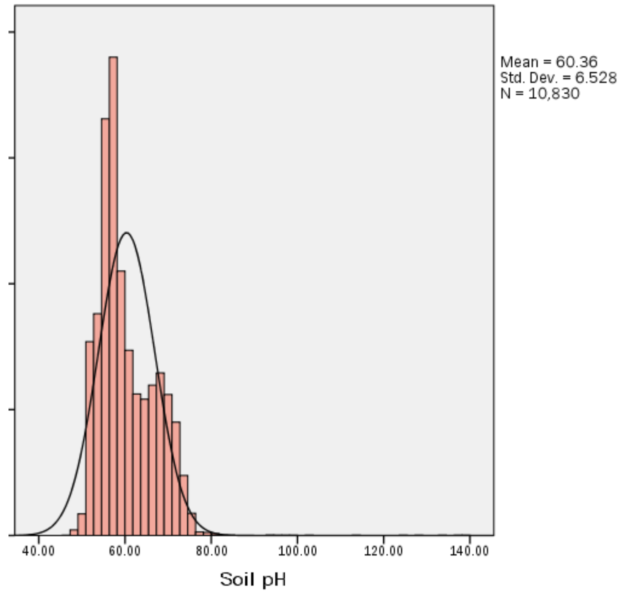

Mean = .01
Std. Dev. = .358
N = 10,833
Frequency
Aspect North


Mean = 9.05
Std. Dev. = 4.866
N = 10,833
Slope

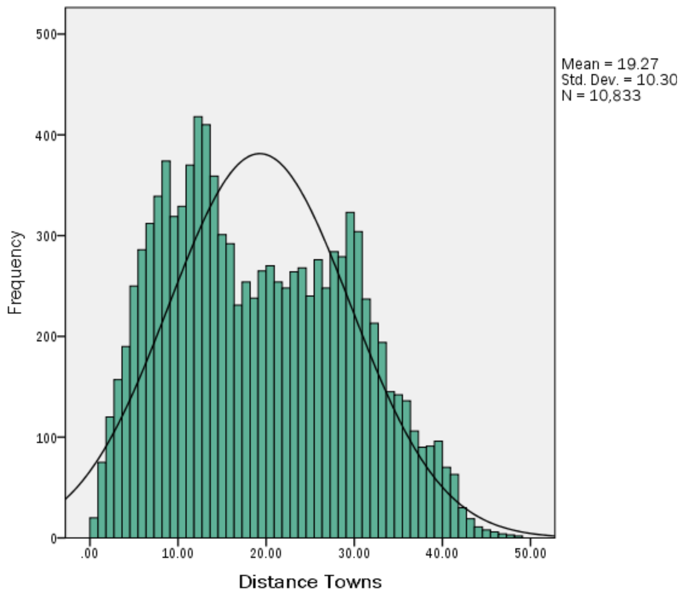

Mean = .00
Std. Dev. = .13
N = 10,833
Frequency
Curvature

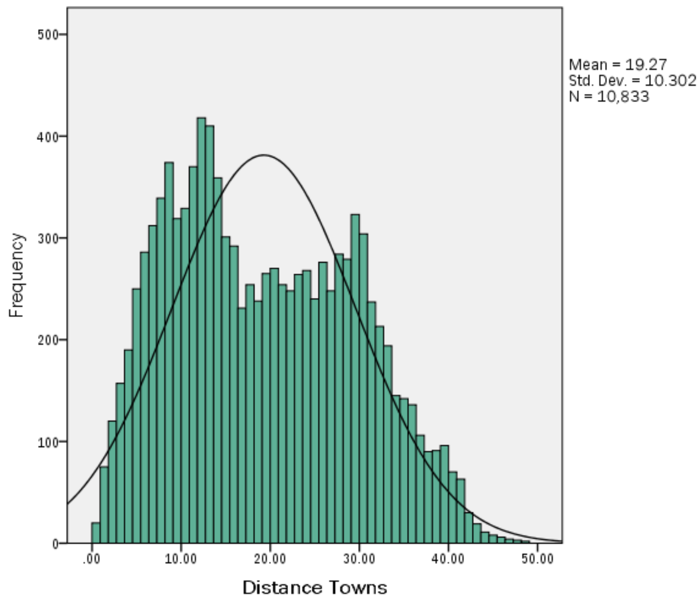

Mean = 6.36
Std. Dev. = .416
N = 10,833
Frequency
TWI

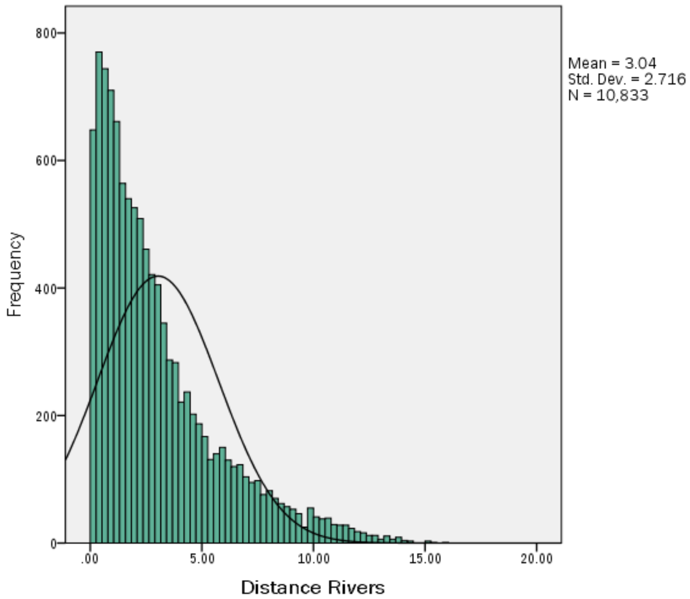


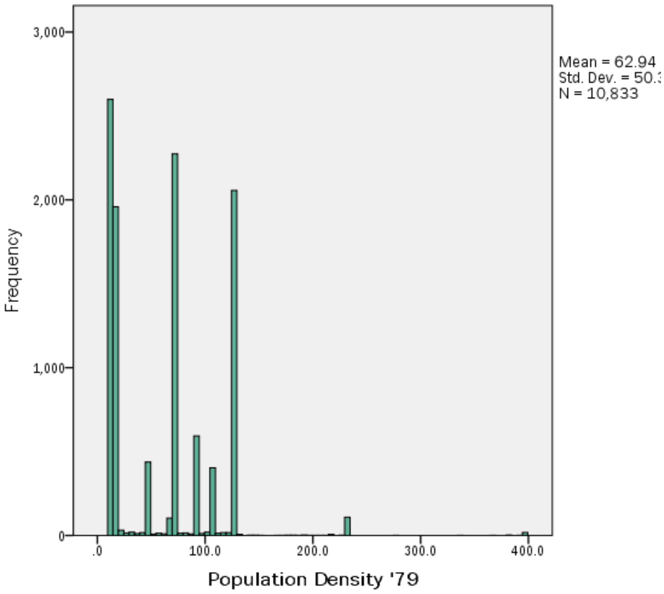


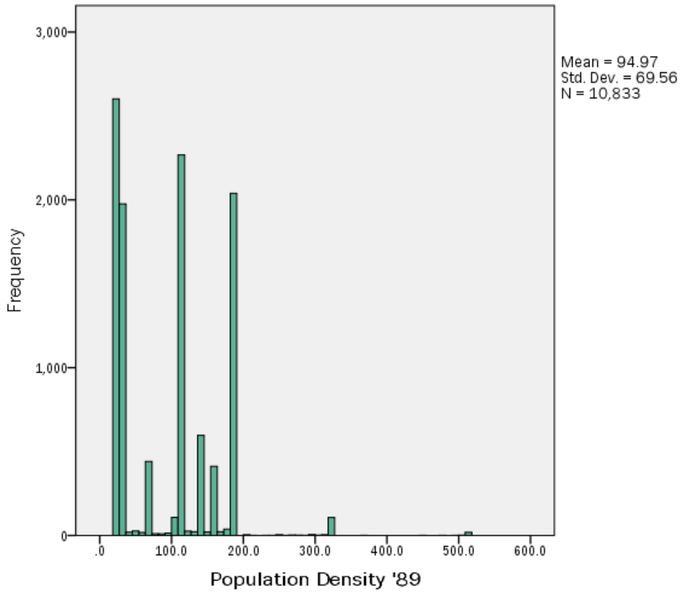
Socio-economic variables

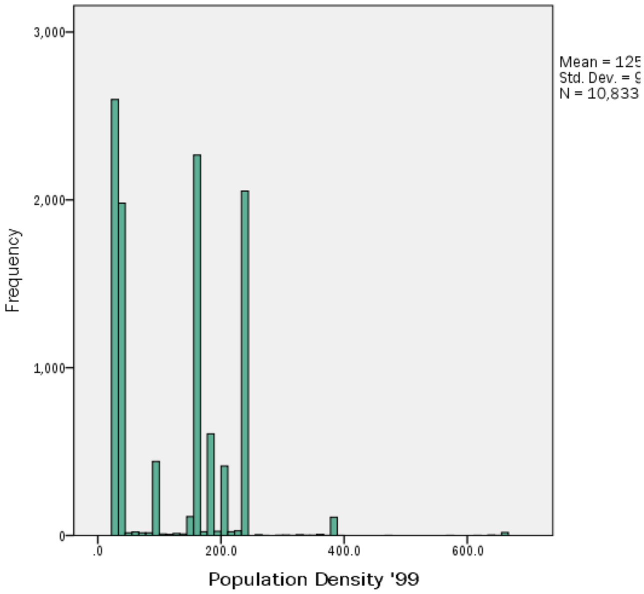


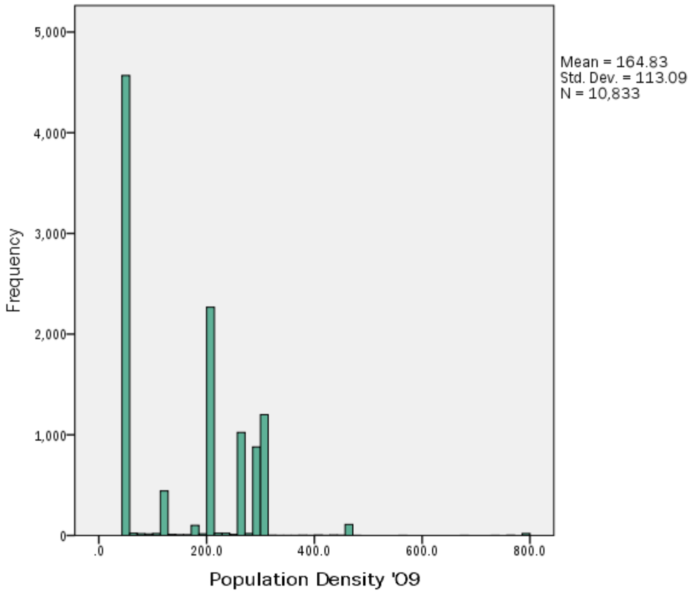





Mean = 62.94
Std. Dev. = 50.3
N = 10,833
Frequency
3,000
2,000
1,000
0
.0
100.0
200.0
300.0
400.0
Population Density '79


Mean = 94.97
Std. Dev. = 69.56
N = 10,833
Frequency
3,000
2,000
1,000
0
.0
100.0
200.0
300.0
400.0
500.0
600.0
Population Density '89


Mean = 125
Std. Dev. =
N = 10,833
Frequency
3,000
2,000
1,000
0
.0
200.0
400.0
600.0
Population Density '99


Mean = 164.83
Std. Dev. = 113.09
N = 10,833
Frequency
5,000
4,000
3,000
2,000
1,000
0
.0
200.0
400.0
600.0
800.0
Population Density '09

APPENDIX H – NATURE OF LAND USE CHANGES

Table H1: Change matrix indicating the nature of land use change period 1

1973 – 1984 (IN KM²)		1973										
		F	SA	CI	TE	TP	R	I	WB	WL	TOTAL	GAIN
1984	Forest	5197.7	32.4	00	0.0	0.6	6.0	0.0	0.0	0.0	5236.8	39.1
	Smallholder Agriculture	275.8	9829.9	0.0	0.0	4.3	242.2	0.0	0.0	0.0	10352.2	522.3
	Irrigated Commercial Agriculture	0.0	0.0	602.0	0.0	0.0	9.4	0.0	0.0	0.0	611.4	9.4
	Tea Estates	6.1	0.0	0.0	287.3	0.0	0.0	0.0	0.0	0.0	293.4	6.1
	Tree Plantations	**134.3**	107.5	0.0	2.2	383.7	50.9	0.0	0.0	0.0	678.7	295.0
	Rangeland	110.0	10.0	0.0	0.0	27.0	6357.5	0.0	0.0	0.0	6504.5	146.9
	Infrastructure	0.0	1.4	0.0	0.0	0.0	0.0	59.3	0.0	0.0	60.8	1.4
	Waterbodies	0.0	0.0	0.0	0.0	0.0	0.0	0.0	156.0	0.0	156.0	0.0
	Wetland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	207.6	208.0	0.4
TOTAL		5724.0	9981.2	602.0	289.5	415.6	6666.1	59.3	156.4	207.6	24101.7	
LOSS		526.3	151.3	0.0	2.2	31.9	308.6	0.0	0.4	0.0		

Table H2: Change matrix indicating the nature of and use change period 2

1984 - 1994 (IN KM²)		1984										
		F	SA	CI	TE	TP	R	I	WB	WL	TOTAL	GAIN
1994	Forest	4936.2	6.2	0.0	0.0	0.0	8.0	0.0	0.0	0.0	4950.5	14.2
	Smallholder Agriculture	266.1	10289.7	0.0	0.0	34.7	359.6	0.0	0.0	0.0	10950.2	660.5
	Irrigated Commercial Agriculture	0.0	5.7	611.4	0.0	0.0	8.5	0.0	0.0	0.0	625.5	14.1
	Tea Estates	11.1	3.4	0.0	284.9	0.0	2.2	0.0	0.0	0.0	301.5	16.7
	Tree Plantations	1.5	13.0	0.0	8.6	612.8	70.3	0.0	0.0	0.0	706.2	93.4
	Rangeland	21.8	11.5	0.0	0.0	31.2	6055.8	0.0	0.0	0.0	6120.4	64.5
	Infrastructure	0.0	**22.6**	0.0	0.0	0.0	0.0	60.8	0.0	0.0	83.3	22.6
	Waterbodies	0.0	0.0	0.0	0.0	0.0	0.0	0.0	156.0	0.0	156.0	0.0
	Wetland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	208.0	208.0	0.0
TOTAL		5236.8	10352.2	611.4	293.4	678.7	6504.5	60.8	156.0	208.0	24101.7	
LOSS		300.5	62.4	0.0	8.6	65.9	448.6	0.0	0.0	0.0		

Table H3: Change matrix indicating the nature of and use change period 3

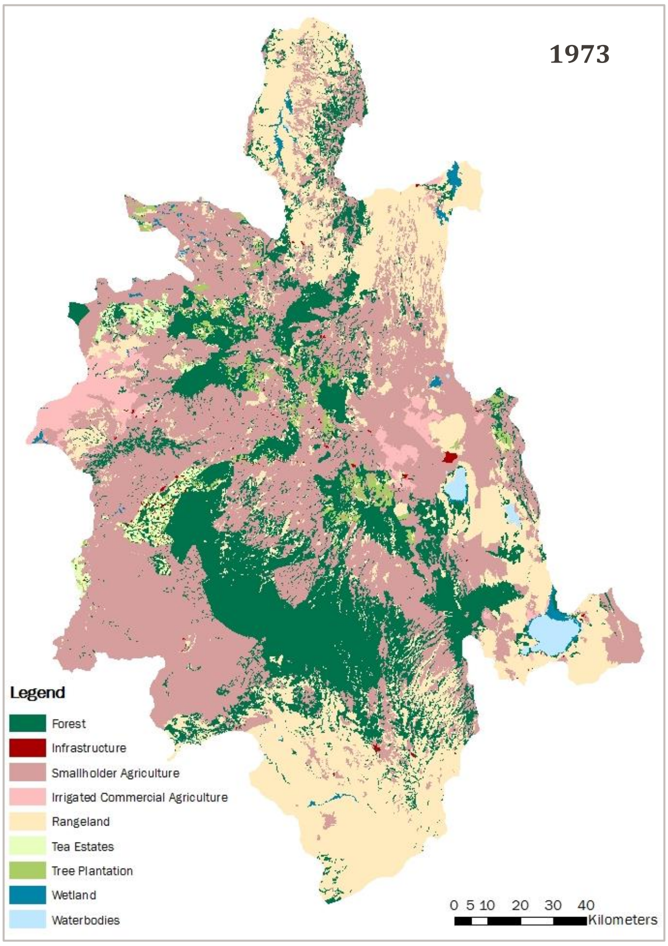
1994 - 2003 (IN KM²)		1994										
		F	SA	CI	TE	TP	R	I	WB	WL	TOTAL	GAIN
2003	Forest	4561.6	0.7	0.0	0.0	14.5	10.1	0.0	0.0	0.0	4586.9	25.3
	Smallholder Agriculture	304.6	10867.1	0.0	0.3	**288.5**	220.9	0.0	0.0	0.0	11681.5	814.4
	Irrigated Commercial Agriculture	0.4	**50.2**	625.5	0.0	0.0	12.5	0.0	1.5	0.0	690.2	64.6
	Tea Estates	22.1	4.7	0.0	301.2	8.8	1.4	0.0	0.0	0.0	338.2	37.0
	Tree Plantations	0.0	0.0	0.0	0.0	345.6	6.7	0.0	0.0	0.0	352.4	6.8
	Rangeland	61.7	7.1	0.0	0.0	44.8	5863.9	0.0	0.0	0.0	5977.5	113.6
	Infrastructure	0.0	**20.4**	0.0	0.0	4.0	4.8	83.3	0.0	0.0	112.6	29.2
	Waterbodies	0.0	0.0	0.0	0.0	0.0	0.0	0.0	154.5	0.4	154.9	0.4
	Wetland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	207.6	207.6	0.0
TOTAL		4950.5	10950.2	625.5	301.5	706.2	6120.4	83.3	156.0	208.0	24101.7	

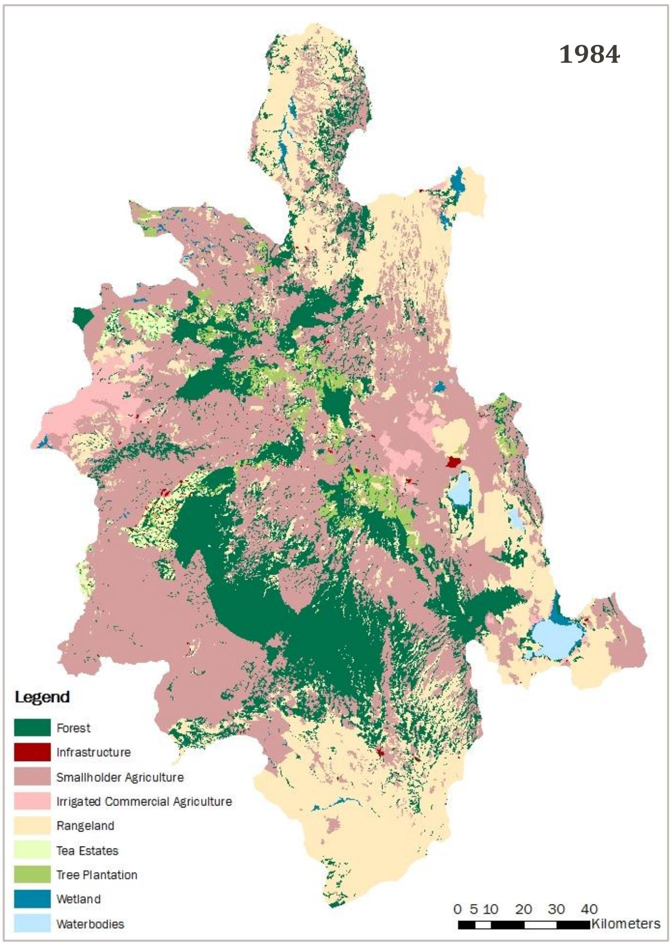
LOSS	388.9	83.1	0.0	0.3	360.6	256.5	0.0	1.5	0.4

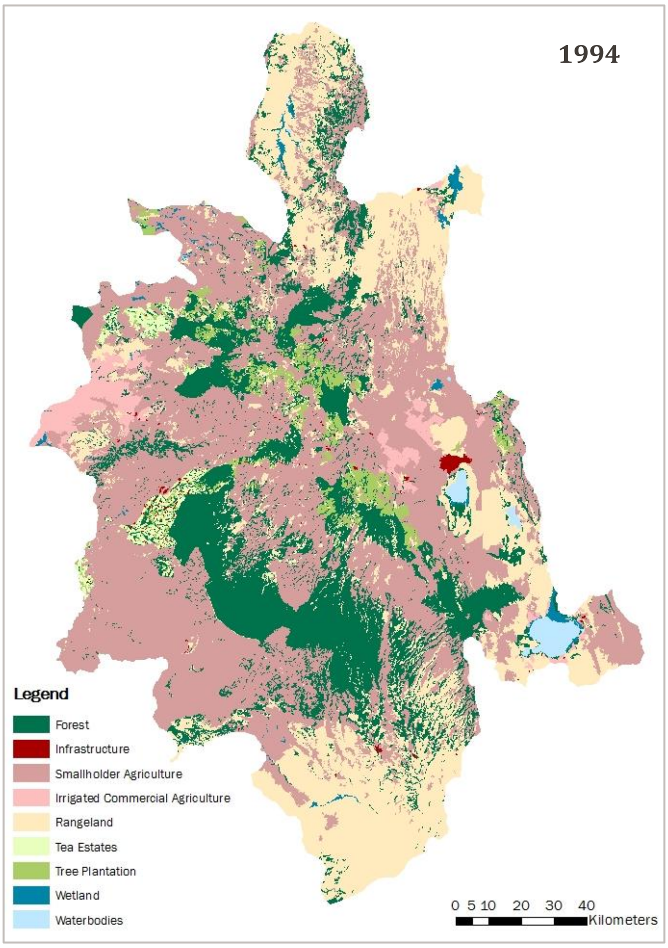
Table **H4: N**ature of and use change period 4

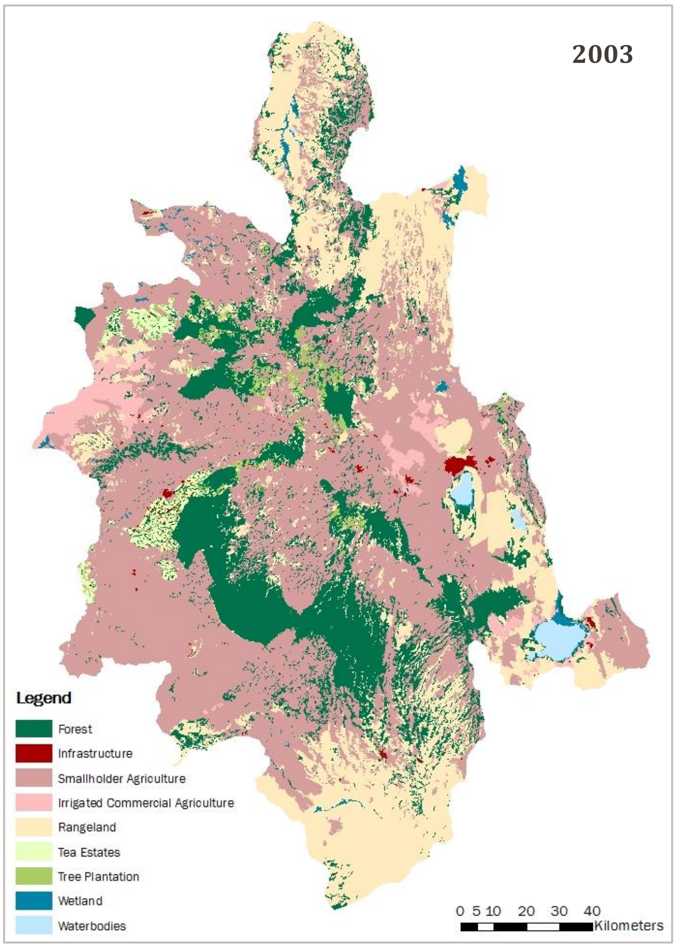
2003 - 2013 (IN KM²)												
		2003										
		F	**SA**	**CI**	**TE**	**TP**	**R**	**I**	**WB**	**WL**	**TOTAL**	**GAIN**
2013	**Forest**	4183.8	51.4	0.0	0.2	2.1	27.9	0.0	0.0	0.0	4265.4	81.7
	Smallholder Agriculture	382.0	11553.0	0.0	0.8	20.3	260.3	0.0	0.0	0.0	12216.4	663.4
	Commercial Agriculture	1.9	2.7	690.2	0.0	0.0	2.9	0.0	0.0	0.0	697.7	7.5
	Tea Estates	0.4	0.0	0.0	326.0	0.5	0.0	0.0	0.0	0.0	326.9	0.9
	Tree Plantations	2.4	41.2	0.0	11.2	322.8	27.1	0.0	0.0	0.0	404.7	81.9
	Rangeland	16.4	23.4	0.0	0.0	6.6	5658.7	0.0	0.0	0.0	5705.0	46.3
	Infrastructure	0.0	9.9	0.0	0.0	0.0	0.5	112.6	0.0	0.0	123.0	10.4
	Waterbodies	0.0	0.0	0.0	0.0	0.0	0.0	0.0	154.9	0.0	154.9	0.0
	Wetland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	207.6	207.6	0.0
TOTAL		4586.9	11681.5	690.2	338.2	352.4	5977.5	112.6	154.9	207.6	24101.7	
LOSS		403.1	128.5	0.0	12.2	29.5	318.8	0.0	0.0	0.0		

APPENDIX I – SPATIAL PATTERN LAND USE CHANGES

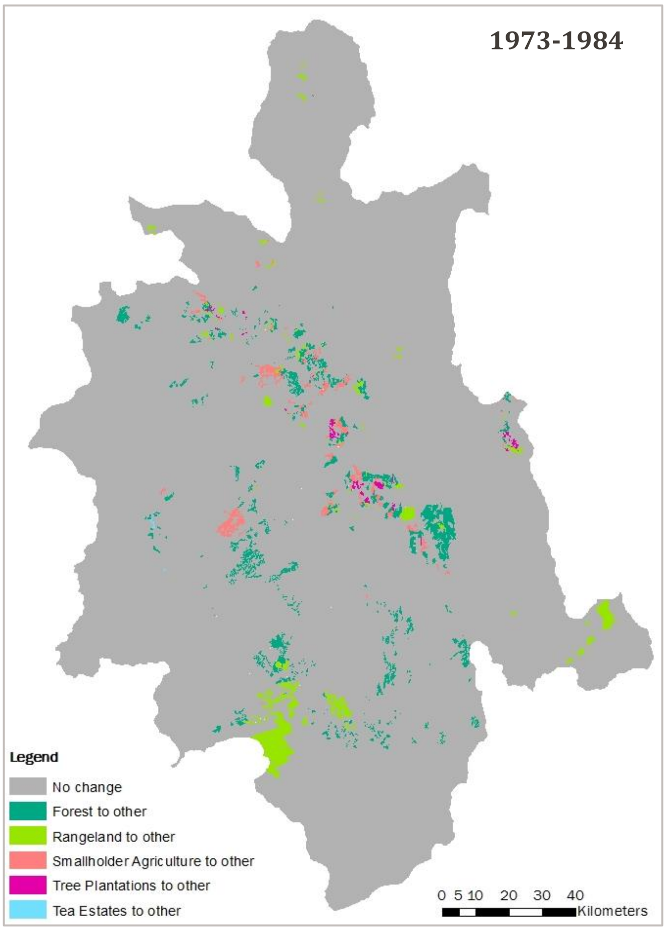


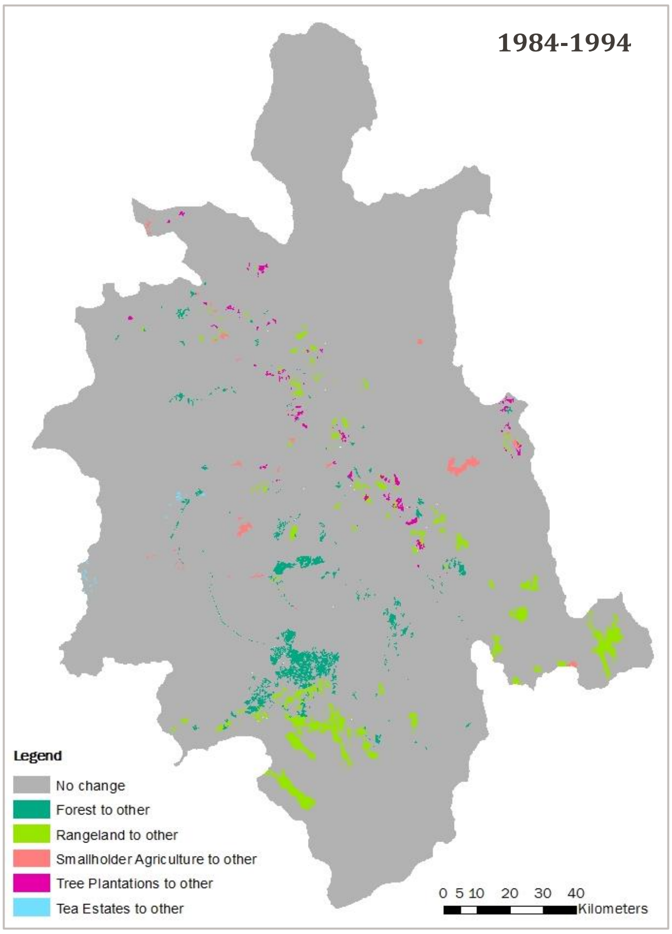


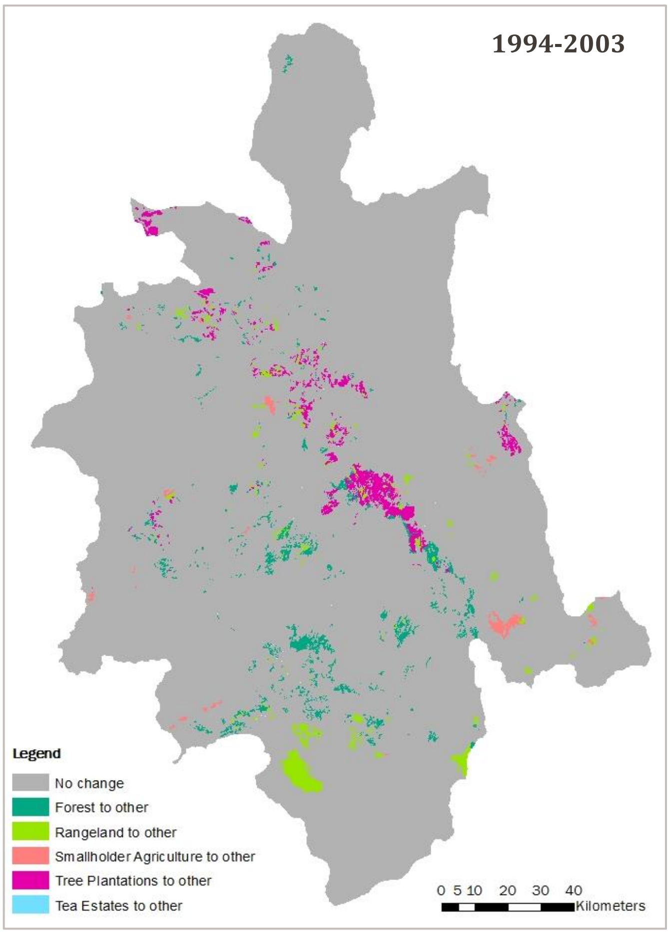


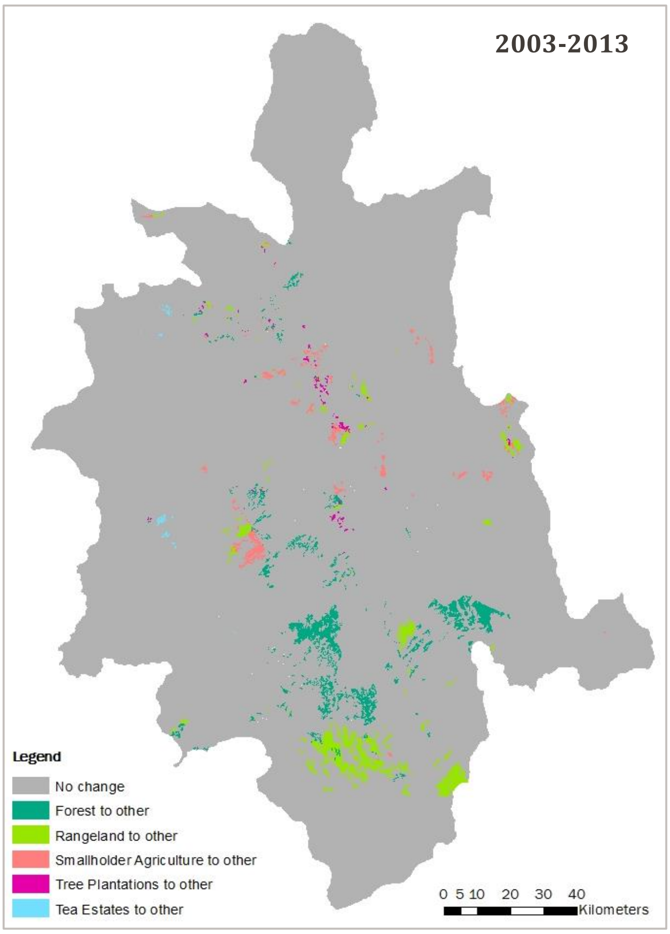


APPENDIX J – LAND USE CHANGE MAPS









APPENDIX K – LUC MAPS MAIN CONVERSIONS