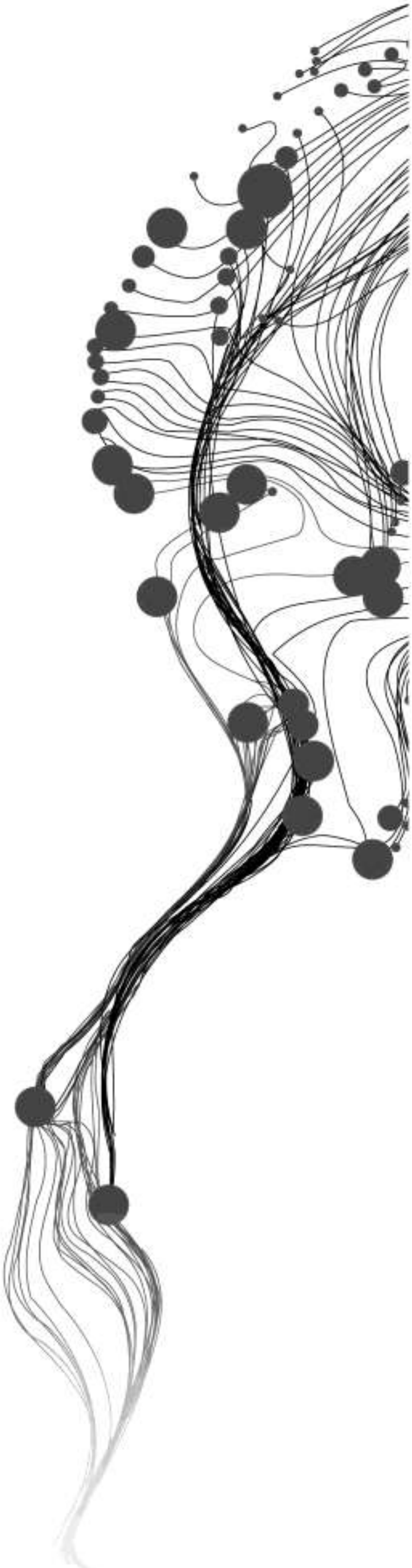
RELATING LAND, PEOPLE AND ENVIRONMENT in Lake Naivasha's Watershed, Kenya

DESTA JULA BEKALO
March, 2011

SUPERVISORS:
Drs. Robert Becht (First Supervisor)
Prof. Dr. Anne van der Veen (Second Supervisor)



RELATING LAND, PEOPLE AND ENVIRONMENT in Lake Naivasha's Watershed, Kenya

DESTA JULA BEKALO
Enschede, The Netherlands, March, 2011

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Governance and Spatial Information Management (GSIM)

SUPERVISORS:
Drs. R. (Robert) Becht (First Supervisor)
Prof. Dr. A. (Anne) van der Veen (Second Supervisor)

THESIS ASSESSMENT BOARD:
Prof. Dr. A. (Anne) van der Veen - Chairman
Dr. Frans H. J. M. Coenen- External examiner - University of Twente (CSTM)
Dr. ir. P. R. (Pieter) van Oel – Internal examiner
Drs. R. (Robert) Becht- (First Supervisor)

ABSTRACT

Linking remote sensing and social science is a new approach to understand human impact on the biophysical environment and to respond to environmental impacts of human economic activities. Integration between social sciences and natural sciences is vital for a better understanding of the Naivasha economy in Kenya that changed so drastically. Human population in general and population density in particular often used as proxy measures for land use and land cover changes and other spatial changes. Thus, estimating the population densities inhabited on the different landscapes, as well as examining the factors affecting population distribution, is essential in formulating appropriate management strategies for sustainable use of natural resources and it helps in examining the risk of natural resource degradation. Therefore, in this paper, an integrated approach developed to reveal the multifaceted relationship between human and environmental interactions using remotely sensed data and socioeconomic data in the Lake Naivasha's watershed. An innovative approach presented to estimate populations for different land use/land cover types and grid cells based on a Weighted Areal Interpolation Algorithm coupled with geographically weighted regression. Land cover information that had derived from satellite imagery was use as weighing factor and population census data for the final estimation. In addition, the study also assessed land fragmentation, examining the relationship between population distribution and other physical variables (slope, soil, rainfall, altitude, and distance from roads). The population estimate for land cover/land use types show that land cover/land use data contain sufficient information to infer population distribution and to remodel the census data in grid cells to understand the spatial pattern of population distribution in the watershed. Land fragmentation and roads are some of the factors indicating population distribution. Furthermore, the analysis result shows that rainfall and soil are the most influential variables in population distribution in the Lake Navisha's watershed.

Keywords: *Land cover, Population estimation and distribution, Weighted areal interpolation, Grid, Land fragmentation, physiographic variables*.

.

ACKNOWLEDGEMENT

Always special thanks go to the Almighty God for His guidance and faithfulness. Sincere gratitude is also due to The Netherlands Fellowship Program (NFP) through funding, afforded an opportunity to broaden my academic, social, and cultural horizons; pursuit of graduate studies in an international environment to have an enriching experience, indeed.

More so, I express deepest appreciation to my supervisor Drs. Robert Becht for the role that he played in the development of the proposal and efficient organization, coordination of fieldwork logistics and the comments he gave me. My deepest appreciation also to Professor Anne van der veen for their able guidance and patient amendment of the many mistakes that I made in the course of the research. I would like to express my appreciation, also, to Drs. Johan de Meijere, for his invaluable ideas and support during my study at ITC.

My generous thanks also go to the World Wildlife Fund (WWF), Naivasha, Kenya who supplied vehicles for field mobility and for their consistent support during the fieldwork. I am, also, greatly indebted to the ITC staffs for their indispensable support of expertise and knowledge. Further, I salute GSIM class for their friendship and special thanks to Alicia Follosco for her companionship and limitless support. To all my friends from ITC courses that always left me feeling at home; you *guys* were simply wonderful!

Lastly, but not the least, I thank my family and special thanks to my beloved wife for supporting my quest for knowledge and for their unwavering love and affection that, always, gave me a cause to keep going.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYM

ASTER	Advanced Space borne Thermal Emission and Reflectance Radiometer
DEM	Digital Elevation Model
ECW	Enhanced Compression Wavelet
ESRI	Environmental System Research Institute
FAO	Food and Agricultural Organization
GIS	Geographic Information System
GPS	Global Positioning System
GWR	Geographic Weighted Regression
ILRI	International Livestock Research Institute
ITC	International Training Centre (Faculty of Geo-information Science and Earth Observation)
KML	Keyhole Mark-up Language
KNBS	Kenya National Bureau of Statistics
KSS	Kenya Soil Survey
NDVI	Normalized Difference Vegetation Index
OLS	Ordinary Least Square
PASW	Predictive Analytics Software
PCA	Principal Component Analysis
PSM	Population Surface Modelling
R^2	A symbol for the coefficient of determination of a linear regression
RMS	Root Mean Square Errors
ROC	Rate of Change
SPSS	Statistical Package for the Social Sciences
SRTM	Shuttle Radar Topographic Mission
TTA	Training and Test Area
UN	United Nations
UTM	Universal Transverse Mercator
WGS	World Geodetic System

1. INTRODUCTION

The basic link of population with different land uses such as agriculture is that every increase and distribution of human population entails a higher minimum requirement of food and other goods (UN, 1997). Growth of the population demands various commodities, which are meet through intensive or extensive cultivation of land that lead to the encroachment of agricultural land in the forest, bush lands, water bodies, and swamps. This results to land fragmentation since scarce resources are split-up between communities and families. Land encroachment and fragmentation adversely affect the ecosystems and thereby cause environmental degradation (Di Giulio et al., 2009).

Worldwide land use and land cover changes are being drive by the need to provide food, shelter, roads, industry, recreation etc. to serve the growing global population. Land use such as cropland, pastures, plantation, and built-up areas have expanded to fulfil human needs (Foley et al., 2005). Land use is mainly the result of human activities and its changes are the result of the complex interplay between socio-economic, and demographic variables, such as migration, population size and density, and institutional and environmental factors (Lesschen et al., 2005). The growing demand for space for human settlement, agricultural production, and grazing, industrial, and commercial purposes is increasingly becoming a great concern for planners, natural resource managers, and decision makers. Investigating the land and estimating the population density inhabited on the different landscape, as well as examining the factors affecting population distribution, are essential in formulating appropriate management strategies for sustainable use of natural resources and helps in examining the risk of natural resource degradation.

In this study, populations is estimated in grid cells using land cover information as weighing factors and census data based on Weighted Areal Interpolation Algorithm with the help of Geographic weighted Regression (GWR), land cover/land use that was derived from satellite imagery and population based on census data. Furthermore, the study examine land fragmentation in relation to population distribution, the character and the spatial scale of land fragmentation in Lake Naivasha's watershed, and the influence of physical factors such as, soil, rainfall, slope, altitude and distance from roads in relation to the population distribution.

1.1. Organization of the Paper

The research organized into six chapters. The first chapter contains, the problem statement, the significance of the study, objectives of the study, research questions, and methodological approach to the study. The second chapter summarizes similar previous researches and which was used as a theoretical background for the study. Chapter three presents a brief introduction of the study area, data collection and data preparation.

The fourth chapter presents a detailed description of the methods carried out to accomplish the research task, including the research techniques, uses, method of data analysis, results and analysis of the data. Chapter five discusses on the results of image classification, population estimate for land cover/land use types and grid cells. The impacts of roads, land fragmentation, and physiographic variables on population distribution were also discussed.

Finally, chapter six concludes by presenting the issues discussed in this paper as well as by providing recommendations for further improvement.

1.2 Problem Statement

Population data are commonly collect in irregular reporting census block and aggregated for analysis at administrative level. However, such data cannot sufficiently represent specific spatial scale for the area of interest, entirely dependent on irregular political administrative unit. It hinder analysis and may be incompatible with other information sources (Wu and Murray, 2005). Recently estimating population using satellite imagery with the help of census data emerged as a new technique. It complements the limits posed by census data and the results can be applying and used for analysis with other socioeconomic variables in temporal and spatial scales. To overcome the problems imposed by census data, the census populations are remodel to land use/land cover types and regular grid cells. The model offers a view of population distribution in the area of interest, which is independent of administrative limits. The results will give valuable information for every location in every spatial scale.

The Lake Naivasha drainage basin is an important source of agricultural production, tourism, and electric power, triggering migration to the area through the years. Population around Lake Naivasha has registered a rapid increase from 43 867 in 1969 to 158 679 in 1999 and it is estimated to reach more than 250,000 by the year 2005 (Mireri, 2005). This brought about the growth of small-sized rural and urban human settlement in the watershed and intensifying economic activities associated with the functional use of land. The growing population in the area is creating enormous pressures on the natural resource base in the watershed. Understanding the spatial distribution of the lake Naivisha's watershed population and investigating the factors affecting their distribution is an important aspect for proper land use management as well as for sustainable development of the area. Human population in general and population densities in particular often used as proxy measures to estimate land use change and other spatial changes. In this regard, the spatial population distribution and its link with environmental variables and the factor affecting population distribution are not yet explored in the Lake Navisha'a watershed. Therefore, the study will classify the image to derive land cover map to estimate population for land cover/land use types and modelling at 1km grid cell. Furthermore, the study assesses land fragmentation and investigates the relationship of population distribution with other physical variables (slope, soil, land use, rainfall, altitude, and distance from the roads). This study will be using as an important step in the detailed evaluation for various activities and to consider population densities in the area of interest as alternative means to measure and model other socio-economic variables using remote sensing data.

1.2 General Objective

In View of the above-mentioned problems, the following objectives are formulated: (a) to derive updated land cover/land use map and estimate populations for different land cover/land use types (b) to model population in grid cells, assess land fragmentation in relation to population density and examine the relationship of population distribution with physical variables in the Lake Naivasha's Watershed

1.2.1 Specific Objectives

- To classify the image and update the land use/land cover maps of the study area using satellite imagery.
- To estimate population for aggregated land use/land cover types
- To model population in grid cells and generate population surface map
- To assess the reliability of the model for population distribution
- To assess and examine land fragmentation in the study area
- Examine the relationship of population distribution with physiographic variables.

1.2.2 Research Questions

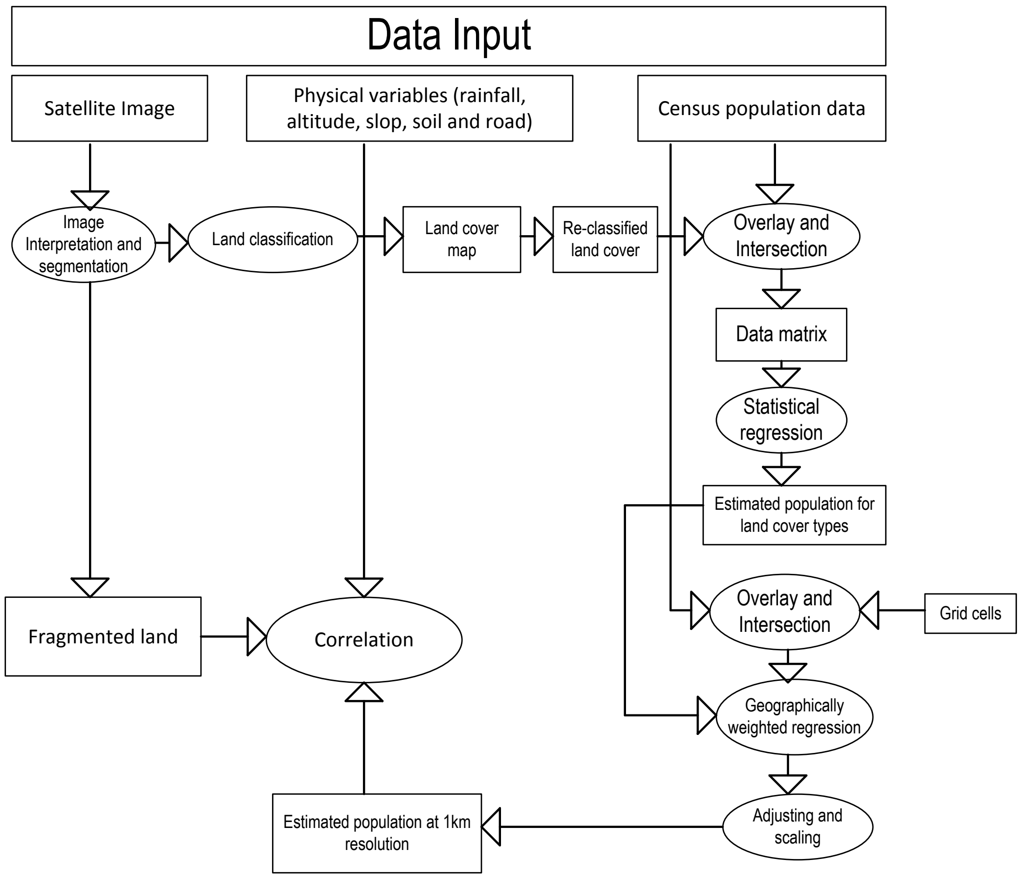
- ✓ What is the spatial distributions and extent of land use/land cover in the watershed?
- ✓ What are the population estimates for different land use/land cover types in the watershed?
- ✓ What are the spatial distributions of population in Lake Naivasha's watershed?
- ✓ How reliable is the population estimate for different spatial location?
- ✓ How can land fragmentation in the study area is described? What is its relation to population density?
- ✓ Which physiographic variables can best explain population distribution?

1.3 Research approach

1.3.1 Methodology

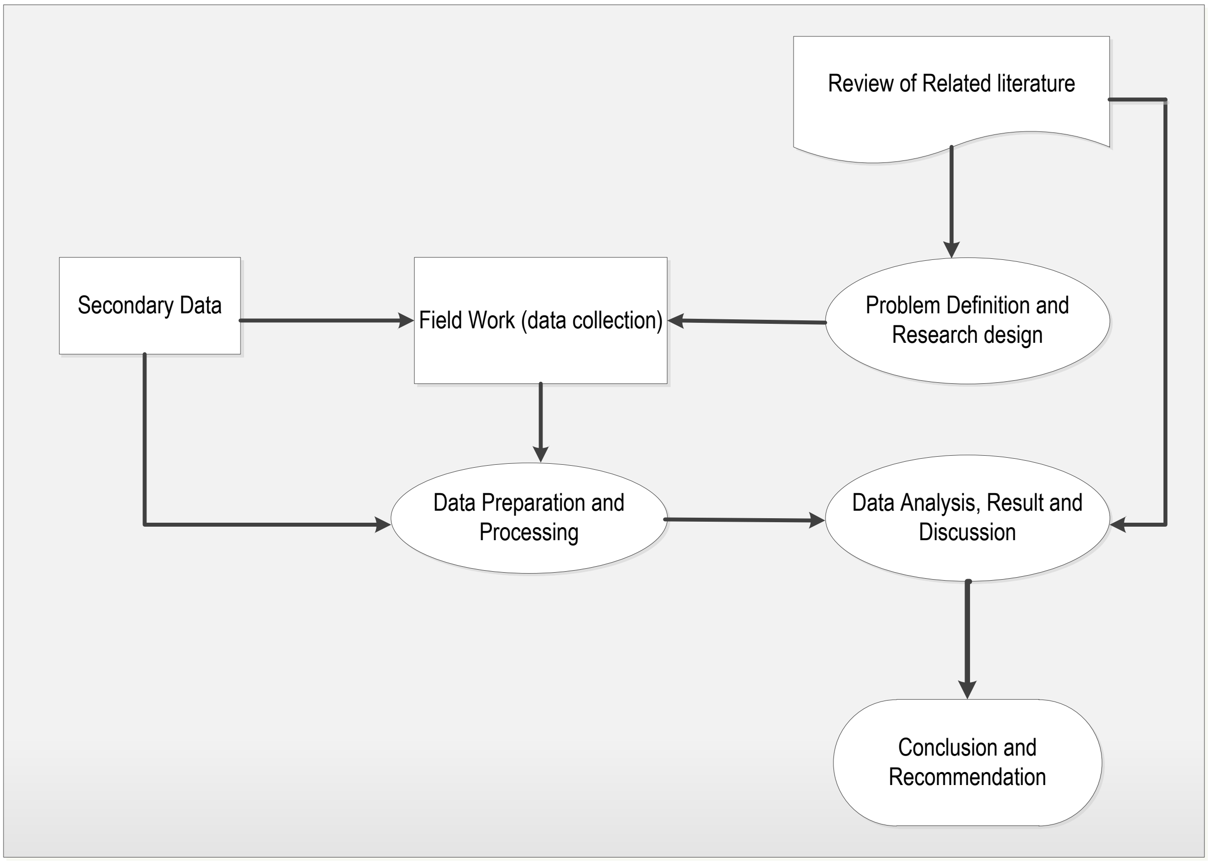
The workflow in this study is illustrated in Figure 1. The first step is to classify the image to get land cover/land use maps for the watershed. The second is to reclassify the land cover/land use maps. The third is to overlay the sub-location population polygon and the re-classified land cover/land use maps and apply areal weighted interpolation, statistical regressions to estimate the population for land cover/land use types and finally to derive weighing factors. The fourth element is to construct 1km grid cells. The fifth is to overlay the sub-location population polygon and grid cells. The sixth is applying weighted areal interpolation; the regression coefficient (land cover estimate) used as weight variables using geographically weighted regression (GWR) to derive population estimate for each grid cells. The seventh is to locally fit the estimate of regression for each sub-location and grid cell sub unit with scaling techniques. The eighth is validating the population estimate. The final step is analysing land fragmentation with population density and examining the relationship between population distribution with other physical variables (slope, soil, rainfall, altitude, and distance from the roads).

Figure 1 Diagram Showing Methodological Approach



1.3.2 Research Framework

Figure 2 Flowcharts Illustrating the Stages of the Research



2. THEORETICAL BACKGROUND

2.1 Population Distribution, Density and Environment

Population distribution is the pattern of where people live. All kinds of physiographic factors have influence on population density with different extents, levels of significance, and scales. Researches that focused on population distribution reveal that people usually concentrate in broad areas with low elevation and excellent allocation of water and soil (Song et al., 2007). Population distribution is uneven on the land surface. Sparsely populated area tend to be difficult places to live, these are usually hostile environments. On the contrary, places, which densely inhabit areas, are habitable environment. Population density is an indicator for the number of people in an area, give average number calculated by dividing the number of people by area. It is usually denote as the number of people per square kilometre.

There are wide ranges of human and natural factors that affect population density. Some of the factors that discouraging and encouraging settlements are: Relief, altitude, climate, soils, natural vegetation, natural resources, natural hazards and other human factors such as economy, culture etc. The following summarizes the argument above.

Table 1 Natural and Human Factors Affecting Population Distribution

Factors	High Density	Low Density
Relief (shape and height of land)	Low land which is flat	Highland that is mountainous
Resources	Areas rich in resources (e.g. coal, oil, wood, fishing etc.) tend to densely populated	Areas with few resources tend to be sparsely populated e.g. The Sahel
Climate	Areas with temperate climates tend to be densely populated as there is enough rain and heat to grow crops	Areas with extreme climates with hot and cold tend to be sparsely populated
Human Factors	**High Density**	**Low Density**
Political	Countries with stable governments tend to have a high population density	Unstable countries tend to have lower population densities as people migrate
Social	Groups of people want to live close to each other for security	Other groups of people prefer to be isolated
Economic	Job opportunities encourage high population densities, particularly in large cities	Limited job opportunities cause some areas to be sparsely populated

Source: (Internetgeography)

The relation between human population distribution and land use change is much debating (Ningal et al., 2008). However, several studies have concluded that the spatial distribution of human population on the land surface is considered as a fundamental determinant of land use impacts on natural ecosystems (Small, 2004). In developing countries, high rural human population density is a frequent concern in terms of overpopulation and pressure on the environmental carrying capacity (Smailes et al., 2002). On a large spatial scale, human presence is positively related to biodiversity (Di Giulio, et al., 2009), suggesting that people contribute to biodiversity improvement by species introduction and habitat diversification. On a smaller spatial scale, however, such as urbanization it alters the land cover and affects the natural habitat. Populated areas are characterized by land fragmentation (small patches), isolated by roads, settlements or intensively managed agricultural lands (Niroula and Thapa, 2005). Human activities both modify and model landscapes; and landscape structure, function and processes are related to land-use types and land-use intensities (Hietel et al., 2004). This implies that population density can be a proxy measure in a range of other factors, like labour costs, and it is a direct driver of land use change.

Several studies have also conducted in relation to land use/land cover change, vegetation abundance and population densities. Pozzi and Small (2002) consider the relationships of population density and vegetation as principal demographic and physical characteristics in urban and suburban areas of the USA. They found that the amount of vegetation diminishes with increasing population density, and suggest that it is possible to characterize land use (rural, urban, and suburban) based on population density. A recent study of migration and environment in Ghana identifies significant but weak correlation at large scale between demographic variable, migrations in this case, and vegetation covers and trends (Gesest et al., 2010). They argued that out-migration shows a more positive Normalized Difference Vegetation Index (NDVI) trend and migration flows could explain by vegetation dynamics and more strongly by population densities. The two aforementioned studies directly or indirectly examine the linkage between population densities with land use types and investigate the effects of population distribution on the environment. Some other studies also suggest that the correlation between population distribution and land cover and land use types are high and reliable (Yuan et al., 1998).

To wind up, the theoretical concept and the research reviewed in relation to population distribution and environment linkage mentioned above were used as a theoretical background in the forthcoming result and analysis chapter. Specifically for the sub-section that examine the factors of natural as well as social on population distribution and density.

2.2 Land Fragmentations (patches) Causes and Consequences

Land fragmentation is defined as the existence of a number of spatially separate plots of lands which are formed as a single units (Tan et al., 2006). Land fragmentation has often been seen an only ecological problem, but fragmentation also has a societal perspective, it is how human perceives landscape fragmentation and in how landscape fragmentation potentially influence human wellbeing (Di Giulio, et al., 2009). The inter-relationship of ecological and human dimension of landscape fragmentation becomes more evident when looking at the landscape where most people live, in sub-urban and urban areas.

In agricultural areas land fragmentation can be indicated as area per landowner, the most commonly used indicator of fragmentation (Dijk, 2003). However, ownership distribution alone does not give a complete picture of fragmentation; do not always correspond with the actual parcelling of the landscape. The actual use of agricultural lands by individuals can be quite similar, so differentiating individual parcels may be difficult. As a result, land fragmentation may not be associated with individual parcels. Dijk (2003) suggests that in some countries, the land use structure is much better than the ownership statistics, he took an example from previous socialist countries such as Czech republic, Slovakia and etc. In other cases, private landowners join forces and form family association, so, the number of users in this case an indicator of land fragmentation of agricultural land. The other indicator of land fragmentation mentioned is the overlap of the two previously mentioned indicators.

The advantage and the disadvantage of land fragmentation have analysed and discussed in various researches. At the farm level, several operational disadvantages emerge. The negative effect of land fragmentation impose significant cost in terms of output, create difficulties to work on like drainage, irrigation and soil improvement (Nguyen et al., 1996). Regardless of limited farm production, if each farm is physically separate from another by fence, ditches or hedgerows, this element together with farm infrastructure, contribute to a loss of productive land and aggravate land fragmentation. Some, however appreciated because it reduces the chance of the entire crop or natural vegetation being destroyed by floods, disease or hailstorms (Dijk, 2003).

The relationship between land and people is intense. People's standard of living, wealth, social status, and aspiration are all closely link to land. The size of land holdings gradually decreases when they fragmented to parcels, most probably because of population growth. Land fragmentation is a state of division of holdings into discrete parcel that are disperse over a wide area and usually formed as a single unit. Fragmentation occurs into parallel directions: vertical subdivisions and horizontal dispersion. The vertical subdivision leads to a gradual reduction in the size of holdings; the distance between farmsteads, where farm households are located and agricultural inputs are stored. Land parcel gradually increase because of horizontal dispersion, where land holding comprises more than one parcel of land in different spatial location (Niroula and Thapa, 2005). When land holdings and land parcels are fragmented, they got gradually smaller and disperse widely, they also argue that population growth is one of the major factors triggering land fragmentation. However, several other factors are also playing the roles.

In most developing countries like Africa, following the tradition reinforced by laws, land holding and some land parcels equally divided among sons, when they decided to live separately. Land fragmentation has thus become an on-going process, resulting the land holdings and land parcels getting smaller and smaller, and dispersed over successive generations (Mearns and Sinha, 1999). The high population growth in Africa induced land fragmentation because of the inheritance system, which generate land subdivision and aggravate land fragmentation. The research output focuses on forest ecology and management confirm that fragmentation in forest areas caused by human land use activities and thus population is an important determinant of fragmentation. Rapid economic development, as well as pressure on development from increasing human population, are also among the major underlying factors for tropical forest fragmentation (Abdullah and Nakagoshi, 2007).

In this part of the review, especially, those research works related to population distribution or density in relation to land fragmentations have used as a theoretical background during the investigation of population and land fragmentation. While some of the concepts mentioned above were not directly and broadly used in this paper, they are indirectly related to the topic under investigation and the review might be used to give insight into further research especially in examining whether land fragmentation can be used as a proxy measure for population estimate or not.

2.3 Review of Population Estimation Techniques

Several studies have focused on population estimates by relating socioeconomic and physiographic data. Factors often indicated for modelling population distribution in these studies are: elevation, slope, temperature, distance to transportation and land cover (Wu and Murray, 2005). An example for such a model was used to calculate China's population density at one kilometre grid cells (Tian et al., 2005). Representing data in grid form has at least three advantages: "(1) regular grids can be easily re-aggregated to any area arrangement required; (2) producing population data in grid form is one way of ensuring compatibility between heterogeneous data sets; and (3) converting data into grid form can provide a way of avoiding some of the problems imposed by artificial political boundaries" (Yue et al., 2003). Furthermore, estimating population based on grid cells has an advantage in which each grid contains an estimate of total population that is representative for that particular location. Meanwhile, Wu and Murray (2005) applied impervious surface fractions that derived from satellite imagery and a cokriging method to interpolate population density by modelling spatial correlation and cross correlation. Because of the introduction of new technology and the demand for high resolution population data, many recent studies focused on the use of digital simulation technology such as Dasymetric mapping to estimate raster based human population distribution (Tian, et al., 2005). This study also investigated the correlation between land cover data and other factors that affect population distribution and argued that land cover data contain sufficient information to infer population distribution. Dasymetric mapping is referred to as a process of disaggregating spatial data to a finer unit of analysis, using additional data to help refine locations of population or other phenomena (Maantay et al., 2007). The disaggregation process will result in areas of homogeneity that take into account the actual phenomena being modelled, rather than areal units based on administrative or other arbitrary boundaries. Another technique is Population Surface Modelling (PSM) and is used as a density estimation technique to convert vector point objects into a raster surface of values and has been applied using census data to model population density for all location (Harris and Chen, 2005), assuming population census block unit as point location to represent the population in the block.

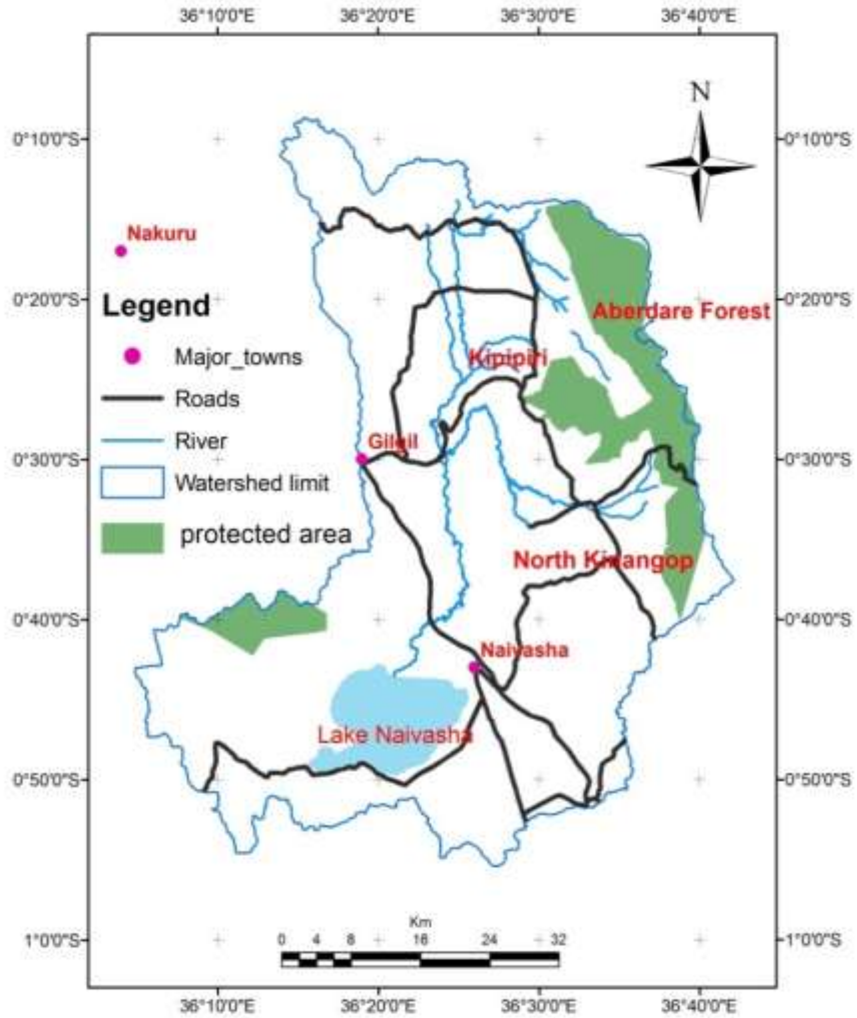
The methods of population redistribution and estimation techniques mentioned above are used as background information to achieve one of the objectives of the paper and helped to figure out the appropriate method to remodel population in the study area, considering the availability of data. For the population estimate in this study, weighted areal interpolation with the help of WGR is found to be appropriate. GWR is a local form linear regression considering the spatial varying relationships and produce a verity of parameter estimate (regression coefficient).

3. STUDY AREA, DATA COLLECTION AND PREPARATION

3.1 Study Area

Lake Naivasha watershed is located in the central southwest part of Kenya, approximately 80 km northwest of Nairobi, the nation's capital. It lies in the rift part of the Kenyan Rift valley and is found between latitudes 00° 10' to 00º 55' S and longitudes 36º 09' and 36º 40' E. It covers an area of 3400km² with a climate that is predominantly semi-arid (Becht et al., 2006)

Figure 3 Map of Lake Naivasha's Watershed (Study Area)



The lake Naivasha dominates the southern part of the Naivasha's watershed. It has an altitude of 1,890 m above sea level, lies on the floor of Africa's Eastern Rift Valley; it has a mean surface area of 140Km² and it is the second-largest freshwater lake in Kenya; its watershed is mainly a semi-arid environment with scarce surface and underground water resources. The area around the lake has witnessed major land use transformation following colonization of Kenya. At the beginning of 1900s the land use in the watershed changed from pastoral economy to large scale white settler farming and since independence in 1963 the area has registered rapid land subdivision. The land use changes since independence have led to rapid growth in population, human settlement, intensive commercial farming, tourism and geothermal production (Mireri, 2005). The higher part of the land is used mainly for wheat and cattle, around lake Naivasha vegetables are grown, closer to the lake edge mainly irrigated floriculture ventures can be found. The area to the west and east of the lake are occupied mainly by small holders growing maize, vegetables and pyrethrum by large grain farms (Becht, et al., 2006). Due to land use transformation since the independence of Kenya, much of the catchment areas around Kingnop and Kipipiri were settled by indigenous Kenyans. The transformation has continued over the years as large farms are sold to land buying companies, which later subdivided the land into small holdings. In the catchment households own up to a maximum of 10 acres of land (Mireri, 2005).

The population of the area according to the 1999 census is 568,481 and after ten years, increased to 654,973 for the water basin (taking into consideration, administrative boundary change in the study area). The majority of the watershed is in the valley province, it covers three districts, namely Naivasha, Nyandarua south and Nyandarua north containing a total population of 972, 511 and a total household number is 249,197(KNBS, 2010). The study area comprises of 29 locations and 62 sub-locations which population data were aggregated. Sub-locations are the smallest administrative unit the census count first aggregated.

To conclude, the information discussed about the study area has been considered in the study. In particular, the census population data for the study area was used to redistribute population at 1km^2 grid cells. The grid cells population distribution has an advantage over the census, because comparing census population over a long term is highly problematic due to boundary changes. This is because the only traditional response to census count has been to aggregate data in the administrative unit (Gregory, 2002). The population estimate of 1km^2 (more detail in the coming analysis section) using areal interpolation technique allows the comparison of data from many dates. Lastly, familiarity with the study area allowed the comparison of results for realistic values.

3.2 Sampling for Field Data Collection

The Lake Naivasha's watershed was stratified based on the existing (Afri-cover made by FAO in 1999) was produced from satellite imagery. This land cover/land use vector datasets was used to assign the sampling areas to record sample point and description of land cover. In addition the current satellite imagery (ASTER) was used, interpreted and segmented based on colour and texture interpretation keys using eCognition software. The existing land cover/ land use vector datasets was overlaid on segmented image. The segmented image polygons that match vector polygon datasets by at least 75% or fall inside the polygon were used for training. For field data collection purposes, some of these polygons were selected for ground survey for training data and for accuracy assessment. However, during the actual fieldwork it was impossible to implement entirely this sampling strategy due to mobility constraints. As a result most sampling were collected in a clustered manner along the roadside, fairly covered the watershed (Appendix_1). Samples were collected based on approximately 50m X 50m resolution considering the image object segmentation.

3.3 Spatial Data Coverage and Data Preparation

For this study both primary and secondary data were gathered. The major data set used were remote sensing, population, rainfall, altitude, slope, road network, soil and other ancillary data. Details follow in the subsequent sub-sections:

3.3.1 Remote Sensing Data

The geo-referenced, ortho-rectified and mosaic satellite image (ASTER 02/02/2008) was collected from the ITC geo-database. The image was re-projected to UTM zone 37 south map projections. The image is 99 per cent cloud free and it could thus be used for land cover classifications.

3.3.2 Population Data

For this study population data is basic requirements to achieve the objectives of the study. The recent 2009 Census reports (population of the watershed at sub-location level) and the census map (paper copy) of three district: the Naivasha, Nyandarua south and Nyandarua north covering the watershed were collected from the Kenya National Bureau of Statistics (KNBS). The Census maps of the watershed were scanned and geo-referenced to UTM zone 37 south (map projections) based on the control points collected from the map copy. To have a digital copy of the administrative units, on screen digitizing was employed and polygons were created in shape file format using Arc Map/Arc GIS software. The census populations were encoded for each sub-location (smallest administrative unit) polygon. In order to have a sub-location polygon perimeter and a shape area, file geo-database was created and the sub-location shape files were imported to the geo-database, as a result the perimeter and the area were automatically created for each sub-locations.

3.3.3 Physiographic and Road Network Data

3.3.3.1 Soil Data

Download from International Livestock Research Institute's (ILRI) database. It has prepared by Kenya soil survey (KSS) in 1982 and revised in 1997. The following soil types have used for analysis, the description is from the metadata:

Clayey: sandy clay, silty clay and clay texture classes.
Loamy: loam, sandy clay loam, clay loam, silt, silt loam and silty clay loam.
Sandy: loamy sand, sandy loam texture classes
Very clay: more than 60% clayey

3.3.3.2 Rainfall Data

Data on annual rainfall distribution in millimeters for ten years average (1990-1998 and 2004) for the watershed have used. The data has collected from the ITC archive.

3.3.3.3 Digital Elevation Model (DEM)

The DEM of 90m resolution downloaded from The NASA Shuttle Radar Topographic Mission (SRTM) website. The data has used to extract elevation and to calculate percent slope of the study area.

3.3.3.4 Road Network:

The road network data was also downloaded from (ILRI) website. It was originally digitize from topographic map sheets of Kenya, at a scale 1:50,000. The type of road class used in this study is a primary road type, which has classified by speed as 80km/h. The data was updated using Google earth image.

All the data mentioned above were originally in a geographic projection and at national level. First the data were clipped based on the study area and then re-projected to UTM Zone 37 and WGS84 datum in order to fit with other data.

3.3.4 Ancillary Data

The existing land cover data set (Afri-cover land use map produced by FAO for Kenya, 1999) were re-projected to UTM Zone 37 south map projection and WGS84 (datum ellipsoid) was used for sampling for field data collection. Geo-referenced and ortho-rectified high resolution (20cm) aerial photography image of 2009, covering an area of 300km^2 around Lake Naivasha were collected from Romaine Company in Kenya, which were used to collect some points for land cover classification and visual evaluation of the classification. The quick bird image covering an area of 100km^2 from the ITC geo-database and high resolution image of Google earth (spot-image, 2009 and 2010) were used for image interpretation and land fragmentation analysis.

3.3.5 Field Data Collection

Sample points and land cover attributes were collected from various sources for image classification and accuracy assessment. A field survey was conducted between 16.09.2010 and 07.10.2010, which yielded a set of 221 sample points. An additional 106 survey points were taken from the 2008 sample sets collected by Were (2008) for the same area. Moreover, 56 more points acquired from aerial photographs and high resolution satellite imagery. From these, a total of 383 sample points were collected.

The sample points from the field survey were collected using the hand-held GARMIN 12X Global Positioning Systems (GPS) and a mobile GIS (IPAQ 4700). The ASTER satellite image was converted to Enhanced Compression Wavelet (ECW) and uploaded on Mobile GIS and used for navigation during field data collection. These points were collected to determine the land cover and land use types, to assess the field data with the image characteristics, to assess land fragmentation for specific areas for later image segmentation and land fragmentation analysis and for the validation of land cover map. The 50m by 50m visual estimate of land cover and land use attribute (percent cover cropland, grass, trees, settlement, horticulture and shrubs) were made and recorded in the field observation sheet (appendix_2) later on prepared in excel and converted to ArcGIS shape file format.

4. RESULTS AND ANALYSIS

4.1 Image Classification

In this study the image classification was done using object oriented classification method. In the case of object oriented classification, the analysis is not focused on single pixel values, but on group of pixels, such groups of pixels are called object, the object is created in the course of the segmentation process, and then follows the classification (Lewiński et al., 2007). Image objects can be identified on the basis of Patches of spectrally similar pixels, this often referred to as segmentation (Aplin and Smith, 2008). Segmentation of the image into homogeneous areas takes into account spatial attributes such as shape, compactness, size, and smoothness in addition to the inherent spectral information (Zhoua and Troya, 2008). The object based approach to classify the ASTER satellite image of 15m spatial resolution was implemented in eCognition Version 8 software. The first step was the creation of false color composites using green red, and near-infra-red bands and the execution of multi-resolution segmentation using the parameter specified (Table 2). The scale parameter was determined using a method proposed by Drăguţ et al,(2010). It is used to estimate for multi-resolution image segmentation of remotely sensed data. Local variance of the image object increases with the decrease in scale parameter as the homogeneity of objects in the scene decreases. The peak of Rate of Change (Figure 4) indicates where object reached meaningful organization in terms of their variation of homogeneity. The segmented image produced 55,943 polygonal objects with similar spectral and neighborhood Characteristics.

Figure 4 Estimation of Scale Parameter

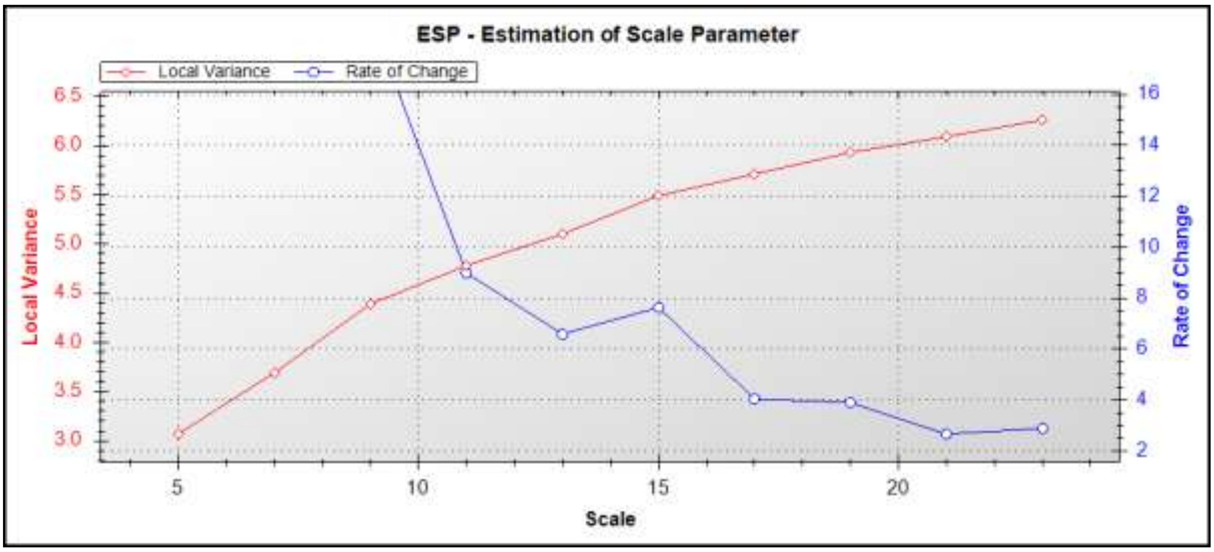


Table 2 Image Segmentation Parameters

Satellite image	level	Scale	Homogeneity criteria		Shape ratio	
			Colour	Shape	Compactness	Smoothness
ASTER	1	15	0.8	0.2	0.3	0.7

Training data was uploaded as a thematic layer in order to have a sample object for image classification. The feature space was defined and the standard nearest neighbour algorithm was used which automatically generates multi-dimensional membership functions based on the sample objects. In this process 90% of the data was used for image classification, that had been collected from field surveys and the sample extracted from the aerial photo and the high resolution satellite imagery. The remaining 10 % of the data was used to create a training and test area (TTA) mask for assessing the quality of the land cover map. Visual inspection and manual editing was done in order to have a good quality map.

4.1.2 Land Cover Map

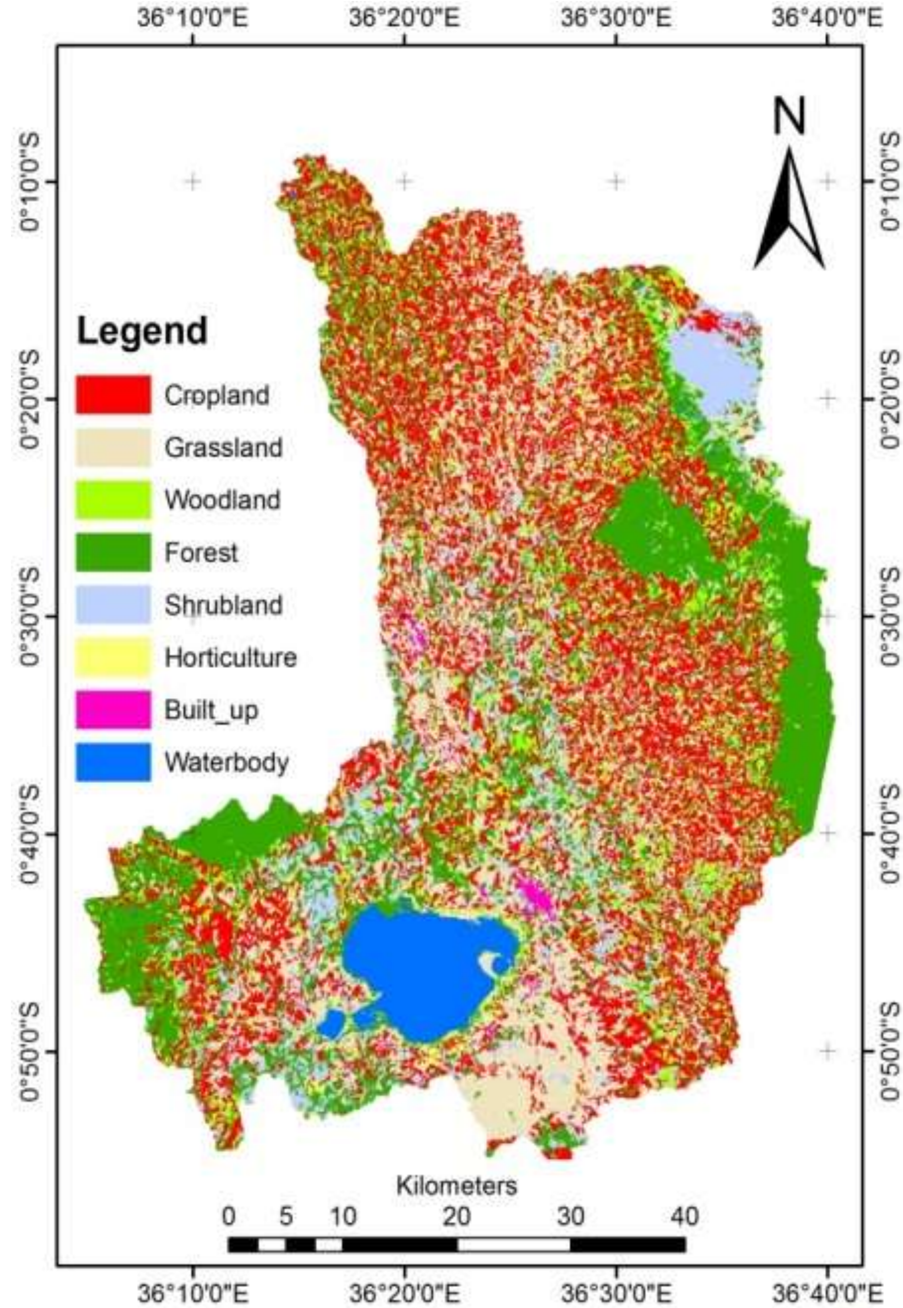
The land cover classification shown in this sub-section are consists of eight land cover/land use types: cropland, horticulture, forest, grassland, woodland, shrub-land, built-up and water-body.

Table 3 Description of Land cover/Land Use Classes

Code	Land cover	Description
01	Croplands	Area dominated by crops > 50% estimate
02	Forests	Area dominated by trees > 70% estimate
03	Grasslands	Areas dominated by Grass > 50% estimate and Shrub or tree mixed
04	Woodlands	Trees 50-60% estimate and shrubs or grass mixed
05	Shrub lands	shrub land > or = 50% estimate and trees or grass mixed
06	built-up	Areas with commercial or residential structures and/ or constructed materials > 60% estimate
07	Water-body	Area covered by open water, lake
08	Horticulture	Areas identified as Flower farm

The classification result (Figure 5) shows that 29.5% of the study area is cropland; 22% of the area is grassland; 24% is forest; 9.2% is woodland; 10.3% is shrub land; 0.4% is horticulture; 0.5% is urban land and 3.9 of the areas are covered by water body. Forest, woodland and Shrub land summed up to 43.6% share of the total area. Almost one third of the areas are served as agriculture.

Figure 5 Lake Naivasha's Watershed Land cover/Land Use Map, 2010



4.1.3 Accuracy Assessment

Accuracy assessment methods can produce statistical output to check the quality of the classification result. It is an important feature of land classification process, indicates map quality, reliability, implication to the users and an insight into the thematic uncertainty (Rogan and Chen, 2004). An assessment to test the accuracy of the results of the classification was performed using the Training Test Area (TTA) mask for the classification result of eight classes. Test areas are used as a reference to check the classification quality by comparing the classification with reference values based on pixels (Definiens AG, 1995-2009). The overall accuracy of the classification was 87%, and the overall kappa statistic equalled 83%. The accuracy assessment result is coherent with acceptable classification accuracy suggested by Congalton (1991). The kappa statistics imply that 83% of the classification agreed with the reference data, leaving only 17% chance for disagreement. The accuracy assessment of the error matrix is based on TTA masks are listed in Table 4 with users and producer accuracy for each class.

Table 4 Error Matrix Based on TTA Mask for Land Cover/Land Use Map

User \ Reference Class	Crop land	Forest	Wood land	Shrub land	Grass land	Horticulture	Built-up	Water body	Sum
Cropland	6591	0	0	0	823	52	0	0	7466
Forest	616	1047	0	0	95	0	0	0	1758
Woodland	298	0	1131	0	0	0	0	0	1429
Shrub land	215	0	0	2821	0	0	0	0	3036
Grassland	641	0	0	0	8074	0	0	0	8715
Horticulture	0	0	0	0	0	188	0	0	188
Built-up	0	0	0	0	1052	0	1454	0	2506
Water body	0	0	0	0	0	0	0	1278	1278
Sum	8361	1047	1131	2821	10044	240	1454	1278	
Producer	0.79	1	1	1	0.80	0.78	1	1	
User	0.84	0.60	0.79	0.93	0.93	1	0.70	1	
KIA Per Class	0.70	1	1	1	0.71	0.78	1	1	
Overall Accuracy	**0.87**								
KIA	**0.83**								

The producer accuracy showed the percentage of the reference pixels that had been explained by the extracted pixels; a user's accuracy indicated the percentage of pixels that had been correctly extracted and; kappa statistics (the measure of reproducibility) assessed the probability of chance agreement between the reference dataset and the classified land cover map. The matrix essentially yields the indication of land classification quality in relation to the TTA mask. For instance, in the cropland category 8361 pixels contained in the TTA mask were cropland. Out of these 6591 pixels were correctly classified as cropland, however, 616, 298, 215 and 641 pixels were misclassified as forest, woodland, shrub land and grassland respectively. Similarly, out of 1047 pixels that were classified as forest, all are correctly classified, equally true for woodland, Shrub land, built-up and for water body. Cropland and grassland shows the poorest result, they confused each other and with other land cover types.

In the next sections, the classified land cover/land use map is used to produce the aggregated land cover/land use for the population estimates for different land cover types, as well as to derive a weighting factor based on aggregated land cover information. Finally it is used to estimate population at 1km^2 grid cells.

4.2 Land Cover/Land Use Data for Population Estimate

Land cover data contain sufficient information to estimate population distribution and it can be used to model spatial pattern of population density, because it is highly correlated with many other factors (Tian, et al., 2005). Most studies have used land cover data that was derived from satellite imagery as weighing factor to map population distribution using areal interpolation (Yuan, et al., 1998), they argued that areal interpolation using land cover data as weighing factor have proven to be relatively accurate and makes land cover weighing the normative approach to areal interpolation. “Areal interpolation refers broadly to technique that assign data from one or more set of geographic areas to which the data are aggregated (the source) to another incompatible and super imposed set (the target) using spatial algorithms."(Reibel and Agrawal, 2007). Areal interpolation is a well-designed and, with the use of Geographic Information System (GIS), gives solution to matching two spatially not coincide data set (such as land cover map and census block map). In this section of the study, land cover data that was derived from the ASTER satellite imagery of 2008 were used to reclassify land cover and to calculate weighing factors for later use in population estimate in 1km² grid cells.

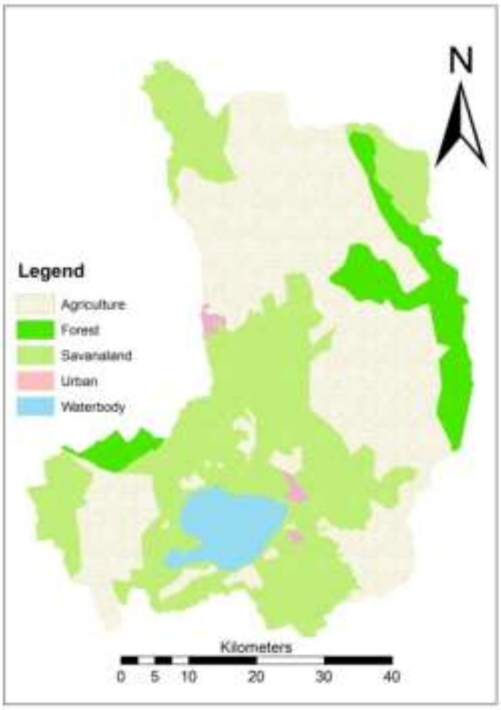
4.2.1 Reclassification of Land Cover Data

For statistical regressions fewer categories and more observations generally provide more reliable results. As we seen in the classification above some categories of land cover are mixed with each other such as, cropland with grassland, woodland with shrub land and horticulture with savanna land and cropland. In the case of horticulture predominantly surrounded by cropland and in most areas located near to settlement, and it takes the characteristics of the nearby land cover, as a result some categories of land cover may not be significantly different with respect to population distribution. Therefore, based on the similarities of possible population densities, the land cover map above reclassified into five categories:

1. Agriculture – including cropland and horticulture
2. Savanna land –including woodland, shrub land and grassland
3. Settlement - classified as built-up
4. Forest - classified as forest
5. Water body – classified as water body

The reclassification (Figure 6) was based on the dominance of the land cover category and areas were delineated to derive land cover polygon to facilitate an overlay operation with census sub-location map. Areal interpolation algorithm in combination with statistical regression is employed. The algorithm is used to estimate population for the reclassified land cover/land use types and to get a weighing factor for grid population estimate.

Figure 6 Reclassified Land Cover/Land Use Map, 2010



4.2.2 Deriving Weighing Factor and Estimating Population for Land Cover/ Land use Type

The data used in this part of analysis are the digital census map (sub-location level) and the reclassified map. The areal interpolation methods are employed; it consists of a series of steps, mostly performed in a GIS environment using the ArcGIS 10 platform. A simple method of areal interpolation is to weight the variable's values by a ratio derived from the relative areal measurement of the two types of zones (source and target). Areal weighting is based on the assumption that population is distributed homogeneously throughout the source zone (census unit). The amount of population estimated to be in target zone is assumed to be proportional to the amount of area in the source zone versus the target zone. The ratio of the area of source zone to target zone is then applied to population in the source zone to yield the population total to target zone (Maantay, et al., 2007). The following section gives details.

In ArcGIS, the census population map at sub-location level is superimposed on the reclassified land cover map. When the census sub-location map (source unit) is overlaid by the reclassified land cover map (target unit) and intersects using intersection operation in ArcMap, each source unit is broken to smaller subunit. This subunit is the combination of a source unit ***(i)*** and a target unit ***(j)***. The total area of ***(Z_{ij})*** for a certain land cover type within a source unit are the sum of all subunits of the same ***(ij),*** where ***i*** is the ***i^{th}*** sub-location and ***j*** is the ***j^{th}*** land cover type. From the result of the operation a data matrix (appendix_3) is prepared to facilitate regression analysis. The data matrix contain ***m*** rows and ***n+ 1*** columns, where ***m*** is a number of source subunits in a working area and ***n*** is the number of land cover types. Each element of the matrix from the second column onwards is the total area ***(Z_{ij})*** of the ***j*** land cover type within ***i^{th}*** sub-location. The first column is the total population counts ***(p_i)*** for ***i^{th}*** sub-location, ***m*** varies from sub-location to sub-location, and there are 102 subunits for the whole study area. The number of land cover types is restricted to 3, leaving out water and forest as being uninhabited. The land cover types used for this analysis are agriculture land, savanna land and urban land cover.

The data matrix (102 X 4) is imported into PASW statistics 18 and analysed by OLS regression model with the non-constant options. Without constant is chosen because there should be no population if there is no area. The Ordinary Least Square (OLS) regression model is a well acknowledge formula to represent counts (Xie, 1996). The model can be written as:

$$P_i = \sum_{j=1}^{3} (b_j * Z_{ij} + \varepsilon_i) \qquad \text{eq.1}$$

Where ***Pi*** is the total population in ***ith*** sub-location, ***bj*** is regression coefficient (weighing factors) for ***j^{th}*** land cover type and ***Zij*** is area of ***j^{th}*** land cover type in ***i^{th}*** sub-location. The sub-location populations were then regressed on the areas of inhabited land cover types.

Table 5 Regression Result for aggregated Land Cover/Land Use Types

Parameter	Regression coefficient *(b)*	*Sig.*
Agriculture	229.3	.000
Savanna land	161.4	.000
Urban	5861.1	.000
	Population	
Water body	0	omitted from analysis
Forest	0	omitted from analysis

The model gives weigh factors for grid population estimate, significant at $p<.001$; the multiple coefficient of determination (R^2) is 0.61, indicates 61% of the total sum of squares deviation in the population distribution in sub-location over all study area can be explained. The regression coefficients are commonly used as weighing factor to estimate population density and can also be considered as population per unit area^2 (Reibel and Agrawal, 2007). The regression derived weigh (estimate) for land cover data are given in Table 5. The regression result reveals that average population estimate for agricultural land, savanna land and urban land cover at 1km^2 area is 229, 161 and 5861 respectively, the forest and water body land cover not included in the analysis were given zero(uninhabitable areas). These results show that land cover data includes most of the population distribution information. In order to use the regression coefficients as a weighing factor for further analysis for grid population estimate, normalizing their values between 0 and 1 is important. Therefore, the normalized value between 0 and 1 is 0.022, 0.016 and 0.586 for agriculture; savanna land and urban land cover types respectively. Once the result is normalized, they are applied to 1km^2 (1km x 1km) grid cells to estimate population and generate a population surface map.

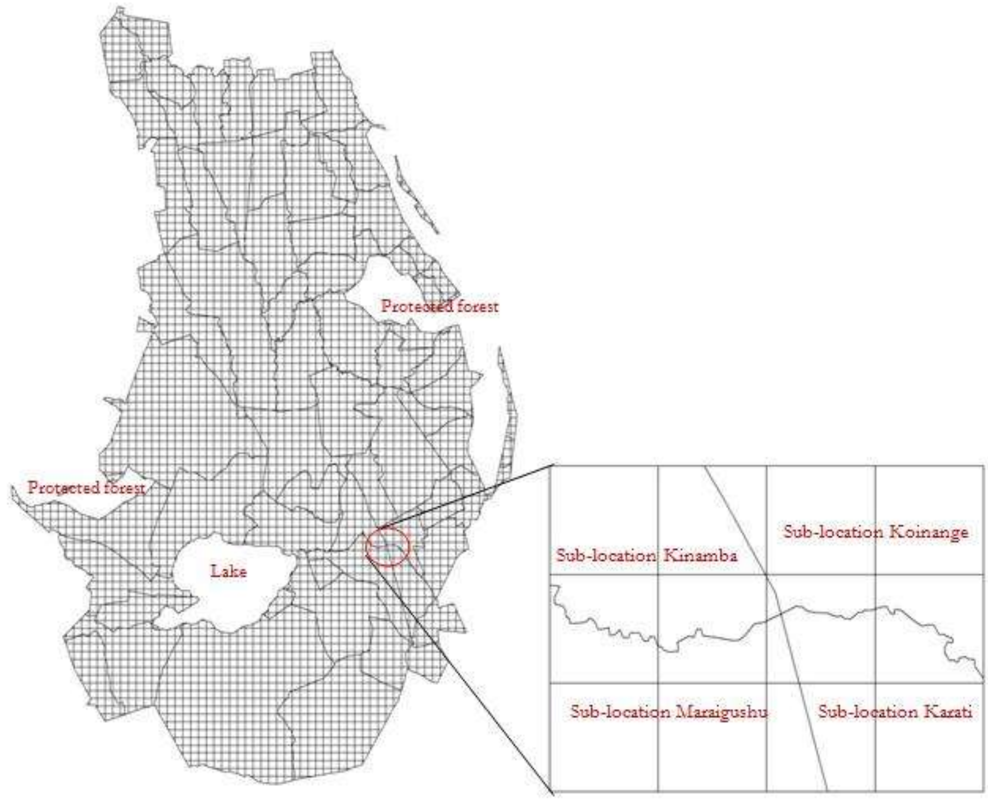
4.2.3 Remodeling Census Population in Grid Cells

Transferring data from one set of geographic unit to another set of non-coincidence unit is often necessary for spatial analysis. For instance, we might have population in census unit and wanted to transfer to another smaller unit, that is the area that possibly intersect (overlap) with the source unit, the overlap area may be some part from one census tract and some other part from another census tract and we may be interested in a number of populations in that unit. In the previous section we have seen some part of the method that how to assign population for different land cover types that are shared by different sub-location, in fact, the analysis was done to derive weighing factors for grid population estimate. In this part of the paper, we try to estimate population in grid cells, using census population data of 2009 for the watershed as source unit; and 1km^2 constructed grid areas as target unit and weighing values that was derived from land cover data is used to refine the population distribution in grid cells. The analysis was employed based on weighted areal interpolation algorithm coupled with Geographically Weighted Regression (GWR) to minimize the errors that could emanate by areal interpolation.

4.2.4 Weighted Population Estimate for Grid Cells

First, one kilometre resolution grids were constructed with ArcMap 10 using grid index tools, the extent of the grid were defined based on the census population map of Naivasha watershed. The census map at sub-location level (the smallest administrative unit in Kenya) with population totals for each sub-location was used to produce population surface showing the estimated number of people residing in each grid cell. One of the key processes in this analysis is map overlay analysis. An overlay operation usually requires two input data layers (a source layer and a target layer) and generates a composite layer. In this section of the analysis, the census map of Naivasha watershed at sub-location level overlaid with 1km^2 grid cells to generate an intersection layer as shown in (Figure 7). From the analysis in Table 5 the population density for the different land use/land cover classes are assigned for each grid unit before the intersection process.

Figure 7 Intersections Map Derived from Sub-location Population polygons and Grid Cells Polygons



The census map was then, sub-divided in to 4957 subunit; the intersection operation produces population attribute for intersected units and grid area attribute in proportion to the area intersected. Note that areas of uninhabitable places were omitted from the analysis. The Geographic Weighted Regression (GWR) of ArcGIS Spatial statistics tools was used to improve population estimate. The keystone of geographic thinking is the assumption that spatial phenomena will vary across a landscape. Regression based models largely ignores this assumption, much to the loss of spatially varying relationships. GWR generate separate regression equation for every feature analyzed in a sample dataset as a means to address spatial variation. Every spatial feature analyzed is due to the influence of the surrounding features. Moreover, GWR is a variation of normal regression where a weighing factor is used. Therefore, WGR is used to calculate a regression coefficient for all 4957 subunit, using population attribute for the intersected units as a dependent variable; intersected grid area attribute as independent variables and the population density of the land use/land cover as weighing factor. The results of this analysis yield an improved estimate population for the intersection layer units than the estimate by simple areal weighted interpolation without GWR, the difference is explained in the subsequent section. The regression coefficients generated for each of the 4957-fragmented unit are the rough estimate. The results should be adjusted to preserve the source zone population. The formula is adopted based on (Gregory, 2002; Martin, 1989) and can be written as:

$$Y_i = \sum_{j=1}^{n} (Z_{ij} * Gij(w) + \varepsilon) \qquad \text{eq.2}$$

Where dependent variable Y_i the total population count for the i^{th} sub-location, G_{ij} is the regression coefficient (estimate) for the j grid in the i^{th} sub-location and independent variable Z_{ij} is the area of the j^{th} grid in i^{th} sub-location, w is weigh factor.

The regression result (appendix_2), shows that the multiple coefficient of determination (R^2) is about 0.72, it indicates 72 % of variation in the sub-location population can be explained by the model and 28% of the population of the sub-location is not explained by the model. From this analysis we have some negative population estimates. To address the negative estimate we have adopted the formula used by Yuan, et al. (1998). All the estimates of the subunit which was a negative number are raised by a negative value of the lowest negative estimate.

$$gij = Gij - Gk \quad \textbf{eq.3}$$

Where gij are the adjusted estimate for i^{th} sub-location in j^{th} grid, Gij are the estimate from the regression for i^{th} sub-location in jth grid, Gk is the lowest negative estimate. The lowest estimate *(gij)* will now be zero, which means that there is no population in that grid subunit or grid. All other adjusted estimates are positive and the ratio between them remains the same as before the adjustment. To account for the proportion of sub-location population not predicted in the model, the ratio of the predicted population ($\sum gij * Zij$) and the census count is used to adjust the estimate (Floweredew and Green, 1989). The formula for scaling the population densities is written as:

$$Rij = \left(\frac{yi}{pi}\right) gij \quad \text{eq. 4}$$

Where Rij is refined population density for i^{th} sub-location in j^{th} grid, yi is population for i^{th} sub-location, pi is the predicted total population for i^{th} sub-location. After scaling, the some of the product of refined population density *(Rij)* and the area *(Zij)* for i^{th} sub-location were proven to be equal to the sub-location population *(yi)* reported in the census 2009.

$$Yi = \sum_{j=1}^{n} (\mathrm{R}ij * \mathrm{Z}ij) \quad \text{eq. 5}$$

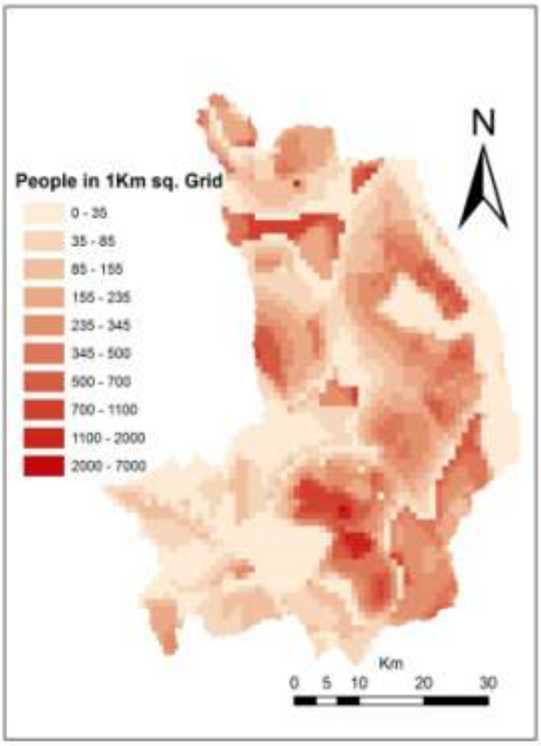
Once the refining is finished, the product of the refined population density (Rij) and the area *(Zij)* gives the population estimate for grid subunit. The remaining operation is adding the result of grid subunit populations to get $1km^2$ grid populations. This operation was done easily using the summarize operation in ArcGIS and joining the summarized table to the grid attribute table. The final output is then the modelled population distribution at $1km^2$ resolution map (Figure 8).

4.2.5 Population Distribution in the Lake Naivisha Watershed

Comparing this modelled grid based population map with census based map on population counts (Figure 9), it is evident that the grid based population map carries much more spatial information. Figure 8, illustrate that those uninhabitable areas such as Lake and protected forest areas shown zero population density, whereas in census map (Figure 9), assuming population distributed equal over the administrative unit and as a result even those uninhabitable areas indicate some density, which is not a representation of the spatial distribution of the population. The rasterized population map (Figure 10) shows a continuous distribution of the population and it has an advantage to integrate with other GIS raster data.

The method used in this part of the section can also be applied to other types of socio-economic variables. The analysis is based on a technique of weighted areal interpolation with the help of GWR using remotely sensed land cover data as an influential factor for population density.

Figure 8 Lake Naivasha's Watershed Population Distribution Map at 1km Resolution



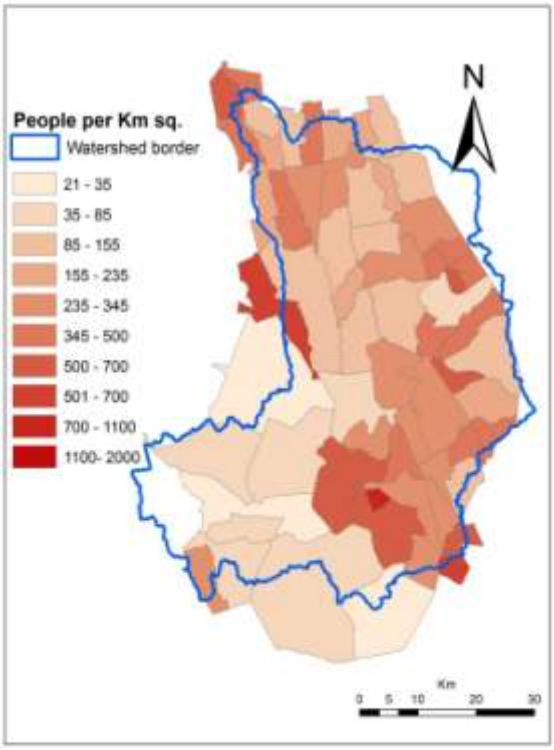
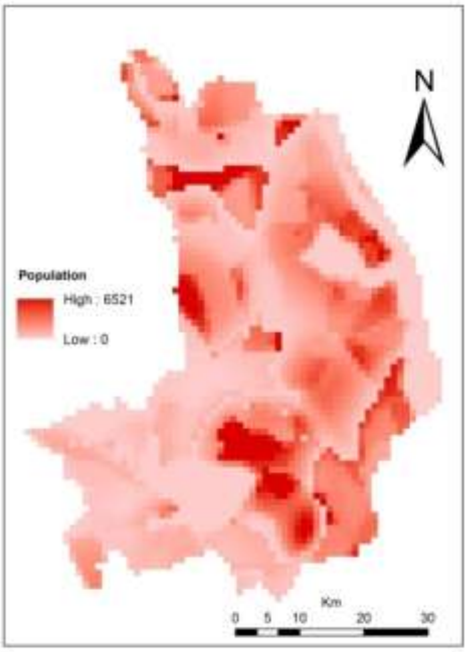


Figure 9 Lake Naivasha's Watershed Population Map based on Census Count by Sub-location.

Figure 10 Lake Naivasha's Watershed Population Density Surface



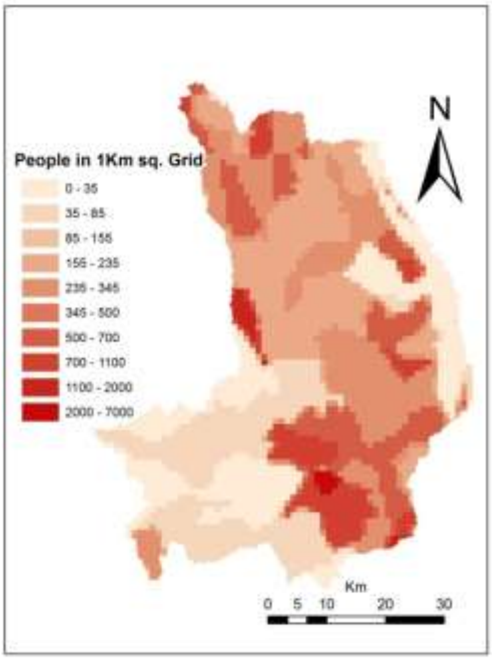


Figure 11 Population Density Map using Areal Weighted Interpolation without WGR

Population for each intersection unit from the composite layer can easily be calculated using areal weighting interpolation technique without using GWR, which is by multiplying the sub-location population by the ratio of the area of the intersected unit to the area of the sub-location. The population estimate for a given target zone (grid) is the sum of these weighted counts across the set of sub-location intersected unit that exhaust the territory of the grid.

$$Gp = \sum Gij * Zij/As \qquad eq.6$$

Where ***Gp*** is grid population estimate; ***Gij*** is population of i^{th} sub-location in ***j*** grid; ***Zij*** is area of i^{th} sub-location in ***j*** grid and ***As*** is the area of the i^{th} sub-location. Finally, the intersected unts population estimate summed across to have grid population. The result is seen in Figure 11. Majority of the grid that had not intersected by sub-location boundary have got similar values and the error margin is very large. However, weighted areal interpolation technique coupled with GWR (Figure 8) produce better result. GWR improve the estimate by considering spatial verifying relationship to produce separate regression coefficient for each intersection unit. Kernel function is embedded in the WGR to fit a smoothly tapered surface by cross validation and correlation of geographic features. So as, regression coefficient produced by WGR are smoothly distributed toward to the low density locations, in most cases holds the reality interns of population distribution. The coefficient surface (Figure 10), generated using the GWR tool is also helpful for identifying the spatial pattern noticeable in the study area.

4.2.6 Assessing the Validity of the Model

Since a sub-location population is remodelled using a weighted areal interpolation technique within its own boundaries for grid cells, the sum of the estimated grid population in a sub-location and the true population for that sub-location is always equal. This means that the official total census population is preserved; population is neither reproduced nor diminished. However, this might not indicate the actual population distribution in the study area. As a result, assessment of the reliability of the result against the real world population is important.

In order to assess the reliability of the result, artless but useful image and photo interpretation technique was introduced. First, a sampling of grid cells was conducted. According to the standard formula for necessary sample size, the sample size was determined. The formula can be written as:

$$\mathrm{SS} = z^2 * \frac{p(q)}{c^2} \qquad eq.7$$

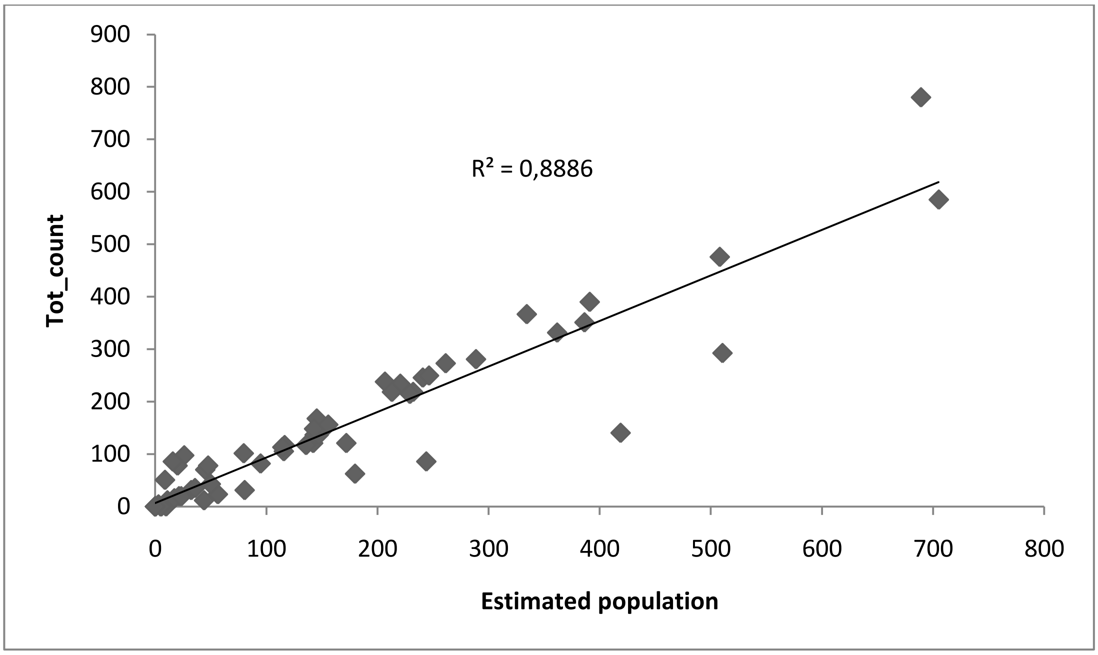
Where ***SS*** is sample size; **z** is confidence level; ***P*** is the percentage picking a choice (sampling ratio); ***q*** is ***1-p*** and **c** is the confidence interval. When the sampling precision is more than 90%, the errors is less than 10% and sampling ratio is 0.5, the necessary sample size is 68 (Thompson, 2002). Therefore, a total of 72 (1km x 1km) sample areas were selected from the study area using the constructed grid that has been used for population estimate. The selection of the area was done by random selection method specifying minimum distance of 2km using ArcGIS environment. The selected areas (grid cells) were converted to Keyhole Mark-up Language (KLM) in order to facilitate superimposition over Google earth image. For a total of 72 grid cells building is counted assuming every dispersed building/house is considered to be a household. Only spot image 2009/2010 and areal photo of 2009 were used for interpretation (counting). The counted building is tabulated in respective of the estimated population and multiplied by average household size (3.9) of the study area. The average household was calculated: dividing total census population of the study area by household number of the same. The population derived from Google earth using the above mentioned method considered approximately as actual population for grid cells. It is the only sources of information and closest substitute for ground validation of this study to compare the estimated population of the grid cells with real population distribution in the study area.

To quantify the errors introduced in the population estimate the following formula of Root Mean Square (RMS) error was used. The formula is based on the standard equation. The result relies on the average difference between the estimated value and actual value (population count based on Google earth image).

$$E^{RMS} = \left[\frac{1}{m} \sum_{m} (Ye - Ya)^{2} \right]^{1/2} \qquad \text{eq.8}$$

Where $\boldsymbol{E^{RMS}}$ is the *RMS* errors $\boldsymbol{m}$ is the number of grid cells, $\boldsymbol{Ye}$ is estimated population for grid cells and $\boldsymbol{Ya}$ is actual population in that grid cell derived by counting. The calculated result shows that only 11 % errors is observed, this indicate that 89 % of the estimated grid cell population is matched with real population distribution in the watershed. The result was also assessed using scatter plot (Figure 12). The result shows, the multiple coefficient of determination (R^2) is 0.888, it means that 89 % of actual population distribution is explained by the model for different spatial locations at 1km^2 grid cells; only 11% of the population distribution not explained which is almost identical to the RMS error result. The association between estimated population density and population density derived by counting are graphically well represented, the majority of the dots very close to the theoretical line, indicating most places have got a perfect match between estimated population and real world population that had been derived by counting, only few dots are far apart, indicating only in a few places mismatch have been observed.

Figure 12 Scatter Plot Showing the Relationship between Estimated Population and Population Count



In summary, the result calculated by RMS errors and Scatter plot function indicates the reliability of the model and therefore, the population estimate in grid cells produces much more detail spatial information if we compare to census map. As a result the population estimate in a grid cell in this study can also be used for further analysis. The next sections of the analysis will correlate land fragmentation, road and physiographic variables with population distribution based on the result of the model.

4.3 Land Fragmentation Analysis

The interrelationship between human population distribution and land fragmentation becomes evident when looking at the landscape in areas with a high population density. Ecologists agree that fragmentation changes the landscape and being measured by interior to edge ratios, patch shape, total patch boundary length, isolation, connectivity and patch numbers. Despite the increased use of measures of landscape structures experts do not agree on how to measure landscape patterns (Bogaert, 2000; Geneletti, 2004). This section of the study tries to focus on characterizing land fragmentation (patch) in different landscapes using fragmentation size and numbers of fragmentation for analysis in relation to human population density. This was done for highly populated, moderately populated and low populated areas of the watershed. The correlation between population density and land fragmentation at 1km^2 areas will be examined.

4.3.1 Extracting Land Fragmentation

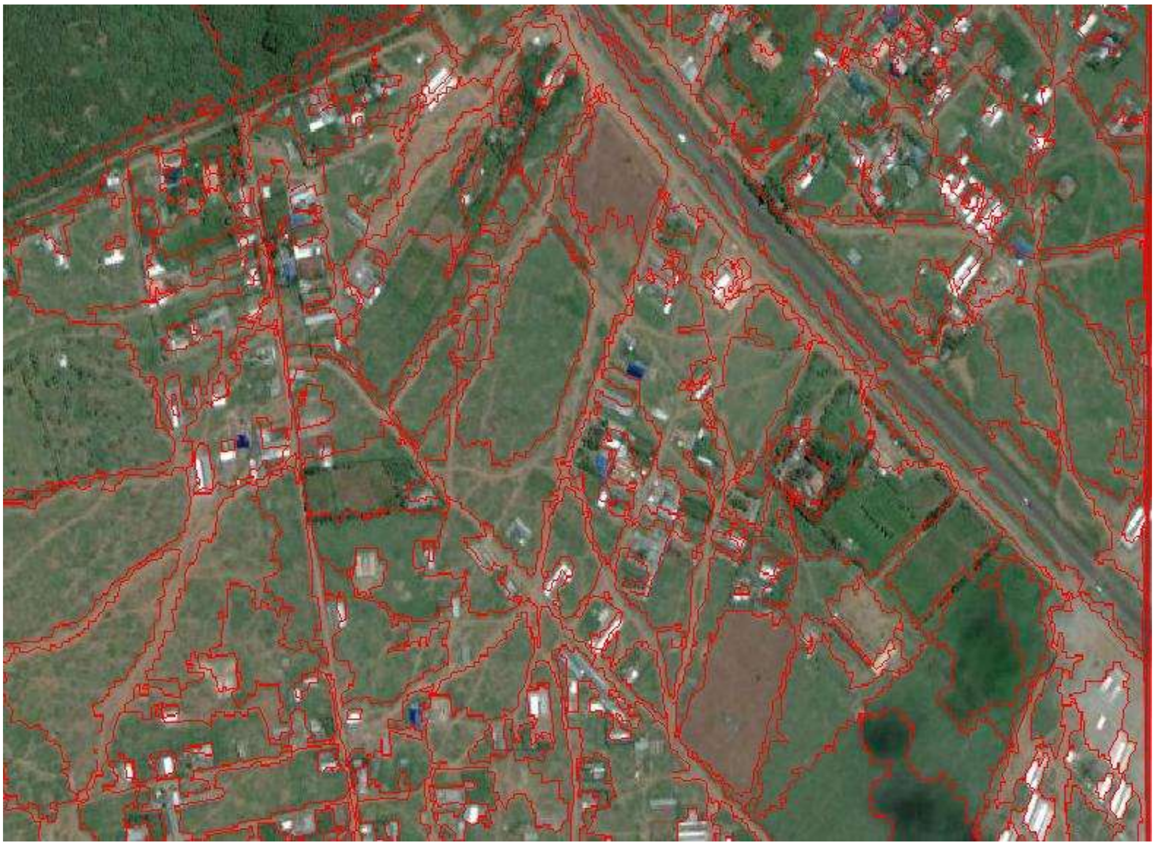
Spot images of 2009 and 2010 (2.5m resolution) from Google earth were used. The 1km^2 grid we have made was converted to KLM to superimpose on Google earth image, the KLM data set were imported into the Google image and a total of 38 (1km x 1km) areas were randomly selected from the area of 2009/2010 spot image coverage and saved for fragmentation analysis. The high resolution spot image covers one third of the study area. The number of samples was determined proportionally to the area coverage. In order to extract land fragmentations or patches, image segmentation was employed using eCognition 8 software. The parameters (Table 6) for image segmentation were determined using try and errors until the optimal were detected for fragmentation.

Table 6 Image Segmentation Parameters to Segment an Image for Land Fragmentation Assessment

Satellite image	**level**	**Scale**	**Homogeneity criteria**		**Shape ratio**	
			Colour	Shape	Compactness	Smoothness
Spot-image	1	80	0.8	0.2	0.3	0.7

The segmentation produced the polygon object of isolated fragmentation or patches based on similar spectral characteristics. The segmented objects were exported to ArcGIS environment for further analysis. The segmentation of the image might be because of roads, individual holdings or other similar activities. Figure 11 shows the segmented image for populated area showing land fragmentations (patches). The land fragmentation shown is mainly due to individual holdings, roads and due to other human interference with the landscape.

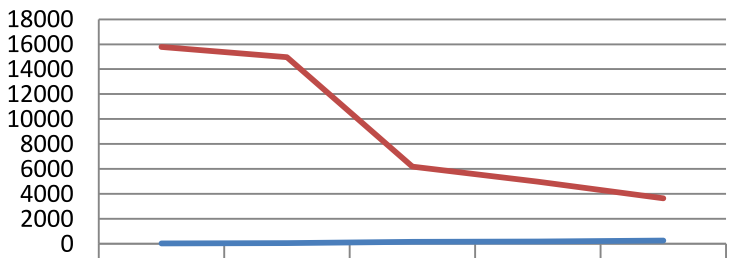
Figure 13 Map Showing Segmentations on Populated Areas



4.3.2 Characteristics of Land Fragmentation in Different Landscape

Fragmentation of land of different landscapes has been assessed considering representative areas in terms of settlement (populated area, and sparsely populated areas), the segmentation result reveals that those landscapes used for settlement show more land fragmentation compare to others. These indicate that human influence on the surrounding environment in terms of road construction, building construction, or cultivation activities and other related human activities. Areas of sparsely populated landscapes show relatively less fragmentation when compared with settlement areas; moreover, areas of non-populated landscape show very little fragmentation.

Figure 14 Land Fragmentations and Average Fragmentation Area in Different Landscape



	1	2	3	4	5
Frag	32	53	172	202	270
Average_frag_area (m2)	15782	14973	6191	4978	3642

As shown in figure 14, the average size of the fragmentation in selected areas decreases when number of fragmentation increase. The average patch size 3642 square metre in the area indicated above is predominantly settlement areas, but the number of fragmentation is about 270, in the area of non-populated area is observed 32 fragmentation, however the average area of the fragmentation is 15782 square metre. Average fragmentation area is influenced by a number of fragmentations; it is also assessed in relation to population distribution.

Figure 15 Frequency Graph for Populated Areas

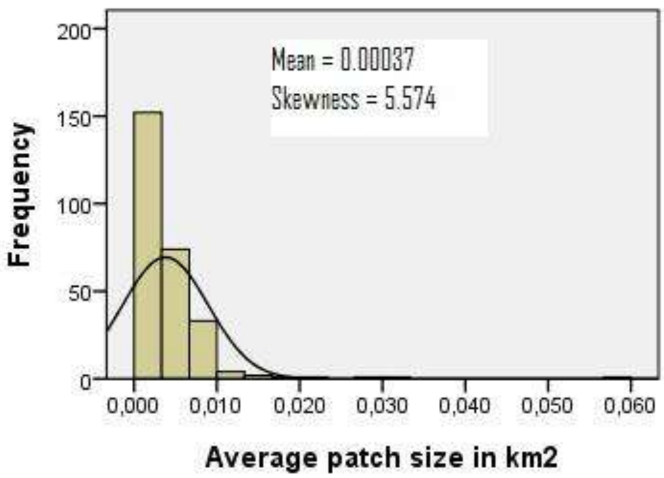
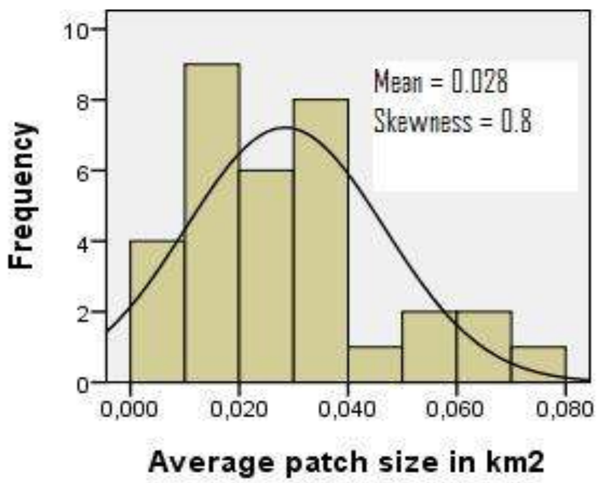


Figure 16 Frequency Graph for Non-populated Areas



The frequency graph 13 and 14 for different landscape (populated and non-populated areas) reveal that those area of settlement (Figure 15), the frequent fragmentation size are clustered at the lower end, the value of skew is positive about 5.574, indicating that the high proportion of the fragmentations are smaller in size and below the average size . However, for the area of non-populated (Figure 16) the frequent fragmentation size are around the average size, the value of skew is about 0.8 nearly showing symmetrical distribution, this indicate that the high proportion of the fragmentation are relatively bigger in size.

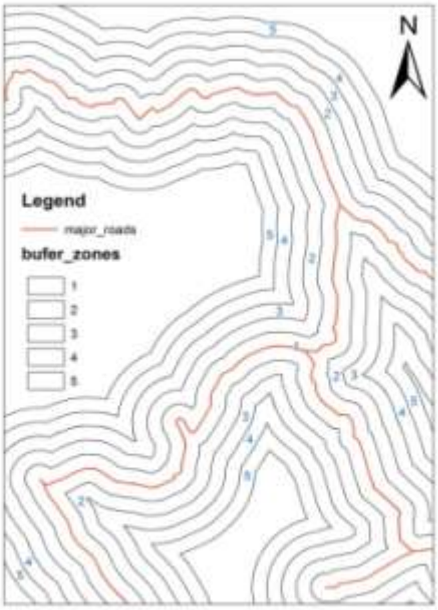
4.3.3 Correlation between Land Fragmentation and Population

Correlations measure indicate how variables or rank order are related (Field, 2009). The correlation of the total population of the area with the number of land fragmentation of the same area was assessed using Pearson's correlation coefficient in the SPSS software. The grid population estimate in the previous section and the segmentation of the respective grid are tabulated to facilitate for the correlation. A total of 38 observations were used covering 38km² area of fairly distributed excluding protected areas were used for analysis. The Pearson's correlation between population and land fragmentation is 0.289, which is significant at 90% confidence ($P < 0.1$ level). This means that population density and land fragmentation have a positive correlation, the more the population density of the area the more land fragmentation or Vice versa. In this kind of analysis we do not expect very high correlations because there are also other factors such as physical factors that influences land fragmentations (patches), this is only just to show whether there is any association between them in the area where human interfered.

4.4 Roads and Population Distribution

Buffer analyses of major roads were conducted to detect the influence of roads on population distribution, because roads are linear features and cannot be directly used to analyse its correlation with populations. For this study, buffer zones of 5 were built at 1km interval (Figure 17) using ArcGIS environment and total population for each zone was calculated by adding grid population that completely fall in that buffer zone and by adding sub-grid population (the product of sub-grid area and population in that grid).

Figure 17 Map Showing Partial View of Road Buffer Analysis



Once the population is known in each zone the relationship of population distribution and roads were computed using linear regression for further analysis (Table 7).

Table 7 Liner Regression Table for Road Buffer Analysis

Coefficients[a]

Model		Un-standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	185434,400	15263,714		12.149	.001
	Buffer_1km_interval	-33409,600	4602,183	-.973	-7.260	.005

a. Dependent Variable: Population ***R^2 = 0.94 Significant at. $P < .05$***

The regression results can be rewritten as:

$$Y = -33409.6 (x) + 185434.4$$

The formula predicted that at every 1km buffer distance from the major roads the population decreases by about 33410 (x values range from 1 to 5), it means that the population decreases as a function of distance from the roads. The result strongly suggests roads are also one of the factors for population distribution in the watershed and can also be a proxy measure to estimate population.

4.5 Population Distribution and Physiographic Variables

Factors that affect population distribution can be divided into two groups; one is Physical (natural) factors such as elevation, slop, temperature, rainfall and other natural factors, which are the basic factors of population distribution. The second are socio-economic factors such as roads, land use, city location and similar others, these have played important role in population distribution (Tian, et al., 2005; Yue, et al., 2003). This part of the study is mainly focused on the correlation of population with physiographic variables and also tried to sort out the most influential physiographic factors on population distribution.

4.5.1 Physiographic Variables

Originally, 90-meter resolution of DEM was first resemble to 1km² resolutions (Figure 18). Per cent slope (flat surface is 0 per cent and 45 degree surface is 100 percent) data of 1km² resolution is calculated from DEM using ArcGIS environment, the calculated slope is shown in Figure 18. The DEM and the percent slope are superimposed over 1km² grids population data that had been estimated in the previous section. Those grid cells matched more than 50% with population grid cells were tabulate to the respective population grid and used for analysis.

Average rainfall data from different rainfall station in the study area was first interpolated (Figure 19) and superimposed over 1km² resolution population data. The attribute was tabulate to the respective population grid.

The soil data for the watershed are mainly four categories according to ILRI geo-database (Figure 20). The data was also transfer to respective population grid using the above principle and the total area covered by different soil types in m² were used for analysis.

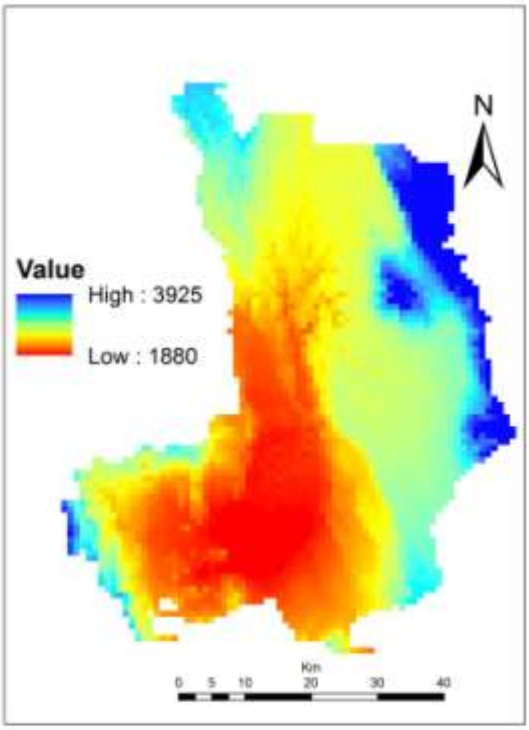
4.5.2 Correlation between Population Distribution and Physiographic Variables

The physiographic data were arranged to facilitate analysis. A total of 2745 observations were used to assess the correlation of physiographic variables with population distribution (Table 8). The result shows that, population distribution is positively correlated with elevation, rainfall and soils particularly with clayey and loamy soils. Slop and soils (sandy and very clayey) are negatively correlated with population distribution. The result is significant at $P < 0.01$ level.

Table 8 Pearson's Correlation Coefficient for Population Distribution and Physiographic Variables

Pearson's	**Elevation**	**Percent slop**	**Average rainfall**	Soils			
				Clayey	**Loamy**	**Sandy**	**Very clayey**
correlation coefficient	.069	-.045	.089	.152	.073	-.106	-.183
Sig	001	.042	.000	.000	.000	.000	.000
N	2745	2745	2745	2745	2745	2745	2745

Figure 18 DEM for the Lake Naivisha Watershed



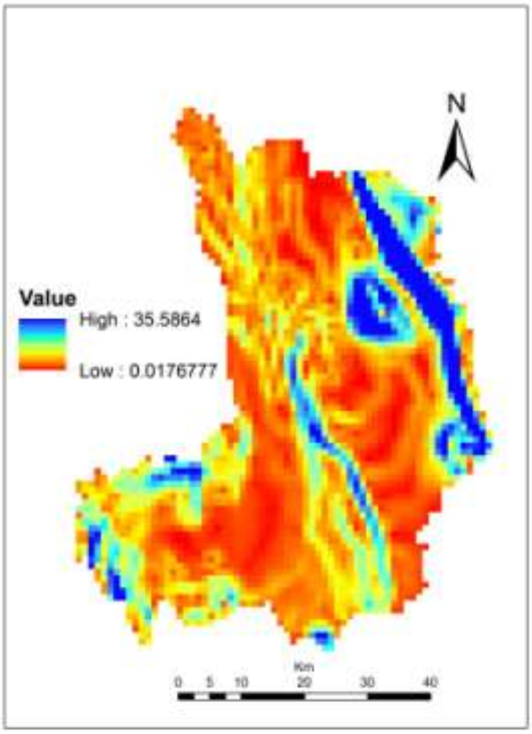
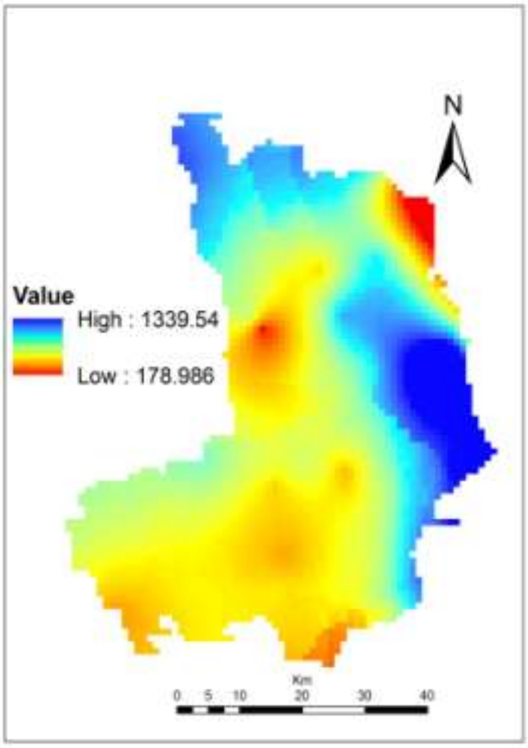


Figure 19 Percent Slop for the Lake Naivisha Watershed

Figure 20 Rainfall Distribution for the Lake Naivisha Watershed



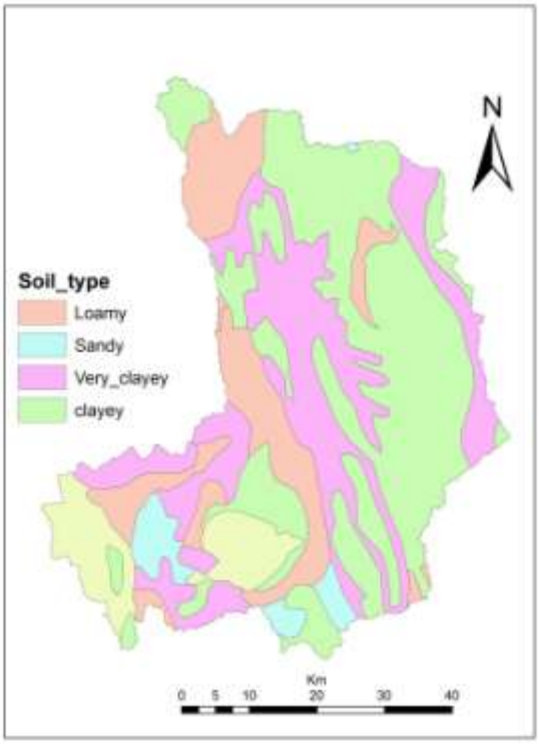


Figure 21 Soil Type Lake for the Naivisha Watershed

The result stipulate majority of the watershed population is concentrated in upper catchment in high rainfall areas. Sloppy, sandy and very clayey areas are not attractive for population settlement. The Pearson correlation coefficient may not give sufficient information about the physiographic variables that most influence on population distribution; as a result the Principal Component Analysis (PCA) was used to identify the most influential factor and it is helpful to examine the redundancy of information in the factor affecting population distribution (Tian, et al., 2005). A principal component analysis based on 1km^2 grid cells was conducted on physiographic variables (Table 9). The analysis result shows that the first principal component contains 27 percent of information about population distribution, according to the matrix, the factor highly correlated with the first principal component is an elevation, the coefficient is about 0.846, and however, the other two components score for elevation is less correlated with the first component score. Average rainfall is scored second in the first principal component; the coefficient is about 0.839, the two respective components of rainfall score relatively better correlated with the first principal component when compared to the elevation coefficient. Note that also in previous Pearson correlation (Table 8), rainfall and elevation are positively correlated with population; elevation and average rainfall in the intra-correlation matrix (appendix_5) also shows that their correlation coefficient is 0.696, significant at $p < 0.001$ it means that elevation and rainfall are highly correlated to each other. As a result rainfall can be taken as a major factor correlated to the first principal component. Very clayey soil is also highly correlated with the second principal component, which contain about 18 percent information on population distribution information. However, clayey soil is highly correlated with the third principal component. Thus, clayey soil is more influence on population distribution than very clayey soil.

Table 9 PCA Table for Factors Affecting Population Distribution in Lake Naivasha's Watershed

Principal component	**1**	**2**	**3**
Contribution percent	**28**	**18**	**15**
Elevation	.846*	-.035	-141
Percent slop	,333	,502	,078
Average rainfall	,839*	-,087	-,147
clayey	,341	-,647	-,473
loamy	-,106	-,069	,902*
sandy	-,508	-,061	-,304
very clayey	-,014	,824*	-,161

* ***Highest correlations***

The remaining component is correlated to the loamy soil containing 15 percent of population distribution information. Generally, the principal component analysis reveals that rainfall/elevation and soil includes most of the population distribution information for the Lake Naivasha's watershed, it is about 60 percent of the information.

5. DISCUSSION

The major outputs of the thesis are a land cover map of lake Naivasha's watershed generated through image classification; population estimation of the same area for land cover/land use types and grid cells; and correlation results of land fragmentation and other physical variables with population distribution of the same. The following discussion elaborates on these outputs.

5.1. Image Classification for Land Cover Mapping

Using the object oriented classification method on an ASTER image dated February 2008, eight land cover classes were generated. The image was visually interpreted and evaluated before the field work to decide on the appropriate classification method. Per-object classification (based on groups of pixels) was found to be more appropriate rather than per-pixel classification (based on single pixel values). This is because an object oriented classification based on segmentation algorithm does not only rely on the pixel values, but also consider pixel spatial continuity based on texture and topology. The ASTER image was taken during the dry season in the study area; as a result the areas of grass and cropland, as well as woodland and shrub lands, had similar image characteristics. On the other hand, the field work was in September which is the wet season and do not coincide with the image date. Considering this, a per-pixel classification was assumed to be more complicated and will create spectral confusion, thus the object-oriented classification was employed for the task. It was also considered that in the chosen method, the problem of classification resulting from spectral confusion is reduced, because classification is not by single pixels, but rather image object.

The outcome of the object oriented classification is adequate as it has achieved an overall accuracy of 87 percent and a kappa statistic of 83 percent (Table 4). It means that only 17 percent disagreed with the reference data and the remaining 83 percent was in accordance with the reference data. The lower kappa per class is cropland and grassland, 70 and 71 percent respectively. These results might be due to spectral confusion since their spectral diversity decreased during the dry season where crops were harvested and grass is drying out, so would take similar spectral values. The accuracy assessment also shows a similar confusion of grassland with built-up areas; this problem might emanate from the mismatch of the field data collection time and image date. During the dry season, grasslands become bare and take a similar reflectance with built-up surfaces. Far more refinement is possible for mapping quality by gathering additional ground truth information over time and increasing the temporal accuracy of an image.

5.2. Estimating Population for Aggregated Land Cover/Land Use Types

The problem of combining spatially incompatible data can be solved by re-aggregation and estimation by areal interpolation. In this part of the study, an areal weighted interpolation technique was used to estimate population for aggregated land cover. In the past, researchers have tried to solve the problem of combining spatially incompatible data from re-aggregating area unit from two super-imposed zone systems to the zone coverage containing the smallest possible compatible zones. Compatible re-aggregation zones are an artificial zone system consisting only of one or more whole area units in both original zone systems to which data are aggregated. However, with the use of GIS areal interpolation, it is much easier than re-aggregation. Any properly executed areal interpolation, even a relatively crude area weighting technique, will better preserve the scale and exhaustiveness of the zones used for spatial data being processed (Reibel and Agrawal, 2007). Note that areal weighted interpolation relies on the assumption that there is no internal variation in population distribution in the source zone; this assumption is generally the weakest part and is rarely held in the real world (Xie, 1996). It is also believed that the amount of errors introduced by areal interpolation decreases as the number of targets is reduced (Gregory, 2002). The areal interpolation technique used in this study produced good results, because the land cover data for the watershed are re-aggregated to have a few targets and reducing the errors introduced by interpolation. This study provide a well-documented and generally accurate areal interpolation technique using land cover/land use data produced from remotely sensed image processing and classification with the help of census data for the watershed. The result of the analysis in population estimate for aggregated land cover/land use type indicates that the correlation between population distribution and land cover/land use types are high and reliable, land cover/land use data have a lot of information to estimate and model population for different spatial scales and interests.

The land cover/land use data were aggregated, as mentioned earlier, and some land covers were discontinuous in nature while some were considered similar in terms of population distribution. During areal interpolation operation, an aggregated land cover/land use type produces better estimates than using spatially discontinuous land cover types and the result can be used as a weighing factor for further estimates for grid cells.

The methodology used in this study to estimate population distribution for different land cover/land use types can be more refined by obtaining a better land cover map and using finer resolution remote sensing imagery, or by conducting image classification which uses methods that are focused on criteria related to the population variable which is being modelled. Furthermore, the uninhabitable areas can also be easily detected in high resolution imagery and therefore, can easily be omitted from the analysis during areal interpolation and which can greatly contribute to better results.

5.3. Modeling Population in Grid cells using Land Cover and Census Data.

In the above discussion, the technique used to remodel population data from a census for different land cover/land use types using a simple areal weighted interpolation was addressed. This part of the discussion gives attention to weighted areal interpolation coupled with GWR using the composite layer (Sub-location population and Grid polygon) and land cover/land use estimate as weighing factor to estimate population in grid cells. A key function of a weighted areal interpolation model is to distribute census counts to cells based on population probability coefficients with the help of ancillary data. Although many factors affect population distribution such as elevation, slope, roads and others, all may be used as input variables in addition to land cover/land use variable to refine the estimate. One of the operations used in this study is an overlay operation wherein two polygons, namely sub-location population polygon as a source unit and grid cells polygon as a target unit, were intersected for further statistical analysis. This function was used as a basic tool for spatial analysis and it is regarded as an operation in statistical modelling (Unwin, 1996). Generally, the use of weighted areal interpolation technique with the help of ancillary data using GWR will produce the most accurate results. Land cover/land use data as weighing factor to estimate population is a normative approach to weighted areal interpolation (Gregory, 2002; Maantay, et al., 2007). The weighted areal interpolation used in this study is a kind of cyclic approach; uses census data in conjunction with land cover estimate produced by areal interpolation mentioned above. In turn, the result of the areal interpolation was used to employ weighted areal interpolation using GWR to estimate 1km x 1km grid cell population. This approach will help in providing a more precise picture of where people actually live.

In this section of the study, a modified method was applied by identifying the correlation between socio-economic variables (census population), physiographic data (land cover) and target location (1km resolution grid cells). This correlation employed to remodel the census population counts using the model local fitting technique to offset influences other than land cover. The result of this analysis is subject to statistical analysis and in fact, it is also influenced by land cover classification. However, this kind of grid-based population is readily integrated with other GIS data and very elemental to a variety of GIS applications. Based on the application of weighted areal interpolation with the help of different ancillary data using WGR, the utilization of population distribution information can be improved, while also creating a more realistic model of real world conditions. The method employed in this analysis is a kind of holistic approach, starting from image classification up to population estimate to grid cells using sub-location population distribution as a major input. One can realize the relevance of the technique by comparing the result of raster population surface with census mapping. The model result helps to integrate the socio-economic data into GIS and serves as a fundamental role to geographical analysis. It is of even greater significance to GIS which offers the potential for complex analysis and modelling of these data.

Some refinement may be possible to the technique outlined here and the method adopted for its evaluation; for example, the use of a more sophisticated distance model for population distribution from urban centre (Martin, 1989). Developing a model based on the criteria related to population distribution for appropriate use of other ancillary data such as slope and roads. Identifying non-settlement areas with the help of high resolution satellite imagery or aerial photography of the same time will give a reasonably accurate surface. Moreover, using census block count as source unit for estimation is far better. For the evaluation of the results, ground truth is more appropriate to compare against the real world population distribution. The method employed in this paper is applicable to other countries and regions; and it is also more appropriate to rural agricultural areas.

5.4. Population Distribution and Land Fragmentation

Densely populated areas are both ecological and social entities, characterized by small fragmentations created as a result of human activities (Di Giulio, et al., 2009). The fragmentation analysis employed in this paper highlight the relationship between human population density and land fragmentation. The size of land fragmentation (patches) decreases when areas are populated, indicating that human activity is one of the sources of fragmentation. Landscapes are fragmented or isolated by roads, settlement, and intensive agriculture and as a result of other human activities. The correlation coefficient 0.289 mentioned above indicates that there is an association of population distribution and land fragmentation. The analysis results also show that land fragmentation can also be used as a proxy measure for population distribution in an area. This approach to land fragmentation analysis may lead to further research. In particular, using land fragmentation coupled with other physical variables to estimate the population of an area is a potential research aspect. However, a more detailed understanding of densely populated landscapes requires integration of natural and social sciences (McIntyre et al., 2008). So far, only little research has been conducted in relation to population density and land fragmentation. Further research is needed that integrates both the societal and ecological aspects of fragmentation in relation to population distribution for the purpose of population estimate.

Lastly, land fragmentation analyses in this study relied on high resolution satellite imagery from Google Earth, but the imagery has its own limitation. It does not preserve image characteristics (proper resolution and bands) when saved for analysis. Furthermore, only some parts coincide with census date (+/- one year) that limits the number of samples taken for analysis.

5.5. Road Network and Population Distribution

Road buffer analysis was employed to detect the relationship between population distribution and road network. The result shows that the population density in Lake Naivasha's watershed is strongly correlated with distance from the roads, indicating that this variable has an important influence on population distribution. Population density changes dramatically with the distance from roads. The multiple coefficient of determination (R^2) is 0.94 and significant at $p < 0.05$ level (Table 7), indicating that 94 percent of the variation of population distribution is explained by the model and therefore, there is a reliable correlation between watershed population and road network. The watershed population prefers to settle near roads for accessibility to other areas for their social and economic benefit. The road buffer analysis in this study was employed only using major roads, all accessible roads were not considered in order to avoid complication during analysis since some roads are near to each other.

5.6. Physiographic Variables and Population Distribution

Many factors affect population distribution such as elevation, slope, rainfall, roads, land cover, city location and other natural factors. In this part of the discussion, only the natural/physiographic factors that affect population distribution will be discussed. As it was indicated in the previous section, land cover has a lot of information about population distribution and may be the best choice in population distribution modelling, because it is highly associated with many other factors. Considering this, the one-kilometre grid cells raster surface map was produced with the help of land cover and census sub-location map. The land cover polygon map was used as weighing factor for population estimation in a grid cell. The weigh from land cover/land use type were determined by regression coefficient and transferred the land cover information to model $1km^2$ grid cell populations. The correlation between population distribution and physiographic variables (Table 8) indicate rainfall and elevation have a positive impact on population density, whereas sloppy areas, sandy and very clayey soils have a negative impact on population density in the Lake Naivasha's watershed. The principal component analysis (Table 9) shows that rainfall and soils have the most influential variables for population distribution in the watershed and taking about 60 percent of the information. The variable, elevation which scored the highest in the first principal component analysis may not be used to model population density because of its correlation with other factors such as rainfall, slope and temperature.

6. CONCLUSION AND RECOMMENDATIONS

6.1. Conclusion

Based on the results obtained and their analyses, the following conclusions are drawn:

(a) In Lake Naivasha's watershed, agriculture is the dominant area coverage, accounting almost one third. Forest is the second largest area coverage, which is about one fourth, followed by grassland. The natural vegetation (forest, woodland and shrub land) constitute about 43.6 percent. Majority of the area under this coverage are uninhabitable, where some are protected and others reserved for natural or wildlife conservation, thus are inaccessible for settlement.

(b) Majority of the population in the lower catchment is concentrated in urban and in the surrounding settlement areas particularly in Naivasha's urban centre. Some concentrations of the population were also observed near Gilgil town. However, the majority of the population in Lake Naivasha's watershed are largely dispersed to cropland areas. Cropland cover is the second largest populated per km^2 next to urban. Savanna land has the lowest populated per km^2. Protected forest and water body (Lake Naivasha) are uninhabitable areas.

(c) Population estimation at 1km grid resolution shows the spatial distribution of population in Lake Naivasha. Areas of non-populated, sparsely populated, moderately populated and densely populated areas can be seen from the population surface map. Densely populated spots are detected in few settlement areas and much more population is concentrated in the northeast and eastern parts of the watershed adjacent to protected areas. Protected areas and the water body show zero population density. Raster surface population map represent the reality better than the census population map since population is hardly distributed uniformly throughout the space.

(d) The population estimation in 1km grid resolution is reliable; the validation result shows that 89 percent of the population estimates in grid cells represent the real spatial distribution of the population. Therefore, the model is reliable and can be used to estimate population at different spatial scales as well as for particular areas of interest.

(e) Land fragmentations (patches) have a positive correlation with population density, and therefore, it can be used as a proxy measure for population estimation or other socio-economic variables. However, it needs careful investigation for the heterogeneous landscape like as Lake Naivasha's watershed.

(f) Population density changes dramatically with the distance from roads. There is a reliable correlation between watershed population and road network. The watershed population prefers to settle near roads for accessibility to other areas for their social and economic benefit. Rainfall and soils have the most influential variables for population distribution in the watershed and taking about 60 percent of the population distribution information. Elevation scored the highest in the principal component analysis but may not be used as a variable to model population density because of its correlation with other factors such as rainfall, slope and temperature.

6.2. Recommendations

(a) Many factors affect population distribution such as elevation, slope, roads, and settlement locations and similar others. This paper also identified rainfall and soil, the most influential factors in population distribution in the watershed and therefore, subsequent researches may developed a model that will consider more than one weighing factors, and develop a criteria to use rainfall and soil values as weighing factor (considering, multicollinearity effect during estimation) to refine population estimates in grid cells in addition to land cover information. Moreover, estimating population density at better resolution of less than 1km may give sufficient details.

(b) Land fragmentation as a proxy variable for population estimate particularly to rural agricultural areas may be an interesting approach. This study also indicates its correlation with population density; and it is a reliable method to use land fragmentation as a proxy measure to estimate population density with the help of remote sensing data. However, it needs a careful investigation to isolate natural and social aspect of land fragmentation. Use of appropriate image characteristics, and spatial resolution needs to be considered and a detailed research follow-up is recommended if relying on land fragmentation for the population estimate.

LIST OF REFERENCES

Abdullah, S. A., and Nakagoshi, N. (2007). Forest fragmentation and its correlation to human land use change in the state of Selangor, peninsular Malaysia. *Forest Ecology and Management, 241*(1-3), 39-48.

Aplin, P., and Smith, G. M. (2008). Advances in Object-based Image Classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII*(B7), 3.

Becht, R., Odada, E. O., and Higgins, S. (2006). Experience and Lessons Learned Brief.

Bogaert, J. (2000). Landscape fragmentation assessment using a single measure. *In: Wildlife Society Bulletin, 28(2000)4, pp. 875-881.*

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

Definiens AG. (1995-2009). Definiens understanding image, eCognition.

Di Giulio, M., Holderegger, R., and Tobias, S. (2009). Effects of habitat and landscape fragmentation on humans and biodiversity in densely populated landscapes. *Journal of Environmental Management, 90*(10), 2959-2968.

Dijk, T. v. (2003). Scenarios of Central European land fragmentation. *Land Use Policy, 20*(2), 149-158.

Dragut, L., Tiede, D., and Levick, S. R. (2010). ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science, 24*(6), 12.

Field, A. (2009). *Discovering Statiatics Using SPSS*.

Floweredew, R., and Green, M. (1989). In Accuracy of spatial databases, Goodchild, M. F., and Gopal, S. (eds.)239-247. London:Taylor and Francis Ltd.

Foley, J. A., DeFries, R., Asner, G. P., and Barford, C. (2005). Global Consequences of Land Use. *Science, 309*, 574.

Geneletti, D. (2004). Using spatial indicators and value functions to assess ecosystem fragmentation caused by linear infrastructures. *In: International Journal of Applied Earth Observation and Geoinformation JAG, 5(2004), pp. 1-15.*

Gesest, K. V. d., Vrieling, A., and Dietz, T. (2010). Migration and environment in Ghana: a cross-district analysis of human mobility and vegetation dynamics. [Journal]. *Environment and Urbanization, 22*(1), 124.

Gregory, I. N. (2002). The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems, 26*(4), 293-314.

Harris, R., and Chen, Z. (2005). Giving dimension to point locations: urban density profiling using population surface models. *Computers, Environment and Urban Systems, 29*(2), 115-132.

Hietel, E., Waldhardt, R., and Otte, A. (2004). Analysing land-cover changes in relation to environmental variables in Hesse, Germany. [Research article]. *Landscape Ecology, 19*(5), 489.

ILRI. from http://www.ilri.org/gis/

Internetgeography. from http://www.geography.learnontheinternet.co.uk/topics/popn1.html

KNBS. (2010). *The 2009 Kenya Population and Housing Census: Population Distribution by Administrative Units*.

Lesschen, J. P., Verburg, P. H., and Staal, S. J. (2005). Statistical methods for analysing the spatial dimension of changes in land use and farming systems. *LUCC report*(7).

Lewiński, S., Geodezji, I., and Modzelewskiego, u. (2007). Object-oriented classification of Landsat ETM+ satellite image *Water and land development, 10*, 15.

Maantay, J. A., Maroko, A. R., and Herrmann, C. (2007). Mapping Population Distribution in the Urban Environment: The Cadastral-based Expert Dasymetric System(CEDS). [Journal]. *Cartography and Geographic Information Science, Vol. 34*(No. 2), 25.

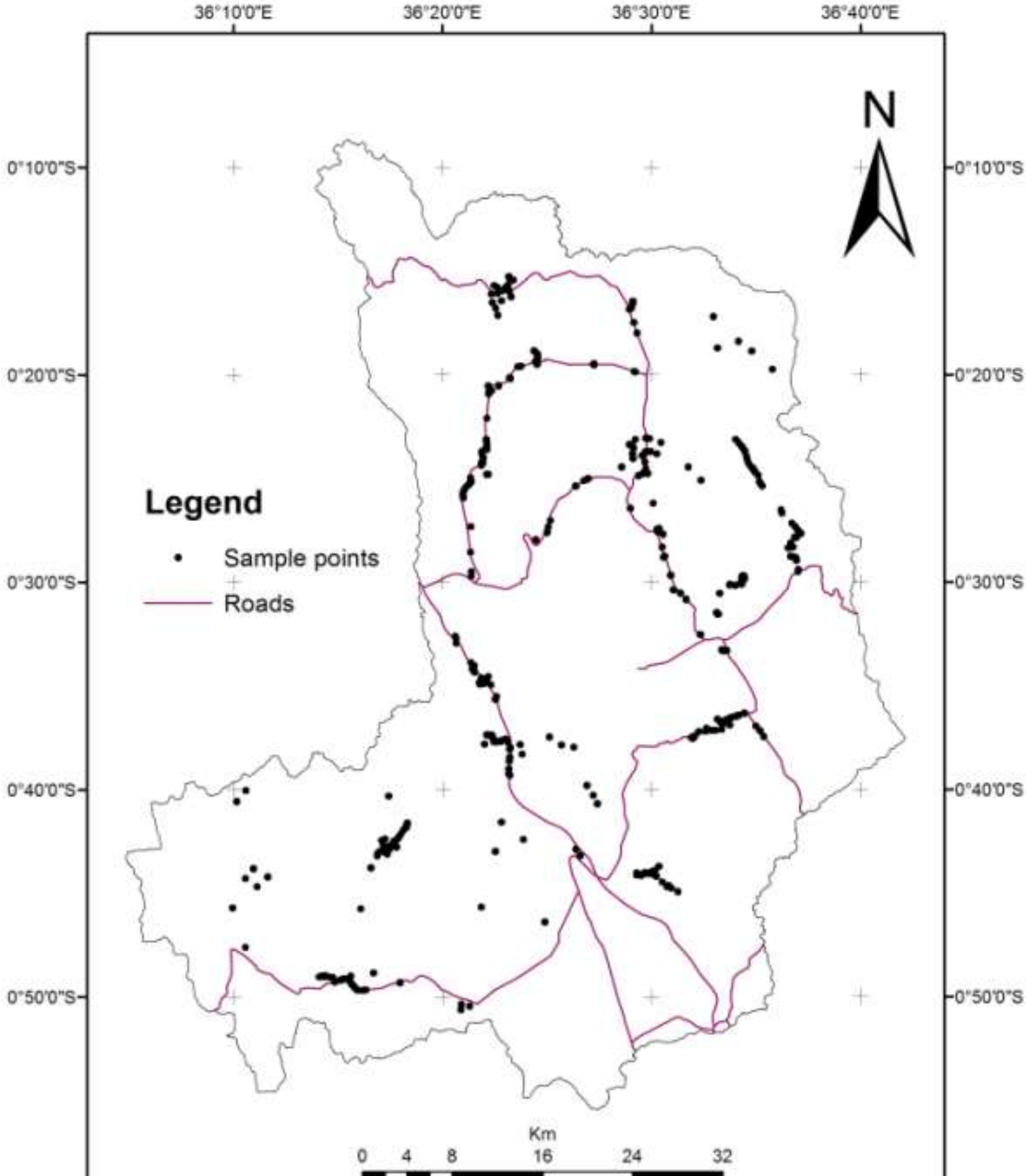
Martin, D. (1989). Mapping Population Data from Zone Centroid Locations. *Transactions of the Institute of British Geographers, 14*(1), 90-97.

McIntyre, N. E., Knowles-Yánez, K., and Hope, D. (2008). Urban Ecology as an Interdisciplinary Field: Differences in the use of “Urban” Between the Social and Natural Sciences. In J. M. Marzluff, E. Shulenberger, W. Endlicher, M. Alberti, G. Bradley, C. Ryan, U. Simon & C. ZumBrunnen (Eds.), *Urban Ecology* (pp. 49-65): Springer US.

Mearns, R., and Sinha, S. (1999). *Social Exclusion and Land Administration in Orrisa, India. Policy Research Working Paper 2124.*

Mireri, C. (2005). Challenges Facing the Conservation of Lake Naivasha, Kenya. Proceedings of Integrated Watershed Management
Nguyen, T., Cheng, E., and Findlay, C. (1996). Land fragmentation and farm productivity in China in the 1990s. *China Economic Review, 7*(2), 169-180.
Ningal, T., Hartemink, A. E., and Bregt, A. K. (2008). Land use change and population growth in the Morobe Province of Papua New Guinea between 1975 and 2000. *Journal of Environmental Management, 87*(1), 117-124.
Niroula, G. S., and Thapa, G. B. (2005). Impacts and causes of land fragmentation, and lessons learned from land consolidation in South Asia. *Land Use Policy, 22*(4), 358-372.
Okello Were, K. (2008). *Monitoring spatio - temporal dynamics of land cover changes in lake Naivasha drainage basin, Kenya.* Unpublished MSc, ITC, Enschede.
Pozzi, F., and Small, C. (2002). Vegetation and Population Density in Urban and Suburban Areas in the U.S.A., 8.
Reibel, M., and Agrawal, A. (2007). Areal Interpolation of Population Counts Using Pre-classified Land Cover Data. *Population Research and Policy Review, 26*(5), 619-633.
Rogan, J., and Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning, 61*(4), 301-325.
Smailes, P. J., Argent, N., and Griffin, T. L. C. (2002). Rural population density: its impact on social and demographic aspects of rural communities. *Journal of Rural Studies, 18*(4), 385-404.
Small, C. (2004). Global Population Distribution and Urban Land Use in Geophysical Parameter Space. *Earth Interactions, 8*(8), 18.
Song, G., Li, Z., Bao, Y., Lü, H., Gao, J., Wang, H., et al. (2007). Spatial distribution regularity and influence factors of population density in the LRGR. *Chinese Science Bulletin, 52*(0), 90-97.
SRTM. from http://www.SRTM.csi.cgiar.org/
Tan, S., Heerink, N., and Qu, F. (2006). Land fragmentation and its driving forces in China. *Land Use Policy, 23*(3), 272-285.
Thompson, S. k. (2002). *Sampling- Wiley series in probability and statistics*. New york: Johon Wiley &Sons.
Tian, Y., Yue, T., Zhu, L., and Clinton, N. (2005). Modeling population density using land cover data. [doi: DOI: 10.1016/j.ecolmodel.2005.03.012]. *Ecological Modelling, 189*(1-2), 72-88.
UN. (1997). Population and Environment Dynamics, Poverty and Quality of life in Countries of ESCAP Region. *Asian population studies series No.147*.
Unwin, D. (1996). Integration through overlay analysis. In Spatial Analytical Perspectives on GIS, Fischer, M.; Scholten, H.J. and Unwin D.(eds.)129-138. London:Taylor and Francis Ltd.
Wu, C., and Murray, A. T. (2005). A cokriging method for estimating population density in urban areas. [doi: DOI: 10.1016/j.compenvurbsys.2005.01.006]. *Computers, Environment and Urban Systems, 29*(5), 558-579.
Xie, Y. (1996). The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems, 19*(4), 287-306.
Yuan, Y., Smith, R. M., and Fredrick Limp, W. (1998). Remodeling census population with spatial information from LandSat TM imagery. *Computers, Environment and Urban Systems, 21*(3-4), 245-258.
Yue, T., Wang, Y., Chen, S., Liu, J., Qiu, D., Deng, X., et al. (2003). Numerical Simulation of Population Distribution in China. *Population & Environment, 25*(2), 141-163.
Zhoua, W., and Troya, A. (2008). An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing, 29*(11), 16.

APPENDICES



Appendix_1: Sample points for image classification

Appendix_2: Field data collection format

ID	X	Y	ALT_(m)	LAND_FORM	LAND COVER	LAND USE	ESTIMATES OF LAND COVER ATTRIBUTE	PHOTO No.	REMARKS
1	221083	9955812	2442	plain	settlement and grassland	residential	>70% settlement with mixed grass		fragmented/Miharat
2	220650	9955829	2442	plain	settlement and cropland; m	residential and agricultur	60% agriculture and 40% settlement		fragmented/Miharat
3	220076	9955639	2378	plain	cropland; maize and grassla	agriculture	90% agriculture and 10% grassland		fragmented
4	220025	9956057	2392	plain	cropland; maize and grassla	residential and agricultur	60% agriculture and 40% settlement		fragmented
5	222181	9956057	2410	plain	cropland; maize and grassla	agriculture	80% agriculture and 10% settlement		fragmented
6	223455	9953462	2420	plain	settlement	residential	90% settlement with mixed grass		fragmented/Miharat
7	223115	9953814	2448	plain	cropland; maize and grassla	residential and agricultur	70% agriculture and 30% settlement		fragmented
8	221629	9956317	2468	gradient land	cropland; maize and grassla	grazing	90% grassland and 10% cropland		fragmented
9	222589	9957087	2472	gradient land	grassland, trees, cropland	agroforestry	65% open trees, 20% grassland and 15% agri.		fragmented
10	220896	9955925	2415	plain	cropland; maize and potato	agriculture	70% agriculture, 20% grassland and 10% trees		fragmented
11	221366	9954357	2406	plain	cropland; maize and grassla	agriculture	85% agriculture, 10% grassland and 5% trees		fragmented
12	219051	9954921	2397	plain	cropland; maize, grassland a	agroforestry	70% trees, 10% agriculture and 5% grassland		fragmented
13	220298	9957339	2350	plain	cropland; maize, grassland a	agroforestry	60% trees, 10% grassland and 10% agriculture		fragmented
14	221252	9957500	2341	plain	cropland; maize and potato	agriculture	60% grassland, 30% agriculture and 10% grassland		fragmented
15	208484	9970594	2342	plain	settlement and cropland; m	agriculture	60% grassland, 30% agriculture and 10% settlement		fragmented
16	208797	9970869	2337	plain	cropland; maize and grassla	agriculture	60% grassland, 40% agriculture		fragmented
17	209029	9971104	2343	gradient land	cropland; maize and grassla	agriculture	60% agriculture, 30% grassland and 10% settlement		fragmented
18	209129	9971644	2357	plain	cropland; maize and open tr	agriculture	60% agriculture, 20% trees and 20% grassland		fragmented
19	209465	9971511	2369	gradient land	cropland; maize	agriculture	100% agriculture		fragmented
20	209075	9970580	2347	plain	cropland; maize and open tr	agriculture	60% agriculture, 40% trees		fragmented
21	209226	9970046	2337	plain	cropland; maize	agriculture	100% agriculture		fragmented
22	208094	9970304	2342	plain	settlement	residential	>90% settlement		fragmented
23	208062	9970871	2357	plain	grassland	grazing	100% grassland		
24	207759	9971038	2370	plain	cropland; maize and grassla	grazing	95% grassland, 5% agriculture		
25	207440	9970633	2393	plain	cropland; maize and settlem	agriculture and settlemen	>95% settlement and 50% agriculture		fragmented
26	207572	9969530	2351	ridges land	cropland; maize and grassla	agriculture (grazing) and s	50% agriculture, 40% grassland and 10% settlement		fragmented
27	209042	9971847	2340	gradient land	grassland, trees, cropland	agroforestry	70% trees, 20% grassland and 10% settlement		fragmented
28	208404	9969667	2360	gradient land	cropland; maize and settlem	agriculture and settlemen	70% agriculture and 30% settlement		fragmented
29	207891	9969021	2357	gradient land	cropland; maize and grassla	agriculture	70% agriculture and 30% grassland		fragmented
30	208077	9968416	2359	plain	cropland; maize and grassla	agriculture	80% agriculture and 20% grassland		fragmented
31	206847	9935672	1920	plain	grassland	grazing	>95% grassland		

Appndix_3: Data Matrix

1	DISTRICT	LOCATION	Land_class	Shape_Area	S_LOCATION	POP_TOTAL	Agriculture	Savannaland	Urban
2	NAIVASHA	LONGONOT	Savanaland	24.37	LONGONOT	4715	0.00	24.37	0.00
3	NYANDARUA SOUTH	MAGUMU	Agriculture	2.45	BAMBOO	11855	2.45	0.00	0.00
4	NAIVASHA	MAIELA	Agriculture	27.07	MAIELLA	9220	27.07	0.00	0.00
5	NAIVASHA	MAIELA	Savanaland	3.30	MAIELLA	9220	0.00	3.30	0.00
6	NAIVASHA	LONGONOT	Agriculture	22.33	MUNYU	8750	22.33	0.00	0.00
7	NAIVASHA	LONGONOT	Savanaland	3.33	MUNYU	8750	0.00	3.33	0.00
8	NAIVASHA	HELLS GATE	Savanaland	114.88	OLKARIA	25437	0.00	114.88	0.00
9	NAIVASHA	HELLS GATE	Agriculture	1.16	OLKARIA	25437	1.16	0.00	0.00
10	NAIVASHA	HELLS GATE	Agriculture	7.11	OLKARIA	25437	7.11	0.00	0.00
11	NAIVASHA	HELLS GATE	Agriculture	11.97	OLKARIA	25437	11.97	0.00	0.00
12	NYANDARUA SOUTH	MAGUMU	Agriculture	14.78	MATURA	8216	14.78	0.00	0.00
13	NAIVASHA	LAKEVIEW	Agriculture	1.10	LAKEVIEW	20082	1.10	0.00	0.00
14	NAIVASHA	LAKEVIEW	Savanaland	9.00	LAKEVIEW	20082	0.00	9.00	0.00
15	NAIVASHA	HELLS GATE	Agriculture	22.29	MIRERA	39209	22.29	0.00	0.00
16	NAIVASHA	HELLS GATE	Savanaland	64.65	MIRERA	39209	0.00	64.65	0.00
17	NAIVASHA	HELLS GATE	Settelement	2.92	MIRERA	39209	0.00	0.00	2.92
18	NAIVASHA	HELLS GATE	Settelement	0.09	MIRERA	39209	0.00	0.00	0.09
19	NAIVASHA	HELLS GATE	Agriculture	1.06	MIRERA	39209	1.06	0.00	0.00
20	NYANDARUA SOUTH	MAGUMU	Agriculture	28.42	KARATI	7283	28.42	0.00	0.00
21	NYANDARUA SOUTH	MAGUMU	Savanaland	1.99	KARATI	7283	0.00	1.99	0.00
22	NAIVASHA	MOI NDABI	Agriculture	50.80	KIPKONYO	2524	50.80	0.00	0.00
23	NAIVASHA	MOI NDABI	Savanaland	0.97	KIPKONYO	2524	0.00	0.97	0.00
24	NAIVASHA	MOI NDABI	Savanaland	13.48	KIPKONYO	2524	0.00	13.48	0.00
25	NAIVASHA	NAIVASHA EAST	Agriculture	17.04	MARAIGUSHU	12134	17.04	0.00	0.00
26	NAIVASHA	NAIVASHA EAST	Savanaland	34.58	MARAIGUSHU	12134	0.00	34.58	0.00
27	NAIVASHA	NAIVASHA EAST	Settelement	0.52	MARAIGUSHU	12134	0.00	0.00	0.52
28	NYANDARUA SOUTH	NYAKIO	Agriculture	21.22	GITHABAI	5879	21.22	0.00	0.00
29	NYANDARUA SOUTH	NYAKIO	Savanaland	2.86	GITHABAI	5879	0.00	2.86	0.00
30	NYANDARUA SOUTH	NYAKIO	Agriculture	6.64	KOINANGE	6323	6.64	0.00	0.00
31	NYANDARUA SOUTH	NYAKIO	Savanaland	18.84	KOINANGE	6323	0.00	18.84	0.00
32	NAIVASHA	NDABIBI	Agriculture	42.09	NDABIBI	8398	42.09	0.00	0.00
33	NAIVASHA	NDABIBI	Savanaland	15.18	NDABIBI	8398	0.00	15.18	0.00
34	NAIVASHA	NDABIBI	Savanaland	31.12	NDABIBI	8398	0.00	31.12	0.00
35	NYANDARUA SOUTH	NJABINI	Agriculture	18.44	MURUAKI	6174	18.44	0.00	0.00
36	NYANDARUA SOUTH	NJABINI	Savanaland	4.94	MURUAKI	6174	0.00	4.94	0.00
37	NAIVASHA	NAIVASHA EAST	Savanaland	34.18	KINAMBA	9135	0.00	34.18	0.00
38	NAIVASHA	KARATI	Savanaland	29.92	KARATI	8302	0.00	29.92	0.00
39	NAIVASHA	KARATI	Agriculture	7.67	KARATI	8302	7.67	0.00	0.00
40	NYANDARUA SOUTH	NJABINI	Agriculture	15.59	TULAGA	7880	15.59	0.00	0.00

Appendix_4: GWR result

Shape_Length	P_namber	Shape_Leng	gij	gij_zij	Sumgij_zij	Rij	Gp	AreaKm2	Shape_Area
2038.868062	2807	2038.868062	10011.416285	1511.446439	948435.202078	480.739149	72.57829	0.150972	150972.289655
1493.103487	2853	1493.103487	11457.673843	1439.484577	948435.202078	550.187128	69.122747	0.125635	125634.975848
3070.943314	2854	3070.943314	11200.798718	6770.194942	948435.202078	537.852217	325.098634	0.604439	604438.586229
3432.120333	2855	3432.120333	10937.701777	4878.780225	948435.202078	525.21854	234.274611	0.446052	446051.677452
1422.992566	2856	1422.992566	10859.160578	1152.174377	948435.202078	521.447063	55.326371	0.106102	106101.605958
1765.41964	2857	1765.41964	10789.665431	1036.372176	948435.202078	518.109969	49.765654	0.096052	96052.299591
584.672991	2858	584.672991	10135.650906	73.363596	948435.202078	486.704783	3.522853	0.007238	7238.173132
1910.656036	2859	1910.656036	9708.832626	1355.703609	948435.202078	466.209356	65.09966	0.139636	139636.108831
883.732175	2908	883.732175	9911.482762	170.287528	948435.202078	475.940432	8.177053	0.017181	17180.832807
609.896932	2957	609.896932	8197.437655	94.956029	948435.202078	393.633537	4.559703	0.011584	11583.623159
2332.341401	2958	2332.341401	8760.900136	407.719995	948435.202078	420.690495	19.578345	0.046539	46538.596349
1094.115853	3005	1094.115853	7881.228534	173.931256	948435.202078	378.449461	8.352021	0.022069	22069.053708
869.774212	3052	869.774212	5796.864444	237.645624	948435.202078	278.360184	11.411528	0.040996	40995.546215
2318.239872	3053	2318.239872	6126.329943	1645.388325	948435.202078	294.180819	79.010058	0.268577	268576.511544
3597.022821	3099	3597.022821	4864.753775	3744.777087	948435.202078	233.601073	179.820806	0.769777	769777.312481
3827.404415	3100	3827.404415	4973.208258	4124.331261	948435.202078	238.808959	198.046654	0.82931	829309.983974
1987.256225	3101	1987.256225	4957.337125	1175.22383	948435.202078	238.046842	56.433185	0.237068	237067.562725
3248.422019	3144	3248.422019	2922.668955	1281.179607	948435.202078	140.343918	61.521085	0.438359	438359.467428
3973.997653	3145	3973.997653	3522.018845	3518.119836	948435.202078	169.124157	168.93693	0.998893	998892.961983
3824.552473	3146	3824.552473	3861.843876	3658.326562	948435.202078	185.442248	175.66953	0.9473	947300.481094
1520.953915	3147	1520.953915	4278.455421	374.021712	948435.202078	205.447557	17.960184	0.08742	87419.798729
1647.835371	3190	1647.835371	685.776487	62.778041	948435.202078	32.930366	3.014545	0.091543	91543.006311
3913.105298	3191	3913.105298	1507.023179	1472.297507	948435.202078	72.365889	70.698394	0.976957	976957.439533
3634.399418	3192	3634.399418	2153.218768	1731.674852	948435.202078	103.395616	83.153459	0.804226	804226.15571
2083.06738	3193	2083.06738	2916.386169	537.265506	948435.202078	140.042224	25.799003	0.184223	184223.033317
2213.307256	3240	2213.307256	9287.260293	2493.254742	948435.202078	445.965834	119.723836	0.26846	268459.660167
3114.311167	3241	3114.311167	324.417391	173.050866	948435.202078	15.578229	8.309746	0.53342	533420.435366
821.951854	3242	821.951854	1120.262214	31.455787	948435.202078	53.793978	1.510478	0.028079	28078.949964

OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	402	
2	ResidualSquares	8998592274.90921	
3	EffectiveNumber	78.621996	
4	Sigma	1358.155707	
5	AICc	85625.797401	
6	R2	0.720757	
7	R2Adjusted	0.716314	
8	Dependent Field	0	POP_TOTAL
9	Explanatory Field	1	Areakm2

Appendix_5: Intra-correlations among physiographic variables

Correlations								
Variables	Population	Elevation	Slope	Rainfall	clayey	Loamy	sandy	Very clayey
Population	1							
Elevation	.069**	1						
Slope	-.045*	.151**	1					
Rainfall	.089**	.696**	-.069**	1				
Clayey	.152**	.277**	-.127**	.359**	1			
Loamy	.073**	-.135**	-.047*	-.022	-.384**	1		
Sandy	-.106**	-.219**	-.075**	-.166**	-.173**	-.099**	1	
Very clayey	-.183**	-.048*	.154**	-.171**	-.412**	-.236**	-.106**	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).