FINE RESOLUTION MODELLING OF MALARIA RISK FACTORS AND POTENTIAL MALARIA RISK PREDICTION.
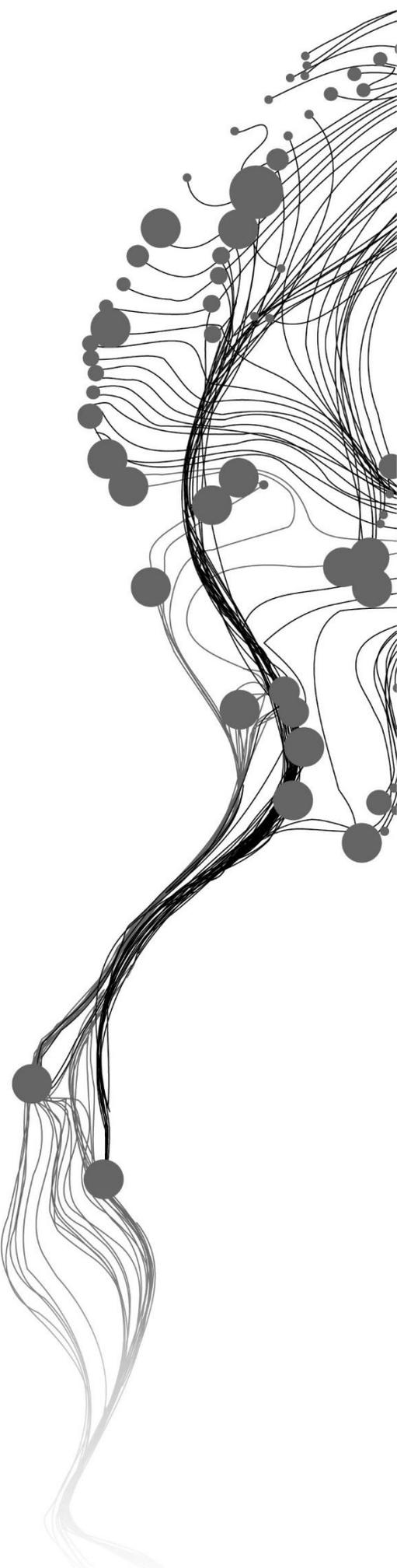# A CASE OF HOMA BAY COUNTY, KENYA.

KENNETH MICKEY OCHIEN'G KASERA
 [February 2016]

SUPERVISORS:
Dr. S. Amer
Dr. J. A. Martinez

FINE RESOLUTION MODELLING OF MALARIA RISK FACTORS AND POTENTIAL MALARIA RISK PREDICTION.
A CASE OF HOMA BAY COUNTY, KENYA.

KENNETH MICKEY OCHIEN'G KASERA
Enschede, The Netherlands, [February, 2016]

ABSTRACT

Malaria remains to be one of the major killers in the world; with governments spending billions of USD dollars on control measures yet malaria still poses a threat to 3.2 billion people globally. In Kenya, twenty-five million people are at risk of malaria, of which approximately 428,000 cases were reported in Homa Bay County in the year 2014. These measures range from vector control to malaria diagnosis and treatment. However, the operational challenge facing present-day elimination of malaria is the need for high-resolution location-based surveillance and targeted prevention responses. Geographic mapping has traditionally played a great role in diseases surveillance but its full potential has not yet being achieved. Moreover, previous malaria risk models are based on species presence data and malaria household surveys, which is expensive to acquire. This research uses malaria cases from the health records and readily available remote sensing (satellite imageries) and GIS datasets to model malaria risk factors and generate potential malaria risk map.

Various remote sensing datasets were generated from Landsat 8 satellite (land surface temperature, normalized difference vegetation cover, land cover, water hyacinth, and topographical wetness), sentinel1 (wetlands), moderate resolution imaging spectroradiometer (evapotranspiration), and climate hazard group infrared precipitation with station data (rainfall). Soil drainage, poverty, population dataset, and altitude were sourced from Kenya soil survey, World resource institute, and NASA respectively. Additionally, the malaria occurrence data for each health facility was sourced from health sub-county headquarters in Homa Bay County. Raster based surface travel time method based on multiple layers (slope, land cover, road and rivers) was used to generate health catchments for calculation of malaria infection rate per health facility. Moreover, identification and categorisation of malaria risk factors in Homa Bay County was done using factor analysis model. The association between factors and malaria infection rate was done using correlation analysis, and collinearity between factors assessed using the variance inflation factor model. Overlay index model was then used to create the potential risk map using the correlation coefficient between the risk factors and malaria infection rate as factor weights.

Results from factor analysis reveal that malaria-causing factors in Homa Bay County are categorised into three, namely, biophysical (rainfall, normalized difference vegetation index, land cover, evapotranspiration, land surface temperature, distance to hyacinth and topographical wetness), topographical (altitude, slope and soil drainage) and socio-economical components (poverty, and distance to wetlands). In addition, rainfall, altitude, temperature, and normalized difference vegetation index are considered as very significant risk factors with land cover as the least. Results from correlation analysis and collinearity also reveal a weak linear association between risk factors and malaria infection rate, and that the factors are not correlated respectively.

High-resolution remote sensing datasets and health records can be successfully combined to model and predict malaria risk. The potential risk map generated is 64% accurate using the malaria infection rate as the reference dataset for validation. The zones close to Lake Victoria are of high potential malaria risk with zones of high altitude and far from the lake considered as low risk. Moreover, moderate potential risk is experienced in more than half of the county. Approximately 287,0000 cases out of 428,000 reported malaria cases in the year 2014, occurred within 1km from wetlands and within 1km from water hyacinth; this makes wetlands and water hyacinth locations key actions areas apart from other potential risk areas within Homa bay county. Poverty stricken zones also have high infection rate; incorporating this complex aspect of human life into malaria prevention is highly needed in Homa Bay county. However, more investigation is needed to fully ascertain the risk since the risk map is a potential risk map. Future research on multi-temporal analysis of malaria risk in Homa bay is however recommended to fully understand and ascertain the risk.

Keywords: Spatial modelling, Malaria risk, Malaria risk factors, Anopheles habitat, Potential risk, Homa Bay County.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

ACT-Artemisinin based Combination Therapy

AIDS- Acquired immunodeficiency syndrome

CHRIPS- Climate hazard group infrared precipitation with station data

EPR- Epidemic Preparedness and Response.

GIS- Geographic Information System

HIV- Human Immunodeficiency Virus

IPT-Intermittent Presumptive Treatment

IR100-Infection rate per 100 persons

IRS- Indoor Residential House Spraying

ITN-Insecticide Treated Nets

KMO- Kaiser-Meyer-Olkin

LST-Land Surface Temperature

MIP-Malaria in Pregnancy

MODIS- Moderate resolution imaging spectroradiometer

NASA – National Aeronautics and Space Administration

NDVI-Normalized Difference Vegetation Index

PCA- Principal Component Analysis

PHC-Primary Health Care

RS- Remote sensing

UNICEF-United Nations Children Funds

USAID-United States Agency for International Development

VIF- Variance Inflation Factor

WHO- World Health Organization

1. INTRODUCTION.

Malaria is the major cause of mortality in Africa; it is the leading cause of under-five deaths in many African countries (Kleinschmidt, Bagayoko, Clarke, Craig, & Le Sueur, 2000). It is also ranked as one of the top ten killers in low economic countries (Abuelezam, Buckee, Childs, Dye, Gupta, Murray, and Williams, 2015). According to Kenya Malaria Fact Sheet (2015), twenty-five million out of forty million Kenyans are at risk of malaria. Moreover, malaria accounts for 30-50% of all outpatient attendance, 20% of all admissions to health facilities, coupled with the loss of estimated 170 million working days to the disease each year, not to mention, it is the causes of under-five deaths in Kenya, estimated at 20% (Kenya Malaria Fact Sheet, 2015). This study uses remote sensing information and statistical methods and tools to model malaria risk factors and develop potential malaria risk map based on environmental, socio-economic factors, and malaria reported cases (health records). *This chapter contains background information, research problem, study objectives, research questions, assumptions, and limitation.*

1.1 Background information.

Malaria is a vector-borne disease caused by plasmodium parasite transmitted to humans by the bite of a female anopheles mosquito (Benali, Nunes, Freitas, Sousa, Novo, Lourenço, and Almeida, 2014). There are four species of malaria human plasmodium namely: plasmodium falciparum, plasmodium malaria, plasmodium ovale and plasmodium vivax. Malaria diseases is common in the tropical and subtropical regions with 3.2 billion people at risk globally (WHO, 2015). In addition, it is estimated that more than one million people in Africa die every year from malaria, children being the most vulnerable (Armstrong Schellenberg, Smith, Alonso, & Hayes, 1994).

In the recent past, the debate on disease epidemiology concerns the importance of information relating to exposure and host factors (Bundy, Barker, Grenfell, Hoti, Michael and Ramaiah, 2001); mosquito acts as a vector-host for malaria plasmodium parasite and man as the exposure agent. The issue of exposure and host factor interplay in the rise or decrease of malaria transmission in space, and therefore, spatial extent have been investigated based on vector prevalence trends, exposure agent, and the interaction between the aforementioned (Cox, Hay, Myers, Shanks, Stern, Snow, Randolph and Rogers, 2002). Consequently, the aforementioned interaction in space has also led to increase or decrease in malaria prevalence in zones earlier known and not known to be malaria hotspots (Bayoh, Hightower, Mueke, Mutuku and Walker, 2009).

Moreover, successful malaria programmes move towards elimination of residual transmission and therefore, vector target in both high and low-risk areas, needs be identified, and mapped (Cohen, Dlamini, Novotny, Kandula, Kunene, Simon and Tatem, 2012). This has to be supported by the generation of high-resolution maps of malaria risk for periodic surveillance. In addition, periodic surveillance takes into account continuous changes in the state of the earth in terms of habitat environmental aspects (Reiter, 2001). Various ways of modelling changes in the state of the earth in terms of disease transmission have been developed by researchers, which includes species habitat modelling, regression, Bayesian models, and risk factor analysis (Reiter, 2001).

The Outbreak of malaria disease in the world poses a threat to human population, this has geared the implementation of prevention, and control measures by various governments to safeguard the life of its citizens (Attaway, Bennett, Falconer, Jacobsen, Manca, and Waters, 2014). These measures include the use of remote sensing and geographic information system (GIS) technology in diseases surveillance (Koram, Bennet, Adiamah, & Greenwood, 1995). Upon introduction and development of GIS in present time, the act of mapping in malaria control and reduction has grown (Kelly, Tanner, Vallely, & Clements, 2012).

Thus, ecological concepts have been linked to the geospatial domain, sampling frameworks, and data collection standards established (Peterson, 2003).

In addition, a major advancement in the remote sensing field has led to a virtual explosion in ecological investigations (Cohen & Goward, 2004). Temporal-spatial risk analysis is possible as yearly, monthly, and daily satellite images of specific ecological sites under surveillance availed. Regional, national, and local risk analysis has also been facilitated with the launching of earth observation satellites with wide imaging swath (Cohen & Goward, 2004). Consequently, image-processing techniques have been developed to extract spatial habitat related information from satellite images to be applied in the machine learning techniques. Moreover, the need for spatial information on environmental, socio-economic variables and malaria infection datasets have become urgent mostly in endemic zones (Rincón-Romero, Edilberto and Londoño, 2009), and therefore, geospatial information pertaining to such known factors needs to be generated and applied in malaria risk assessment. (Hirzel & Le Lay, 2008).

Vulnerability and risk concept has long been the discussion of every urban planning system, with the aim of discovering how much and who has been affected by the changes in the state of the earth (Institute for Environment and Human Security, 2015). Consequently, frameworks, scientific methods, and tools have been developed to assess socio-economic vulnerability and risk in the context of natural hazards (Institute for Environment and Human Security, 2015). Despite multiple frameworks established and research done in the identification of malaria risk areas; better tools and methodology still need to be developed for fine resolution malaria risk mapping (Rincón-Romero, Mauricio Edilberto and Londoño, 2009).

Most risk modelling frameworks and tools (regression, artificial neural networks, species distribution models, and Bayesian models) use household health surveys and species presence data, which is not readily available in most cases and very expensive to collect (Onchiri, 2014). However, health facilities in Kenya record monthly malaria cases for both under five and over five years of age, this data is readily available and therefore, malaria risk modelling tools and methodologies need to be developed taking into consideration the aforementioned data. In Kenya, health facility catchments to attribute the malaria cases do not exist and therefore, this also calls for the development of various ways of creating the catchments to present the recorded malaria data in space.

Various measures ranging from mosquito vector control to malaria diagnosis and treatment have been put in place by medical domain to help reduce malaria transmission. The recent measures being: creation of genetically modified mosquitos which do not transmit malaria parasite, use of mosquito nets, insect repellents, in-house spraying, draining stagnant water, and traps to kill mosquitos (Daily Nation, 2015). This research is part of this great initiative of finding a solution to reduce malaria risk.

1.2 Research problem.

Reduction of malaria is a social good in itself (Heggenhougen, Hackethal, & Vivek, 2003). In Kenya various malaria control initiative and policies have been implemented (see figure2 in chapter 2) leading to a reduction of malaria (see figure 3 in chapter 2) but still high transmission is recorded in epidemic zones. However, studies indicate that the Kenyan government (see figure2 in chapter 2) has not implemented the larval control measure. This mainly focuses on larva as a development stage for mosquito, and therefore, estimation of malaria risk infection governing vector control is necessary (Bangs, Maguire, & Barcus, 2002). However, the integration of environmental variables and malaria reported cases using remote sensing and statistical tools to locate high-risk potential zones could provide decision makers in the health domain with information for implementing the larva control strategy.

Epidemiological surveillance is necessary for developing any multi-dimensional malaria control strategy, currently malaria risk maps are generated from health surveys (Onchiri, 2014). This is costly, approximately USD120 million was used by Kenyan government in 2013 for upscale (Malaria World Report, 2014).This research uses health malaria records (malaria reported cases) instead of surveys together with readily available remote sensing datasets to generate malaria risk maps. In addition, generation cost can be greatly reduced by applying the proposed method. Moreover, the operational challenges facing malaria reduction is

the need for the high-resolution based surveillance in time and space (Kelly, Tanner, Vallely, & Clements, 2012), therefore, updated yearly risk information can be made available for yearly risk and vulnerability monitoring by the geospatial planning domain.

1.3 Research Objectives.

1.3.1 General research objectives.

The main objective of the study is to model spatial malaria risk factors, predicting potential malaria risk areas based on remote sensing derived environmental and socio-economic variables and malaria health records.

1.3.2 Specific research objectives.

1 To identify potential spatial environmental and socio-economic factors leading to malaria risk.

2 To derive spatial environmental and socio-economic factors from available earth observation satellites and RS/GIS techniques for Homa Bay County.

3 To determine the association between spatial environmental and socio-economic factors and malaria infection rate in Homa Bay County.

4 To identify significant spatial environmental and socio-economic factors leading to malaria infection in Homa bay county.

5 To develop a potential malaria risk map for Homa bay county.

1.4 Research questions.

The following research questions were answered to achieve the specific objectives.

Research question for objective 1. To identify potential spatial environmental and socio-economic factors leading to malaria risk.

 ✓ What are the spatial factors environmental and socio-economic leading to high and low malaria risk?

Research question for objective 2. To derive spatial environmental and socio-economic factors from available earth observation satellites and RS/GIS techniques for Homa Bay County.

 ✓ Which remote sensing imageries and digital image processing procedures are used to derive the spatial environmental factors?

Research question for objective 3. To determine the association between spatial environmental and socio-economic factors and malaria infection rate in Homa Bay County.

 ✓ What is the association between malaria infection rate and environmental and socio-economic factors?

Research question for objective 4. To identify significant spatial environmental and socio-economic factors leading to malaria infection in Homa bay county.

 ✓ What are the significant spatial factors leading to malaria infection in Homa Bay County?

Research question for objective 5. To develop malaria risk potential map for Homa bay county.

 ✓ Which locations or areas inhibit high potential in malaria risk?

 ✓ Which settlements fall in the high and low-risk zones?

1.5 Assumptions.

 ✓ Malaria cases are all reported at the health facility (not treated at home).

1.6 Limitation.

The limitation of the study are.

- ✓ Incompleteness of geospatial data for Homa bay county (no health facilities catchment for Homa bay County).

- ✓ Lack of accurate population data for settlements.

1.7 Thesis outline.

There are five chapters in the thesis namely, introduction, literature review, methods, results, discussion, conclusion and finally recommendation.

Chapter 1. Introduction: This section entails the background information of the study, research problem, objectives, research questions, hypothesis, limitations, and lastly assumptions.

Chapter 2. Malaria risk modelling: Concepts definition and current knowledge about the study is found in this section. Conceptual framework under which this study is based is also found in the literature review section.

Chapter 3. Methodology: This chapter contains the area of study and its characteristics (economic, health system, demographic, topographic, transport and climatic).Various methods used in data collection, generation (both primary and secondary) and analysis. The methodological framework on how the results were achieved is also part of this section.

Chapter 4. Results and discussion: The result of data analysis and modelling are found in this section. In-depth debate on the results and critical analysis on various variables in relation to malaria risk are found in this section.

Chapter 5. Conclusion and recommendation: This section entails main discoveries, interpretation of the results in summary.

2 MALARIA RISK SPATIAL MODELLING.

In this chapter, emphasis is laid on mosquito, habitat suitability, malaria risk, remote sensing, malaria risk factors, malaria control, and spatial modelling. The aforementioned elements create the conceptual framework under which this study is based.

First, the key concepts used in this chapter and entire study are defined below.

Anopheles mosquito.
"Mosquito is a slender long-legged fly with aquatic larvae"(Oxford Dictonary, 2015). Mosquito belongs to family Culicidae with a lifespan of 10 days. According to Freudenrich (2015), there exist more than 2,700 species of mosquitos, which includes culex and anopheles among others. Mosquitos are responsible for transmitting most of the devastating diseases in the world today, they are very efficient vectors of human beings (Beck-Johnson et al., 2013). The abundance of female adult mosquitos is key in determining the occurrence of vectors diseases to human population.

Anopheles mosquito habitat.
Habitat is the environment inhabited by a particular species of organism. This is the natural ecological living environment where organism finds food, shelter and reproduce. It is composed of physical factors such as biotic and abiotic factors, interplaying to create a favourable condition for organism development.

Fine spatial resolution.
Spatial is defined as any phenomenon or observation relating to space or having the character of space and geographic position. Fine is defined as high quality while resolution is defined as the ability to make features distinguishable. Fine spatial resolution, therefore, refers to high-quality observation of distinguishable features relating to space.

Malaria occurrence.
Female anopheles mosquito transmits life threating disease to humans known as malaria through biting (WHO, 2015). The number of people diagnosed with the disease is the malaria occurrence (confirmed and clinical counts). Confirmed counts are individuals diagnosed with malaria from laboratory testing. Clinical malaria is based on fever and positive blood film in less endemic zones. However, in high asymptomatic parasitaemia endemic zones it is common to assume that individuals with fever and parasitaemia suffer from malaria hence they are also included in malaria occurrence (Armstrong et al., 1994). Consequently, parasite density determination is, however, necessary for the correct diagnosis of clinical malaria in endemic zones(Peelman, Trape, & Morault-Peelman, 1985).
For the purpose of this study, both *clinical and confirmed cases* are included in the malaria occurrence data.

Malaria infection rate.
Malaria infection rate is defined as the number of people per 100 diagnosed with malaria. It is calculated by dividing the number of malaria occurrence by the total population and multiplying by 100.

Potential risk.
According to Oxford Dictonary (2015), vulnerability is defined as the possibility of being exposed to illness, harm or risk either physically or emotionally. Potential vulnerability includes risk, illness, diseases, or situations that cannot be fully verified. At least one important condition for the vulnerability has to be detected (Qualys Community, 2015). Further investigation is required to determine if the risk is present or not.

Risk maps.

This is the outcome model of potential disease risk based on spatial environmental and socio-economic data and malaria infection rate.

Environmental variables.

Environmental variable is defined as physical, chemical, biological and socio-economic elements whose interaction affects an organism or group of organisms, either negatively or positively. Variables can also be referred to as factors or constraints.

Poverty.

For the purpose of this thesis, poverty is defined as being in a state of need and lack of resources. Absolute poverty and relative poverty concepts are both applied in this study. The absolute poverty concept is based on minimum standards all over the world that no human being should fall below while relative poverty is based on a comparison between one society and another. Anybody living below 1.90 dollars per day is considered as poor in this research (World Bank, 2015).

Modelling.

Modelling refers to the identification, selection, and providing proof of relation between relevant segments of a system. The main aim of modelling is to make an observation or particular segment of the world more understandable.

2.1 Mosquito habitat.

Mosquito inhabits forests, marshes, tall grasses, weeds and wet grounds (Mosquito World, 2015). According to Mosquito World (2015), culex and anopheles are the most common water mosquitos, they lay eggs on clumps, rafts and hyacinth on the surface of stagnant water ponds and lakes. Mosquitos also lay eggs in moist soils commonly known as flood water mosquitos, they withstand drying out of water (University of Florida, 2015). In addition, temperatures ranges between 25 to 30 degrees Celsius has been proved to be favourable for mosquito development, at more than 30 degrees Celsius the abundance of potentially infectious mosquitos reduces drastically (Beck-Johnson et al., 2013). Beck-Johnson et al.(2013) further state that pools with water temperatures warmer than air temperature are more conducive for mosquito breeding. Mosquito habitat conditions are outlaid in details in *Section 2.4*.

2.2 Habitat suitability.

Habitat suitability constitutes a good tool for decision making (Garzón et al., 2006). It is based on ecological theory that species occupy locations within the environment that are most suitable for reproduction and development (Hongoh, Berrang-Ford, Scott, & Lindsay, 2012). Moreover, habitat modelling is important in our understanding of ecosystem dynamics, relationship between biota and its ecological niche (Australia Government, 2015).

The presence of a species in a location is determined by three components namely, a local environment which allows the population to grow, the interaction between species within a given locality (example predation and competition) and lastly accessibility given the dispersal ability (Hirzel & Le Lay, 2008). Hirzel & Le Lay (2008) further explain the identification of key environmental variables determining the habitat (niche) as crucial in habitat suitability modelling. In addition, environmental locations with similar characteristic with those in which malaria parasite are known to survive can be easily predicted and mapped (Gwitira, Murwira, Zengeya, Masocha, & Mutambu, 2015).
The succeeding section explains the concept of habitat modelling.

Habitat modelling.

The development of mathematical models in the study of ecology has proven to be very useful, however, there is a scarcity of models which take into consideration the spatial linkages between the environment and the species (Lourdes Torres-Sorando, 1997). Therefore, several habitat suitability frameworks have been

developed namely habitat species distribution models, resource selection functions, and ecological niche models addressing similar concepts using different tools (Hirzel & Le Lay, 2008).

According to Hirzel & Le Lay (2008), diverse tools namely regression, envelope modelling, classification trees, fuzzy logics, Bayesian models, artificial neural networks and factor analysis have been used in habitat modelling depending on data availability. These tools are appropriate depending on data availability see figure 1.

Modelling method	appropriateness	References.
Species distribution modelling, Bayesian and envelope modelling	Species presence data available	(Stevens & Pfeiffer, 2011)
Regression modelling	Survey data (mainly binary data)	(Tuyishimire, 2013).
Factor analysis	Selection of composite factors.	(Nardo et al., 2008)
Artificial neural networks	Training of datasets on machines	(Özesmi, 1999)

Figure 1. Habitat modelling tools.

Various habitat modelling studies have been done based on correlation between the habitat factors mainly temperature and water presence to malaria prevalence or incidence ignoring other habitat modifying factors. However, the final results of this models are simple and needs more caution in applying hence the necessity to incorporate the modifying factors into habitat modelling to improve the accuracy and details of the result (Benali et al., 2014).

Species distribution modelling approach in habitat modelling.

Species distribution models can be either rule-based or quantitative. It's based on collected presence (organism) data which is unavailable in many cases (Stevens & Pfeiffer, 2011). It is also referred to as niche modelling, commonly applied in the field of epidemiology. In addition, species distribution encompasses the integration of environmental variables and biological data in a modelling scenario (Cossio et al., 2012). It also has three main parts namely, ecological, data and statistical models within the framework of space and time. It uses the principle of maximum entropy, which is a machine learning technique (defined as the best probabilistic distribution representing the current state or scenario of a system). Consequently, various states which the system may or can exist must be identified and parameters known (Singh, 2003).

2.3 Malaria risks in Kenya.

Malaria being an acute febrile disease, its symptoms appear on individuals seven to ten days after infective mosquito bite (WHO, 2015). According to WHO (2015), the latest estimates released in September 2015 indicates that there were 214 million cases of malaria and 438000 malaria deaths in 2015. Sub-Sahara Africa accounts for 80% of the cases and 78% of deaths globally (WHO, 2015). More than 70% of malaria deaths occur in the age group of under-five. However, between the year 2000 and 2015, the malaria incidence reduced by 37% averting deaths of approximately 6.2 million people globally (WHO, 2015). Despite the reduction in malaria cases, the diseases pose danger to 3.2 billion people globally.

A Huge amount of monetary fund's goes into malaria risk reduction. Over 2.5 billion USA dollars is raised every year by countries and global partners to fight malaria in epidemic countries, additionally, the World Bank through its program for malaria control in Africa allocates 700 million USA dollars to priority countries every year for the same (Malaria No more, 2015).

Apart from HIV and AIDS, malaria is the other leading cause of morbidity and mortality in Kenya. Twenty-five million out of forty million Kenyans are at risk of malaria states Kenya Malaria Fact Sheet (2015). In addition, it accounts for 30%-50% of all outpatient attendance, approximately 20% of all admissions to health facilities, coupled with the loss of estimated 170 million working days to the disease annually, not to mention, it is the causes of under-five deaths in Kenya, estimated at 20% (Kenya Malaria Fact Sheet, 2015).

Malaria infection rate around Lake Victoria (Homa Bay, Kisumu, and Migori counties) is particularly high (USAID, 2015). The percentage of plasmodium falciparum parasite is most dominant in Homa Bay,

Kisumu and Migori counties, at above 20%. However, high prevalence is also experienced in the northern western and southeastern part of the country, with low prevalence rate experienced in the central, eastern parts of Kenya (as low as 0.5%). Most part of the county experience prevalence between 1 to 20%. Figure 4 shows the plasmodium falciparum prevalence map for Kenya by the year 2009.

As a malaria control measure, the country has been stratified into four epidemiological zones to address varied risk (USAID, 2015): Epidemic areas classified as zones with stable malaria prevalence above 20%, highland prone zones having prevalence between 5 to 20%, seasonal transmission areas experiencing prevalence between 1 to 5%, and lastly low malaria zones with prevalence lower than 1%. About 26% of the Kenyan population live in a malaria epidemic zone.

The World Health Organization approved a malaria control strategy in 1978 based on the principles of primary health care (PHC). Large vertical programmes were replaced by community-based, integrating primary health care, and community participation (National Malaria control Program, 2015). Since then diverse measures have been put in place by the Malaria Control Initiative in Kenya namely: management of malaria in pregnancy (MIP), vector control, epidemic preparedness and response (EPR), awareness raising, monitoring and evaluation (Kenya Malaria Fact Sheet, 2015). See figure2 on control strategies.

In 2001, the government of Kenya launched a 10-year national malaria strategy consisting of intervention measures, vector control and diseases diagnosis (Malaria Control, 2010). This led to a reduction in malaria burden in Kenya. Figure 2 shows the intervention measures implemented by the Kenyan government. Malaria transmission intensity has tremendously reduced in most parts of the country between the year 2006 and 2008 with the adoption of ITN (insecticide treated nets) policy (See figure 3). However, the malaria data in Kenya is not consistent and, therefore making it hard to visualize the trend from the year 2008 onwards (see figure 3).

According to World Health Organization (2014), malaria death has reduced in Kenya from 160 per 100,000 in the year 2010 to below 40 per 100,000 in the year 2013. With the main sources of funds from the Kenyan government, Global fund, USAID, WHO, UNICEF and World Bank, the number of malaria cases is perceived to be on the decline (Meyrowitsch et al., 2011). By the year 2009, 60 % of Kenyan population had an access to ITN as compared to 10% in 2003 (World Health Organization, 2014). Despite this, moderate-to-high levels of transmission persist in certain endemic zones. In addition, malaria survey conducted by Malaria Control and Ministry of Public Health (2010) confirmed that malaria prevalence remains high in rural areas at 12% compared to 5% in urban areas.

Intervention	Policies/strategies	Yes/No	Year adopted
ITN (insecticide treated nets)	ITNs/LLINs distributed free of charge. ITNs/LLINs distributed to all age groups	Yes yes	2006 2010
IRS	IRS is recommended (indoor residential house spraying)	yes	2003
Larval control	Use of larval control recommended	No	-
IPT	IPT used to prevent malaria during pregnancy (intermittent presumptive treatment)	Yes	2001
Diagnosis	Patients of all ages should receive diagnostic test. Malaria diagnosis is free of charge in the public sector.	Yes	2009-
Treatment	ACT is free for all ages in the public sector. Artemisinin-based monotherapies withdrawn A single dose of primaquine used as gametocidal medicine for P. falciparum. System for monitoring adverse reactions to antimalarial	Yes Yes No Yes	2006 - --

Figure 2. Malaria control strategies/initiatives in Kenya.
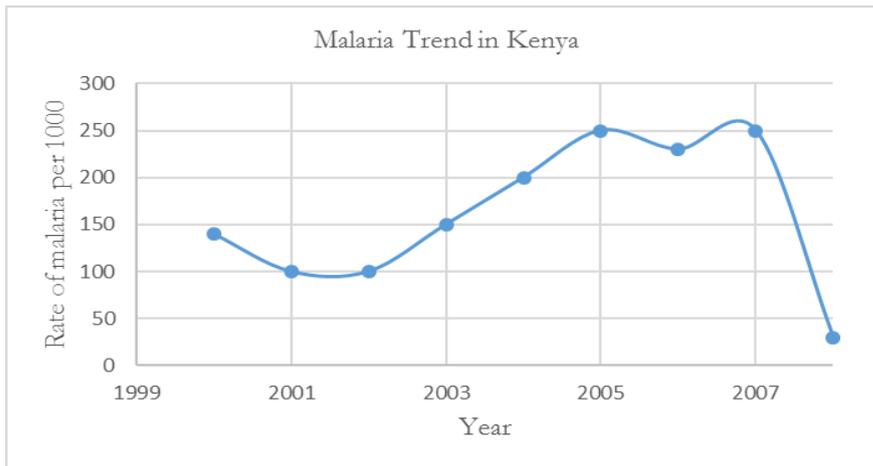
Figure 3. Trend in the rate of malaria. Source (WHO, 2009).data beyond 2008 not available.

Malaria risk maps are produced accurately at the national level in Kenya, they are less usable for high-resolution surveillance (see figure 4). This makes it hard to conduct monitoring and evaluation at the local level. The integration of satellite-based data with in situ data for surveillance as indicated by Midekisa et al. (2012) can assist in the generation of fine resolution malaria risk maps.
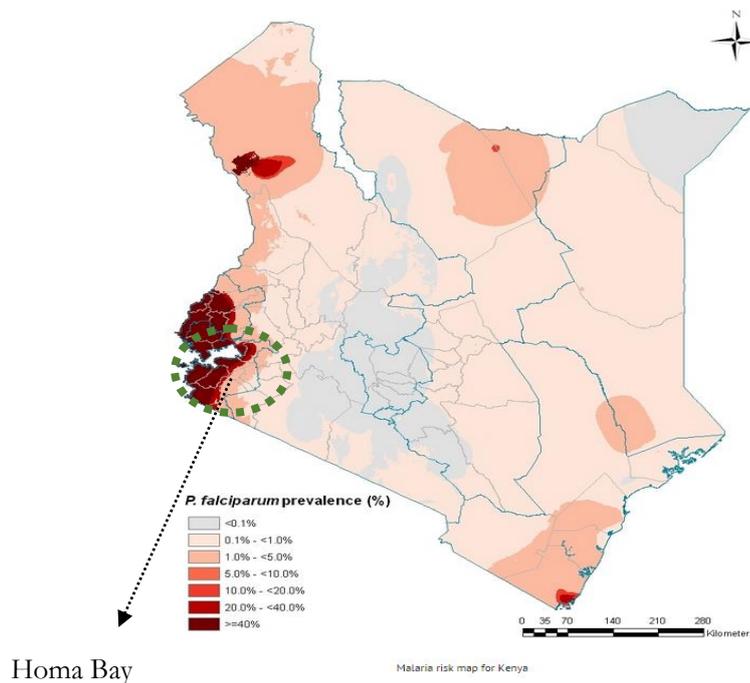
Figure 4. Plasmodium falciparum prevalence for Kenya, source (Noor et al., 2009).

2.4 Malaria risk modelling.

Successful control of malaria depends on detailed knowledge of its epidemiology (Koram et al., 1995). Modelling is, therefore, necessary for mapping the spatial patterns of malaria and generating knowledge for malaria elimination (Tuyishimire, 2013). In the recent past, various data-driven modelling frameworks have been developed by logically combining statistics and geographic information systems, namely geo-statistics models (Noor et al., 2009). They include, factors analysis model, regression models and habitat suitability models. However, habitat suitability models are based on species presence data which is unavailable and

costly in most cases (Cossio et al., 2012). In this study, factor analysis is used in modelling malaria risk due to lack of binary malaria infection data and mosquito presence data.

Risk factors and disease infection rates have been combined in a modelling environment to determine the spatial clusters and patterns within a given area (Tuyishimire, 2013). In order to achieve this, geostatistical mapping models are used in determining the relationship between malaria spatial distribution and environmental data, example temperature, rainfall, altitude, slope, distance to hyacinth, distance to wetlands and, and land cover among others (Kelly et al., 2012). Socio-economic characteristics (poverty levels) of a given population has also been included in malaria risk modelling to bring on board the coping capabilities of the population, the poor are more vulnerable as compared to the rich (Koram et al., 1995).

Many researchers use straight line relationship between the aforementioned without testing, this leads to incomplete risk analysis as explained by Austin (2002) in his book on species distribution and ecological theory. Consequently, correlation analysis between environmental factors (climatic, ecological, topographical, and demographical factors) and infection rates has been proposed to reduce the effect of straight-line relationship (Curtis & Carey, 2012).

According to Kelly et al. (2012), advanced GIS-based analysis (for example overlay index method and spatial multi-criteria evaluation) have been adopted to identify malaria risk zones at various spatial scales taking into consideration spatial relation among interplaying factors. Additionally, health facility catchments spatial extents are considered elaborate and appropriate in modelling disease risk as this is the lowest level of interaction between patients and health systems (Noor et al., 2006).

Traditional malaria risk modelling involved the use of health surveys interpolated to create the risk map (Onchiri, 2014). Onchiri (2014) further explains that the method uses interpolation process that introduces arithmetical errors in the analysis; the variability in environmental data used in data driven models is also lost. Moreover, in traditional environmental based risk models, representation of continuous risk factors like temperature and rainfall has being difficult, the meteorological stations are not distributed evenly in space introducing errors in the data during interpolation (Phillips & Marks, 1996). In addition, the method is also considered expensive and not accurate for high-resolution risk mapping (Kelly et al., 2012). Malaria risk modelling in this study comprises of three stages, namely *identification of risk factors, generation of health catchments, and application of data-driven models on identified risk* (see figure 5).

Steps	Method	Reason	References
1	Malaria risk factors	To identify the risk factors	(Stresman, 2010).
2	Health catchment delineation	To elaborately present malaria occurrence data, since patient origin data is not available.	(Alegana et al., 2012).
3	Factor analysis	Explain the variation between interplaying risk factors.	(Nardo et al.,2008),
4	Correlation analysis	Correlation measures the degrees of strength to which two or more variables are linearly related. To test straight line relationship	(Curtis & Carey, 2012)
5	Overlay index analysis	To logically combine the factors to generate risk index.	(Gogu & Dassargues, 2000)

Figure 5. Steps in malaria risk modelling.

Malaria risk factors.

Various factors both environmental and socio-economic contribute to malaria risk (Githeko et al., 2006). According to Stresman (2010), malaria risk factors can be divided into two, namely main factors and modifying factors. The main factors directly affect malaria risk as they tend to affect mosquito development directly (example temperature, land cover and rainfall), the modifying factors are indirect in their effect; they contribute to a more conducive environment for mosquito breeding and development. In addition,

Gwitira et al. (2015) also list malaria risk factors including temperature, rainfall, altitude, slope, evapotranspiration, presence of wetlands and water hyacinths.

These factors vary from place to place in terms of significance in risk contribution (Homan et al., 2016). Stryker & Bomblies (2012) reports that land cover plays a key role in mosquito development. Cropland mainly maize, sorghum, millet, and rice enhance larval development. This is confirmed by Ye-Ebiyo, Pollack, Kiszewski, & Spielman (2003) in the study on effects of maize proximity to larval development, where mosquitos near the aforementioned land cover were found to be bigger than mosquitos far off. Big size mosquitos live longer hence more malaria transmission and more risk (Ye-ebiyo, Pollack, Kiszewski, & Spielman, 2003).

The full life cycle of mosquito depends on favourable temperature, this includes mosquito population dynamics and malaria transmission as discussed by Beck-Johnson et al.(2013). According to Malone et al.(2003), a temperature range of 25° to 30°C is an optimum condition for mosquito development increasing its density and malaria infection rate. Malone et al.(2003) further state that temperature above 30°C and below 25°C drastically reduces mosquito population rate leading to low malaria infection rates in such zones. Rainfall is also a main contributing factor. Low rain intensity is associated with high larvae presence as excess rain flashes out the premature larvae (Illinois Education., 2015). Areas with low slopes and slow in draining water experience high infection rates. Illinois Education institute (2015) further explains that water is held for many days in these zones creating a favourable breeding site for mosquitos after rainy seasons.

Topography has a great influence on mosquito development; locations with high elevation (altitude) value (above 1800m above sea level) are considered unsuitable for mosquito development hence low malaria risk. A Study conducted by Sambasivarao (2013) on participatory risk mapping of malaria confirms that low malaria transmission is experienced in these locations. Topographic wetness described as the spatial distribution of moisture saturation and a component of soil hydrological condition is another risk factor. Flood water mosquitos lay eggs in moist soils with high topographical wetness and optimum temperature in the absence of water ponds (Illinois Education., 2015). This renders zones without wetlands but with high topographical wetness as potential threat zones (Cossio et al., 2012). In addition, Stresman (2010) records that zones with evapotranspiration levels lower than 800mm per year experience high malaria transmission, as this increases the topographical wetness (suitable for flood water mosquitos to lay eggs).

Koram et al.(1995), includes poverty as a factor for malaria risk. In modelling malaria risk in Gambia, poverty is considered of great significance as a positive association is revealed between poverty levels and malaria risk; high poverty rates (which was indicated by poor housing) leads to high malaria risk. (Koram et al.,1995). Moreover, zones close to wetlands and water hyacinth are considered to be of high malaria risk, within the flight range of mosquitos (1-3 miles approximately 1-4km) the infection rates are high (American Mosquito Control Association, 2015). American Mosquito Control Association (2015) further states that water hyacinths provides breeding mats for mosquitos and, therefore, a risk factor.

Migration of people from one place to another has also been considered as a risk factor, human mobility exposes non-immune people to new malaria transmissions or risk (Heggenhougen et al., 2003). Heggenhougen, Hackethal, & Vivek (2003) further mention that cultural behaviour of the people (like attitude and perceptions towards ITNs and time of use of ITNs) and malaria parasite resistance to anti-malarial drugs fuels malaria infection. This aspect is of malaria risk is costly and complex to bring under surveillance explains Onchiri (2014) as it requires periodic household surveys.

The next section explains various data-driven modelling procedures in malaria risk mapping.

Factor analysis model.

Factor analysis describes set of analogous methods rather than single techniques; it's basically a way of describing a large number of variations with a small number of latent (Kahn, 2006). It attempts to identify factors that explain the pattern of correlation within a set of observed factors. Kahn (2006) further mention that both data types are usable in factors analysis, continuous and dichotomous. There exist two types of factors analysis namely, explanatory and confirmatory. Explanatory factors analysis is applied in identifying

complex relationship between variables without setting any predefined structure while confirmatory factor analysis is used to test hypothesis, reconfirm or validate already defined structures (PAI, 2015).

Uses of factor analysis include data reduction, structuring of data, classification (clustering), scaling of data, exploration, mapping and hypothesis testing (Rummel, 2015). In ecological studies, factor analysis is used to explain the interplay of components used in mapping to create risk or habitat maps (Kaplunovsky, 2005). According to Nardo et al.(2008), sets of rules of thumb exist for selection of variables. These rules include, Kaiser criterion (it drops all the factors with eigenvalues below 1.0 as they explain less variance), Scree plot (plots the eigenvalues, selecting factors that sharply drop before levelling off) and variance explained criteria.

The factors analysis model is given by a set of variation and covariation variables x (j=1to p), the function of factors η (k=1 to m) and residuals ε (j=1 to p).

$$x_{i1} = v_1 + \lambda_{11}\eta_{i1} + \lambda_{12}\eta_{i2} + \ldots + \lambda_{1k}\eta_{ik} + \ldots + \lambda_{1m}\eta_{im} + \varepsilon_{i1} \qquad (2.4)$$
$$x_{ii} = v_i + \lambda_{i1}\eta_{i1} + \lambda_{i2}\eta_{i2} + \ldots + \lambda_{ik}\eta_{ik} + \ldots + \lambda_{im}\eta_{im} + \varepsilon_{ii}$$
$$\ldots$$
$$x_{ip} = v_p + \lambda_{p1}\eta_{i1} + \lambda_{p2}\eta_{i2} + \ldots + \lambda_{pk}\eta_{ik} + \ldots + \lambda_{pm}\eta_{im} + \varepsilon_{ip}$$

Where v_i are intercepts, λ_{ik} are factor loadings, η_{ik} are factor values, and ε_{ii}.

Several approaches for factor extraction exist namely generalised least squares, maximum likelihood, alpha rationing, and PCA axis method. According to Nardo et al. (2008), the most common used method is the PCA as it is recommended for developing composite indicators. In addition, various data rotation types exits after the initial extraction methods, this includes orthogonal rotations (varimax and equimax), and oblique rotation example promax (Institute for Digital Research and Education, 2015). Orthogonal rotation imposes the restriction that the factors cannot be correlated while oblique rotation allows correlation of factors. However, the issue of which extraction method to use in retaining most information is still under discussion (Nardo et al., 2008). For this study, principal component analysis is used to analyse composite indicators because PCA assumes that the initial communality is 1. According to Kaplunovsky (2005), this means that equal variance is awarded to all the factors before extraction.

Correlation analysis.

Correlation measures the degrees of strength to which two or more variables are linearly related (Brutlag, 2015). With its application dating back to 1850 in biological fields, correlation analysis has since being applied in other fields like urban planning to show the relation among urban phenomena (Páez & Scott, 2005). Additionally, Pearson correlation coefficient is used in testing linearity between variables (Laerd Statistics, 2015). The correlation values range from 1 to -1, with value 1 as perfect positive correlation and -1 as perfect negative correlation (MetaStock, 2015). Zero denotes no linear association between the variables. Equation 2.5 shows the correlation analysis formula.

Correlation analysis equation: $= r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$ \qquad (2.5)

Where r is the correlation coefficient,
N is the number of variables or observations.
$\bar{x}$ Represents the mean
x represents the data point in question.
S is the standard deviation for variables x and y.
If there is a relationship between the variables, then as one deviates from the mean the other should also deviate in either the same direction or different direction (Field, 2012).

Health catchment spatial modelling.

Health care utilization is affected by several factors which include geographic accessibility, therefore, empirical data is required to understand accessibility concept (Alegana et al., 2012). According to Alegana et al.(2012), understanding how population utilize health care and defining the spatial extent of health catchment is important for efficient planning and distribution of health service. In African countries, adequate information on demographic characteristics and economic power of a given population is rarely

available to help develop health catchment models. In addition, few countries have complete and reliable spatial database of health service and providers (Alegana et al., 2012).

Various health catchment models have been proposed based on data availability. The first being straight-line distance model also known as Euclidian distance method. It is based on establishing the extent of catchments by calculating the distance from the facility to the patient resident, using the straight-line distance (Euclidian). According to World Health Organization (2015), this approach assumes that people visit the closest facility as distance overrides other factors. This method also fails to account for different topographical features, road networks and the difference in utilization rates (Alegana et al., 2012).

The second model is drive time method based on the road network. This method is preferable in developed countries with widespread vehicular transport; it is rarely used in developing countries where large population walk to health facilities and transport networks data is not available (Alegana et al., 2012). The last method is the raster based cost surface method (based on travel time on multiple factors which include slope, roads, rivers and land cover). According to Soediono (1989), slope influence the route selection of human beings. In addition, the speed at which human beings walks is influenced by the land cover type, with rivers and wetlands impassable by foot (Alegana et al., 2012). The type of road selected also determines the speed of travel as explained by Sturrock et al.(2014) in fine scale malaria risk mapping study. "These forms of distance measurements are used to analyses utilization by metrics, using the number of facilities within a certain distance to other health facilities and gravity model"(Alegana et al., 2012).

The gravity model first applied in the field of economics is comparable to Newtons law of gravity (Anderson, 2010). It states that forces between two bodies varies proportionally to the product of their masses and inversely to distance between the two bodies. In this model, patient interplay with health facility is denoted by flow from patient origin and masses represented by utilization effects example health facility capacity (Alegana et al., 2012).

Moving speed is calculated either by using the anisotropic or isotropic principle. Tobler's hiking function; referred to as the exponential function determining the speed considering the slope angle, is applied in the calculations (see equation 2.6 for Tobler's speed formula). In summary, the walking speed decreases with an increase in slope angle. (See figure 6, Tobler's hiking function graph).

Tobler's equation. *$V = 6*exp(-3.5 [(S+0.05)])$* (2.6)
Where V is the calculated speed

   S is tan θ; θ is the slope in degrees calculated by elevation difference divided by cell distance (spatial resolution)
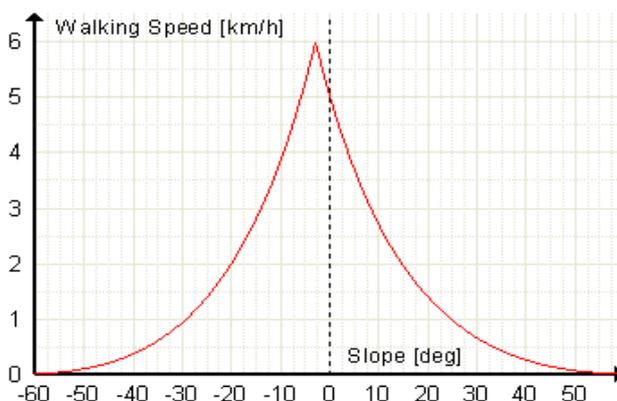
Figure 6. Tobler's hiking function graph. (Source from Tobler (1993).

Overlay index model.

First applied in risk assessment of groundwater by Gogu & Dassargues (2000) in the year 2000, its application in risk mapping has far grown since then. Its entails overlaying of various risk factor maps showing potential zones based on known thresholds under which risk occur (Gogu & Dassargues, 2000). Gogu & Dassargues (2000) further mentions that overlay index method mainly relies on quantitative and visual interpretation of mapped data. In addition, factors are rated from 1 to 10, depending on perceived and known significance acting as weight to be applied in indexing (Gogu & Dassargues, 2000). The risk index is the weighted sum of factors, its computed by using the formula in equation 2.7.(equation adapted from Gogu & Dassargues (2000).

$$\text{Overlay index} \sum_{j=1}^{n} (W_j * R_j) \quad (2.7)$$

Where n, is the number of factors.

Wj represents the weight factor.

Rj is the rating factor.

2.5 Remote sensing application in malaria risk analysis.

National Oceanic and Atmospheric Administration (2015) defines remote sensing as the science of obtaining information about the phenomenon from a distance. Remote sensing, geographic information systems and modelling combined have contributed towards clear understanding and investigation of ecology (Cohen & Goward, 2004). It enables scientist to model the phenomena in question spatially showing illustrations of its various attributes.

Remote sensing innovations has proved useful in public health and epidemiological studies (De Oliveira, Dos Santos, Zeilhofer, Souza-Santos, & Atanaka-Santos, 2013). The ability to combine thematic set of layers in space and time for risk analysis has been made possible with products from various satellites example Landsat and moderate resolution imaging spectroradiometer. According to Cohen & Goward (2004) regional and local applications relying on temporal data sets enabling explicit ecological modelling has been made possible and more accurate (spatial accuracy). However, in order to reduce the burden of malaria occurrence, various stakeholders need to be aware of the risk in time and space (Adu-Prah & Kofi Tetteh, 2015). In addition, identifying malaria risk zones and information retrieval from imaging satellites is important in prioritizing action areas and strategies; relevant information is extracted and used towards timely prevention and vector control (Nath et al., 2013).

2.6 The Relation between mosquito and malaria.

Malaria transmission in a given region is dependent upon the presence of susceptible anopheles mosquito feeding on man (Mutuku et al., 2009). Moreover, a study conducted in Nyanza province (Kenya) by Hightower et al. (1998) indicate positive associations between malaria occurrence in humans and mosquito distribution. This indicates that the higher the malaria transmission and occurrence the higher the mosquito presence. Therefore, malaria prevalence or occurrence in a given region can be used as an indicator of mosquito presence.

2.7 Malaria control conceptual framework

The conceptual framework consists of five parts, namely, environmental variables (1), anopheles cycle (2), parasite cycle (3), socio-economic characteristics, (4) and finally malaria control (5). See figure 7.

Environmental variables: Constitutes of climatic, ecological and topographical factors that promotes the development of mosquito. Climatic factors include temperature, rainfall and evapotranspiration. Ecological factors are wetlands, soil drainage, vegetation cover and hydrology. Topographical factors include slope and altitude. However, ecological factors are influenced in one way or another by human activities leading to variation in malaria risk within a given region (Adu-Prah & Kofi Tetteh, 2015).

Anopheles cycle: Suitable breeding conditions for complete vector cycle leads to high mosquito population and malaria density; reduction in the same leads to more mortality rates in the mosquito population. This is directly affected by control measure in place.

Parasite cycle: Cycle begins with a mosquito bite, followed by an incubation period of 7 to 10 days. The individual can either be cured or succumb to the disease.

Demographic and socio-economic characteristics: Children under the age of five are most vulnerable to malaria (Kleinschmidt et al., 2000). However, the vulnerability depends on the composition and poverty level of the population (Koram et al., 1995).

Malaria control: This includes various measures put in place to prevent transmission of malaria. The measures are vector control (natural), diagnosis (treatment) and transmission prevention (artificial).

This study focuses only on the 1st component (environmental variables), 3rd component (Parasite cycle) replacing the cured and deaths with malaria occurrence, and 4th component (Demographic and socio-economic characteristics). The interaction between the three components is used in the creation of risk map.
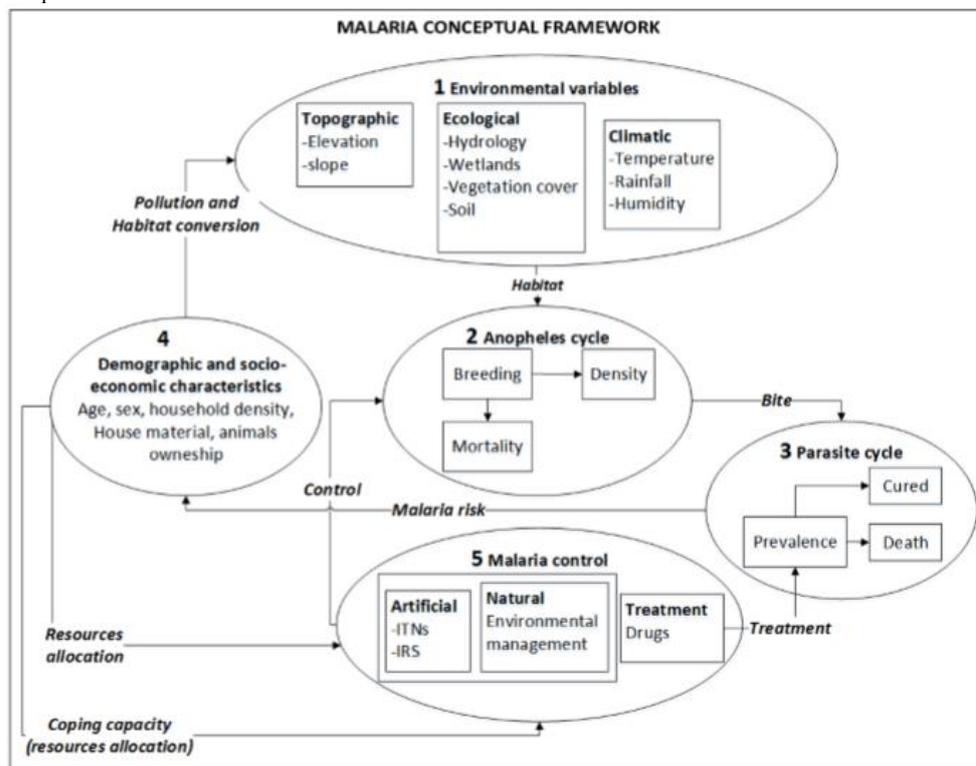
Figure7. Conceptual framework adopted from Tuyishimire (2013).

2.8   Link to methods and data.

Correlation and factor analysis models were used in malaria risk analysis as applied by Malone et al.(2003) in studying malaria risk assessment in Eritrea. This is due to lack of mosquito presence data for Homa Bay County. In addition, malaria infection rate derived from malaria occurrence and population was used as a measure of risk. The factors were categorised into climatic (temperature, evapotranspiration and rainfall), hydrological (wetlands and soil drainage), topographical (slope and altitude), ecological (land cover, vegetation cover, topographical wetness and hyacinth) and socio-economic (poverty) based on literature review; the factors were derived using remote sensing image analysis and GIS techniques.

Raster based cost surface travel method was then used to create health catchments as it takes into consideration the topographical features, which either hasten or impede movement. Finally, overlay index method was used to logically combine the factors to generate risk index.

3 METHODOLOGY.

This chapter entails information about the study area (Homa Bay County), data preparation, data collection, data integration, and data analysis methods.

3.1 Study area.

Homa Bay County is located along the shores of Lake Victoria, in the western part of Kenya. The county covers an area of 3,161 sqkm, administratively subdivided into 6 sub counties, 40 wards, 19 divisions, 116 locations and 226 sub-locations. Figure 8 shows the area of study.
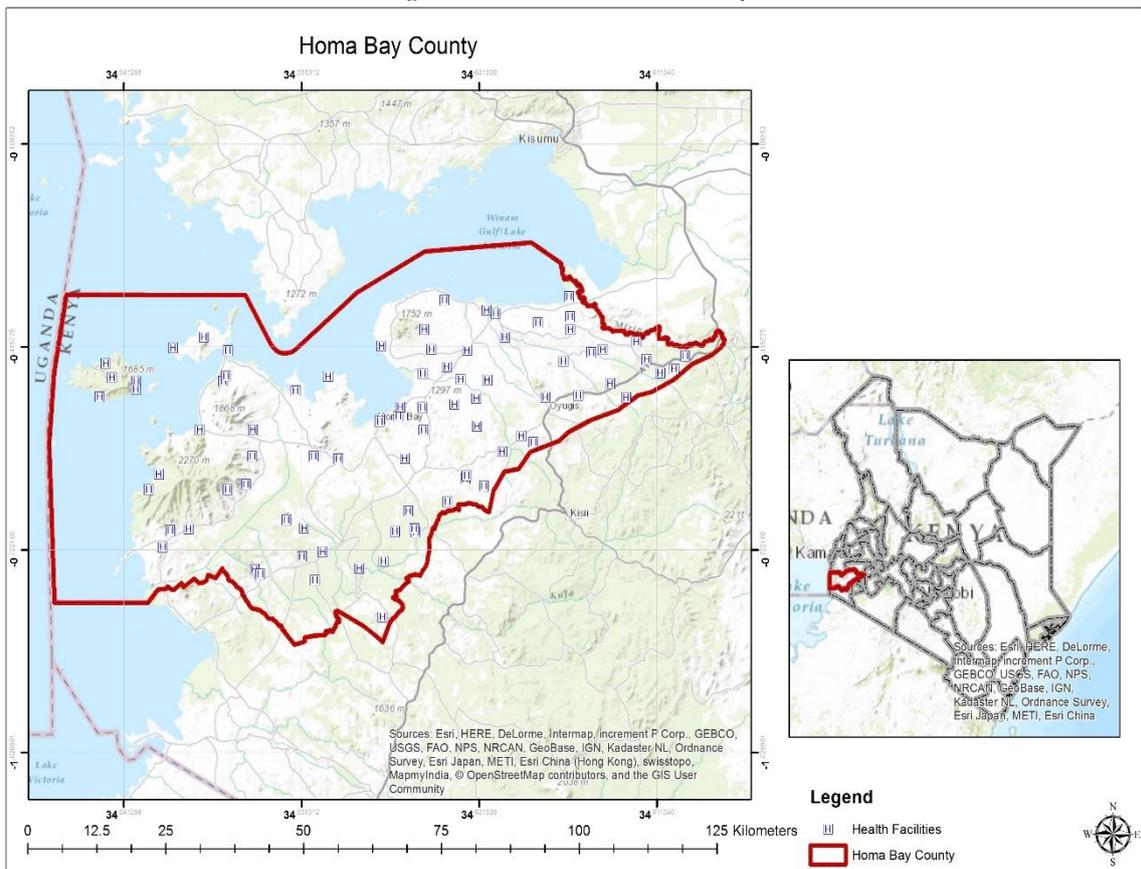
Figure 8. Study area.

Demographic and economic characteristics.

According to Homa Bay County Government (2015), approximately 1.06 million people resided within Homa bay county in the year 2014; the population is estimated to be 1.7 million by the year 2017. The number of females is more than the male at 540,000 and 500,000 respectively (Homa bay County Government, 2015). In addition, population density varies from one sub-location to the other within the county.

Located along the shores of Lake Victoria, the main economic activities are fishing and agriculture accounting the largest share of household livelihood. The main cash crops grown in Homa Bay County includes maize, millet, sorghum, potatoes, sugar cane, pineapples and ground nuts (Kenya Investment Authority, 2015). However, World Resources Institute (2015) poverty data indicate that 713,000 people out of 906,000 in 1999 belonged to the absolute poverty group. In addition, only 193,000 were above the absolute poverty. The proportion of poor individuals has remained the same since 1999 as reported by Answers-Africa (2015). Based on this, the population under absolute poverty is approximately 867,000 people out of 1.06 million people in the year 2014.

Climate and topography

The county has a semi-arid climatic condition, with an annual temperature ranging between 17°C (minimal) and 34°C (maximum). With an estimated annual rainfall of 1,100mm (minimum of 250mm and maximum of 1200mm), the county experience two rainy seasons March to June (long rains) and September to November (Kenya Information guide, 2015). Elevation ranges from 1130m to 2270m above sea level. (Elevation values generated from digital elevation model by Shuttle radar topographic mission).

Transport.

Homa Bay County is accessible by road through Kisumu in the north-eastern part, on the south-eastern part through Kisii and Migori. In addition, it is accessible by lake through Mbita, Jinja (Uganda) and Bukoba (Tanzania). Homa Bay County is also accessible by air travel through the newly built Kabunde airstrip. However, the main means of transport are walking and motorcycling (UN-Habitat, 2008).

Healthcare system.

The health system in Homa Bay County is hierarchically organised: (1) dispensaries, (2) health centres, and (3) district hospitals. It includes nine 3rd-tier hospitals, 29 health centres, and 46 dispensaries which are connected to 127 community health units (County Government of Homa Bay, 2013). This includes both public and private facilities. However, primary health care services are provided at the dispensary level with secondary and tertiary services provided at the health centres and nine 3-tier hospital level respectively. In addition, malaria occurrences recorded by the community units are reported at dispensaries level, which further reports to the Ministry of Health malaria control department. Finally, the doctor-patient ratio in the county is 1: 40,000 and nurse patient ration of 1:1,500 (County Government of Homa Bay, 2013).

Homa Bay County is one of the malaria epidemic zones in Kenya (see figure 4). Out of 6.2 billion Kenyan Shilling (USD 60M approved budget for the year 2015-2016), 1.4 billion Ksh (USD 13M) was allocated for health sector (Standard Digital News, 2015). This accounts for 22% of the county's budget. Figure 9 shows the location of 84 geocoded health facilities within Homa Bay County.
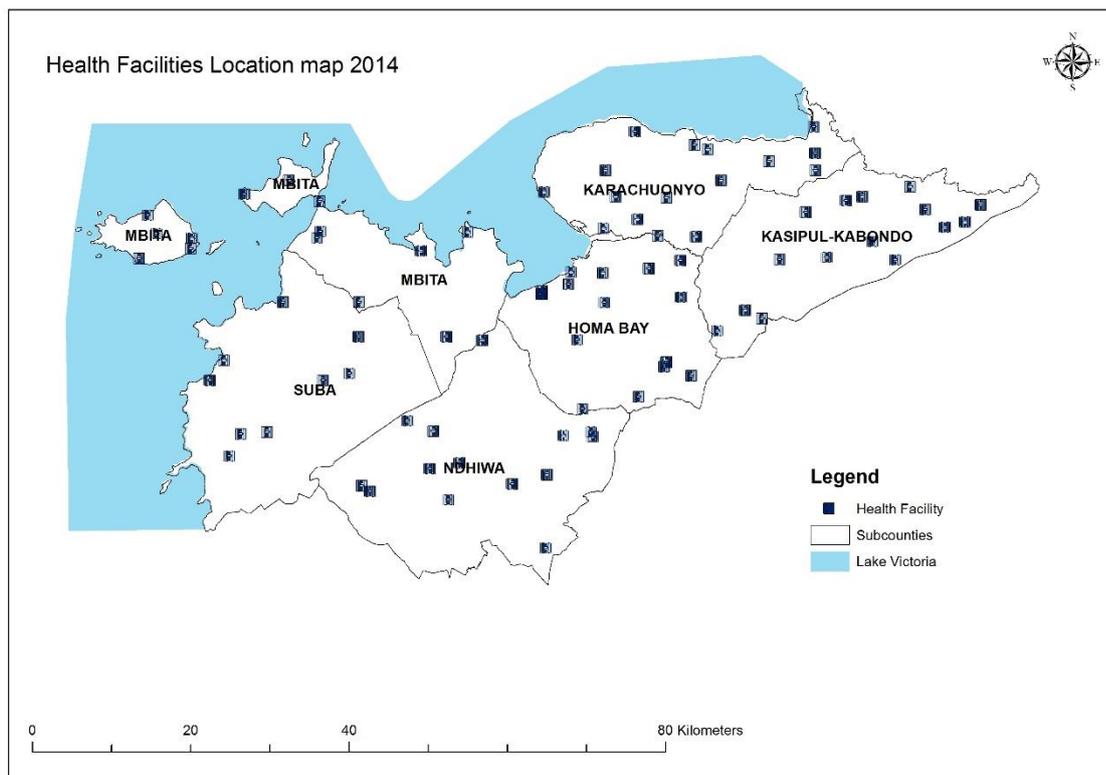
Figure 9. Health centres location in Homa bay County.

Various control measures have been initiated by the health department of Homa Bay County since the year 2001, see figure 2 (Malaria Control, 2010). They include insecticide-treated nets (ITN), indoor residential house spraying (IRS) and intermittent presumptive treatment (IPT). However, a single dose of primaquine used as gametocidal and larval control measures have not been implemented in Homa Bay County.

3.2 Data preparation.

Satellite imageries from Landsat, Sentinel 1 and moderate resolution imaging spectroradiometer (MODIS) for the year 2014 were acquired from United States Geological Surveys (USGS) and sentinel hub respectively. The base year for all the imageries was 2014. Landsat imageries were for the months of February, May, September, and December. Sentinel 1 radar imageries for months of May, June and July (based on image availability).

Monthly *evapotranspiration and rainfall* product for 2014 were sourced from MODIS and climate hazard group infrared precipitation with station data respectively. List of datasets generated from Landsat were *land surface temperature, normalized difference vegetation index, water hyacinth, topographic wetness and land cover*. Additionally, *wetlands* dataset was generated from sentinel 1. Land cover was classified into six main classes for field validation process and, health facilities excel sheet prepared with various months to record malaria data from the health departments. Moreover, random sampling was applied in fieldwork; each sub-county allocated 10 random points for land cover validation. Finally, expert questionnaires were developed for purposive sampling applied in land cover standardization (See appendix 1, 2 and 3 respectively). Figure 10 shows the list of data for generation, sources, and spatial resolution.

1.Environmental variable	Source	Spatial resolution.
Land cover	USGS Landsat.	30m
Land surface temperature	USGS Landsat	100m
Evapotranspiration	NASA Earth explorer (MODIS)	1km
Water Hyacinth	USGS Landsat	30m
Slope	SRTM NASA	30m
Water ponds/wetlands	Sentinel 1	10m
Elevation (altitude)	SRTM NASA	30m
Rainfall	CHIRPS	5km
Soil drainage	Kenya Soil Survey	1km
Topographical wetness	USGS Landsat	30m
Population	Census 1989 and 1999	30m
2. Malaria occurrence/cases	Ministry of Health	none
Poverty density	World resource institute	1km

Figure 10. List of data generated and sources.

3.3 Data generation and collection.

This research was primarily based on both primary and secondary data comprising of environmental variables from satellite images. Malaria occurrence data sourced from the Ministry of health Homa Bay County and socio-economic data from 1989 and 1999 Kenyan census.

3.3.1 Environmental and socio economic variables.

This subsection describes the methodology used to derive the factors from satellite images and other sources.

Land cover.

The data was generated from radiometrically corrected 30m spatial resolution Landsat 8 imageries by creating a layer-stacked image of every month (February, May, September, and December). The imageries were selected based on the cloud cover, leaving out imageries for the months with more than 10% cloud

cover. Radiometric calibration was done for each spectral band by converting image digital number to top of atmosphere reflectance as discussed in the Landsat 8 handbook (USGS, 2015). The stacked imageries were then co-registered to allow pixel-to-pixel overlay and used in creating a multi-temporal image, which was further classified using supervised classification (maximum likelihood classifier) into six classes. The classes were settlements, bare ground, cropland, grassland, forest, and water.

Land surface temperature.

Based on the split window algorithm, the land surface temperature was generated from Landsat 8.(Rajeshwari & Mani, 2014). For each month (February, May, September, and December) thermal bands 10 and 11 were converted to radiance and then used in the calculation of temperature brightness. Corrected near infrared and red bands (Top of atmosphere reflectance) were used to derive normalized difference vegetation index, the maximum vegetation value and soil values picked and used in the generation of fractional vegetation cover.

The fractional vegetation cover was then used to generate land surface emissivity for bands 10 and 11, to be applied in the split window. Atmospheric water vapour value used in the split window algorithm was sourced from Aerosol robotic network (Goddard Space Flight Center, 2015). The final output of the four months were averaged to give the annual land surface temperature in degrees Celsius.
The spatial resolution of Land surface temperature is 100m with a mean bias ranging from -1°C to +2.2°C (annual bias).

$$LST = TB10 + C1\ (TB10\text{-}TB11) + C2\ (TB10\text{-}TB11)2 + C0 + (C3+C4W)\ (1\text{-}\ \varepsilon)$$
$$+ (C5+C6W)\ \lrcorner\ \varepsilon \qquad\qquad (3.1)$$

Where,
LST is the land surface temperature (K)
C0 to C6 - Split window coefficients values (Rozenstein, Qin, Derimian, & Karnieli, 2014), see appendix 4.
TB10 and TB 11- brightness temperature for bands 10 and 11.
ε is the mean LSE (land surface emissivity) for bands 10 and 11.
W is the water vapour content.
Δ ε – Difference in LSE.

Evapotranspiration.

Evapotranspiration data was sourced from National Aeronautics and space administration (NASA) project for estimating global evapotranspiration from the land surface using moderate resolution imaging spectroradiometer (MODIS) satellite. The product is referred to as MOD 16. It includes global evapotranspiration (ET), latent heat influx (LE) and potential LE datasets at 1km² spatial resolution regularly. With time intervals of 8 days, monthly and yearly, the datasets covers the period between the years 2000 to date.

The algorithm used in the generation of MOD 16 is based on Penman_Monteith equation described by Mu, Zhao, and Running (2011). It was applied in evaporation estimation study for lake Nasser by Hassan (2013) and evaluation of two end member based surface evapotranspiration study by Tang & Li (2015). The mean absolute bias (MAE) ranges from 0.31mm to 0.40mm per day (Mu et al., 2011). Monthly datasets from January to December were averaged to achieve the annual evapotranspiration for Homa Bay County. The data was converted to points and interpolated to generate 100m spatial resolution dataset. (The units are in mm/year).

Water ponds (wetlands).

Water ponds were derived from processed sentinel 1 radar imageries with a spatial resolution of 10m (ground range resolution). The processes included geometric correction, calibration, speckle filtering, co-registration and finally sub-setting/clipping. The stacked image for three months (May, June and July) was classified into two classes (water and others) using support vector machine classifier (see appendix 5 for classified wetlands result). Radar image was used in water points mapping due to its backscatter property on smooth surfaces). Consequently, wetlands layer was converted to points.

Euclidean distance to wetlands points was computed from the points. Distance to wetlands was used instead of wetlands to incorporate the 1km flight range of mosquitos. (The same reason applies to water hyacinth).

Topographical wetness.

Topographical wetness was generated from radiometric corrected Landsat 8 data. Six bands for each corrected imageries were used in deriving the wetness data for each month (bands 2 to 7). Tasselled cap transformation was applied to the bands based on the coefficients discussed by Baig, Zhang, Shuai, & Tong (2014). The values were then standardized by a rescaling factor creating a range between 0 and 1. 1 represents areas of high topographical wetness value and 0 area with low wetness value.

Topographic wetness= *(coeff2*band2) + (coeff3*band3) + (coeff4*band4)*
$\qquad\qquad$ *+ (coeff5*band5) + (coeff6*band6) + (coeff7*band7)* $\qquad$ (3.2)

Where, band represents corrected bands 2 to 7 and coefficients represents tasselled cap transformation coefficient values (see appendix 6). The spatial resolution of topographic wetness product is 30m.

Normalized difference vegetation index (NDVI).

The normalized difference vegetation index dataset was generated from reflectance bands 4 and 5 of Landsat 8, band 4 being the red band and band5 the infrared. NDVI for the four-month imageries were generated and averaged to create annual 30m spatial resolution NDVI data set. Equation (3.3) shows the NDVI generation formula.
The NDVI equation is as follows.
NDVI = *(Infrared band −red band)/ (Infrared band+ red band)* $\qquad$ (3.3)

Water hyacinth.

Water hyacinth dataset was generated from radiometric corrected near infrared bands of Landsat 8. Infrared bands for the four months imageries (February, May, September and December) were stacked together to create a colour composite. Areas along the lakeshore with high infrared reflectance in all the four months indicated stagnant water hyacinth (based on visual interpretation). Unsupervised classification was performed on the image and other classes (except water hyacinth class) combined to create a two class classified image. Water hyacinth raster dataset was extracted and converted to points. The points were used in creating the Euclidian distances from water hyacinth. Hexagon tessellations were created to analyse the number of cases within the flight range of mosquito (same for the distance from wetlands, see appendix 7 for 1km effect analysis flow chart).

Altitude (elevation) and slope.

Elevation dataset was sourced from NASA, shuttle radar topographic mission (SRTM) project. The dataset has a spatial resolution of 30m. The slope dataset was then derived from elevation sourced from shuttle radar topographic mission. The slope inclination was measured in degree rise with 30m spatial resolution.

Rainfall.

Rainfall dataset was sourced from climate hazard group infrared precipitation with station data (CHIRPS). It's a 30-year partial global rainfall dataset with 5km spatial resolution, spanning 50°S to 50°N (Climate Hazard Group, 2015). Monthly datasets were averaged, converted into points, and interpolated to achieve 100m spatial resolution. (The units are in mm per year).

Soil drainage.

Sourced from Kenya soil survey, the soil drainage data was classified into four classes depending on drainage capacity. The classes are extremely slow, slow, well and rapid. The data was then resampled to 100m spatial resolution.

Population data.

Population data was sourced from World resource institute; it entails Kenya census data for 1989 and 1999 (World Resources Institute, 2015). The rate of population change was calculated by determining the percentage growth above the base year population (1989 that was equated to 100%). The exponential growth model was selected over the geometric model for 2014 population projection. This is because population is a continuous phenomenon with overlaps from one generation to the other. See equation (3.4) for the exponential growth model.

Exponential growth model = $Pt(e^{(r*n)})$ (3.4).

Where Pt represents the population at the base year, *r* is the rate of growth, *n* is the number of years from the base year to projection year and *e* is an exponential function.

Poverty density.

Poverty density is defined as the number or poor people per sqkm. Poverty data was sourced from World resource Institute for the year 1999. The rate of change in poverty was calculated and used to estimate the population under poverty in 2014. Applying the assumption that the poverty proportion is the same (Answers-Africa, 2015), Out of 1,060,700 people, 867,000 are estimated to be poor in the year 2014.

3.3.2   Malaria occurrence data.

Malaria occurrence data was sourced from Ministry of Health Homa Bay County. Monthly confirmed and clinical malaria cases from health facilities were recorded into excel and summed to generate a monthly and yearly total. The total number of malaria cases in Homa Bay county for the year 2014 is approximately 428,000 people out 1.06 million population. The highest cases were recorded during the months of April, May, and June, with the lowest in January and December. (See figure 11). This is because April to June fall in the long rain season and therefore, more suitable habitats for mosquito breeding translating into high malaria transmissions (Mutuku et al., 2009). The data recorded at the community health care level is reported at the dispensary level. Eighty-four health facilities report the malaria cases to their various sub-county health departments headquarters every month (Eighty-four is the total number of facilities from dispensary to third tier facility). See figure 9 for health facility map.
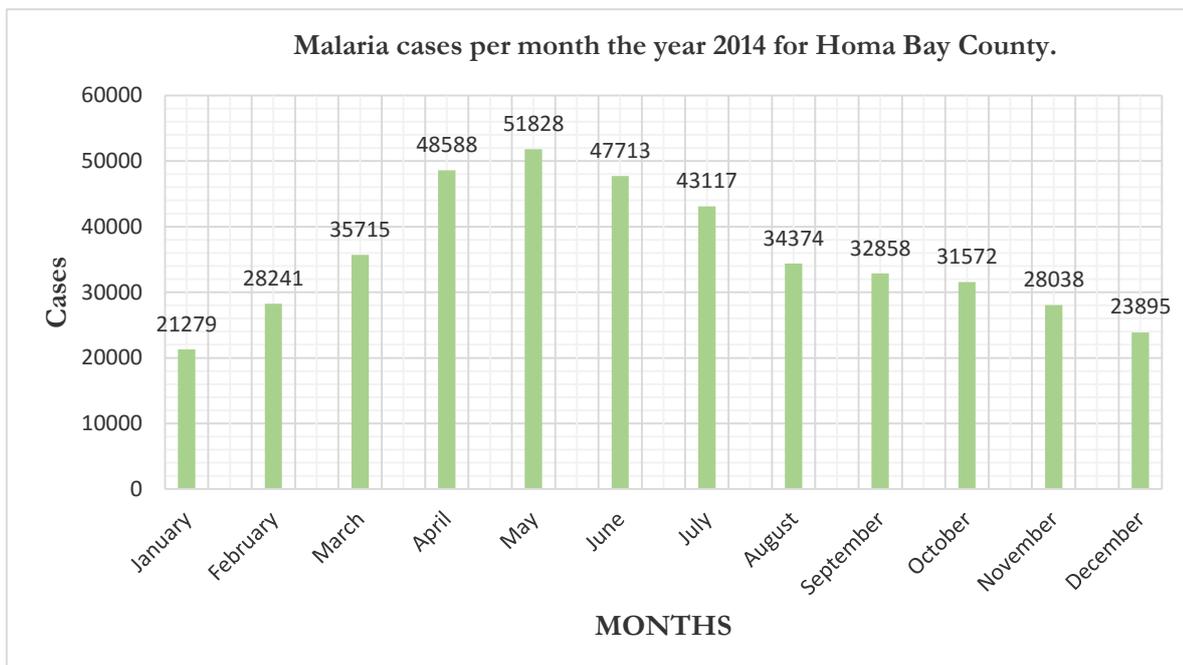
Figure 11. Malaria cases per month for Homa bay county the year 2014 (source Ministry of Health Homa Bay County).

3.4 Data integration.

The datasets were derived and collected from various sources and therefore, the need for data integration. The creation of health catchments to incorporate the malaria occurrence and its conversion to infection rate data is found in this section. This section also entails various integration procedures applied to standardizing the datasets.

3.4.1   Health catchment delineation.

In order to distribute malaria occurrence data across space, health facility catchments were created for each health facilities. The raster-based cost surface method was used in creating the catchments as explained by Alegana et al. (2012). Elevation data converted into slope in degree rise, land cover, rivers, wetlands and roads were combined to create the surface travel layer. The speed of movement on various slopes was calculated using Tobler's equation (see equation 2.6) in spatial analyst raster calculator. For steep slopes, the speed ranges from 0 to 1 while flat or low slopes have higher movement speed (above 3 km/h). (See figure 12, slope speed map). Various walking speed as shown in figure 13, adopted from Alegana et al.(2012) was allocated to land cover classes.
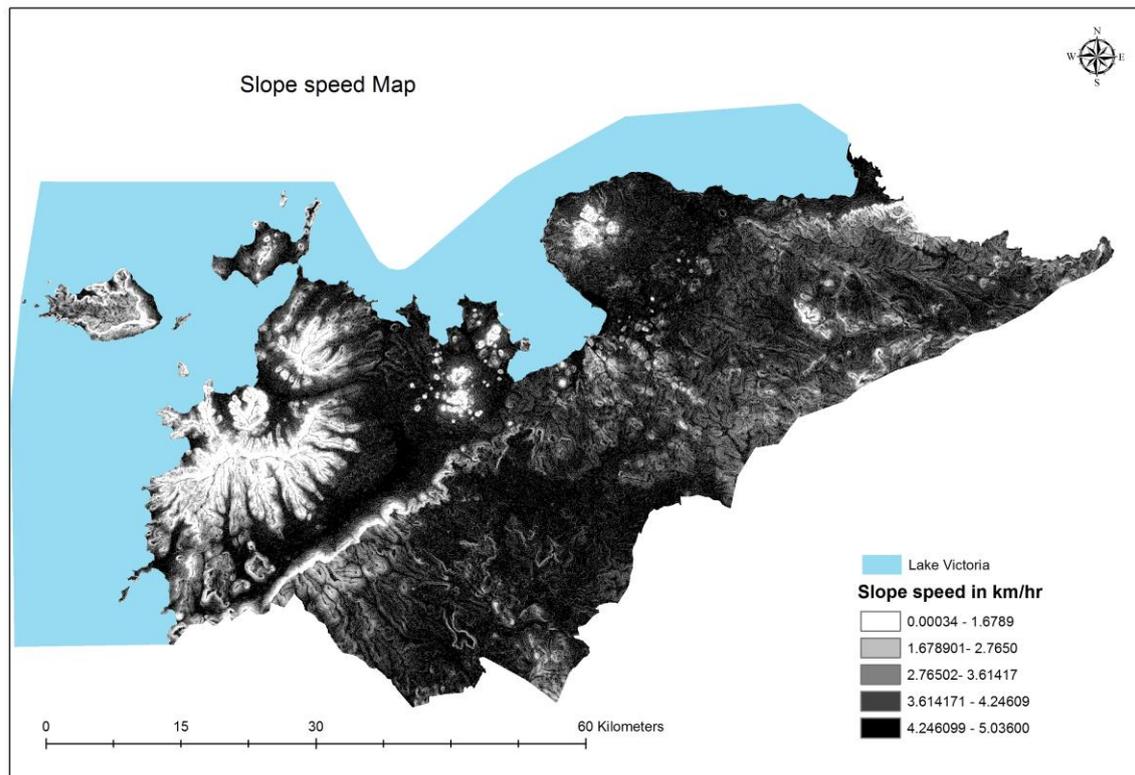
Figure 12. Slope speed map.

Description	Classification	Speed in Km/hr	Model
Spatial representation of land cover classes. Spatial resolution of 30m	Forest	1	Walking
	Water	0	None
	Cropland	2	Walking
	Grassland	4	Walking
	Bare ground	5	Walking
	Settlement.	3	Walking

Figure 13, assumed travel speed for each land cover class.(adopted from Alegana et al. (2012).

Barriers to movement (rivers and wetlands) were allocated a speed of zero, rasterized and mosaicked with the land cover layer. Areas that both land cover and river overlap, the river speed overruled the land cover speed and therefore allocated zero. The same was done to wetlands and land cover. Road layer was

classified into three categories namely, primary (class A), secondary (class B) and tertiary (class C) according to the road surface. Speed in km/hr was awarded to each class: 30km/hr to tertiary roads, 40km/hr to secondary roads and 60km/hr to primary roads based on motorcycling mode of transport (adopted from Alegana et al. (2012).

The road dataset was rasterised and combined with the land cover. Areas of overlap between the two layers, roads speed overruled the land cover speed assigning the road value to that particular pixel. Roads speed also overruled locations where it overlapped with barriers of movement speed (see figure 14a, harmonized land cover map with speed ranging from 0 to 60km/hr).

The harmonized land cover map was combined with the slope speed layer using the mosaic function in ArcGIS software. Based on visual interpretation, zones with steep slope (low speed) lay on forest (in terms of land cover) and were awarded the slope speed. Locations with low slopes (high speed) and high land cover speed (roads) were awarded the land cover speed. Locations with high land cover speed (roads) and steep slopes (low speed) were awarded the land cover speed (road speed). Barriers to movements overruled the slope speed in areas where the two overlapped. The combined slope-land cover speed layer was converted into m/s. see figure 14b.
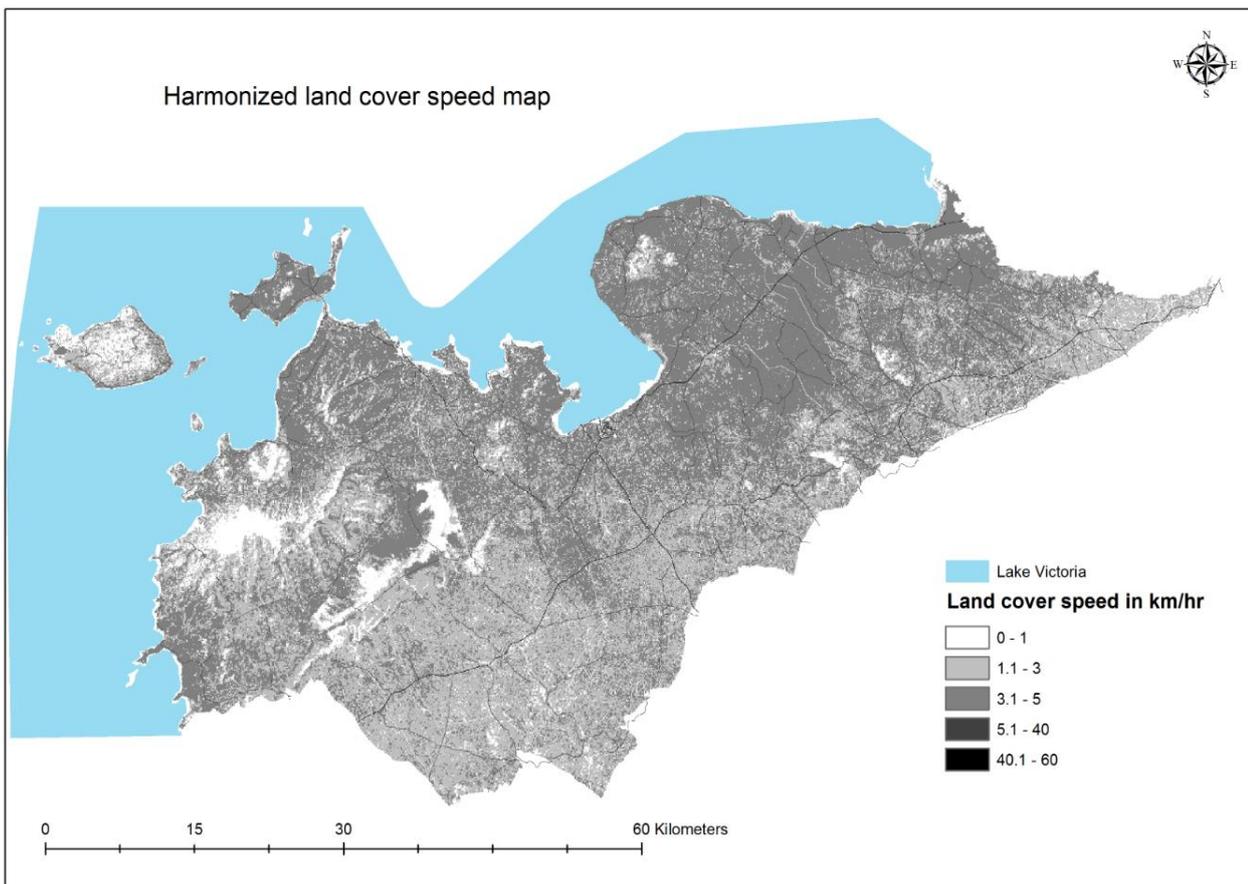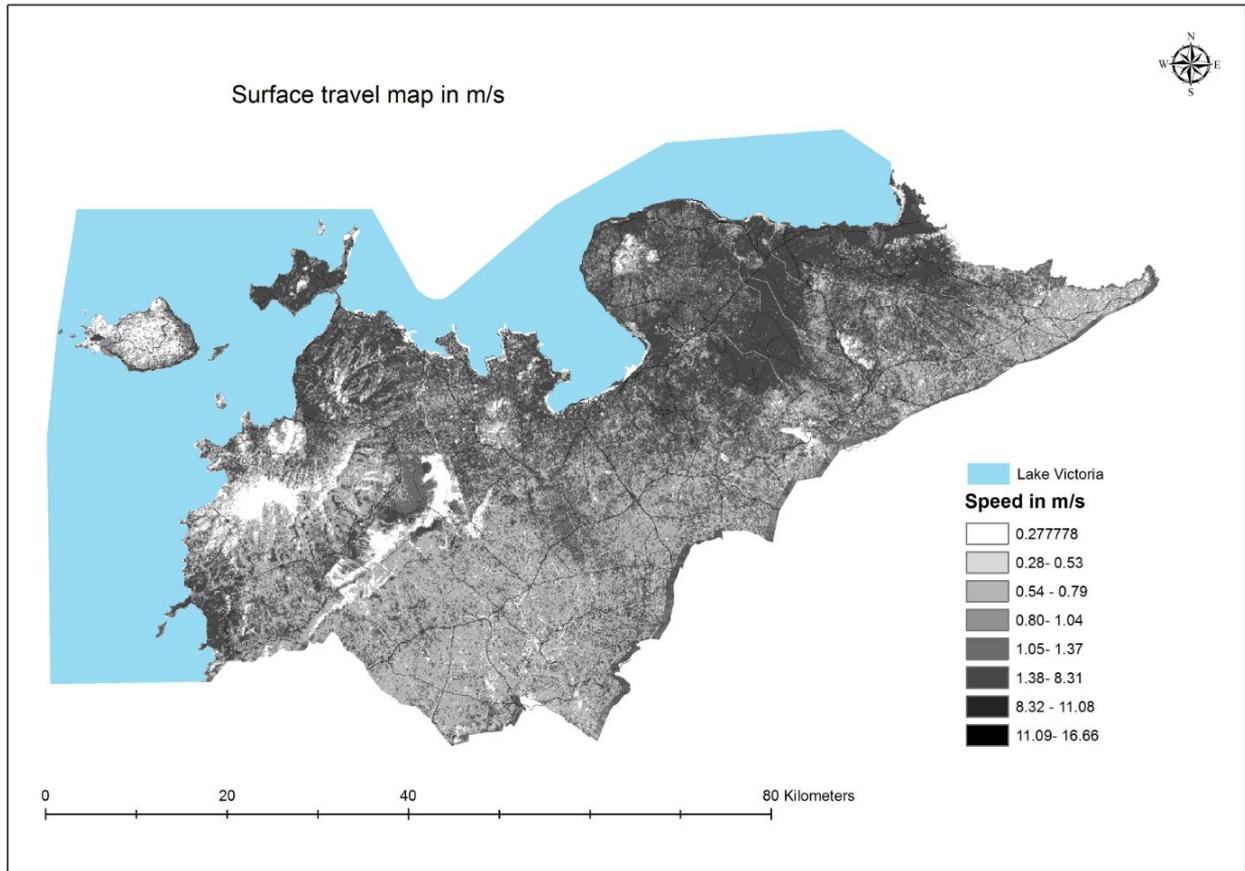
Figure 14 (a) and (b), harmonized land cover and surface travel time in m/s map.

Using cost allocation tool in ArcGIS, the catchment for each type of facility (nine 3-tier, health facility, and dispensaries) were delineated. This is because of different levels of services offered at each facility level. 3rd-tier hospitals have larger catchments than health centres and dispensaries.

The three catchments generated were combined to create one catchment for all facilities. Areas of overlap between dispensaries, health centres, and nine 3-tier hospitals were averaged to create one catchment, as this is the lowest and the first point of interaction between patient and facility (primary health services).
See figure 15, health catchment zones in Homa bay county and appendix 8 for health catchment generation flow chart.

3.4.2   Malaria infection rate data.
Malaria occurrence data was collected from Sub County health ministry departments for calculation of malaria infection rate. There exist 6 sub-county health offices with each recording the occurrences within their jurisdiction. 84 health facilities monthly data were recorded and geocoded to their health catchments (see figure 15). Using 2014 projected population data, the infection rate for each health catchment per 100 persons was calculated using equation 3.5.

The percentage malaria infection rate is 12% for Homa Bay County (see appendix 9 for infection statistics); this is comparable to the rates (12% infection in the rural and 5% in urban) reported by Malaria Control and Ministry of Public Health (2010). See figure 16 for the infection rate map.

Infection rate formula: $= \textit{(Number of cases/Total population)} * 100.$ (3.5)
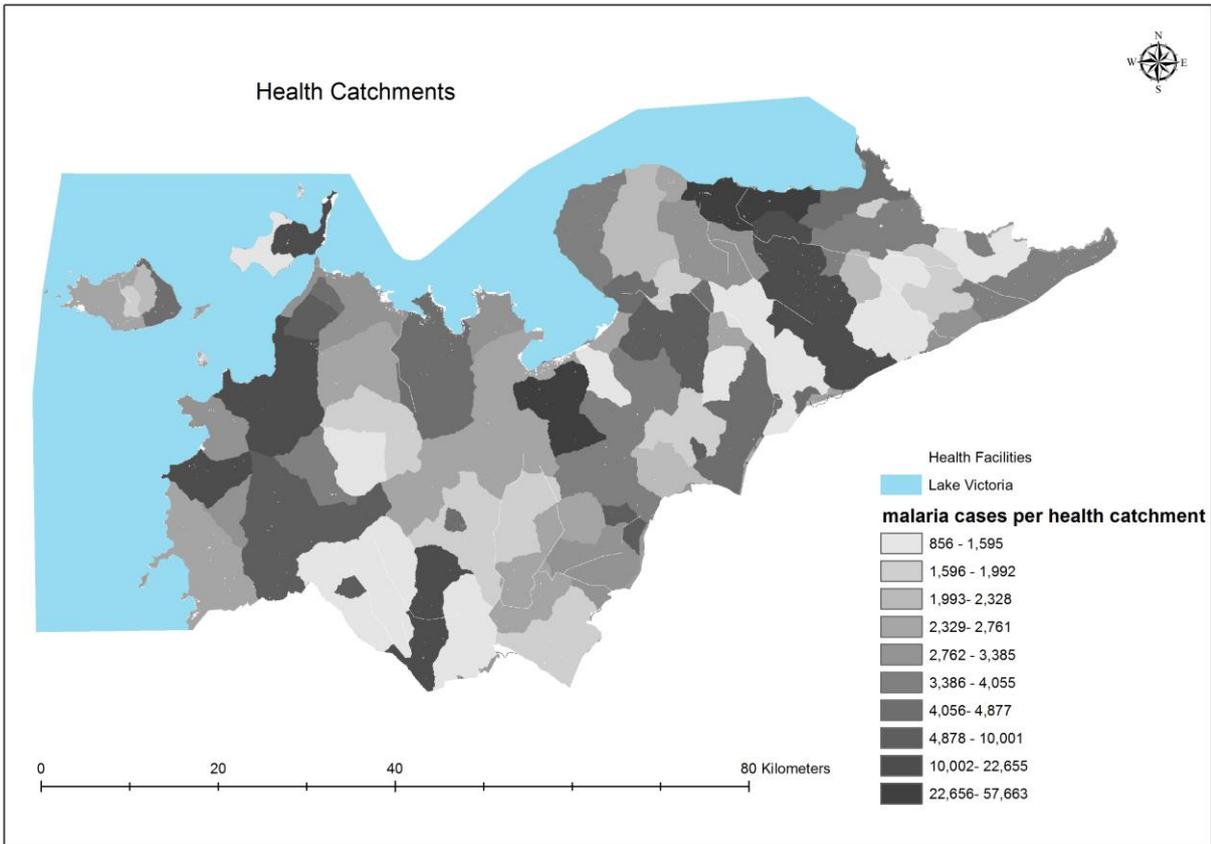
Figure 15. Health catchment.
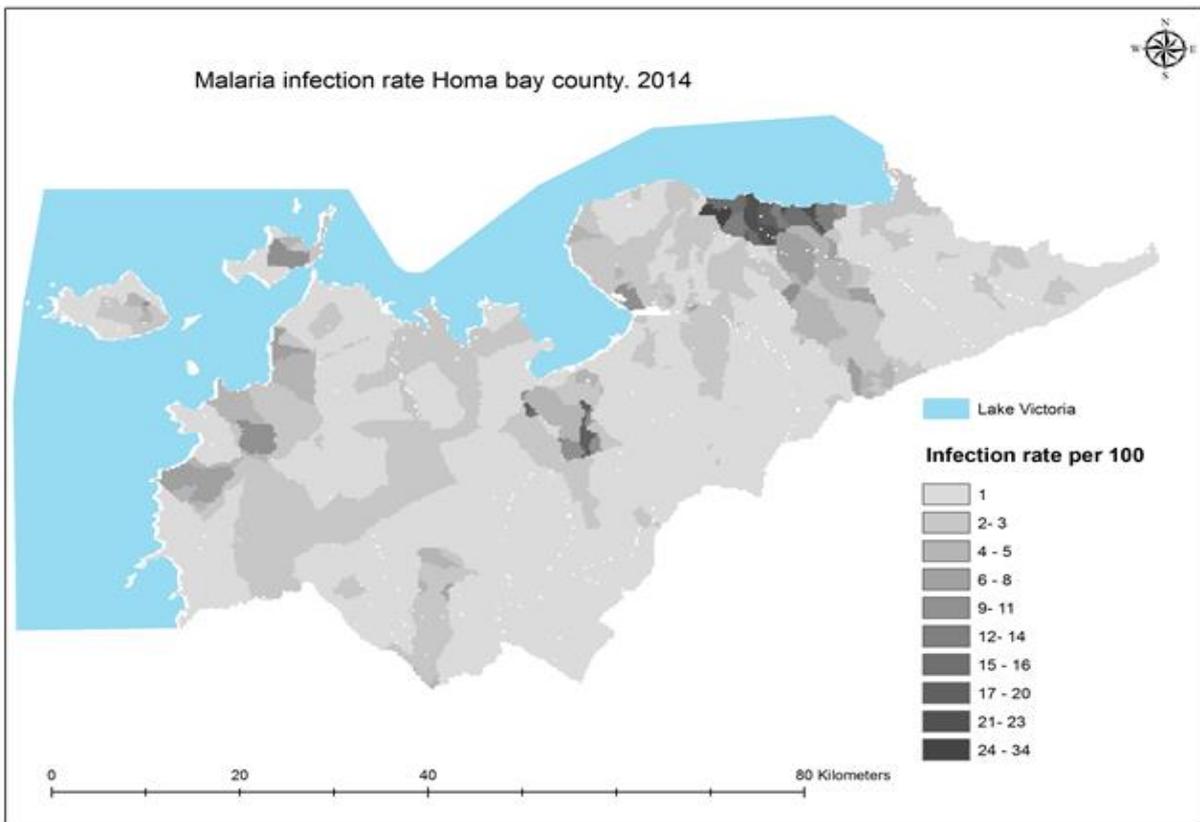
Figure 16. Malaria infection rate map.

3.4.3   Variables standardization.

Standardization is required to bring all variables into proportion with one another or into the same unit (Etzkorn, 2014). Various land cover types have varying effects on mosquito development which further increases or reduces malaria infection rate (Stryker & Bomblies, 2012). This was done to create land cover effects datasets. Additionally, standardization of land cover was done based on expert knowledge derived from administered questionnaires. Each land cover was assigned a value based on its perceived effect on transmission. Finally, pixels were assigned averaged values adding up to 1 (see figure17).

Land cover	Average Weights
Water	0.25
Forest	0.1
Grassland	0.25
Cropland	0.3
Bare ground. (Including roads)	0.05
Settlements.	0.05

Figure 17. Land cover standardization.

Other variables were standardized based on the benefits formula described by Data Mining Research (2015). (See appendix 10, altitude (a), slope (b), evapotranspiration(c), NDVI (d), land surface temperature, (e), rainfall, (f) and poverty (g), distance to water hyacinth (h), and distance to wetlands (i) standardized maps). Equation (3.6) shows the standardization formula.

Xnew = *(X-Xmin)/ (Xmax-Xmin)*   (3.6)

Where X is the value of the pixel being standardized, Xmin is the minimum value while Xmax is the maximum value of the layer.

Each derived variables was resampled to 100m spatial resolution. This is the spatial unit of analysis; thermal bands used in deriving land surface temperature are acquired at a spatial resolution of 100m (USGS, 2014). It implies that variations in land temperature is noticeable from 100m onwards. Temperature is considered a major malaria risk factor by many researchers (explained in malaria risk factors, section 2.4) therefore; it acts as the base for analysis. In addition, a total of 304,920 points were generated from the sub-county layer for each pixel. Raster values from all factor layers were extracted into the points and used in correlation, factor, and validation analysis. Figure 18 shows the list of datasets, their sources, and final spatial resolution. It is similar to figure 10, the final resolution column is the only difference.

1.Environmental variable	Source	Final resolution.
Land cover	USGS Landsat.	100m
Land surface temperature	USGS Landsat	100m
Evapotranspiration	NASA Earth explorer (MODIS)	100m
Distance to water Hyacinth	USGS Landsat	100m
Slope	SRTM NASA	100m
Distance to water ponds	Sentinel 1	100m
Elevation (altitude)	SRTM NASA	100m
Rainfall	CHIRPS	100m
Soil drainage	Kenya Soil Survey	100m
Topographical wetness	USGS Landsat	100m
Population	Census 1989 and 1999	100m
2. Malaria infection rate.	Ministry of Health	100m
Sub-county layer	County Government of Homa bay	100m
Poverty density	World resource institute	100m

Figure 18. List of datasets with the source and final resolution.

3.5 Data analysis

This segment entails data analysis in line with objectives of the study. Environmental and socio-economic variables that were identified to contribute to malaria risk, generated using image analysis techniques and factor analysis are discussed in this section. The methodology used in generation of potential risk map are also part of this section.

3.5.1 Identification of environmental and socio-economic variables leading to malaria risk.

Environmental and socio-economic variables were identified from the literature review. Variables contributing towards malaria risk were categories into four classes namely, topographical, ecological, hydrological, and climatic. The identified variables were either direct or indirect; indirect factors were described as modifying factors. Additionally, the various threshold under which the condition (environmental variables) become suitable or constraint for potential malaria risk were determined. Using image-processing techniques discussed in section 3.3.1 geospatial datasets were created for each variable. Land cover, a direct variable was generated and validated during fieldwork.
Land cover validation.

A total of 60 points were generated with 10 points in each sub-county. The validation points were randomly picked based on accessibility; the land cover class and geographic coordinate of the points recorded. Confusion matrix was then generated to show the overall accuracy of the land cover. The kappa coefficient was used to determine whether the classification is good enough for this particular application. Kappa coefficient below 0.4 represent a poor agreement between the predicted and observed classes, 0.4 to 0.7 represents moderate while above 0.8 represents strong agreement (Foody, 2002)

3.5.2 Determination of linear association among variables.

Correlation analysis was used to test the linear association between variables (both dependent and independent). Bivariate correlation procedure giving correlation coefficient value between two variables without adding a controlling variable was applied. Person's coefficient was then used to check for linear association among variables, its values range from -1 to+1 with significance at 0.01 level (two-tailed test, $p > 0.05$). Continuous data for all the variables and infection rate were used to test for correlation; this was to keep the variability within the various data inputs (correlation results are based on the 60% points randomly selected). See figure 19 for correlation analysis flowchart.
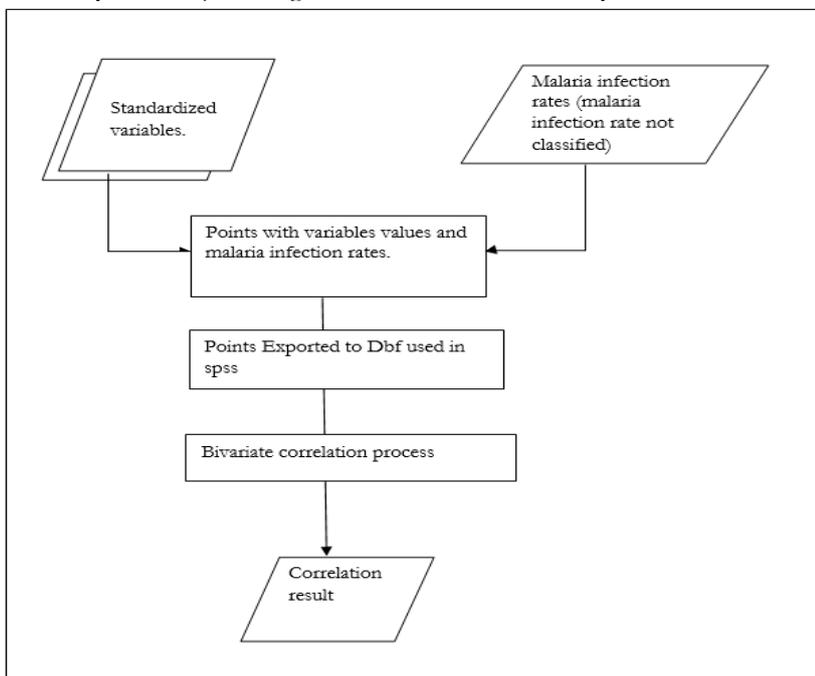
Figure 19. Correlation analysis flowchart.

3.5.3   Identification of significant variables leading to malaria infection rates.

Sixty percent (60% = 182,742 points randomly selected in SPSS) of the generated (see section 3.4.3) points were used in the factor analysis model to determine significant environmental and socio-economic variables leading to malaria risk. The remaining points (40%=122,178 points) were used for potential risk map validation. Part of the process involved collinearity testing; this denotes situations in which one or more independent variables are correlated to a certain degree. According to Benndorf et al. (2011), collinearity causes loss of statistical power in factors. A two-tailed correlation test was then used to determine individual variables association with malaria infection rate. Additionally, VIF (variance inflation factor) was adopted in testing collinearity. Correlation analysis was also done to determine the association between the malaria infection rate and environmental and socio-economic factors

Variance inflation factor

Variance inflation factor quantifies the severity of collinearity. During collinearity process, a complementary VIF value is computed for each variable. The average VIF value should be close to 1, this shows that collinearity is not a problem to the model (Field, and Miles, 2012).

$$VIFn = 1/1-R^2 \tag{3.7}$$

Where R^2 is the coefficient determinant for the nth number of independent variables. As a rule of thumb, independent variables with VIF greater than 10 are excluded from the analysis.

Factor analysis model.

Standardized variables were used in factor analysis procedure. PCA extraction method was applied with orthogonal rotation (varimax); this awards all factors initial communality value of 1 and restricts correlation. Kaiser Meyer –Olkin measure of sampling adequacy was used to test the appropriateness of factors analysis. As a rule of thumb, Kaiser Meyer value of 0.9 is considered as perfect, 0.8 to 0.7 considered as medium, and 0.6 as low while below 0.5 is unacceptable. Bartlett's test of sphericity was then used to affirm the lack of multi-collinearity, normality and reject the null hypothesis that identity matrix exist in the factors. Scree plot, total variance explained, communalities, and rotated factor matrix were used to select significant factors to be used in risk prediction. Factors with values higher than 0.3 in each column in the rotated matrix table were grouped as one component. Finally, various umbrella labels were awarded to this component based on the type of factors therein. Flow chart in figure 20 shows the workflow involved in factor analysis.
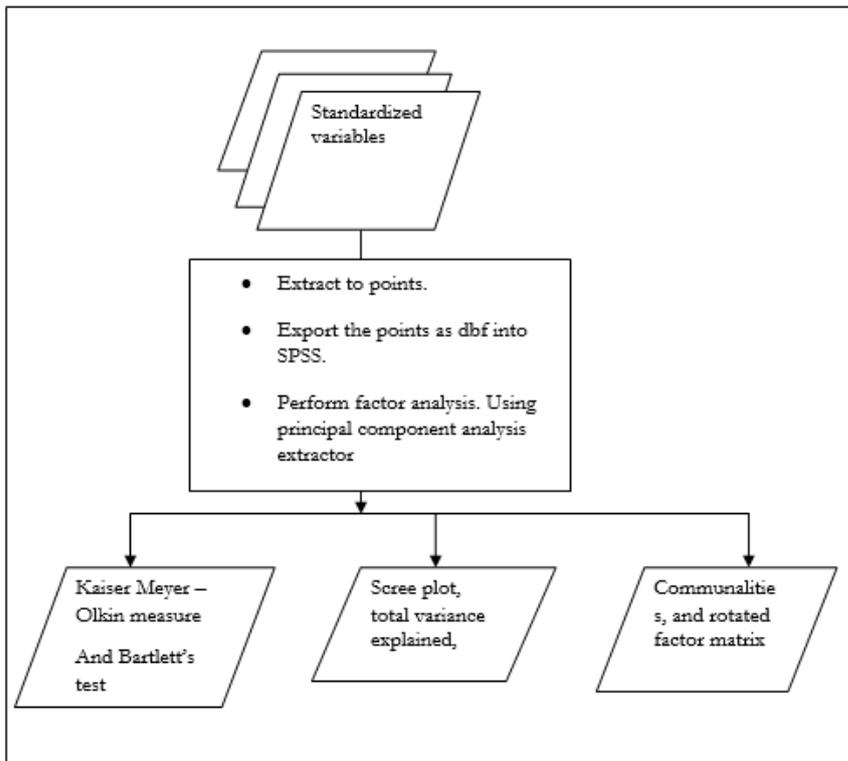
Figure 20. Factor analysis flow chart.

3.5.4   Potential malaria risk map development and validation.

The results from the factor analysis modelling were used to identify significant variables. The model listed significant factors belonging to three component groups in the rotated factor matrix result for Homa Bay County. Overlay index method was used in creating the potential risk map.

Based on variables thresholds from literature review (malaria risk factors, section 2.4), the selected variables were reclassified into values ranging from 1 to 10 (1 for less optimum and gradually increasing to 10 for optimum variable values). Factors weights were awarded based on correlation results between malaria infection and risk factors, this is borrowed from Wang & Mieghem (2007) in the study of constructing overlay network by tuning link weights. See figure 21a for the flow chart for overlay index method and figure 21 (b) for variables reclassification based on thresholds.
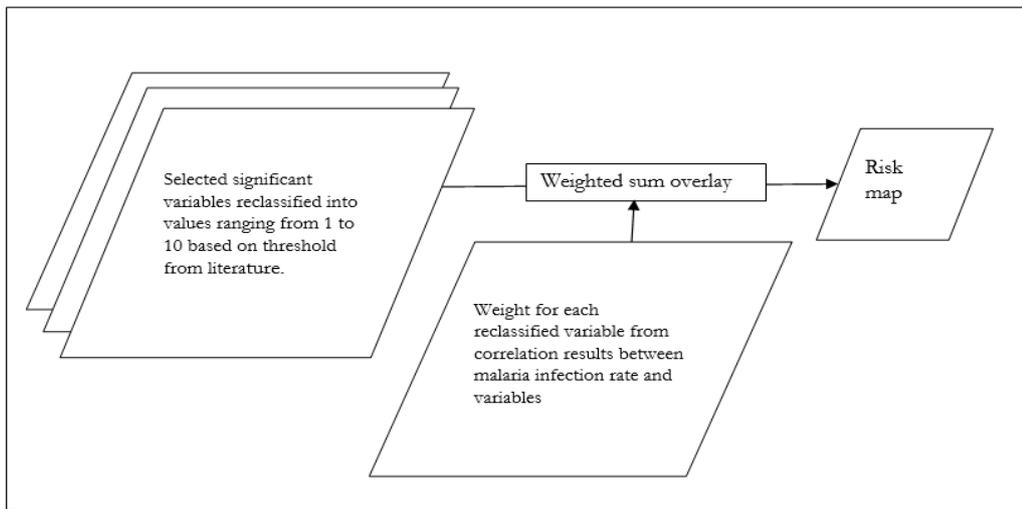
Figure 21a. Overlay index method flow chart

Validation of remote sensing products is a critical concern in earth observation (Li et al., 2013). It is the process of evaluating the integrity and correctness of spatial data. According to Li et al.(2013), the accuracy of the slave data largely depends on the validity and accuracy of the master or reference data and, therefore, ground truth data is normally used as the master data. For this study, malaria infection rate derived from malaria cases was used as the master (reference data) while potential malaria risk map was used as slave dataset.

Malaria infection rate and potential malaria risk map were reclassified into high and low. Firstly, the two layers were both equated to 100% by assigning the highest value to 100 % and lowest 1 % to standardize the layers. As earlier discussed in 3.4.3, standardization helps to make the data comparable. Secondly, zones with infection rate and potential risk below fifty percent were classified as 1 (low infection rate and low potential risk) while those above fifty percent were classified as 2 (high infection rate and high potential risk). This classification was based on Aird, Bangs, Maguire, & Barcus (2002) epidemiological measures of malaria risk in children for hyper-endemic zones, because children are most vulnerable (Armstrong Schellenberg, Smith, Alonso, & Hayes, 1994).

The remaining 40% (122,178 see section 3.5.3) randomly selected points were used to extract the values from the two layers into the points. Using the reclassified malaria infection rate as the reference, points with both low malaria infection rate and low risk, and high infection rate and high risk were considered as correctly classified or matched. Whereas points with high infection rate and low potential risk and vice versa were considered as, a mismatch (not correctly classified). The correctly classified points were summed up, multiplied by 100 then divided by the total points (122,178) to get the risk map accuracy in percentage. See figure 21c for the validation flow chart.

Equation 3.8 shows accuracy calculation formula.

% accuracy = *(x *100)/y*   (3.8)

Where x is the summed correctly classified points,

Y is the total number of points, in this case, it's 122,178.
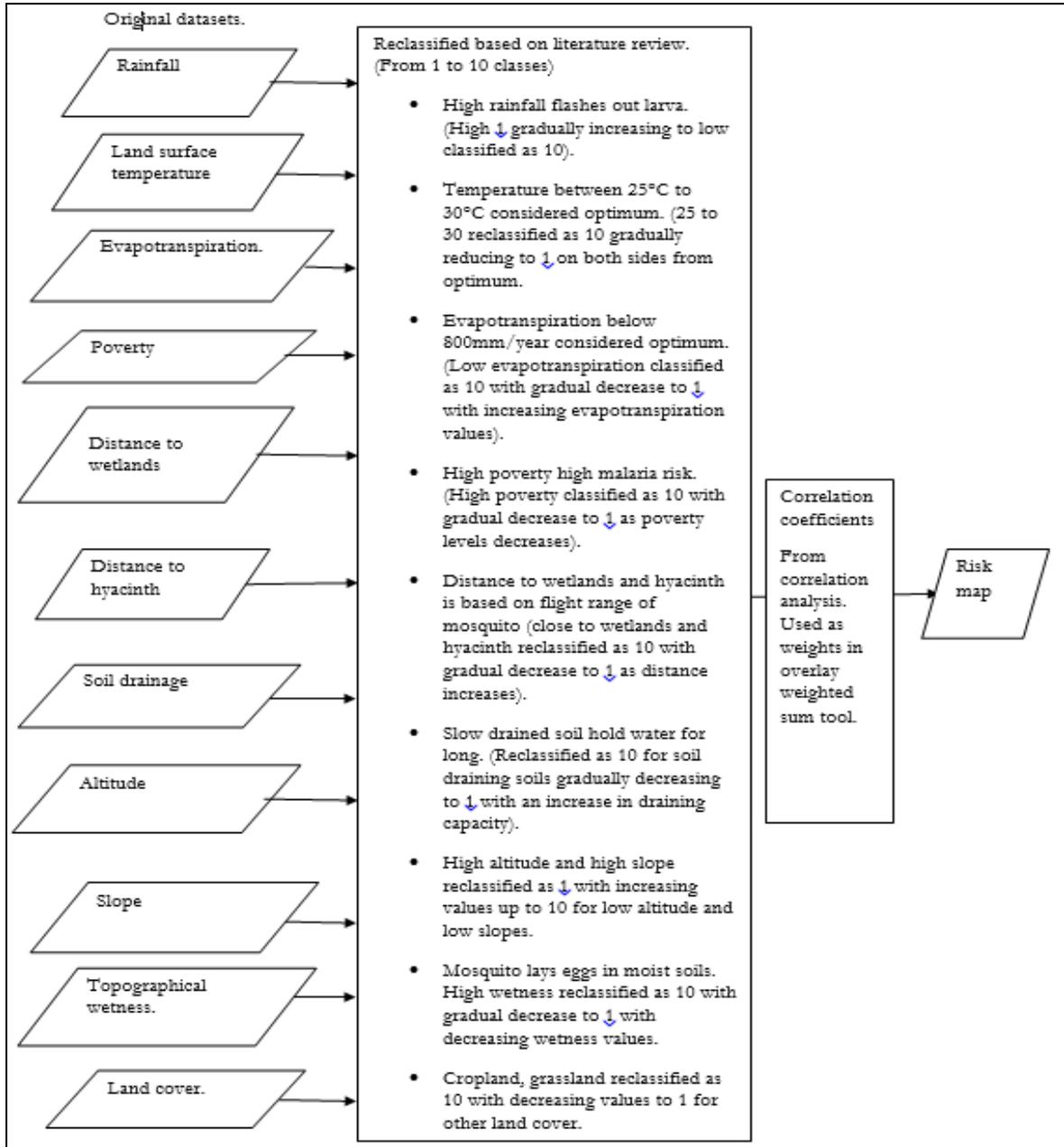
Figure 21(b). Variables classification based on thresholds.
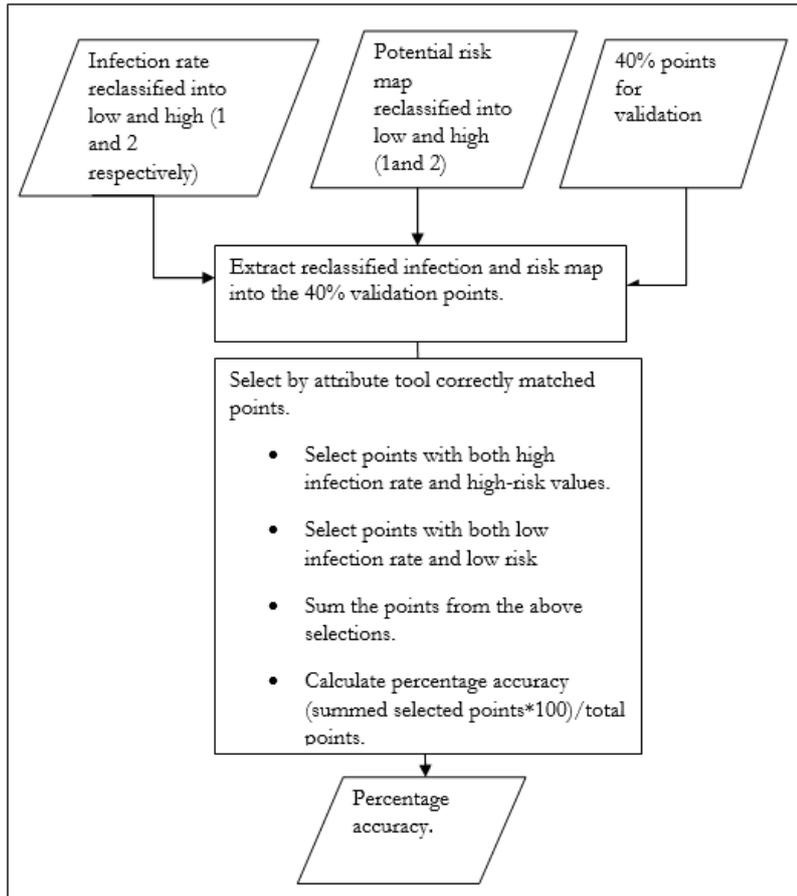
Figure 21c. Potential risk validation flow chart.

3.6 Software packages.

Software packages used in data collection, generation, and analysis are Microsoft Excel, ArcGIS 10.3, Envi 5.0, R-studio, and SPSS.

4 RESULTS AND DISCUSSION.

In chapter four, the results are presented, interpreted, and discussed according to the objectives of the study, research questions, conceptual framework, and methodology. It contains two sections: results, and discussion.

4.1 Results.

This section entails the results of literature review, environmental and socio-economic variables derived from image analysis and GIS techniques, the relation among variables, and potential malaria risk map.

4.1.1 Identified environmental and demographic variables leading to malaria risk.

Based on conducted literature review, environmental variables were categorised into four namely, climatic (temperature, rainfall and evapotranspiration), hydrological (soil drainage, wetlands and topographical wetness), topographical (slope and altitude), and ecological (land cover, NDVI and hyacinth). The factors were further divided into two, main and modifying factor. See figure 19 for the summary of variables
.

Environmental variables.		
• Main factor	• Modifying factor	
Temperature	Altitude/elevation	Topographical wetness
Land cover	Evapotranspiration	Distance from Water ponds
Rainfall	Slope	Soil drainage
Altitude/elevation	Distance from Water hyacinth	vegetation cover
Socio-economic characteristics. Poverty levels.		

Figure 22. Summary of environmental and socio-economic variables leading to malaria risk.

4.1.2 Derived variables using earth observation (remote sensing) and GIS techniques.

This subsection shows the results of environmental variables derived using earth observation image processing and GIS techniques.

Land cover. Six main classes of land cover were detected in the study area (see figure 23). The built-up areas were classified as settlements incorporating both urban and rural built-up areas. Photo 1 shows hyacinth invaded zones along the shores of Lake Victoria.
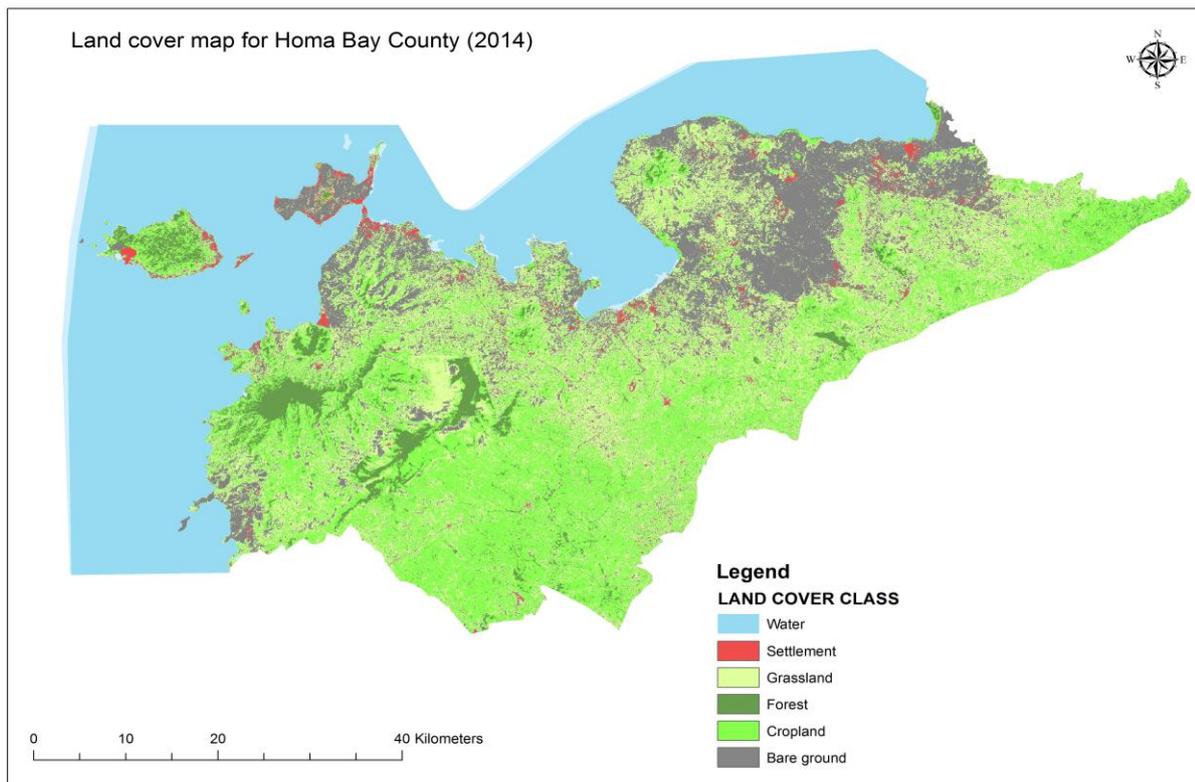
Figure 23. Land cover result, and photo 1 (hyacinth in Homa Bay County)

Land cover accuracy assessment revealed that, out of the 60 points, 48 were correctly classified giving an accuracy of 80 % with a kappa coefficient of 0.7476. A confusion matrix was generated (see figure 24) indicating higher accuracy in cropland, grassland, settlement, bare ground, and water (80%, 80%, 85%, 81%, and 100% respectively). Approximately 1,183sqkm and 861sqkm of Homa Bay County is covered by cropland and grassland respectively. Bare ground covers an area of 737sqkm, settlements cover 48 sqkm and forest cover 18sqkm (any tree bigger than 30m by 30m was classified as forest). Wetlands occupy 36sqkm of the area.

class	Water	Settlement	Bare ground	Forest	Grassland	Cropland	Total
Unclassified	0	0	0	0	0	1	1
Water	3	0	0	0	0	0	3
Settlement	0	12	1	0	0	0	13
Bare ground	0	2	9	0	1	1	13
Forest	0	0	0	2	0	0	2
Grassland	0	0	1	0	12	2	15
Cropland	0	0	0	1	2	10	13
Total	3	14	11	3	15	14	60

Figure 24. Land cover confusion matrix.

Land surface temperature (LST) and normalized difference vegetation cover (NDVI).

The temperature ranges from 13 ° C to 37 °C (see figure 25a). In addition, regions close to the lake experience high annual mean temperatures than the rest of the county. However, the temperature gradually reduces are you move far from the lake towards the Kenyan Highlands (eastern part of Homa bay county). Forest areas experience low temperature while areas with no vegetation (bare ground) experience the

highest temperature values. Forest has the highest NDVI value (0.85) while the lowest value is found on bare grounds. NDVI values range from 0.06 to 0.85 (0.85 indicating high vegetation cover as shown in figure 25b).
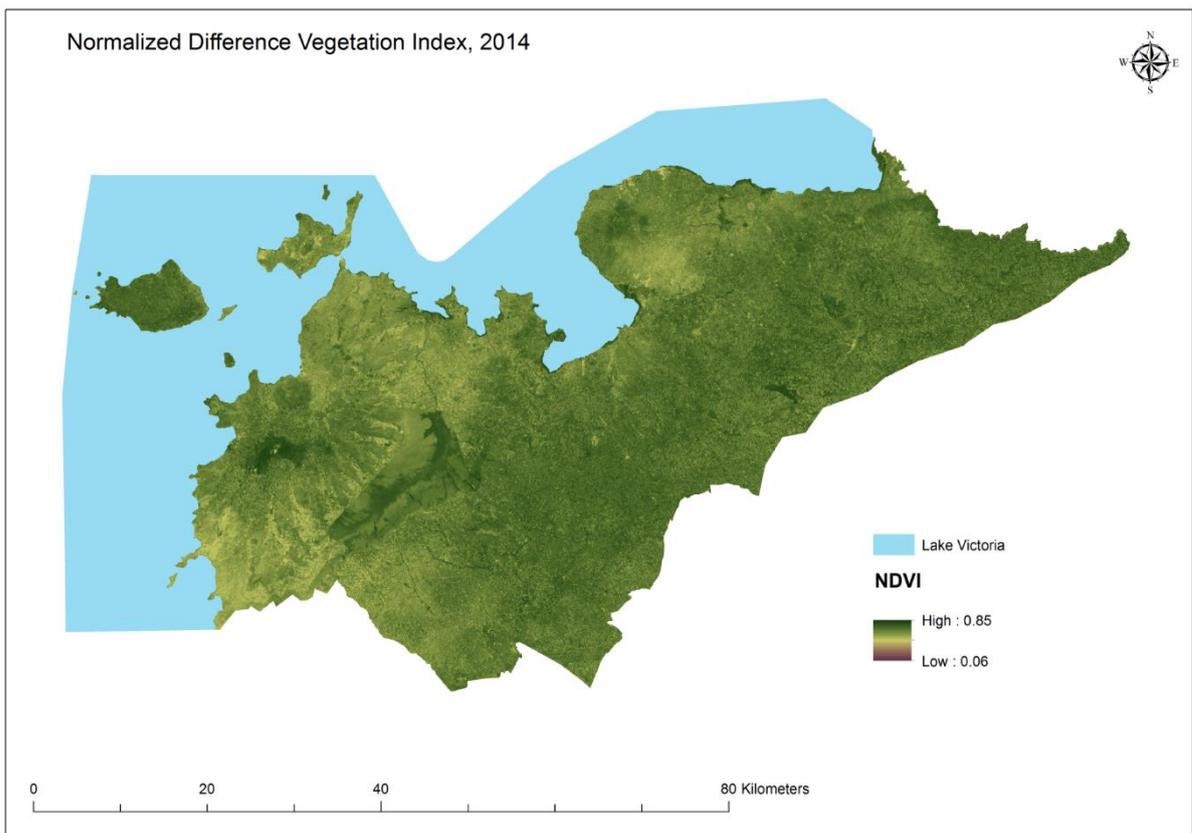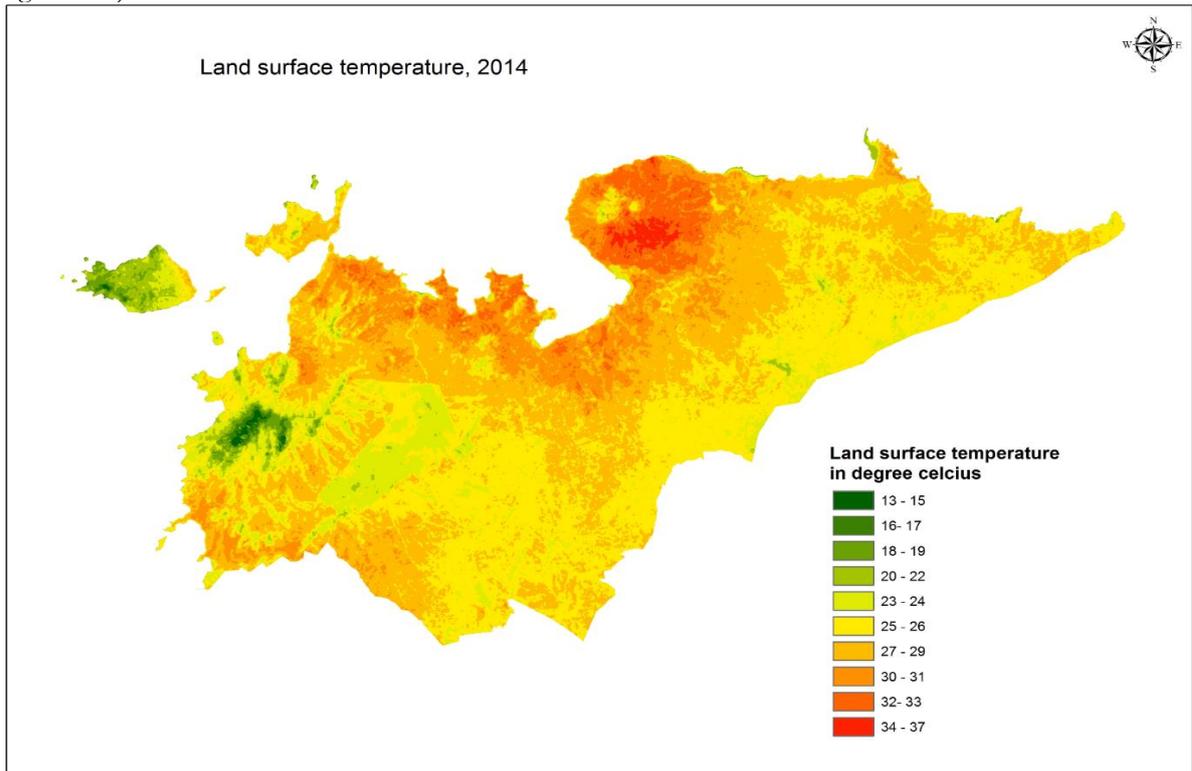
Figure 25a and b. Land surface temperature and NDVI respectively.

The annual temperature as recorded by Kenya Investment Authority (2015), ranges from 17.1° C to 34.8° C, therefore, the remote sensing derived land surface temperature has an error of -3° C and +2.2° for annual minimum and maximum temperatures respectively.

Elevation and slope.

The highest point stands at an altitude of 2270m above the sea level (indicated in black in figure 26 (a), with the lowest point along the shore of lake Victoria. In addition, the elevation increases gradually as you move away from the lake. Steepest slope is located along the Labwe valley indicated in black in figure 26 (b) at 67degrees rise from neighbouring locations.
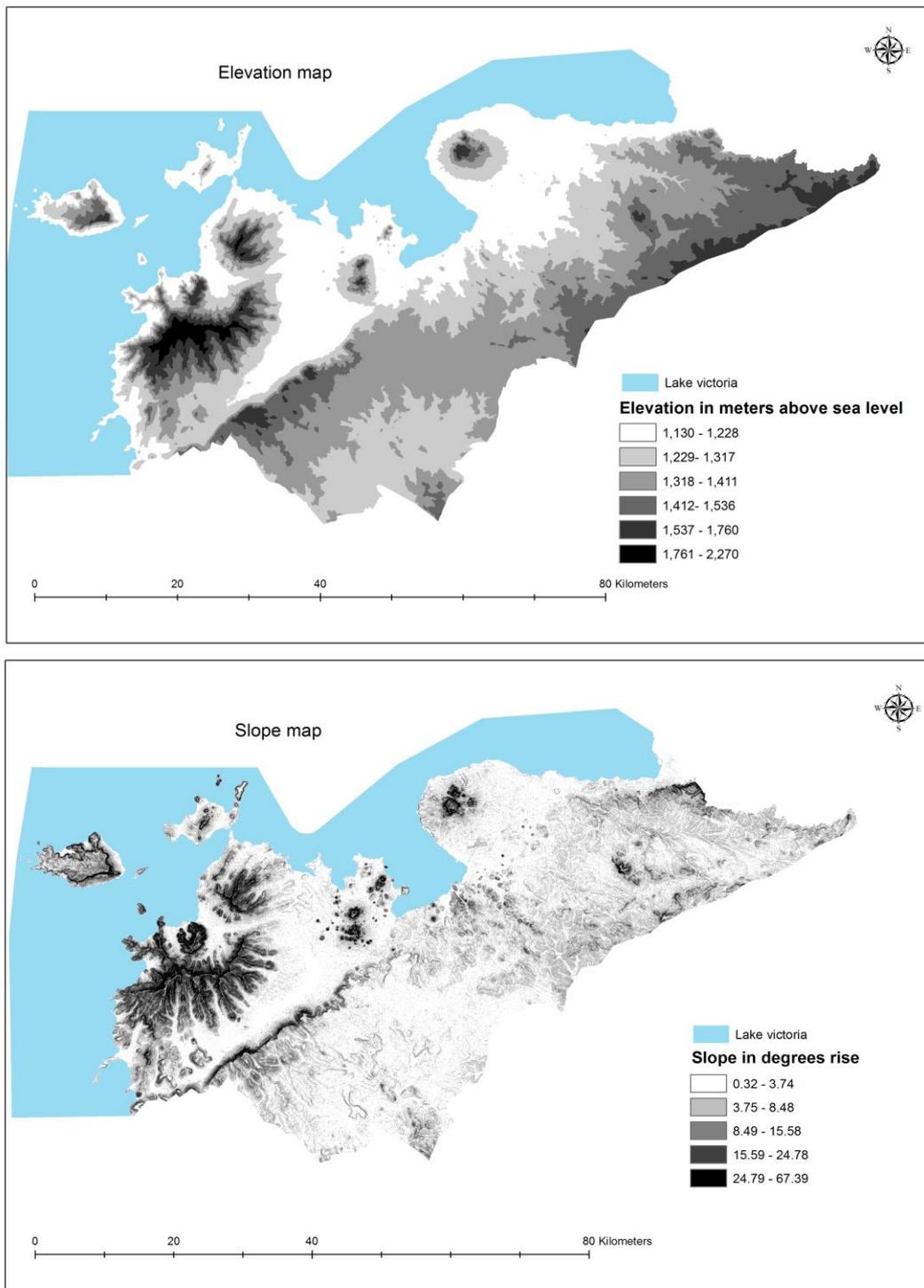
Figure 26 (a) and (b). Elevation and slope map respectively.

Evapotranspiration.

Evapotranspiration values range from 1,500 to 378mm per year, vegetated areas within the county experience high evapotranspiration rates as compared to settlements and bare ground. However, evapotranspiration is high in zones far from the lake as compared to areas close to the lake (see figure 27). Despite these vegetated areas having low-temperature values, they experience high evapotranspiration rates. Finally, mosquito occupies areas of evapotranspiration rate less than 800 (Stresman, 2010), and, therefore, more than half of the county is suitable for mosquito development. Figure 27 shows evapotranspiration map of Homa Bay County.
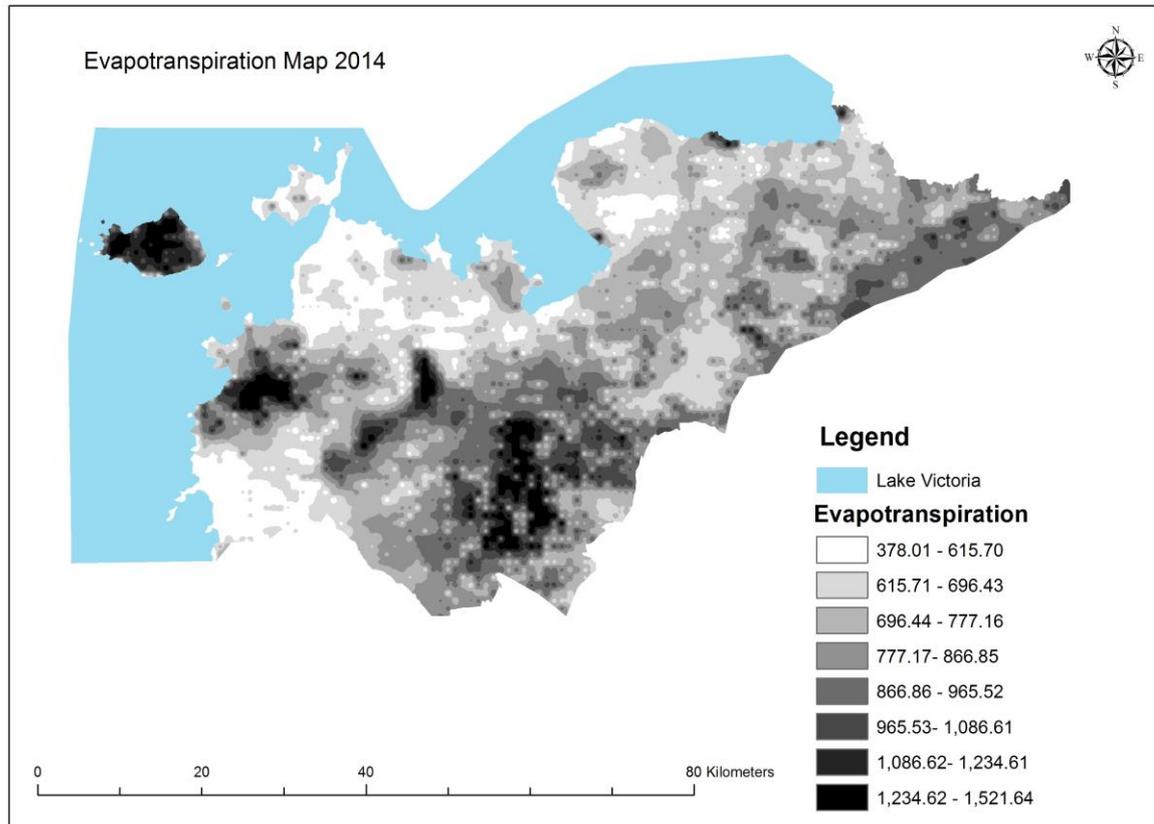
Figure 27. Evapotranspiration map of Homa Bay County.

Water hyacinth.

Mainly found along the shore of lake Victoria (see photo 1). It covers approximately an area of 6 sq.km of the entire shoreline providing a favourable breeding mat for mosquitos. More than 57,000 out of 428,000 malaria cases recorded in 2014 occurred within 1 km distance from the water hyacinth (see appendix 11 (b) for occurrence statistics). Figure 28 shows 1km distance from water hyacinth and 1km effect map of water hyacinth.
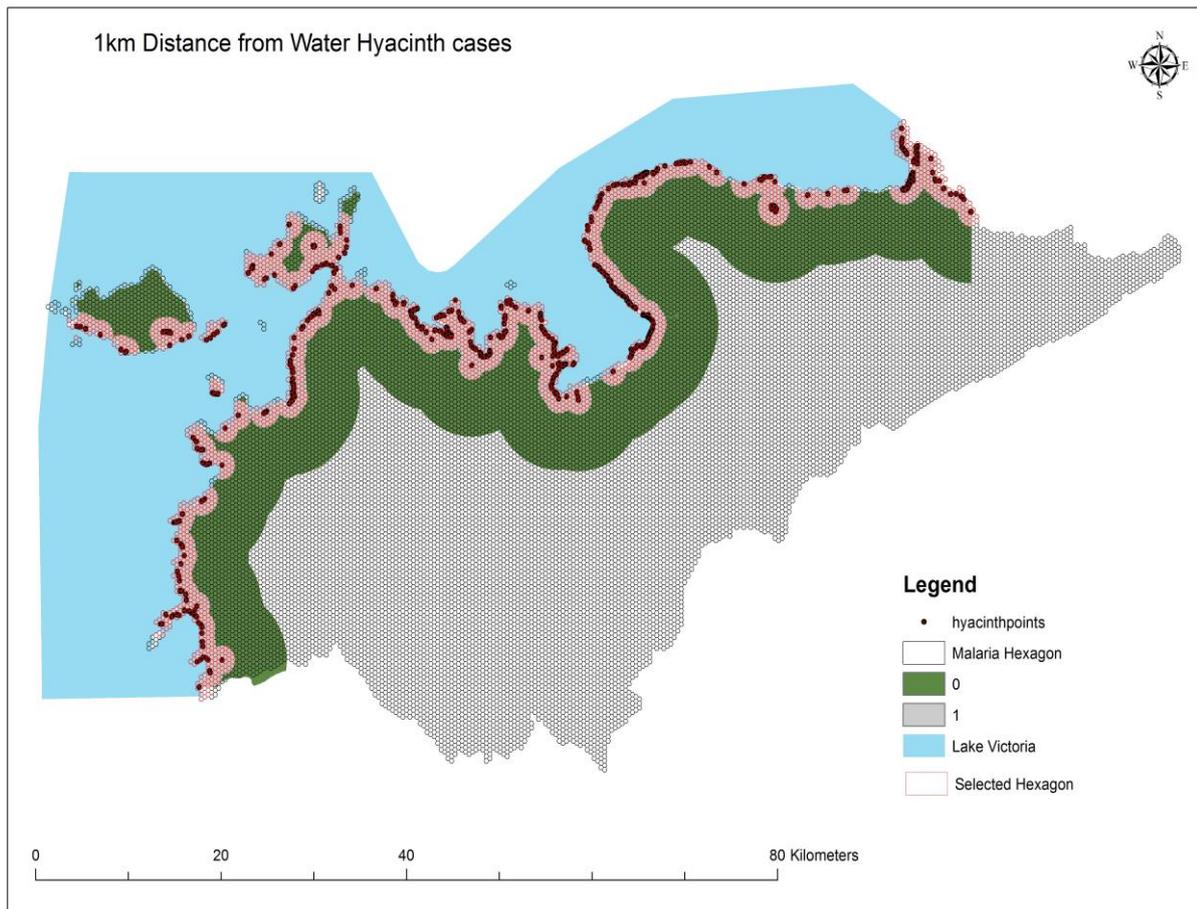
Figure 28. 1km distance from water hyacinth and 1km effect map of water hyacinth.

Topographical wetness.

Forests, cropland and grassland areas experience high topographical wetness. Values ranging from 0 to 1 shows the various level of wetness (black colour represents areas of high topographical wetness and light grey areas with less wetness). Figure 29 shows the topographical wetness result. Mosquito can only inhabit these locations if the soil drainage capacity is extremely slow (Stresman, 2010). Forest and mountainous regions experience high topographical wetness (forest shown in black in figure 29). However, the northeastern part of Homa bay county experience low topographical wetness as compared to western, southern and central parts.

Rainfall

Annual rainfall of 1940mm (maximum) and 943mm (minimum) is recorded as sourced from CHIRPS data. This is way above the 250mm to 1200mm rainfall value as indicated in the literature (Kenya Information guide, 2015). The data was normalized by calculating the difference between the maximum values (1940-1200) and minimum values (943-250), then finding the mean of the result and subtracting from the original CHIRPS data values (see figure 30). However, low rainfall (between 227 and 527 mm/year) is experienced in zones close to the lake, which are low in altitude. In addition, the eastern part of the county receives higher rainfall (between 900 to 1,224 mm/year).
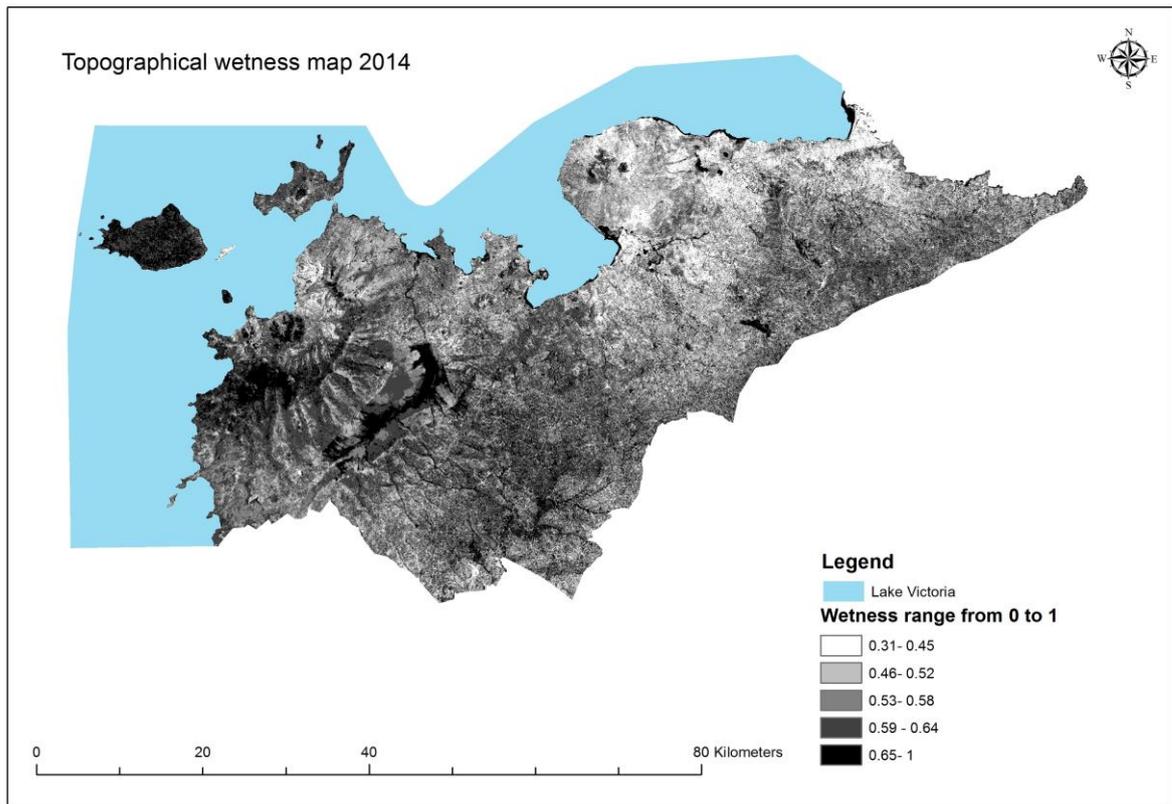
Figure 29. Topographical wetness map.
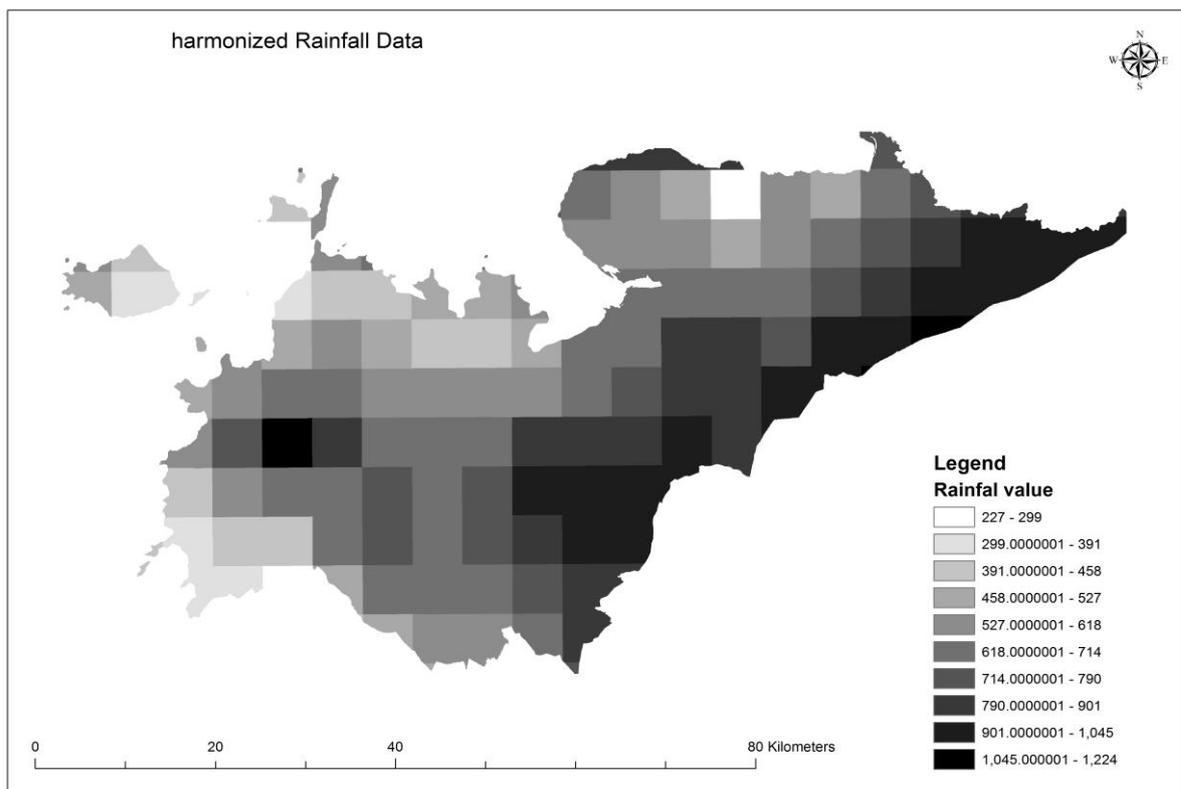
Figure 30. Harmonized rainfall maps respectively.

Soil drainage.

Most part of the county 1708sqkm (about 54%) is well drained. 1276sqkm (40%) is classified as extremely slow in water drainage capacity. However, rapid draining soils only occupy about 7% (221sqkm) while slow draining soils at 1% (45sqkm). In addition, slowly drained soils are majorly found in the southern part of the county, with extremely slow in the western and north-eastern part of the county. Figure 31 shows the soil drainage map.
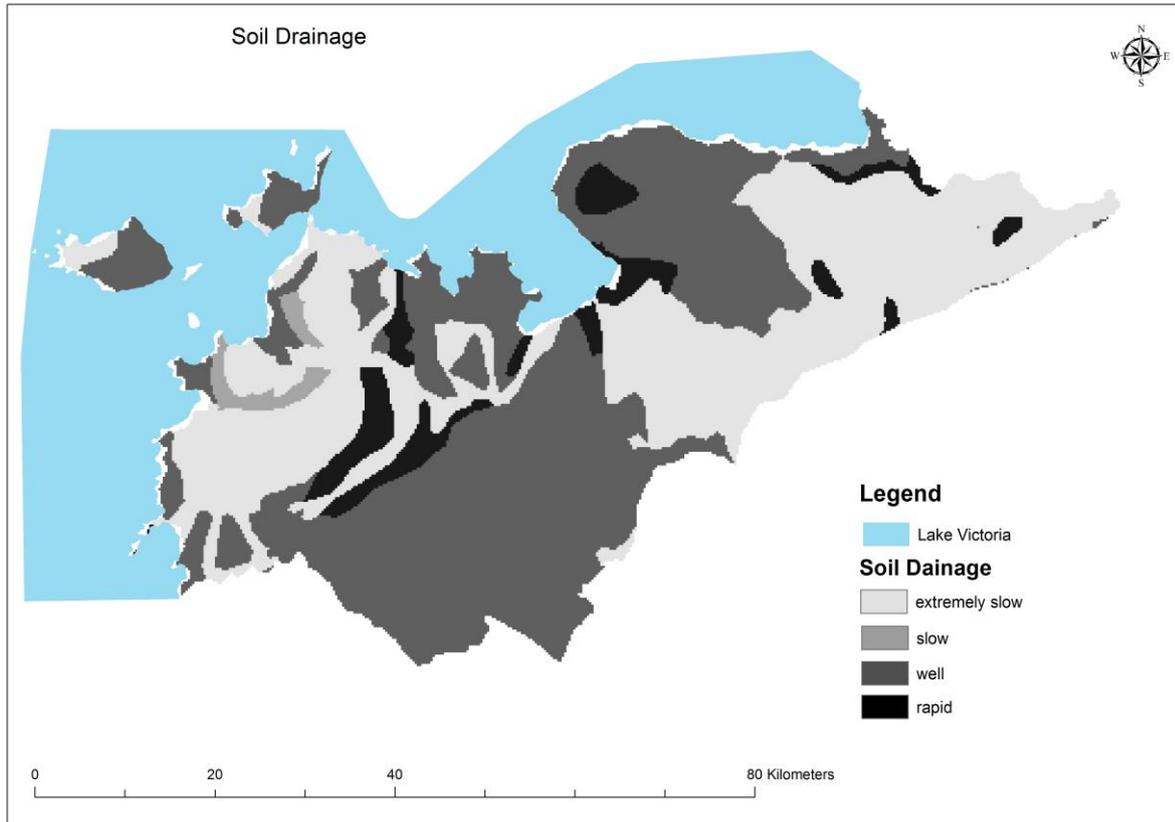
Figure 31. Drainage soil map.

Water ponds/wetlands.

Most wetlands are found in central, northeastern and southwestern part of the study area. In addition, literature reveals that wetlands are suitable habitat for mosquito whose effect can be analysed by the number of cases recorded close to these wetlands. Approximately 230,000 out of 428,000 reported malaria cases occur within a distance of 1km from wetlands (see appendix 11(a) for descriptive statistics). Figure 32 shows the distance to wetland and 1km effects distance map.

Poverty density.
High poverty density is evident in the north-eastern part of the county with more than 350 poor people per sqkm. However, urban centres or towns, for example, Homa Bay town, have low poverty levels. Situated in the central part of the county, Homa Bay town has poverty value of 19 people per sqkm. In addition, southwestern part of the county experience poverty levels ranging from 19 to 290 (see figure 33). Based on the visual interpretation of figure 33, rural areas have high poverty density than urban areas, implying that there are more poor population in rural than in urban (in absolute poverty terms).
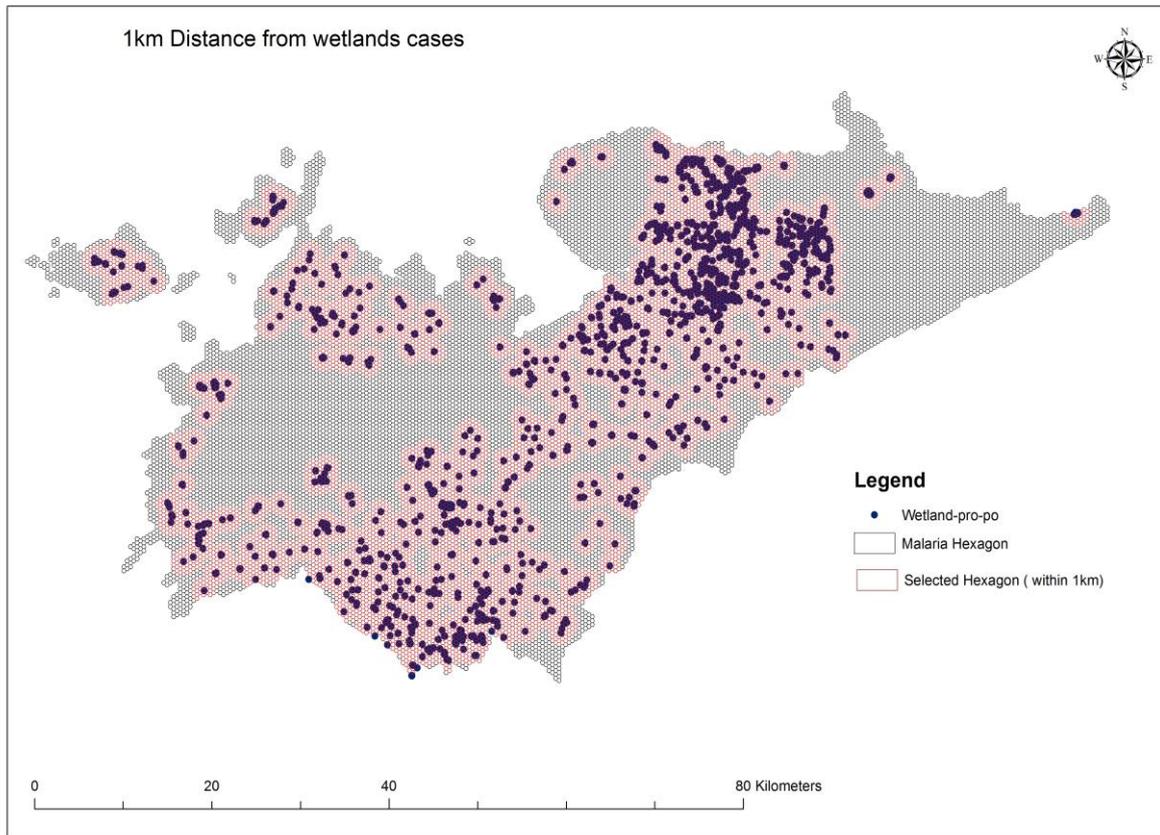
Figure 32. Distance from wetlands and 1km effect distance from wetlands.
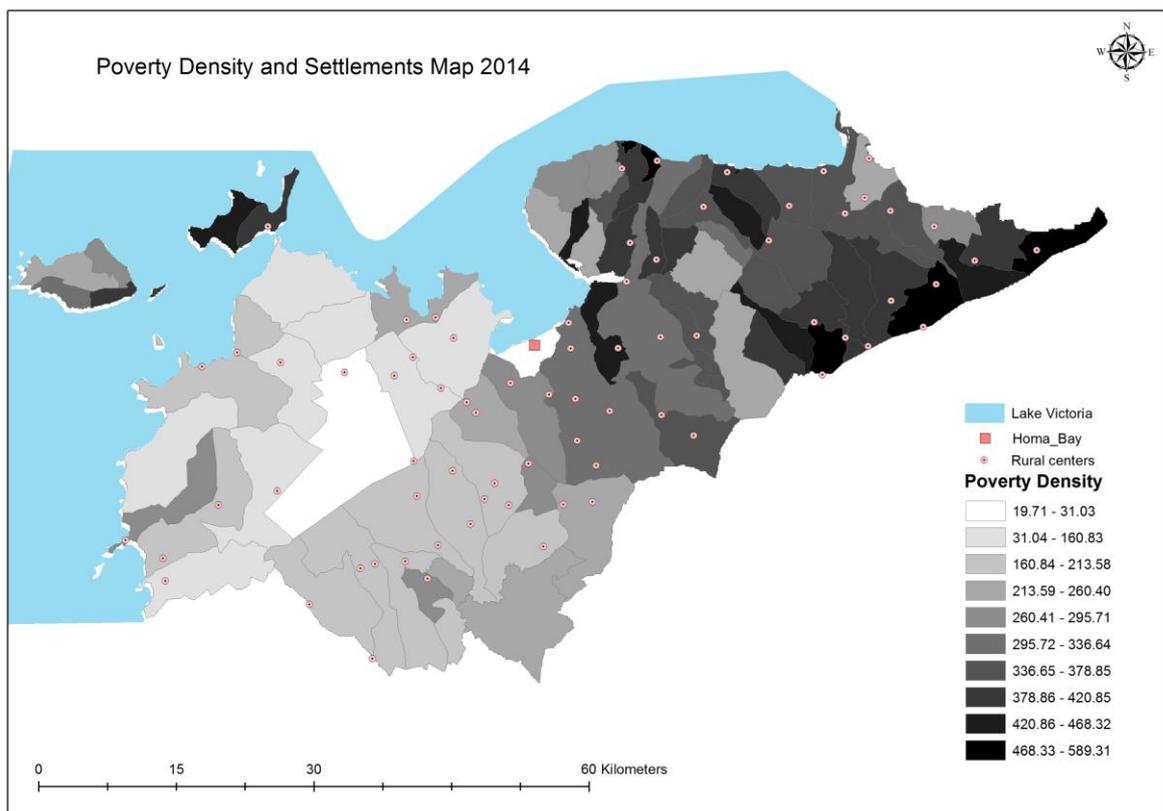
Figure 33. Poverty density and poverty settlement map.

4.1.3 Association among the spatial environmental and socio-economic factors and infection rate.

This subsection shows the *results of correlation analysis* conducted to analyse the association among malaria risk environmental, socio-economic factors and infection rate. Figure 34 (a) shows the linear correlation coefficient results reported in this subsection.

Land cover.

More than 60% of Homa Bay County is covered with grassland, cropland, and forest; therefore, there exist a negative association of 0.154 between land cover and land surface temperature. The more the area in sqkm of the aforementioned cover the lower the temperature. Additionally, a positive association of 0.304 exist between land cover and NDVI, this indicates that the more the land cover (mentioned above) the higher the NDVI value. A positive association of 0.233 also exist between land cover and altitude, the higher the altitude the more the land cover (grassland, cropland, and forest).

Association results of 0.015 exist between slope and land cover; this reveals that no linear association exist between land cover (grassland, cropland, and forest) and slope. However, a weak positive association of 0.187 exist between land cover and evapotranspiration, indicating that the more the land cover the higher the evapotranspiration value recorded. In addition, a positive association of 0.244 exist between land cover and topographical wetness; this indicates that the more the land covered by grassland, cropland, and forest the higher the topographical wetness. There also exist a positive association of 0.274 between rainfall and land cover, the more land covered by grass, forest, and crops the higher the rainfall experienced in such zones.

An association of 0.015 exist between land cover and soil drainage and, therefore, no linear association exist in this relationship. This is because 50% of the soil within Homa Bay County is classified as well drained while the other 50% classified as slow. For accurate results between the two, specific zones of the various soil draining capacity should be investigated against land cover. However, the closer to wetlands the lower the land cover (grassland, cropland and forest), this is depicted by a weak negative association of 0.122 between the two factors. In addition, high poverty density zones have less land cover (grassland, cropland and forest); a weak negative association of 0.173 exist between land cover and poverty. Finally, results on land cover indicate a positive association of 0.383 between land cover and distance to hyacinth. This indicates that regions far from water hyacinth have more land covered by crops, grass, and forest.

Land surface temperature.

A negative association of 0.442 exist between temperature and NDVI revealing that the lower the temperature the higher the NDVI value. In addition, the correlation result between temperature and altitude reveals that there exist a negative association of 0.461. The higher the altitude the lower the temperature and, therefore, less malaria risk expected in this zones. A negative association of 0.32 also exist between temperature and slope, the lower the temperature the higher the slope and vice versa. However, the higher the temperature the lower the evapotranspiration rate, this is because high temperature is experienced in zones closer to Lake Victoria, which is mostly bare and not vegetated in terms of land cover. Consequently, a negative association of 0.544 is realised between the two.

A negative association of 0.272 exist between temperature and distance to hyacinth. This reveals that the further away from water hyacinth the lower the temperature and vice versa. However, the higher the temperature the lower the topographical wetness; a negative association of 0.360 exist between temperature and wetness. In addition, zones with low temperature experience high rainfall. This is because there exist a negative association of 0.235 between temperature and rainfall within Homa bay County. Further analysis reveals that there exist a negative association of 0.221 between temperature and soil drainage, the lower the temperature the higher the drainage capacity of the soil. Additionally, the further the distance from wetlands the lower the temperature, a weak negative association of 0.111 exists between the two factors. High poverty-stricken zones experience high temperatures; a weak positive association of 0.15 also exist between temperature and poverty

Normalized difference vegetation index (NDVI).

There exist a positive association of 0.322 between NDVI and altitude, this depicts that the higher the altitude the higher the NDVI values. Whereas no linear association exist between NDVI and slope, correlation results show an association value of 0.01 between the two aforementioned factors. However, the higher the NDVI the higher the evapotranspiration; the results shows a positive association of 0.611 between NDVI and evapotranspiration. A positive association of 0.502 also exist between NDVI and distance to hyacinth indicating that the further away from water hyacinth the higher NDVI.

A positive association of 0.445 exist between NDVI and topographical wetness. This reveals that the higher the NDVI the higher the topographical wetness. The results also show a positive association of 0.492 between NDVI and rainfall; the higher the rainfall the higher the NDVI values. However, a linear association of 0.042 and 0.025 exist between NDVI versus soil drainage and NDVI versus poverty respectively. This indicates that no linear association exist between the aforementioned. Results also show that no linear association exist between NDVI and distance to wetlands (association value of 0.075).

Altitude.
A positive association of 0.480 exist between altitude and slope, the higher the altitude the higher the slope. Likewise, a positive association of 0.306 exist between altitude and evapotranspiration; zones of high altitude experience high evapotranspiration and vice versa. The results further reveal a positive association of 0.307 and 0.278 between altitude and distance to wetlands and between altitude and distance to hyacinth respectively. This indicates that the higher the altitude the greater the distance to wetlands and water hyacinth.

Correlation test results for the association between altitude and topographical wetness reveals that the higher the altitude the higher the topographical wetness; a positive association of 0.19 exist between the aforementioned. A positive association of 0.499 also exist between altitude and rainfall; this implies that zones of high altitude receive high rainfall. Additionally, high altitude zones receive high rainfall flashing out premature larva; therefore, these zones are not suitable for mosquito habitats as discussed by Illinois Education (2015). The results further reveal a positive association of 0.346 between altitude and soil drainage, the higher the altitude the higher the soil drainage capacity. However, a weak positive association of 0.153 exist between altitude and poverty. This implies that the higher the altitude the higher the poverty density within Homa bay County.

Slope.
A weak positive linear association of 0.167 exist between slope and evapotranspiration, it implies that the higher the slope the higher the evapotranspiration. Further analysis reveals that a positive association of 0.170 exists between slope and topographical wetness; the higher the slope the higher the topographical wetness. The results also reveal a negative association of 0.198 between slope and distance to hyacinth indicating that the greater the distance to hyacinth the lower the slope and vice versa. However, the higher the slope the higher the soil drainage capacity, a positive association of 0.182 between slope and soil drainage reveals this. Finally, the results indicate a negative association of 0.122 between slope and poverty, this implies that the higher the slope the lower the poverty density.

Evapotranspiration.
Correlation analysis shows a positive association of 0.374 between evapotranspiration and distance to hyacinth. This reveals that the higher evapotranspiration values the greater the distance to the hyacinth and vice versa. A positive association of 0.373 also exists between evapotranspiration and topographical wetness; the higher the evapotranspiration the higher the topographical wetness. The results further reveal a positive association of 0.417 between evapotranspiration and rainfall, the higher the evapotranspiration the higher the rainfall experienced in such zones. However, no linear association exist between evapotranspiration and soil drainage (0.043). Finally, association results between evapotranspiration and poverty indicate no linear association between them (value of 0.021).

Distance to hyacinth.
The higher the distance from hyacinth the higher the topographical wetness, a weak positive association of 0.143 between the two factors reveals this. In addition, results between distance to hyacinth and rainfall

reveals a positive association of 0.510; it implies that the higher the distance from hyacinth the higher the rainfall within Homa bay county. The results also reveal a weak negative association of 0.124 between distance to hyacinth and soil drainage, the higher the distance to hyacinth the lower the soil drainage capacity. Consequently, no linear association exist between distance to hyacinth and poverty (association value of 0.042).

Topographical wetness.
A weak positive linear association of 0.101 exist between topographical wetness and rainfall. It implies that zones within Homa bay County of high topographical wetness receive high rainfall. Results also show a negative association of 0.161 between topographical wetness and poverty index, indicating that zones of low topographical wetness have high poverty index. However, a weak negative linear association of 0.148 exist between topographical wetness and distance to wetlands, the closer to wetlands the higher the topographical wetness. In addition, zones of low soil drainage capacity have high topographical wetness; a weak negative association of 0.144 between topographical wetness and soil drainage explains this.

Rainfall.
In Homa Bay County, the higher the soil drainage capacity the higher the rainfall. A positive association of 0.204 between rainfall and soil drainage reveals this. This is because zones of moderate and rapid drained soils area located at high altitude and the higher the altitude the higher the rainfall. Correlation analysis between rainfall and distance to wetlands also reveals a positive association of 0.328, the higher the rainfall the greater the distance to wetlands. The results further reveal a positive association of 0.280 between rainfall and poverty indicating that zones of high rainfall experience high poverty levels.

Soil drainage, distance to wetlands and poverty.
A positive association of 0.285 exist between soil drainage and distance to wetlands. This implies that zones of fast or rapid soil draining capacity are far from the wetlands. The results further revealed a weak positive association of 0.180 between soil drainage and poverty; zones of rapid soil drainage capacity experience high poverty levels. In addition, a positive association of 0.165 exist between distance to wetlands and poverty, the shorter the distance to wetlands the lower the poverty density.

Correlations

		Land_cover	Land_surface_temperature	NDVI	Altitude	Slope	Evapotranspiration	Distance_to_hyacinth	Distance_to_wetlands	Topographical_wetness	rainfall	Soil_drainage	Poverty	Infection_rate
Land_cover	Pearson Correlation	1	-.154**	.304**	.233**	-.017**	.187**	.383**	.073**	.244**	.274**	.017**	-.070**	.133**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Land_surface_temperature	Pearson Correlation	-.154**	1	-.442**	-.461**	-.320**	-.544**	-.272**	-.111**	-.360**	-.235**	-.221**	.150**	.191**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
NDVI	Pearson Correlation	.304**	-.442**	1	.322**	.011**	.611**	.502**	.075**	.445**	.491**	.040**	.024**	.195**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Altitude	Pearson Correlation	.233**	-.461**	.322**	1	.480**	.306**	.278**	.307**	.190**	.499**	.346**	.152**	-.143**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Slope	Pearson Correlation	-.017**	-.320**	.011**	.480**	1	.167**	-.198**	.075**	.170**	-.049**	.182**	-.122**	-.186**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Evapotranspiration	Pearson Correlation	.187**	-.544**	.611**	.306**	.167**	1	.374**	.091**	.373**	.417**	-.045**	.021**	-.148**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Distance_to_hyacinth	Pearson Correlation	.383**	-.272**	.502**	.278**	-.198**	.374**	1	-.104**	.143**	.510**	-.124**	-.042**	-.125**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Distance_to_wetlands	Pearson Correlation	.073**	-.111**	.075**	.307**	.075**	.091**	-.104**	1	.148**	.328**	.285**	.165**	-.127**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Topographical_wetness	Pearson Correlation	.244**	-.360**	.445**	.190**	.170**	.373**	.143**	.148**	1	.101**	.040**	-.161**	.121**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
rainfall	Pearson Correlation	.274**	-.235**	.491**	.499**	-.049**	.417**	.510**	.328**	.101**	1	.204**	.280**	-.171**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Soil_drainage	Pearson Correlation	.017**	-.221**	.040**	.346**	.182**	-.045**	-.124**	.285**	.040**	.204**	1	.180**	-.161**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Poverty	Pearson Correlation	-.070**	.150**	.024**	.152**	-.122**	.021**	-.042**	.165**	-.161**	.280**	.180**	1	.168**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812
Infection_rate	Pearson Correlation	.133**	.191**	.195**	-.143**	-.186**	-.148**	-.125**	-.127**	.121**	-.171**	-.161**	.168**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812	182812

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

Figure 34 (a). Correlation coefficient values among factors.

Figure 34 (a) also shows the linear correlation coefficient between infection rate and predictor variables. The results show weak linear associations among the aforementioned. A scatter plot of infection rate against temperature (one of the main predictors according to the literature review, see section 2.4) visually reveals a weak linear association between them (see figure 34 (b)).

A weak positive correlation coefficient of 0.191 and 0.133 exist between land surface temperature and infection rate, and land cover and infection rate respectively (figure 34 a). The higher the temperature the higher the infection rate, the same goes for the land cover, the more the land covered by grass, crops and forest the higher the malaria infection rate. The results also reveal that the higher the NDVI the higher the infection rate (weak positive association value of 0.195). Further analysis reveals that a weak negative association of 0.143 exist between altitude and infection rate, the higher the altitude the lower the infection rate.

In addition, a weak negative association of 0.186 exist between slope and infection rate, the higher the slope the lower the infection rate. A negative association of 0.148 also exist between evapotranspiration and malaria infection rate indicating that the higher the evapotranspiration the lower the infection rate. However, distance to hyacinth and distance to wetlands, both have a weak negative association of 0.125 and 0.127 respectively with infection rate, this indicates that the greater the distance from wetlands and water hyacinth the lower the infection rate and vice versa. Results also reveal a weak positive association of 0.121 between topographical wetness and infection rate, it indicates that the higher the topographical wetness the higher the infection rate.

Rainfall and soil drainage capacity have a weak negative association of 0.171 and 0.161 respectively with malaria infection rate. They indicate that the higher the rainfall and the faster the soil draining capacity the lower the infection rate. This is because higher rainfall intensity flashes our larva as explained Illinois Education (2015). Finally, a weak positive association of 0.168 exist between poverty and malaria infection rate, this implies that the higher the poverty density the higher the infection rate.
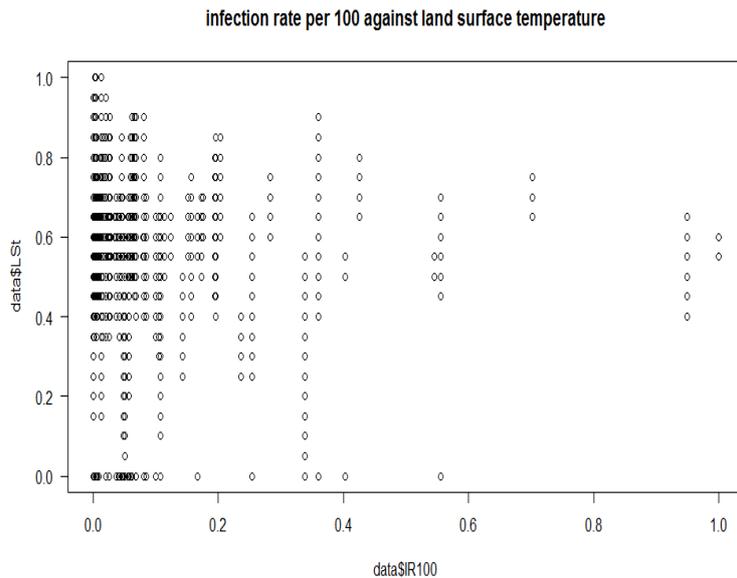
Figure 34 (b). The infection rates against temperature.

4.1.4 Significant spatial environmental and socio-economic factors leading to malaria infection in Homa bay county.

Collinearity test reveals that none of the factors is correlated with the other. This is explained by the values of VIF (variance inflation factor) for each variable in figure 35. The values are lower than 10 with the tolerance higher than 0.2 for each predictor or variables. Figure 35 shows the collinearity test result. In

addition, the average VIF value is 1.836 close to 1 indicating that *collinearity is not a problem for the model*. Equation 4.1 shows the formula for calculating average VIF.

Average VIF = *summation VIF(s)/ number of predictors*.   (4.1)

Average VIF = (1.435+1.383+2.391+2.220+1.270+2.106+2.527+2.024+1.807+1.301+1.351+2.222)/12

Average VIF =1.836

Based on KMO (Kaiser-Meyer-Olkin measure) value of 0.709, the sampling accuracy is acceptable (less than 0.5 unacceptable). This indicates that partial correlation between variables is small and, therefore, factors analysis is applicable in selecting the variables (see figure 36 for KMO results). In addition, Bartlett's test of sphericity indicates the data (variables) inputted into the model have a normal distribution (standardized); see the significance column in figure 36. However, the significance level is lower than 0.05 and, therefore, we accept the null hypothesis that the data has a normal distribution and reject the null hypothesis that the correlation matrix is an identity matrix. (Identity matrix is a matrix with all diagonal elements equal to 1 and off-diagonal elements equal to 0).

The closer the communality value to 1 the better the model in explaining variation within the factors. Based on this fact, variation in rainfall is the most explained, it is followed by altitude, NDVI, slope, evapotranspiration, land surface temperature, distance to hyacinth, topographical wetness, soil drainage, poverty, distance to wetlands and finally land cover in that sequence. This indicates that land cover is the least explained in the components. In addition, adding the communalities and dividing by the total number of variables we generate the average value of communality (0.58), this shows the total variables explained by the three components. The result is the same as the accumulated percentage proportion of variation explained by the first three eigenvalues (see total variance explained table in appendix 13).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.069	.017		-4.086	.000		
	Topographical_wetness	.014	.022	.002	.657	.511	.697	1.435
	Soil_drainage	-.163	.003	-.138	-54.456	.000	.723	1.383
	rainfall	-.707	.008	-.290	-86.763	.000	.418	2.391
	NDVI	.648	.016	.130	40.361	.000	.450	2.220
	Land_cover	.036	.003	.033	13.509	.000	.787	1.270
	Evapotranspiration	.578	.010	.185	58.896	.000	.475	2.106
	Altitude	1.004	.014	.255	74.214	.000	.396	2.527
	Land_surface_temperature	.175	.012	.046	15.084	.000	.494	2.024
	Slope	-.289	.015	-.055	-19.007	.000	.553	1.807
	Poverty	-.242	.005	-.110	-44.589	.000	.768	1.301
	Distance_to_wetlands	-.258	.008	-.086	-34.191	.000	.740	1.351
	Distance_to_hyacinth	-.259	.007	-.124	-38.652	.000	.450	2.222

a. Dependent Variable: Infection_rate

Figure 35. Collinearity results.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.709
Bartlett's Test of Sphericity	Approx. Chi-Square	731395.516
	df	66
	Sig.	.000

Figure 36. Kaiser-Meyer-Olkin Measure results.

Three components or composite factors are achieved out of rotated factor matrix process (see figure 38). However, the components explain 58% of the variance with each having eigenvalues greater than 1 (see

appendix 13 for total variance explained table, complete factor analysis results). *This is based on Kaiser criterion (rule of thumb) which drops all eigenvalues less than 1.* The scree plot also shows the three components dropping sharply before levelling between 4th to 12th-factor numbers. Further results on extracted communalities reveals that 0.228 of the variation in soil drainage, 0.804 of rainfall, 0.651 of NDVI, 0.155 of land cover, 0.556 of evapotranspiration, 0.737 of altitude, 0.552 of land surface temperature, 0.585 of slope, 0.214 of poverty, 0.206 of distance to wetlands, 0.476 of distance to hyacinth, and 0.331 of topographical wetness as individual factors are explained by the three components.

Results for rotated matrix also indicates various variables within the factor components or composite. The first component contains rainfall, NDVI, land cover, evapotranspiration, land surface temperature, distance to hyacinth and topographical wetness; *select factor values of 0.3 and above as this is the rule of thumb by Institute for Digital Research and Education (2015).*

The second component includes altitude, slope and soil drainage; the third contains poverty, rainfall (we eliminate rainfall since its appears in the first component) and distance to wetlands. In addition, various umbrella labels were awarded to these component based on the type of factors therein; the first component is biophysical (climatic, and hydrological), the second is topographical and the third is socio-economical. However, the values in figure 38 not only represents how variables are weighted for each factor (component) but also the correlation between the variables and the factor. NDVI has the highest correlation or weight in the biophysical factor, slope has the highest weight in the topographical factor, and poverty in the socio-economic factor.

Figure 37 shows the summary of selected variables leading to high malaria risk in Homa bay County within each component.

• Biophysical	• Topographical	• Socio economic
rainfall, NDVI, land cover evapotranspiration, land surface temperature, distance to hyacinth and topographical wetness	altitude, slope and soil drainage	poverty, distance to wetlands

Figure 37. Summary of components and selected variables

Rotated Factor Matrix

	Factor		
	1	2	3
Soil drainage	-.028	.358	.314
rainfall	.576	.004	.488
NDVI	.804	.046	.041
Land cover	.380	.016	.101
Evapotranspiration	.734	.126	-.029
Altitude	.357	.649	.434
Land surface temperature	-.558	-.478	.111
Slope	-.035	.761	-.064
Poverty	-.047	-.068	.455
Distance to wetlands	.049	.211	.699
Distance to hyacinth	.641	-.166	.194
Topographical wetness	.472	.221	-.243

Extraction Method: principal component analysis
Rotation Method: Varimax with Kaiser Normalization.

Figure 38. Rotated factor matrix results

The three components were mapped as shown in figure 39, 40 and 41 (biophysical, topographical, and socio-economic respectively). Additionally, a negative sign within the variables in the rotated factor matrix indicate low values and positive sign indicate high values. However, *as a rule, of thumb in the factors maps values above the standard deviation* are used to identify the characteristic of various variables within the components (see figure 39, 40 and 41).

Biophysical component
- • < -2.5 Std. Dev.
- • -2.5 - -1.5 Std. Dev.
- • -1.5 - -0.50 Std. Dev.
- • -0.50 - 0.50 Std. Dev.
- • 0.50 - 1.5 Std. Dev.
- • > 1.5 Std. Dev.
- ▪ Lake Victoria
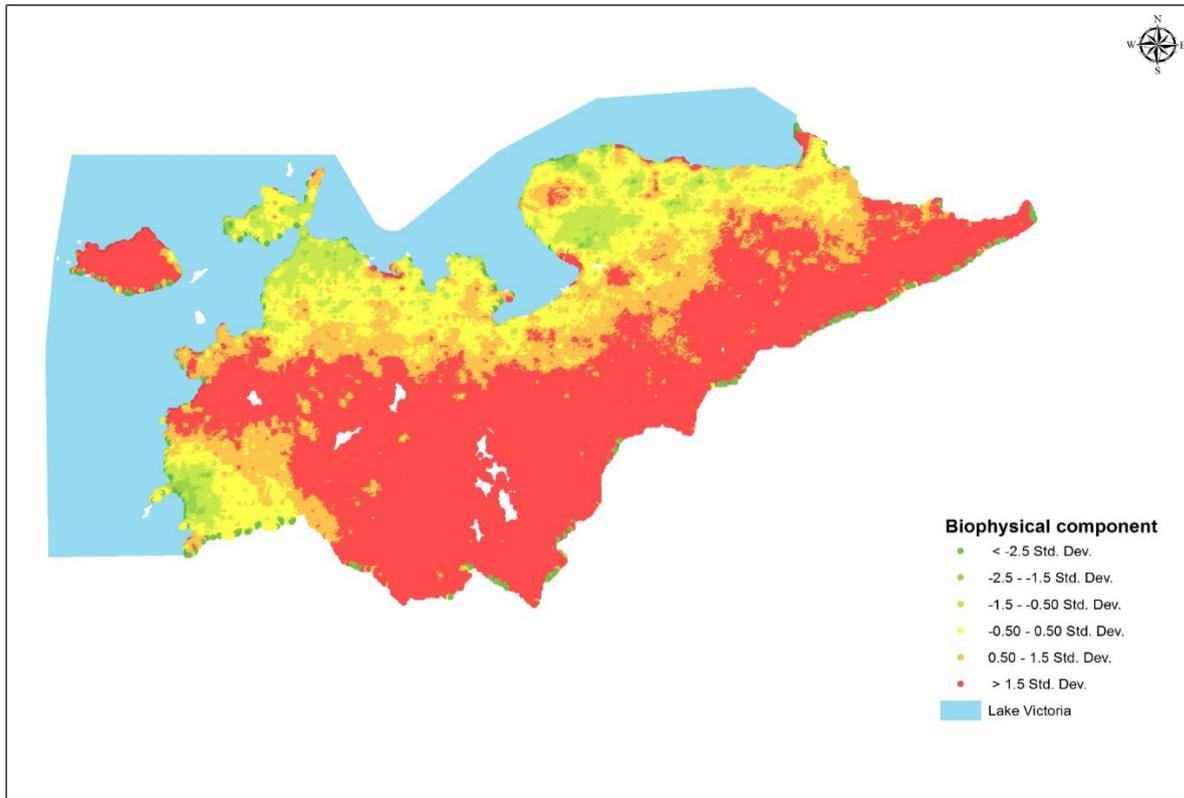
Figure 39. Biophysical component map.

Figure 39 shows the biophysical component map (factor 1). Visually looking at the zones above the standard derivation in figure 39, various characteristic namely high rainfall, high NDVI, more land cover (cropland, grassland, and forest), high evapotranspiration, low land surface temperature, far from wetlands, and finally high topographical wetness are experienced in these zones. Low infection rate is also a characteristic of these zones.

Figure 40 shows the topographical component map (factor 2). Zones with values above the standard deviation reveal various topographic characteristic namely, high altitude and high slope. Visual interpretation between topographical component map and infection rate map reveals that these zones also experience low infection rates.

In addition, figure 41 shows the socio-economic component map, the component entails poverty and distance to wetlands (rainfall and altitude also are significant in this component but since they already appear in the first and the second component respectively, they are assumed). Based on positive values of the standard deviation, these zones experience high poverty levels, high rainfall, high altitude and are far from wetlands. Correlation result between infection rates and distance to wetlands results reveals that zones far from wetlands experience low infection and, therefore, the zones above the standard deviation experience low infection rate.
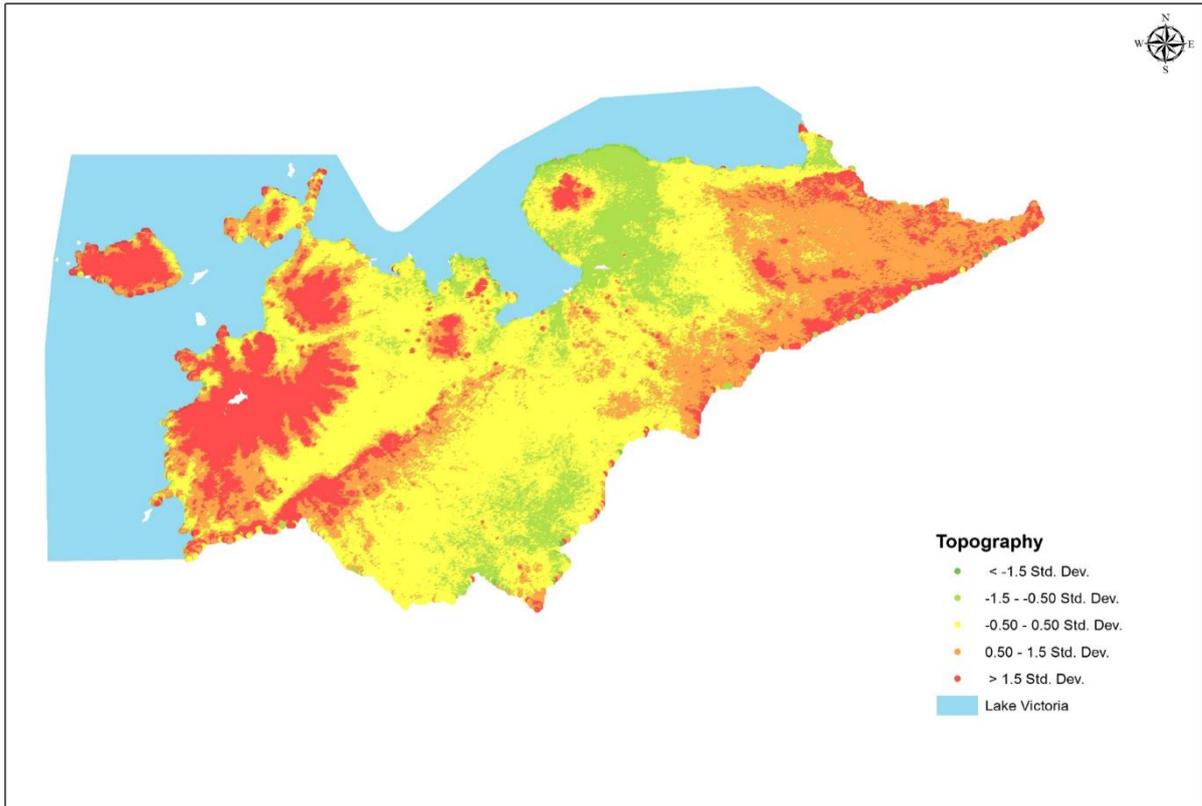
Figure 40. Topographical component map.
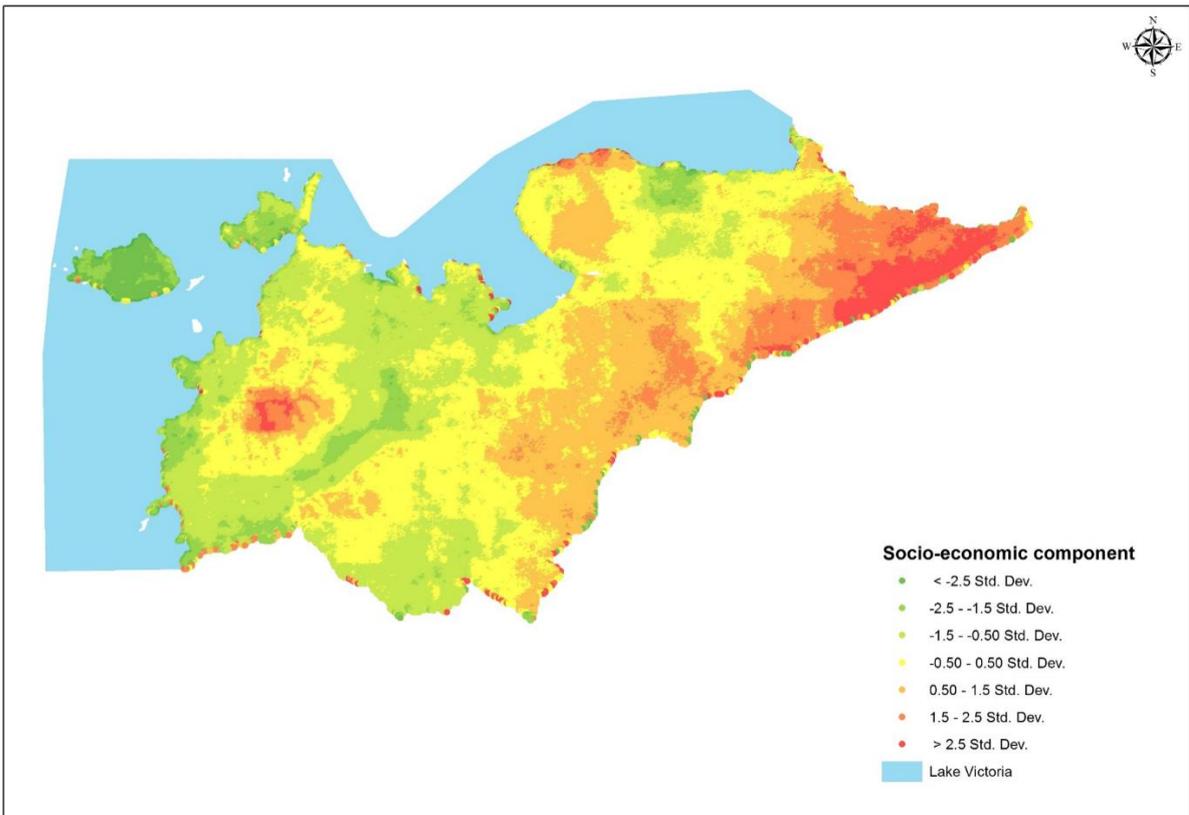
Figure 41. Socio economic component map

4.1.5 Homa Bay County potential malaria risk prediction and validation.

High-risk zones close to 10 (on the risk scale) are located along Lake Victoria, with low-risk areas far away from the lake. The highest altitude point is on the mountain shown in green in the western part of Homa bay county in figure 42, it experience the lowest risk. The risk is recorded as 1 in the risk scale. However, more than half of the county has moderate potential risk levels (represented by yellow colour in figure 42). In addition, the northeastern part of the county experience low potential risk shown in green colour. See figure 42 for the potential malaria risk map of Homa bay county.

Out of 122,178 points for validation, 50,794 points were correctly classified as low infection rate and low malaria risk, 27,787 points were correctly classified as having high malaria infection rate and high potential malaria risk, and 43,597 points misclassified. Using equation 3.8, the percentage accuracy for potential risk prediction was calculated. *The percentage accuracy* is *64.3%* ((50,794+27,787)*100)/122,178.
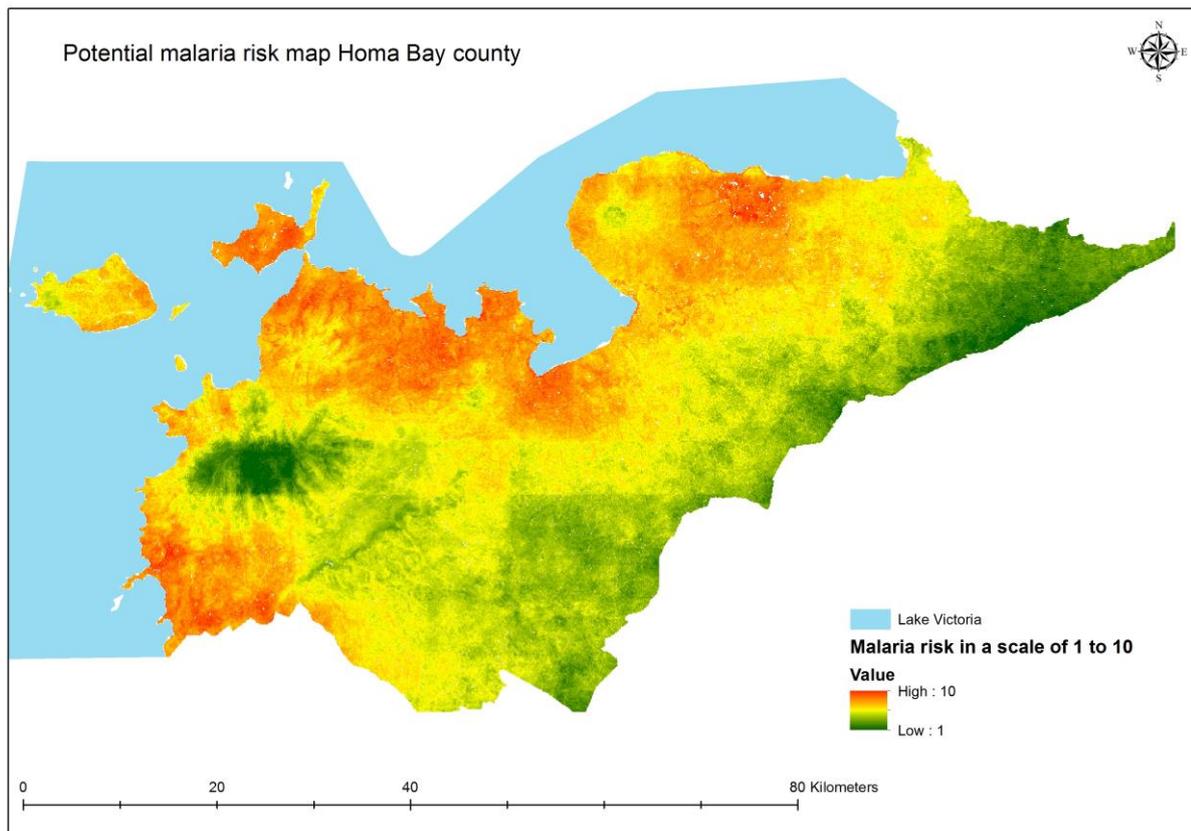
Figure 42. Potential malaria risk map of Homa bay county.

4.2 Discussion.

The overall objective of this study was to model spatial malaria risk factors, predicting potential malaria risk areas based on remote sensing derived environmental, socio-economic variables, and malaria cases (from health records). This subsection contains major findings and result interpretation. Research questions posed in chapter 1 are also answered in this section.

Various environmental and socio-economic factors contribute to malaria risk, this includes topographical, ecological, climatic, hydrological, and socio-economic factors. Topographical factors are elevation and slope, ecological factors include, land cover, vegetation cover and water hyacinth. Climatic factors include temperature, rainfall, and evapotranspiration while hydrological includes wetlands and topographical wetness. Soil drainage and poverty falls under geological and socio-economic factors respectively. The mentioned factors were also identified by Tuyishimire (2013) in studying malaria risk factors in Rwanda, Malone et al.(2003) in studying malaria risk assessment in Eritrea, and Koram et al.(1995) in studying socio-economic factors related to malaria risk in Gambia.

High and moderate resolutions satellites namely Landsat 8, Sentinel1, MODIS (moderate resolution imaging spectroradiometer), and CHRIPS (climate hazard group infrared precipitation with station data) satellites allow large imaging swaths for both local and regional generation and analysis of risk data. Monthly satellite images of various regions of the world are now available dating back to 1972 with the launch of Landsat 1. This advancement allows temporal analysis in disease surveillance at high-resolution mapping. Moreover, the monthly temperature is derived from thermal spectral bands of Landsat 8 at 100m spatial resolution using split window algorithm, this preserves the temperature variation within the data from place to place. Many ecological studies apply the concept of surface interpolation by using thermometer and rain gauge point data from meteorological stations. However, the stations are not evenly distributed within study areas and, therefore, loss of data variation and introduction of arithmetic errors in the data.

Many ecological researchers use the straight-line relationship between malaria infection rate and risk factors without testing. Correlation analysis is therefore used herein testing the association, Curtis, & Carey (2012) used this method in risk assessment practice. The correlation results reveal various linear associations between the aforementioned factors and infection rate. Firstly, between land cover and malaria infection rate, a positive association exists; this implies that the more the land is covered by grassland, cropland, and forest the higher the malaria risk in such zones. The same goes for temperature as a positive association exist between temperature and infection rate, the higher the temperature the higher the risk provided the optimum temperature of 25°C to 30°C be not exceeded. In addition, positive association between vegetation cover (NDVI) and infection rates indicate that the higher the NDVI the higher the malaria infection rate. This is the similar to the case found in Zambia by Nygren et al. (2014).

Zones of high altitude experience low malaria infection (see results on figure 34a; negative association of 0.171). This is because high altitude zones experience low temperature and therefore not suitable for mosquito development. A study conducted by Sambasivarao (2013) on participatory risk mapping of malaria, also confirms that low malaria transmission is experienced in high altitude zones. The correlation results also show a negative association between slope and infection rate. It indicates that the higher the slope in degrees rises the lower the infection rate. This is mainly because high slopes have rapid drainage capacity and, therefore, rainwater does not accumulate in these zones rendering them unsuitable for mosquito breeding.

Zones with high evapotranspiration experience low infection rates (see figure 34a), this is because high evapotranspiration reduces the topographical wetness of the soil and, therefore, flood mosquitos cannot lay eggs in such soils. Stresman (2010) mentions that zones with evapotranspiration above 800mm per year experience low malaria transmission. Additionally, the closer to wetlands and water hyacinth the higher the infection rate (see figure 34a). This is because water hyacinths provide breeding mats to mosquito and, therefore, zones within the flight range of mosquitos (1 -3miles, approximately 1-4 km) from wetlands and water hyacinths experience high infection rate. Virginia Department of Health (2015) and Florida Fish and Wildlife Conservation Commission (2015) also states that in order to reduce malaria risk, these zones should be taken into consideration during malaria eradication planning.

Positive association exist between topographical wetness and infection rate (see figure 34a), the higher the topographical wetness the higher the infection rate. This can be attributed to these zones having moist soils due to more topographical wetness. However, floodwater mosquitos lay eggs in these zones during low rainfall season. Association between rainfall and infection rates show that zones with high rainfall experience low malaria infection. Heavy rainfall intensity flashes out mosquito larva as low rainfall provide suitable habitat for their development. In addition, results indicate a negative association between soil drainage and infection rate; rapidly drained soils do not hold water for long and therefore not suitable for full cycle mosquito development. Morris (1998) in studying drainage consideration for malaria control also recognises the effect of soil drainage capacity in water discharge. Moreover, the higher the poverty density the higher the infection rate (positive association of 0.168). Rural areas lead in poverty density (see figure 33 b); rural areas experience high malaria infections than urban area due to high poverty levels in rural zones. According to a survey conducted by Malaria Control and Ministry of Public Health (2010), malaria

prevalence was found to be higher in rural (12%) as compared to urban (5%). Koram et al (1995) also confirm that the poor are more vulnerable than the rich.

Kaiser-Meyer-Olkin measure of sampling adequacy reveals that factor analysis is appropriate for selecting factors that lead to malaria risk in Homa bay County (see figure 36 for Kaiser-Meyer-Olkin measure of sampling adequacy value). However, three components awarded an umbrella label of biophysical, topographical and socio-economic are derived from the factor analysis result. Biophysical factors include rainfall, NDVI, evapotranspiration, land surface temperature, distance to hyacinth and topographical wetness. The topographic component includes altitude, slope and soil drainage while socio-economic component includes poverty and distance to wetlands.

Communalities indicate the percentage variation explained by the three components for each factor, values close to 1 are desirable. Moreover, results on extracted communalities in appendix 13 shows that variation in rainfall has the highest explained variance, followed by altitude, NDVI, slope, evapotranspiration, land surface temperature, distance to hyacinth, topographical wetness, soil drainage, distance to wetlands, and finally land cover. Rainfall and altitude are therefore the most contributing factor to malaria infection in Homa bay and land cover the least contributor (less variance is explained in the land cover). Rainfall, altitude, and temperature have been noted to be very significant contributors to malaria risk in malaria endemic zones by various studies. Alegana et al.(2013) in studying malaria incidence estimation in Namibia list the aforementioned as very important. Other selected factors are considered as modifying factors, example NDVI, wetlands, water hyacinth, evapotranspiration, slope, topographical wetness, soil drainage and land cover, they create more suitable mosquito development sites leading to higher risk than expected. Stresman (2010) in studying ecological risk factors that modify malaria transmission argues that they should be incorporated in malaria target control for same reason earlier mentioned.

Most of human settlements fall in high potential risk zones (see figure 43). In Rachuonyo (formerly named Karachuonyo), Suba, and Mbita Sub Counties, all the settlement fall within medium (around 5) to high (10) potential risk levels. In Kasipul Kabondo Sub County, two settlements fall within medium to high-risk level with the remaining settlements falling below medium risk level. Seven out of eighteen settlements in Ndhiwa Sub County fall below medium risk levels, eleven fall above medium risk levels. Homa Bay Sub County has 15 settlements with only four falling below medium in terms of risk. Visual interpretation of the risk map, (see figure 43) reveals that zones close to Lake Victoria are the riskiest and zones far from Lake Victoria experience low risks. The low-risk zones area of high altitude, high slope, far from wetlands, far from water hyacinth, high evapotranspiration, high rainfall intensity, rapid and moderate soil drainage capacity, low poverty density and low land surface temperature (vice versa for the high-risk factors). The risk map is 64% accurate using malaria infection rate as the reference data.

The 36% misclassification may be attributed to human behaviour which is complex to model (e.g. time of closing the windows and cultural systems), errors within the remote sensing data used, and malaria control measures in place (see chapter 3, study area on health system). Additionally, only four Landsat 8 satellite images were used for land cover, land surface temperature, topographical wetness, water hyacinth, and NDVI to calculate various yearly averages (images for February. May, September and December). This is due to cloud cover experienced in the remaining months, and, therefore, part of 36% misclassification may be attributed to this. Moreover, the standardization of land cover to portray land cover effects on malaria transmission is based on only six expert responses (due to time constraint in the field); with more respondents the standardization values is bound to change affecting the risk map.
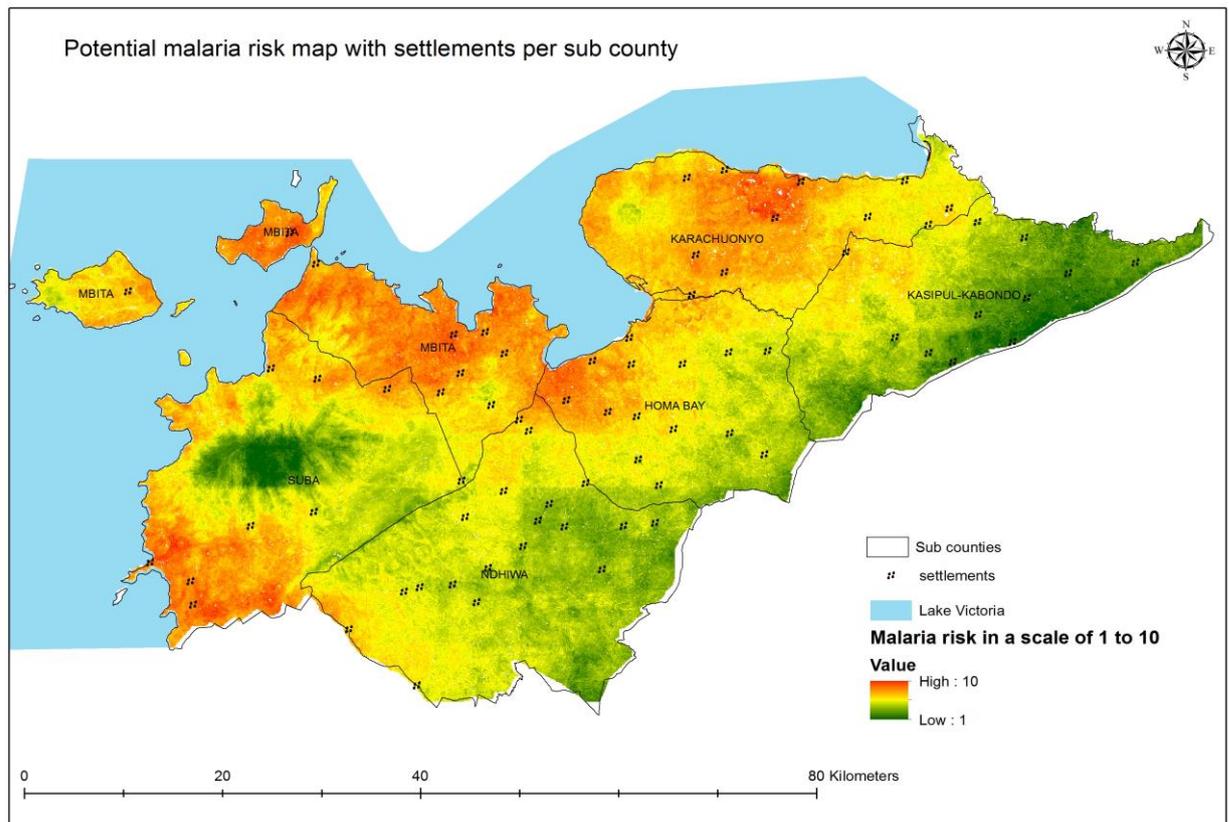
Figure 43. Potential malaria risk map with settlements per Sub County.

The health catchments used to calculate the infection rate is based on the assumption that patients go to the nearest hospital regardless of economic power, all malaria cases are reported at the health facility, and on walking speed of a healthy individual (not patients walking speed). In addition, the infection rate has both confirmed and clinical malaria cases combined; not all clinical cases are malaria cases. However, the rich population within Homa Bay town go to major hospitals regardless of the distance leading to underestimation of the infection rate in health centres close to them and overestimation in the preferred health facilities. This was not captured in this analysis due to lack of comprehensive socio-economic data (income levels of the population). Since more than three-quarters of the population are poor, this analysis is acceptable as poor people go to the nearest facility (approximately 867,000 people out of 1.06 million). Additionally, due to lack of settlement population data, the total population under malaria risk in Homa bay County cannot be estimated.

The results are based on a weak linear association between the factors and malaria infection, this is due to the limitation of data available for modelling. Additionally, higher linear function or order (example the second order of linear function, logistic function) cannot be applied to the malaria infection rate data generated. Figure 34 show the scatter plot of temperature and infection rate visually indicating a weak linear relationship. Moreover, the flight range and flight height of mosquito depend on the wind speed, this study does not consider the wind speed or wind direction in calculating the distance from wetlands and water hyacinth effects of malaria infection. This is because wind speed data for Homa bay is not available. Finally, the percentage contribution of risk factors to malaria infection is not within the scope of this study. This is because the data used cannot provide a reasonable information on this. Preliminary test results on linear regression show that only 12% of the infection rate is explained by the risk factors earlier mentioned in paragraph 2 of this subsection. (see appendix 14 for linear regression result).
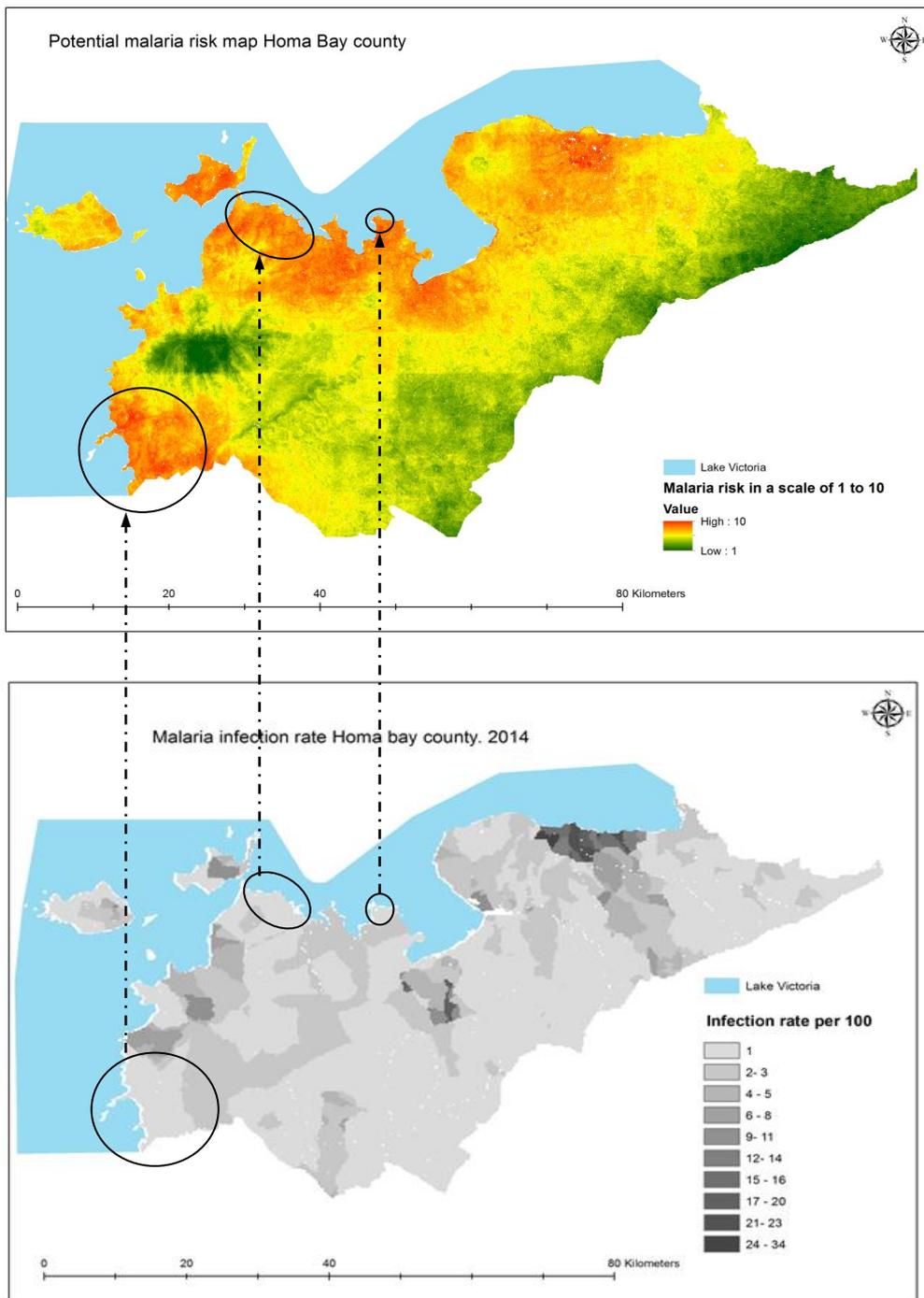
Figure 44 (a) and (b). Regions misclassified in the potential risk based on malaria infection rate.

Circles in Figure 44 (a) and (b) indicates the zones misclassified in terms of potential risk, these zones have low malaria infection rate yet classified as high in the potential risk map. This may be associated to errors in the representation of malaria occurrence (health catchment delineation and infection rate calculation), malaria control measures in place in those specific zones, and errors in the remote sensing data.

5 CONCLUSIONS AND RECOMMENDATIONS.

Malaria risk in Homa bay is attributed to risk factors; this includes biophysical, topographical, socio-economic, institutional, and cultural factors. This chapter contains the summary of the results in line with the objectives of the study. It also contains recommendations for further research, county government authorities, and sub-county health departments.

5.1 Conclusions.

Remote sensing is a good and readily available tool for risk analysis compared to surveys; it has proved useful in public health and epidemiological studies. It allows overlay of the various thematic map to visually present malaria risk. It also has the ability to incorporate a temporal aspect of the risk allowing complete risk analysis. With the advancement in geostatistical models (example factors analysis), malaria risk can be attributed to its cause factors and spatial patterns established. Moreover, epidemiological surveillance is key in developing any multi-dimensional malaria control strategy; the Kenyan Government uses USD 120million to update its risk maps every year, with the use of remote sensing techniques and health records data this can be geared towards other control measures like public awareness.

Malaria-causing factors in Homa bay are environmental, socio-economic, institutional, and cultural. Environmental factors include rainfall, altitude, temperature, slope, soil drainage, land cover, topographical wetness, NDVI, wetlands, water hyacinth, and evapotranspiration. These factors can be generated using remote sensing techniques and applied in risk map generation, which can be further used in malaria control planning and implementation. A socio-economic factor which includes poverty has various definitions and therefore very complex to model.

Human behaviour remains complex to model; this includes time of closing windows, time of using the ITNs, and attitude towards the control measures. However, malaria deaths reduced from 160 per 100,000 in the year 2010 to 40 per 100,000 in the year 2013 with the introduction of free ITNs to all Kenyans in 2008, proving that institutional factors also play a key role in malaria transmission and control. In addition, malaria risk remains high in rural areas than in urban areas; this is mainly due to high poverty levels in rural areas.

Distance to wetlands and water hyacinth is a critical malaria risk concept for Homa bay County. 230,000 out of 428,000 malaria cases in Homa bay County occur within 1km distance from wetlands while 57,000 cases occur within 1km distance from water hyacinth. Consequently, malaria risk reduces with increasing distance from wetland and water hyacinth. Moreover, habitat modifying factors, which includes slope, soil drainage, evapotranspiration, hyacinths, and wetlands, needs to be considered in malaria risk mapping for complete analysis. Correlation results indicate that zones close to wetlands have temperature favourable for mosquito breeding, therefore, the interplay of the two factors lead to high malaria risk in such zones.

The association between malaria infection and risk factors is not a straight line as many researchers take it. The parasite operates and develops within certain threshold example temperature (not only average temperatures but even daily temperatures), a temperature range of 25° to 30°C is an optimum condition for mosquito development increasing its density and malaria infection rate. Temperature below 25°C and above 30°C reduces parasite density drastically. Therefore, higher linear order (logistic) is necessary to fully analysis the relationship between factors and infection rate.

Factor analysis is appropriate for selecting risk factors for Homa Bay County; it investigates concepts that cannot be directly measured. However, results from factor analysis, communality section indicates that rainfall, altitude, NDVI, and temperature play greater roles in malaria risk in Homa Bay County. These are environmental factors that can be temporally analysed using remote sensing techniques to affirm their contributions over time. However, zones around Lake Victoria are of high risk and, therefore, control measures should focus on them since most of the population in Homa bay reside in these zones. Finally,

the risk is a potential risk map, meaning that further investigation is needed to fully ascertain the presence of malaria risk.

5.2 Recommendations.

This study is crucial to malaria reduction in Homa Bay County. Four hundred and twenty-eight thousand (428,000) malaria cases out of 1.06M people in the year 2014, malaria risk is real and needs to be eradicated. The risk factors vary from environmental to socio-economic to human behaviour, this section, therefore, contains proposals for further research, health and county government authorities.

Researchers.
- ✓ This study relied on malaria occurrence data as a measure of risk. Further research is required based on mosquito presence and using species distribution model to generate comparable malaria risk maps for Homa Bay County.
- ✓ The study only shows 2014 malaria risk map, the temporal analysis should be done to show trends for a complete understanding of the risk.
- ✓ The result of this study lacks human behaviour as a risk component. Therefore, there is the need to do further research on risk including human behaviour.
- ✓ Recent population datasets locate people in zones with are not habitable (e.g. mountains). Therefore, there is the need to come up with more accurate ways of representing population distribution.

Health authorities.
- ✓ The health catchments used in this study are generated from surface travel layers. It is also assumed that patient go to the nearest facility of which is not at times true for Homa bay County. This calls for more compressive and accurate health record keeping which incorporates the origin of patients (e.g. patient name, age, village name, medical, and health insurance cards).
- ✓ The results show that malaria risk is high close to Lake Victoria and reduces with increasing distance from the lake. Control measure (e.g. free distribution of insecticide-treated nets) needs to be directed to these zones.
- ✓ 230,000 out of 428,000 malaria cases occur within 1km distance from wetlands while 57,000 cases occur within 1km distance from water hyacinth. This is approximately above half the total cases, control measures (e.g. draining of wetlands not used by the community and removal of water hyacinth from the shore of Lake Victoria) should be directed to these zones.

County government authorities.
- ✓ Results indicate a positive association between poverty and infection rate. The higher the poverty levels the higher the infection rate within Homa Bay County, and therefore, the county government needs to improve the living standard (e.g. housing) of its residence.

LIST OF REFERENCES

Adu-Prah, S., & Kofi Tetteh, E. (2015). Spatiotemporal analysis of climate variability impacts on malaria prevalence in Ghana. *Applied Geography*, *60*, 266–273. doi:10.1016/j.apgeog.2014.10.010

Alegana, V. A., Atkinson, P. M., Wright, J. A., Kamwi, R., Uusiku, P., Katokele, S., Noor, A. M. (2013). Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial-temporal models. *Spatial and Spatio-Temporal Epidemiology*, *7*, 25–36. doi:10.1016/j.sste.2013.09.001

Alegana, V. A., Wright, J. A., Pentrina, U., Noor, A. M., Snow, R. W., & Atkinson, P. M. (2012). Spatial modelling of healthcare utilisation for treatment of fever in Namibia. *International Journal of Health Geographics*, *11*(1), 6. doi:10.1186/1476-072X-11-6

American Mosquito Control Association. (2015). Frequent Asked Questions. Retrieved November 25, 2015, from http://www.mosquito.org/faq

Anderson, J. E. (2010). The Gravity Model. *NBER Working Paper*, *3*(16576). doi:10.1257/aer.102.2.617

Answers-Africa. (2015). Poverty in Kenya: Statistics, Rate and Facts You Should Know. Retrieved December 23, 2015, from http://answersafrica.com/poverty-kenya.html

Armstrong Schellenberg, J. R. M., Smith, T., Alonso, P. L., & Hayes, R. J. (1994). What is Clinical Malaria? Finding Case Definitions for Field Research in Highly Endemic Areas. *Parasitology Today*, *10*(11), 439–442. doi:10.1016/0169-4758(94)90179-1

Attaway, D. F., Jacobsen, K. H., Falconer, A., Manca, G., Bennett, R., & Waters, N. M. (2014). Mosquito habitat and dengue risk potential in Kenya : alternative methods to traditional risk mapping techniques. *Geospatial Health*, *9*(1), 119–130. doi:10.4081/gh.2014.10

Australia Government. (2015). Benthic habitat mapping - Why make habitat maps. Retrieved August 17, 2015, from http://www.ozcoasts.gov.au/geom_geol/toolkit/mapwhymap.jsp

Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, *5*(5), 423–431. doi:10.1080/2150704X.2014.915434

Baird, J. K., Bangs, M. J., Maguire, J. D., & Barcus, M. J. (2002). Epidemiological measures of risk of malaria. *Methods in Molecular Medicine*, *72*, 13–22. doi:10.1385/1-59259-271-6:13

Beck-Johnson, L. M., Nelson, W. A., Paaijmans, K. P., Read, A. F., Thomas, M. B., & Bjørnstad, O. N. (2013). The effect of temperature on Anopheles mosquito population dynamics and the potential for malaria transmission. *PLoS ONE*, *8*(11). doi:10.1371/journal.pone.0079276

Benali, A., Nunes, J. P., Freitas, F. B., Sousa, C. A., Novo, M. T., Lourenço, P. M., Almeida, A. P. G. (2014). Satellite-derived estimation of environmental suitability for malaria vector development in Portugal. *Remote Sensing of Environment*, *145*, 116–130. doi:10.1016/j.rse.2014.01.014

Benndorf, M., Baltzer, P. A. T., & Kaiser, W. A. (2011). Assessing the degree of collinearity among the lesion features of the MRI BI-RADS lexicon. *European Journal of Radiology*, *80*(3), e322–e324. doi:10.1016/j.ejrad.2010.11.030

Brutlag, J. D. (2015). History of Correlation and Association. Retrieved December 10, 2015, from http://www.buttelake.com/corr.htm

Childs, L. M., Abuelezam, N. N., Dye, C., Gupta, S., Murray, M. B., Williams, B. G., & Buckee, C. O. (2015). Modelling challenges in context: Lessons from malaria, HIV, and tuberculosis. *Epidemics*, *10*, 102–107. doi:10.1016/j.epidem.2015.02.002

Cohen, J. M., Dlamini, S., Novotny, J., Kandula, D., Kunene, S., & Tatem, A. J. (2012). Rapid mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malaria Journal*, *11*(Suppl 1), O7. doi:10.1186/1475-2875-11-S1-O7

Cohen, W. B., & Goward, S. N. (2004). Landsat ' s Role in Ecological Applications of Remote Sensing. *BioScience*, *54*(6), 535–545. doi:10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2

Cossio, M. L. T., Giesen, L. F., Araya, G., Pérez-Cotapos, M. L. S., Vergara, R. L., Manca, M., Héritier, F. (2012). *Mapping Mpecies Distribution. Uma ética para quantos?* (Vol. XXXIII). doi:10.1007/s13398-014-0173-7.2

County Government of Homa bay. (2013). *COUNTY GOVERNMENT OF HOMA BAY "A County of Choice" FIRST COUNTY INTEGRATED DEVELOPMENT PLAN.*

Curtis, P., & Carey, M. (2012). Thought Leadership in ERM: Risks Assessment in Practice. *Committee of Sponsoring Organizations of the Treadway Commission (COSO)*, (October), 1–19.

Daily Nation. (2015). Genetically modified mosquitos to help fight malaria - Health, Science and Environment. Retrieved November 25, 2015, from http://www.nation.co.ke/lifestyle/health/genetically-modified-mosquito/-/1954202/2970820/-/i528yrz/-/index.html?utm_source=hootsuite

Data Mining Research. (2015). Standardization and normalization | Data Mining Research. Retrieved November 19, 2015, from http://www.dataminingblog.com/standardization-vs-normalization/

De Oliveira, E. C., Dos Santos, E. S., Zeilhofer, P., Souza-Santos, R., & Atanaka-Santos, M. (2013). Geographic information systems and logistic regression for high-resolution malaria risk mapping in a rural settlement of the southern Brazilian Amazon. *Malaria Journal*, *12*, 420. doi:10.1186/1475-2875-12-420

Etzkorn, B. (2014). Data Normalization and Standardization. Retrieved December 21, 2015, from http://www.benetzkorn.com/2011/11/data-normalization-and-standardization/

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using IBM SPSS Statistics. Statistics* (4th ed., Vol. 58). London: SAGE Publications Ltd. doi:10.1111/insr.12011_21

Florida Fish and wildlife conservation commission. (2015). Water-hyacinth. Retrieved July 31, 2015, from http://myfwc.com/wildlifehabitats/invasive-plants/weed-alerts/water-hyacinth/

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*(1), 185–201. doi:10.1016/S0034-4257(01)00295-4

Freudenrich, C. (2015). Types of Mosquitoes - HowStuffWorks. Retrieved November 3, 2015, from http://animals.howstuffworks.com/insects/mosquito1.htm

Garzón, M. B., Blazek, R., Neteler, M., Dios, R. S. De, Ollero, H. S., & Furlanello, C. (2006). Predicting habitat suitability with machine learning models: The potential area of Pinus sylvestris L. in the

Iberian Peninsula. *Ecological Modelling*, *197*(3-4), 383–393. doi:10.1016/j.ecolmodel.2006.03.015

Githeko, A. K., Ayisi, J. M., Odada, P. K., Atieli, F. K., Ndenga, B. a, Githure, J. I., & Yan, G. (2006). Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control. *Malaria Journal*, *5*, 107. doi:10.1186/1475-2875-5-107

Goddard Space Flight Center. (2015). Aerosol Robotic Network (AERONET) Homepage. Retrieved November 11, 2015, from http://aeronet.gsfc.nasa.gov/cgi-bin/type_one_station_opera_v2_new

Gogu, R. C., & Dassargues, a. (2000). Current trends and future challenges in groundwater vulnerability assessment using overlay and index methods. *Environmental Geology*, *39*(6), 549–559. doi:10.1007/s002540050466

Gwitira, I., Murwira, A., Zengeya, F. M., Masocha, M., & Mutambu, S. (2015). Modelled habitat suitability of a malaria causing vector (Anopheles arabiensis) relates well with human malaria incidences in Zimbabwe. *Applied Geography*, *60*, 130–138. doi:10.1016/j.apgeog.2015.03.010

Hassan, M. (2013). Evaporation estimation for Lake Nasser based on remote sensing technology. *Ain Shams Engineering Journal*, *4*(4), 593–604. doi:10.1016/j.asej.2013.01.004

Hay, S. I., Cox, J., Rogers, D. J., Randolph, S. E., Stern, D. I., Shanks, G. D., Snow, R. W. (2002). Climate change and the resurgence of malaria in the East African highlands. *Nature*, *415*(6874), 905–909. doi:10.1038/415905a

Heggenhougen, H. K., Hackethal, V., & Vivek, P. (2003). The behavioural and social aspects of malaria and its control. *Cdrwww.Who.Int*, 214.

Hightower, A. W., Ombok, M., Otieno, R., Odhiambo, R., Oloo, A. J., Lal, A, Hawley, W. A. (1998). A geographic information system applied to a malaria field study in western Kenya. *American Journal of Tropical Medicine and Hygiene*, *58*(3), 266–272.

Hirzel, A. H., & Le Lay, G. (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, *45*(5), 1372–1381. doi:10.1111/j.1365-2664.2008.01524.x

Homa bay County Government. (2015). County Profile | COUNTY GOVERNMENT OF HOMA BAY. Retrieved November 10, 2015, from http://homabay.go.ke/?page_id=15

Homan, T., Maire, N., Hiscox, A., Di Pasquale, A., Kiche, I., Onoka, K., Takken, W. (2016). Spatially variable risk factors for malaria in a geographically heterogeneous landscape, western Kenya: an explorative study. *Malaria Journal*, *15*(1), 1. doi:10.1186/s12936-015-1044-1

Hongoh, V., Berrang-Ford, L., Scott, M. E., & Lindsay, L. R. (2012). Expanding geographical distribution of the mosquito, Culex pipiens, in Canada under climate change. *Applied Geography*, *33*(1), 53–62. doi:10.1016/j.apgeog.2011.05.015

Illinois Education. (2015). Illinois | Heat, rainfall affect pathogenic mosquito abundance in catch basins | Illinois. Retrieved November 16, 2015, from https://news.illinois.edu/blog/view/6367/205034

Institute for Digital Research and Education. (2015). Annotated SPSS Output: Factor Analysis. Retrieved January 16, 2016, from http://www.ats.ucla.edu/stat/spss/output/factor1.htm

Institute for Environment and Human Security. (2015). Vulnerability Assessment, Risk Management & Adaptive Planning Section (VARMAP) -. Retrieved November 20, 2015, from http://ehs.unu.edu/about/sections/varmap#_

Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, practice: Principles, advances, and applications. *The Counseling Psychologist*, *34*(5), 684–718. doi:10.1177/0011000006286347

Kaplunovsky, A. S. (2005). Factor analysis in environmental studies. *HAIT Journal of Science and Engineering B*, *2*, 54–94.

Kelly, G. C., Tanner, M., Vallely, A., & Clements, A. (2012). Malaria elimination: Moving forward with spatial decision support systems. *Trends in Parasitology*, *28*(7), 297–304. doi:10.1016/j.pt.2012.04.002

Kelly, G., Tanner, M., Vallely, A., & Clements, A. (2012). Malaria elimination: moving forward with spatial decision support systems. *Trends in Parasitology*, *28*(7), 297–304.

Kenya Information guide. (2015). About Homa Bay County in Kenya. Retrieved November 10, 2015, from http://www.kenya-information-guide.com/homa-bay-county.html

Kenya Investement Authority. (2015). Homa Bay County - KenInvest. Retrieved August 4, 2015, from http://www.investmentkenya.com/homa-bay-county?tmpl=component

Kleinschmidt, I., Bagayoko, M., Clarke, G. P., Craig, M., & Le Sueur, D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*, *29*(2), 355–361. doi:10.1093/ije/29.2.355

Koram, K. A., Bennet, S., Adiamah, J. H., & Greenwood, B. M. (1995). Socio-economic risk factors for malaria in The Gambia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *89*, 146–150.

Laerd Statistics. (2015). Pearson Product-Moment Correlation - When you should run this test, the range of values the coefficient can take and how to measure strength of association. Retrieved December 10, 2015, from https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php

Li, X., Wang, S. G., Ge, Y., Jin, R., Liu, S. M., Ma, M. G., Liu, Q. H. (2013). Development and experimental verification of key techniques to validate remote sensing products. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *40*(2W1), 25–30.

Lourdes Torres-Sorando, D. J. R. (1997). Models of Spatial-temporal dynamics in malaria.

Malaria Control. (2010). Republic of Kenya Ministry of Health National Malaria Strategy, (April 2001).

Malaria Control Ministry of Public Health. (2010). 2010 Kenya MALARIA Indicator Survey.

Malaria World Report 2014. (2014). Retrieved August 20, 2015, from http://www.who.int/malaria/publications/country-profiles/profile_ken_en.pdf?ua=1

Malone, J. B., Poggi, E., Igualada, F.-J., Sintasath, D., Ghebremeskel, T., Corbett, J. D., Ford, R. (2003). Malaria environmental risk assessment in Eritrea. *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, *2*(C), 1000–1003. doi:10.1109/IGARSS.2003.1293991

MetaStock. (2015). Correlation Analysis. Retrieved December 10, 2015, from http://www.metastock.com/customer/resources/taaz/?c=3&p=44

Michael, E., Ramaiah, K. D., Hoti, S. L., Barker, G., Paul, M. R., Yuvaraj, J., Bundy, D. A. (2001). Quantifying mosquito biting patterns on humans by DNA fingerprinting of bloodmeals. *The*

American Journal of Tropical Medicine and Hygiene, *65*(6), 722–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11791964

Midekisa, A., Senay, G., Henebry, G. M., Semuniguse, P., & Wimberly, M. C. (2012). Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. *Malaria Journal*, *11*(1), 165. doi:10.1186/1475-2875-11-165

Morris, E. C. (1998). DRAINAGE CONSIDERATIONS FOR MOSQUITO CONTROL Peter Whelan, Department of Health and Community Services, Sept. 1997, 1–8.

Mosquito World. (2015). Mosquito Habitats - Mosquito World. Retrieved November 2, 2015, from http://www.mosquitoworld.net/about-mosquitoes/habitats/

Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800. doi:10.1016/j.rse.2011.02.019

Mutuku, F. M., Bayoh, M. N., Hightower, A. W., Vulule, J. M., Gimnig, J. E., Mueke, J. M., Walker, E. D. (2009). A supervised land cover classification of a western Kenya lowland endemic for human malaria: associations of land cover with larval Anopheles habitats. *International Journal of Health Geographics*, *8*(1), 19. doi:10.1186/1476-072X-8-19

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Giovannini, E., & Hoffmann, A. (2008). *Handbook on constructing composite indicators*. doi:10.1787/533411815016

Nath, M. J., Bora, a. K., Yadav, K., Talukdar, P. K., Dhiman, S., Baruah, I., & Singh, L. (2013). Prioritizing areas for malaria control using geographical information system in Sonitpur district, Assam, India. *Public Health*, *127*(6), 572–578. doi:10.1016/j.puhe.2013.02.007

National Oceanic and Atmospheric Administration. (2015). What is remote sensing? Retrieved November 3, 2015, from http://oceanservice.noaa.gov/facts/remotesensing.html

Noor, A. M., Amin, A. A, Gething, P. W., Atkinson, P. M., Hay, S. I., & Snow, R. W. (2006). Modelling distances travelled to government health services in Kenya. *Tropical Medicine & International Health : TM & IH*, *11*(2), 188–96. doi:10.1111/j.1365-3156.2005.01555.x

Noor, A. M., Gething, P. W., Alegana, V. A, Patil, A. P., Hay, S. I., Muchiri, E., Snow, R. W. (2009). The risks of malaria infection in Kenya in 2009. *BMC Infectious Diseases*, *9*, 180. doi:10.1186/1471-2334-9-180

Nygren, D., Stoyanov, C., Lewold, C., Månsson, F., Miller, J., Kamanga, A., & Shiff, C. J. (2014). Remotely-sensed, nocturnal, dew point correlates with malaria transmission in Southern Province, Zambia: a time-series study. *Malaria Journal*, *13*, 231. doi:10.1186/1475-2875-13-231

Onchiri, F. M. (2014). *Changing Etiologies of Febrile Illness in Areas of Differing Malaria Transmission in Rural Kenya.*

Oxford Dictonaries. (2015). mosquito - definition of mosquito in English from the Oxford dictionary. Retrieved November 3, 2015, from http://www.oxforddictionaries.com/definition/english/mosquito

Oxford Dictonaries. (2015). vulnerable - definition of vulnerable in English from the Oxford dictionary. Retrieved November 3, 2015, from

http://www.oxforddictionaries.com/definition/english/vulnerable

Özesmi, S. L. (1999). An artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecological Modelling*, *116*, 15–31.

Páez, A., & Scott, D. M. (2005). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, *61*(1), 53–67. doi:10.1007/s10708-005-0877-5

PAI. (2015). Types of Factor Analysis | PAI Blog. Retrieved January 12, 2016, from http://info.paiwhq.com/3-types-of-factor-analysis/

Peelman, P., J.F. Trape, & Morault-Peelman, B. (1985). Criteria for diagnosing population exposed. *Tropical Medicine*, (February 1983).

Peterson, A. T. (2003). Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, *78*(4), 419–33. doi:10.1086/378926

Phillips, D. L., & Marks, D. G. (1996). Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. *Ecological Modelling*, *91*, 213–229. doi:10.1016/0304-3800(95)00191-3

Qualys Community. (2015). Difference between Vulnerabilities and Potential Vulnerabilities. Retrieved November 3, 2015, from https://community.qualys.com/docs/DOC-1166

Rajeshwari, A, & Mani, N. D. (2014). Estimation of Land Surface Temperature of Dindigul District Using Landsat 8 Data, 122–126.

Rincón-Romero, Mauricio Edilberto; Londoño, J. E. (2009). Mapping malaria risk using environmental and anthropic variables. *Revista Brasileira de Epidemiologia*, *12*(100), 338–354. doi:10.1590/S1415-790X2009000300005

Rozenstein, O., Qin, Z., Derimian, Y., & Karnieli, A. (2014). Derivation of Land Surface Temperature for Landsat-8 TIRS Using a Split Window Algorithm. *Sensors*, *14*(4), 5768–5780. doi:10.3390/s140405768

Rummel, R. J. (2015). FACTOR ANALYSIS. Retrieved January 12, 2016, from https://www.hawaii.edu/powerkills/UFA.HTM

Sambasivarao, S. V. (2013). Patriciparory Risk Mapping of Malaria Vector Exposure in Northern South America using Environmental and Population Data., *18*(9), 1199–1216. doi:10.1016/j.micinf.2011.07.011.Innate

Singh, V. (2003). Principle of Maximum Entropy. In *Entropy Theory and its Application in Environmental …* (pp. 85–92). doi:10.1007/s10955-006-9121-z

Standard Digital News. (2015). Standard Digital News - Kenya : Homa Bay County approves Sh6.2 billion budget. Retrieved December 13, 2015, from http://www.standardmedia.co.ke/article/2000167713/homa-bay-county-approves-sh6-2-billion-budget

Stevens, K. B., & Pfeiffer, D. U. (2011). Spatial modelling of disease using data- and knowledge-driven approaches. *Spatial and Spatio-Temporal Epidemiology*, *2*(3), 125–133. doi:10.1016/j.sste.2011.07.007

Stresman, G. H. (2010). Beyond temperature and precipitation: Ecological risk factors that modify malaria transmission. *Acta Tropica*, *116*(3), 167–172. doi:10.1016/j.actatropica.2010.08.005

Stryker, J. J., & Bomblies, A. (2012). The impacts of land use change on malaria vector abundance in a water-limited, highland region of Ethiopia. *EcoHealth*, *9*(4), 455–470. doi:10.1007/s10393-012-0801-7

Sturrock, H. J. W., Cohen, J. M., Keil, P., Tatem, A. J., Le Menach, A., Ntshalintshali, N. E., Gosling, R. D. (2014). Fine-scale malaria risk mapping from routine aggregated case data. *Malaria Journal*, *13*(1), 421. doi:10.1186/1475-2875-13-421

Tang, R., & Li, Z.-L. (2015). Evaluation of two end-member-based models for regional land surface evapotranspiration estimation from MODIS data. *Agricultural and Forest Meteorology*, *202*, 69–82. doi:10.1016/j.agrformet.2014.12.005

Tuyishimire, J. (2013). *Spatial Modelling of Malaria Risk Factors in Ruhuha*. Retrieved from http://www.itc.nl/library/papers_2014/msc/nrm/tuyishimire.pdf

UN-Habitat. (2008). *Strategic Urban Development Plan for Homa Bay Municipality, Kenya*.

University of Florida. (2015). Mosquito Information Website - Mosquito Life - Mosquito Habitats. Retrieved November 2, 2015, from http://mosquito.ifas.ufl.edu/Mosquito_Habitats.htm

USGS. (2014). What are the band designations for the Landsat satellites? Retrieved December 17, 2015, from http://landsat.usgs.gov/band_designations_landsat_satellites.php

USGS. (2015). Using the USGS Landsat 8 Product. Retrieved November 11, 2015, from http://landsat.usgs.gov/Landsat8_Using_Product.php

Viginia Department of Health. (2015). Mosquito Breeding Habitats. Retrieved July 31, 2015, from http://www.vdh.virginia.gov/lhd/CentralShenandoah/EH/WNV/mosquito_breeding_habitats.htm

Waldo Tobler. (1993). THREE PRESENTATIONS ON GEOGRAPHICAL ANALYSIS AND MODELING. Retrieved December 18, 2015, from http://www.geodyssey.com/papers/tobler93.html

Wang, H., & Mieghem, P. Van. (2007). Constructing the Overlay Network by Tuning Link Weigts, 139–143. doi:10.1109/CHINACOM.2007.4469347

WHO. (2009). Epidemiological profile Kenya. Retrieved August 23, 2015, from http://www.who.int/malaria/publications/country-profiles/2009/mal2009_kenya_0025.pdf

WHO. (2015). WHO | 10 facts on malaria. Retrieved August 12, 2015, from http://www.who.int/features/factfiles/malaria/en/

World Bank. (2015). Poverty Overview. Retrieved January 25, 2016, from http://www.worldbank.org/en/topic/poverty/overview

World Health Organization. (2014). Kenya: Malaria Country Profile, 1. Retrieved from http://www.who.int/malaria/publications/country-profiles/profile_ken_en.pdf?ua=1

World Health Organization. (2015). WHO | Determining health-care facility catchment areas in Uganda using data on malaria-related visits. Retrieved November 15, 2015, from http://www.who.int/bulletin/volumes/92/3/13-125260-ab/en/

World Resources Institute. (2015). Kenya GIS Data. Retrieved December 21, 2015, from http://www.wri.org/resources/data-sets/kenya-gis-data

APPENDIX

Appendix 1, malaria occurrence spread sheet.

FACILITY_NAME	DIST	Total	Jan-14		Feb-14		Mar-14		Apr-14		May-14		Jun-14		Jul-14		Aug-14		Sep-14		Oct-14		Nov-14		Dec-14	
			confirmne	Clinical	confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clinical	Confirmed	Clincal
ASUMBI MISS H C	HOMA BAY																									
FULL GOSPEL DISP	HOMA BAY																									
GOT KOJOWI H C	HOMA BAY																									
GOT KOKECH FULL GOSPEL DISP	HOMA BAY																									
HOMA BAY DISTRICT HOSPITAL	HOMA BAY																									
KAMRERI (COG) DISP	HOMA BAY																									
KANYAMKAGO OBER DISP	HOMA BAY																									
KIRANDA DISP	HOMA BAY																									
LAMBWE FOREST DISP	HOMA BAY																									
MAGINA H C	HOMA BAY																									
MALELA DISP	HOMA BAY																									
MALELA DISP	HOMA BAY																									
MANYATTA S D A HOSPITAL	HOMA BAY																									
MARINDI HC	HOMA BAY																									
MIROGI MISS H C	HOMA BAY																									
NDHIWA DIST	HOMA BAY																									
NDIRU H C	HOMA BAY																									
NYAGORO H C	HOMA BAY																									
NYAGWETHE DISP	HOMA BAY																									
NYARONGI DISP	HOMA BAY																									
OBER KABUOCH DISP	HOMA BAY																									
OBERA HEALTH CENTRE	HOMA BAY																									
OGANDE C P K DISP	HOMA BAY																									
OMBO KACHIENG DISP	HOMA BAY																									
OYARO DISP	HOMA BAY																									
PALA H C	HOMA BAY																									
RANGWE SUB DIST	HOMA BAY																									
RANGWE S D A DISP	HOMA BAY																									
RAPETH DISP	HOMA BAY																									
ACK HOPE DISP	HOMA BAY																									
WAWARE SINDO DISP	HOMA BAY																									
ADIEDO H C	RACHUONYO																									
ATEMO MISSION HEALTH CENTRE	RACHUONYO																									
GENDIA MISSION HOSPITAL	RACHUONYO																									
GOT BER DISP	RACHUONYO																									
GOT OYARO DISP	RACHUONYO																									
HOMA LIME H C	RACHUONYO																									
KABONDO H C	RACHUONYO																									
KANDIEGE H C	RACHUONYO																									
KENDU BAY MISS. HOSPITAL	RACHUONYO																									
KENDU SUB-DISTRICT HOSPITAL	RACHUONYO																									
KOBUYA DISP	RACHUONYO																									
KOKWONYO DISP	RACHUONYO																									
KOSELE DISP	RACHUONYO																									

Appendix 2, sample of land cover validation points.

Class	Sub_county	X	Y	Value
Settlement	Ndhiwa	34.36608749070	-0.73125330387	2
Cropland	Ndhiwa	34.36716980740	-0.72556700575	6
Cropland	Ndhiwa	34.37293265090	-0.72682879384	6
Cropland	Ndhiwa	34.37282138820	-0.73619054819	6
Grassland	Ndhiwa	34.36586087670	-0.73653730444	5
Grassland	Ndhiwa	34.35855497530	-0.72677573686	5
Bare ground	Ndhiwa	34.36700044540	-0.73154019411	3
Grassland	Ndhiwa	34.36357689460	-0.73056486327	5
Grassland	Ndhiwa	34.37070856690	-0.73021807517	5
Bare ground	Ndhiwa	34.37327713700	-0.73395048656	3
Bare ground	Homa Bay	34.45941396070	-0.53885481075	3
Settlement	Homa Bay	34.46214753970	-0.53861485212	2
Bare ground	Homa Bay	34.46440476240	-0.53442658420	3
Settlement	Homa Bay	34.46143294400	-0.53239325759	2
Cropland	Homa Bay	34.46737949220	-0.54854445677	6
Cropland	Homa Bay	34.45632753630	-0.55452965679	6
Bare ground	Homa Bay	34.45299590220	-0.53909565507	3
Bare ground	Homa Bay	34.45287524570	-0.53143809527	3
Cropland	Homa Bay	34.46260637650	-0.56257934051	6
Bare ground	Homa Bay	34.47618447600	-0.58730840980	3
Settlement	Suba	34.17021221390	-0.54280923457	2
Settlement	Suba	34.16783385260	-0.53754450217	2
Cropland	Suba	34.16716993450	-0.54663906025	6
Bare ground	Suba	34.17572613070	-0.53476679148	3
Settlement	Suba	34.16384139040	-0.54539520819	2
Bare ground	Suba	34.17781951780	-0.54165898200	3
Settlement	Suba	34.16726171810	-0.52921602880	2
Water	Mbita	34.21185130850	-0.43013568295	1
Water	Suba	34.16307908410	-0.53725822279	1
Forest	Suba	34.16964324250	-0.55104219066	4
Bare ground	Suba	34.17372997990	-0.53907505404	3
Settlement	Mbita	34.21854796590	-0.44407276267	2
Settlement	Mbita	34.21246336070	-0.45157894626	2
Bare ground	Mbita	34.19952365570	-0.40501815614	3
Settlement	Mbita	34.20424837160	-0.41597723649	2
Grassland	Mbita	34.22509973140	-0.44404440416	5
Cropland	Mbita	34.22709659190	-0.44375688442	6
Bare ground	Mbita	34.23156208670	-0.45155993309	3
Water	Mbita	34.20314549750	-0.43088712588	0
Grassland	Mbita	34.20736888610	-0.43924937998	5
Bare ground	Rachuonyo Nort	34.66680482020	-0.40276569991	3
Cropland	Rachuonyo Nort	34.67038001530	-0.40521513688	6
Bare ground	Rachuonyo Nort	34.66284887460	-0.39801928314	3
Settlement	Rachuonyo Nort	34.67047673750	-0.41696829445	2
Cropland	Rachuonyo Nort	34.67496218190	-0.40770263630	6
Grassland	Rachuonyo Nort	34.67458219740	-0.40915750356	5

Appendix 3, expert questioner for standardization of land cover and variables.
Expert questioner for Spatial Modelling of Mosquito Habitat Suitability.
A case of Homa bay county, Kenya

Various land cover have various effects on mosquito habitat (either as spatial constrain or factor). Table 1 shows a list of dominant land cover classes within Homa Bay County. Kindly fill a value ranging 0.1 to 1 in the standardize value column with value 1 as the most important and 0.1 as the least important land cover.

Land cover	Weights
Water	
Forest	
Grassland	
Cropland	
Bare ground. (Including roads)	
Settlements.	

Table 1. Land cover classes.
Any other factor…………………………
…………………

Appendix 4, split window coefficient values**.**

Constant	Value
C0	-0.268
C1	1.378
C2	0.183
C3	54.300
C4	-2.238
C5	-129.200
C6	16.400

Appendix 5. wetlands/ water ponds.
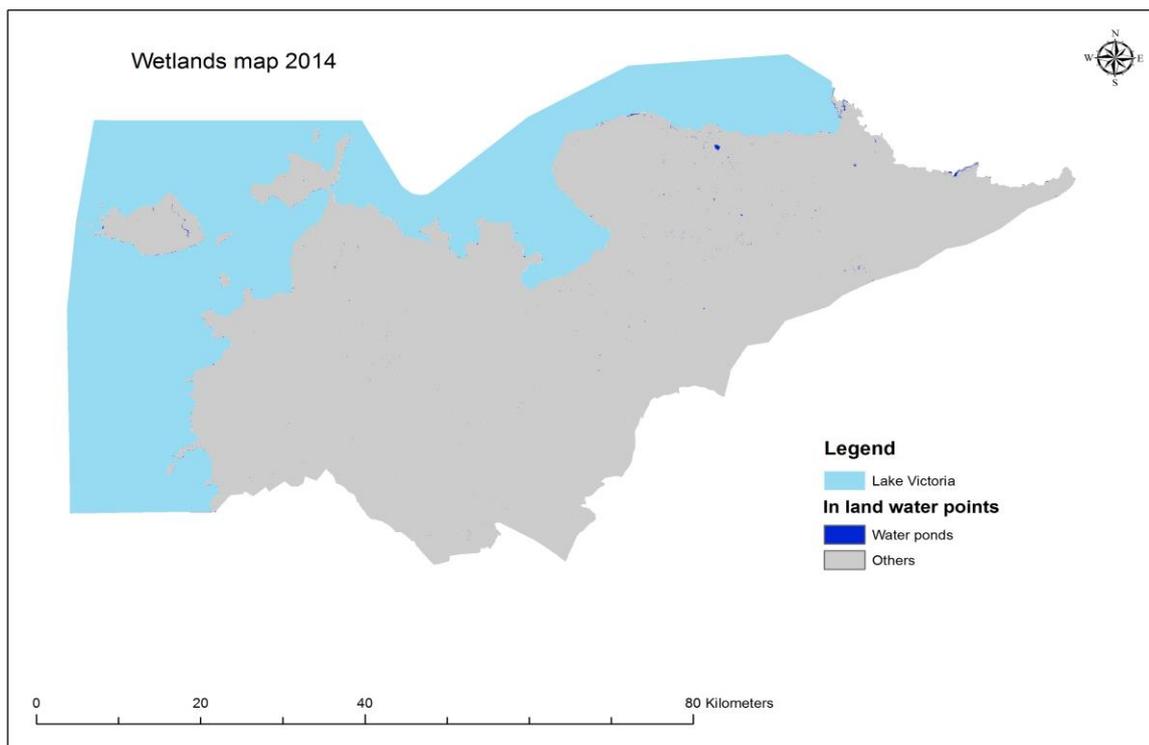
Appendix 6, tasselled cap transformation coefficient values.

TCT	Blue Band 2	Green Band 3	Red Band 4	NIR Band 5	SWIR1 Band 6	SWIR2 Band 7
Brightness	0.3029	0.2786	0.4733	0.5599	0.508	0.1872
Greenness	-0.2941	-0.243	-0.5424	0.7276	0.0713	-0.1608
Wetness	0.1511	0.1973	0.3283	0.3407	-0.7117	-0.4559
TCT4	-0.8239	0.0849	0.4396	-0.058	0.2013	-0.2773
TCT 5	-0.3294	0.0557	0.1056	0.1855	-0.4349	0.8085
TCT 6	0.1079	-0.9023	0.4119	0.0575	-0.0259	0.0252

Appendix 7, 1km effect analysis flowchart.
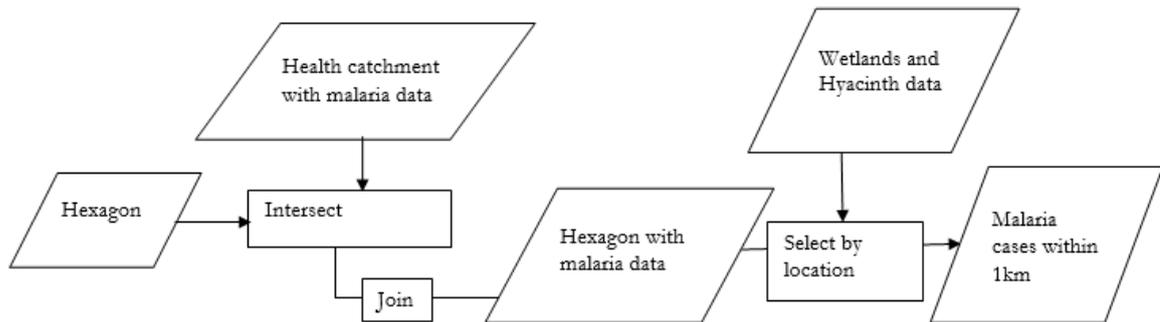
Appendix 8, Health catchment creation flow chart.
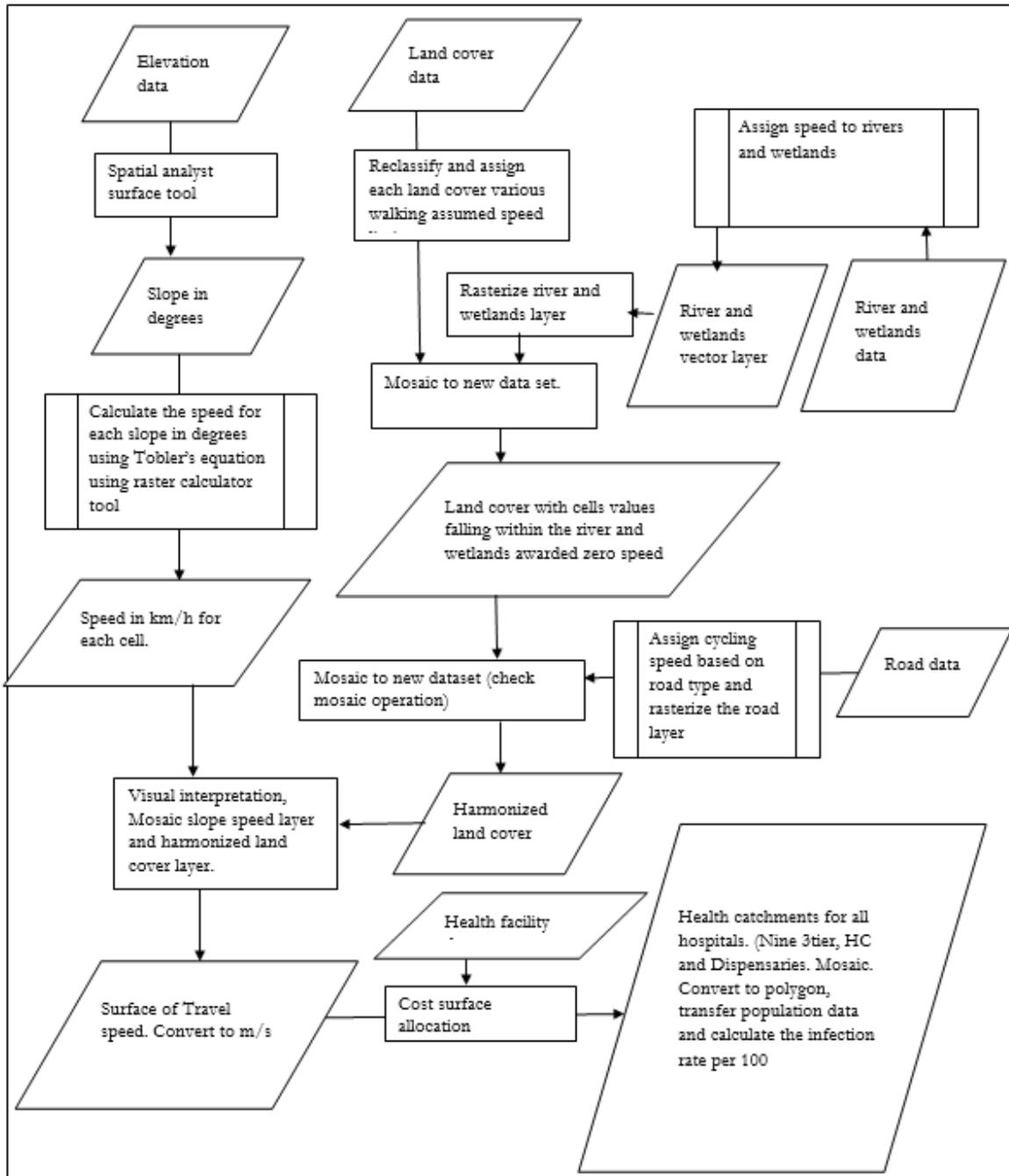
Appendix 9 malaria infection rate (IR) statistics.

Appendix 10 standardized factor maps.

(a)

(b)

(c)

(d)

(e)

(f)

Rainfall standadized map

Lake Victoria

rainfall

- 0 - 0.094023151
- 0.094023151 - 0.207634459
- 0.207634459 - 0.293822348
- 0.293822348 - 0.3878455
- 0.3878455 - 0.481868651
- 0.481868651 - 0.571974171
- 0.571974171 - 0.654244429
- 0.654244429 - 0.728679424
- 0.728679424 - 0.826620206
- 0.826620206 - 0.998995984

0 20 40 80 Kilometers

(g)

Poverty Density standardized Map 2014

Lake Victoria

Poverty standardized values

- 0.00000071 - 0.019608539
- 0.019608539 - 0.247059362
- 0.247059362 - 0.337255378
- 0.337255378 - 0.435294525
- 0.435294525 - 0.486274882
- 0.486274882 - 0.552941503
- 0.552941503 - 0.627451255
- 0.627451255 - 0.701961007
- 0.701961007 - 0.784313891
- 0.784313891 - 1.000000016

0 20 40 80 Kilometers

(h)

Distance to hyacinth map

(i)

Distance to wetlands map

Appendix 11 (a) and (b), number of malaria occurrence 1km from wetlands and water hyacinth respectively.

Field

MC

Statistics:

Count: 10274
Minimum: 13
Maximum: 36
Sum: 232053
Mean: 22.586432
Standard Deviation: 9.468017
Nulls: 0

(a)

Field

MC

Statistics:

Count: 2248
Minimum: 14
Maximum: 36
Sum: 57018
Mean: 25.363879
Standard Deviation: 8.164848
Nulls: 0
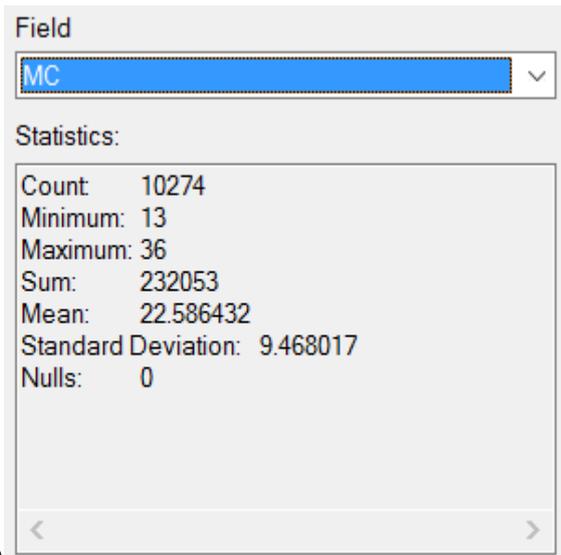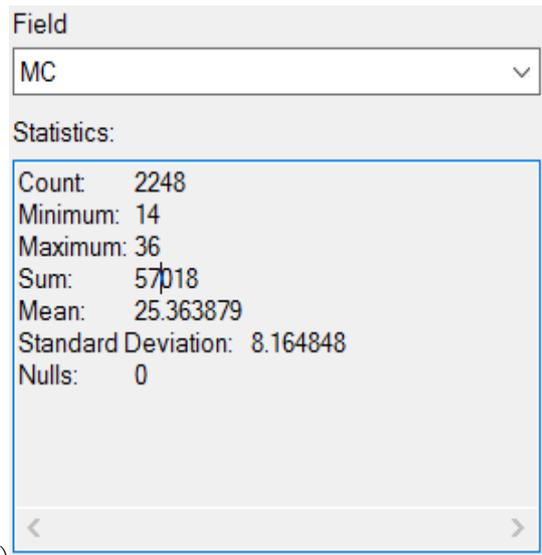
(b)

Appendix 12, Overlay index method data analysis.

Original datasets.

Rainfall

Land surface temperature

Evapotranspiration.

Poverty

Distance to wetlands

Distance to hyacinth

Soil drainage

Altitude

Slope

Topographical wetness.

Land cover.

Reclassified based on literature review.
(From 1 to 10 classes)

- High rainfall flashes out larva. (High 1 gradually increasing to low classified as 10).

- Temperature between 25°C to 30°C considered optimum. (25 to 30 reclassified as 10 gradually reducing to 1 on both sides from optimum.

- Evapotranspiration below 800mm/year considered optimum. (Low evapotranspiration classified as 10 with gradual decrease to 1 with increasing evapotranspiration values).

- High poverty high malaria risk. (High poverty classified as 10 with gradual decrease to 1 as poverty levels decreases).

- Distance to wetlands and hyacinth is based on flight range of mosquito (close to wetlands and hyacinth reclassified as 10 with gradual decrease to 1 as distance increases).

- Slow drained soil hold water for long. (Reclassified as 10 for soil draining soils gradually decreasing to 1 with an increase in draining capacity).

- High altitude and high slope reclassified as 1 with increasing values up to 10 for low altitude and low slopes.

- Mosquito lays eggs in moist soils. High wetness reclassified as 10 with gradual decrease to 1 with decreasing wetness values.

- Cropland, grassland reclassified as 10 with decreasing values to 1 for other land cover.

Correlation coefficients

From correlation analysis. Used as weights in overlay weighted sum tool.

Risk map

Appendix 13, factor analysis results.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.709
Bartlett's Test of Sphericity	Approx. Chi-Square	731395.516
	df	66
	Sig.	.000

Communalities

	Initial	Extraction
Soil_drainage	1.00	.228
rainfall	1.00	.804
NDVI	1.00	.651
Land_cover	1.00	.155
Evapotranspiration	1.00	.556
Altitude	1.00	.737
Land_surface_temperature	1.00	.552
Slope	1.00	.585
Poverty	1.00	.214
Distance_to_wetlands	1.00	.206
Distance_to_hyacinth	1.00	.476
Topographical_wetness	1.00	.331

Extraction Method: principal component analysis .

Rotated Factor Matrix

	Factor		
	1	2	3
Soil drainage	-.028	.358	.314
rainfall	.576	.004	.488
NDVI	.804	.046	.041
Land cover	.380	.016	.101
Evapotranspiration	.734	.126	-.029
Altitude	.357	.649	.434
Land surface temperature	-.558	-.478	.111
Slope	-.035	.761	-.064
Poverty	-.047	-.068	.455
Distance to wetlands	.049	.211	.699
Distance to hyacinth	.641	-.166	.194
Topographical wetness	.472	.221	-.243

Extraction Method: principal component analysis

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.594	29.949	29.949	3.116	25.969	25.969	2.741	22.840	22.840
2	1.824	15.202	45.151	1.233	10.274	36.243	1.501	12.512	35.352
3	1.583	13.195	58.346	1.145	9.538	45.780	1.251	10.428	45.780
4	.948	7.899	66.245						
5	.875	7.289	73.534						
6	.733	6.104	79.638						
7	.710	5.921	85.559						
8	.514	4.286	89.844						
9	.395	3.288	93.133						
10	.330	2.748	95.880						
11	.254	2.114	97.995						
12	.241	2.005	100.000						

Extraction principal component analysis

Scree Plot
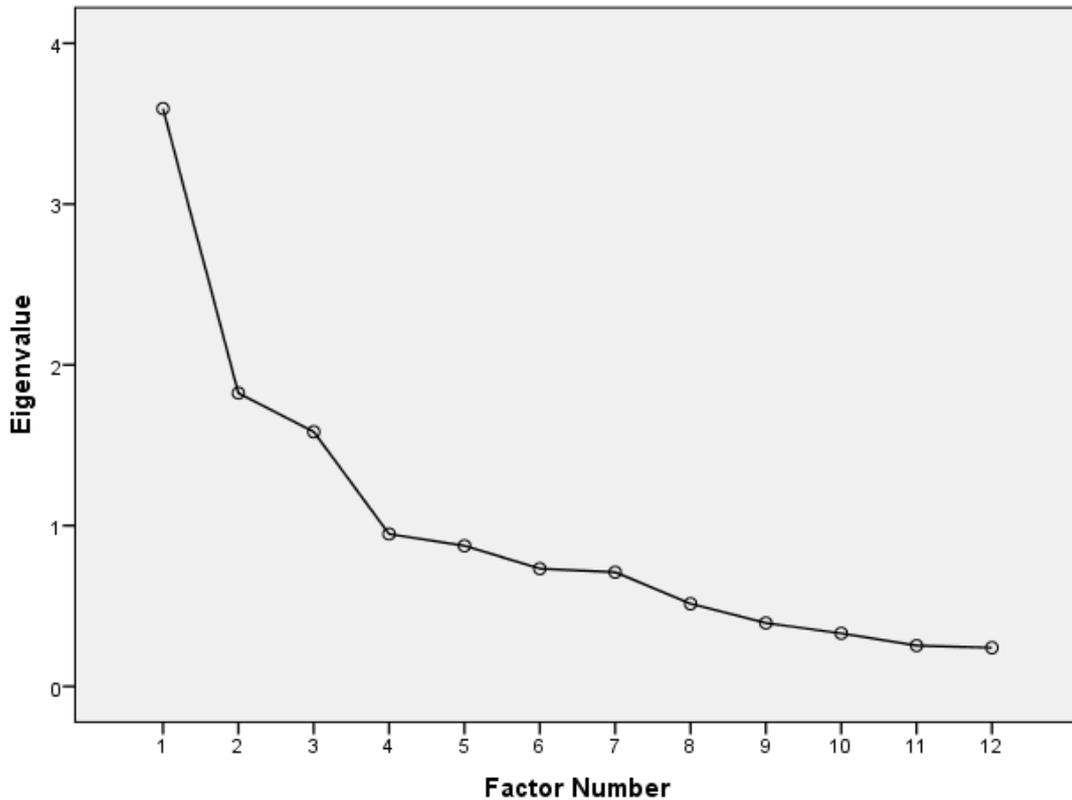
Appendix 14, linear regression model results.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.339[a]	.115	.115	.09874003617645
2	.339[b]	.115	.115	.09874024061344

a. Predictors: (Constant), Poverty, Evapotranspiration, Distance to wetlands, Slope, Land cover, Distance to hyacinth, Soil drainage, topographical wetness , rainfall, LST, ndvi, Altitude

b. Predictors: (Constant), Poverty, Evapotranspiration, Distance to wetlands, Land cover, Distance to hyacinth , Soil drainage, topographical wetness, rainfall1, LST, ndvi, Altitude