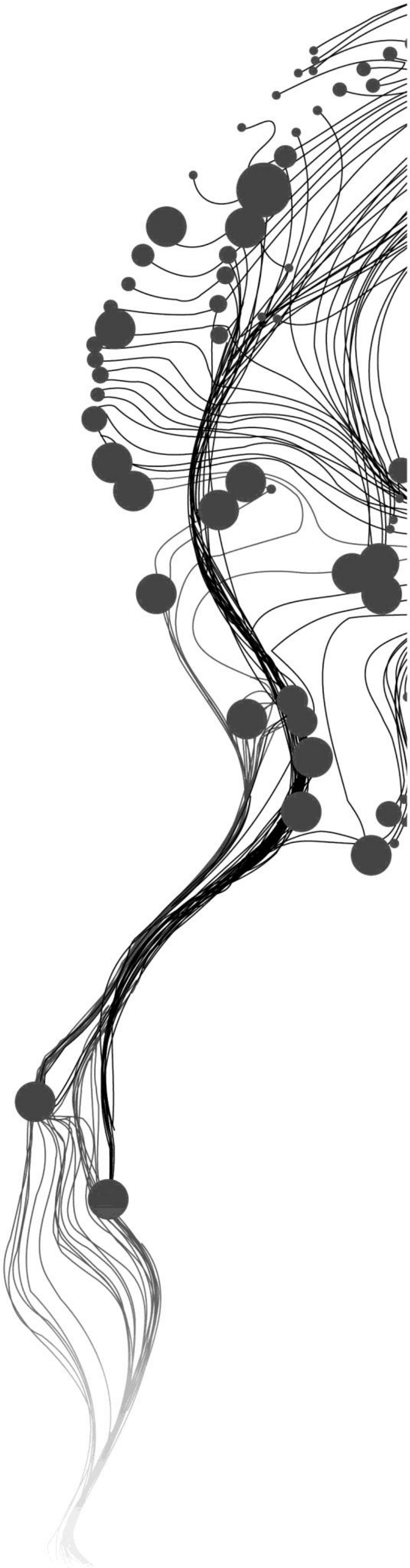
# SPATIAL MODELLING OF MALARIA RISK FACTORS IN RUHUHA SECTOR, RWANDA

JOSEPH TUYISHIMIRE
February 2013

SUPERVISORS:
Drs. J. Looijen
Dr. S. Amer



# SPATIAL MODELLING OF MALARIA RISK FACTORS IN RUHUHA SECTOR, RWANDA

JOSEPH TUYISHIMIRE
Enschede, The Netherlands, January, 2013

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Drs. J. Looijen
Dr. S. Amer

THESIS ASSESSMENT BOARD:
Dr.Ir. C.A.J.M. de Bie (Chair)
Mevr. Dr. P. Mens (External Examiner, KIT Amsterdam)

# ABSTRACT

Malaria is a vector borne disease that is a risk in Sub Saharan Africa and particularly in the East African Valley. Billions of dollars have been invested in the development of its control measures, but it is still killing hundreds of thousands of people per year in developing countries.

Malaria control requires the analysis of the habitat of its vector (*Anopheles Sp.*). Through the combination of high-resolution orthophotograph and fieldwork, malaria vector habitat was identified in one rural Sector of Ruhuha in Rwanda.

Malaria prevalence showed a variation from one household to another and from one administrative unit to another. Malaria clusters were determined using Getis and Ord statistics. This cluster analysis showed that malaria distribution is characterized by zones with high malaria risk, so called hot spots, zones with moderate malaria risk known as not significant spots and zones of low malaria risk known as cold spots.

Malaria causing factors in the study area are classified in three groups: environmental (altitude, proximity to anopheles mosquito breeding sites and land use types), demographic (household size, age, gender, and proximity to household with infected people) and economic (animal ownership and house material).

Malaria control measures that are used in the study area consist of three approaches: natural, artificial and treatment. Natural control measures deal with the removal of potential anopheles mosquito breeding sites. Artificial control measures that are dominated by Insecticide-Treated Nets and consist in breaking the vector-host contact and malaria treatment, which consists in healing infected people.

The relationship between malaria prevalence and malaria causing factors was assessed using logistic regression. The logistic regression proved an increase of malaria infection with a number of environmental, demographic, and economic factors. Malaria infection increases with the proximity to irrigated farmland and to households with infected people. It increases also with household size. It was proved higher in houses made of mud compared to unburnt brick walls. Malaria becomes lower when house walls are made of burnt bricks or cement blocks. If the floor is considered, malaria is higher in earth floor houses compared to cement floor house.

To assess possible differences between malaria clusters, ANOVA and chi-square tests were applied. The ANOVA test proved that the proximity to irrigated farmland and the distance to household with infected people define malaria hot spots. Chi-square test showed significant difference between clusters when house walls and floors are considered.

This study helped to identify and map malaria causing factors in Ruhuha Sector through the use of remote sensing, GIS, and spatial statistics.

**Key words**: Spatial Modelling, Malaria risk, Anopheles habitat, malaria causing factors, malaria control measures, Ruhuha Sector in Rwanda

# ACKNOWLEDGEMENTS

My first sincere gratitude goes to the Almighty God who gave me life and strength during my stay in the Netherlands. Thank you Lord!

I would also like to express my sincere gratitude to the people who helped me in some ways or another during this research process. My sincere thanks are addressed:

To my late father, Martin Ntahobari, to my late mother, Rose Nyirarusuhuko and to my brothers and sisters for their kind love, education and moral support from my childhood, during this research and evermore.

To my supervisors Drs. J. Looijen and Dr. S. Amer for their guidance and supports throughout my research thesis. Their professional experiences and practical comments; critiques and advice inspired me to complete this research.

To the Royal Tropical Institute (KIT) especially Mevr. Dr. P. Mens for the willingness to provide data and Dr. Ente Rood for the technical support during this research.

To Rwanda Biomedical Center/Malaria Research Program especially Mr. Frederic Kateera, Mr. Emmanuel Hakizimana and Mr. Parfait Karinda for the data provision and advice and support during the field work.

To Community Health Workers in Ruhuha Sector for their friendship and technical support during the fieldwork.

To the Netherlands Government through NUFFIC funding (NPT/RWA/071) and ITC for providing me with the opportunity to pursue my Msc studies;

To the National University of Rwanda (NUR) and the Centre for Geographic Information Systems and Remote Sensing of NUR for study leave permission and all other support during my studies;

To my friends, members of ITC-International Christian Fellowship and classmates with whom we shared the good and hard moments of school life.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ABBREVIATIONS

| | |
|---|---|
| ANOVA | Analysis of Variance |
| CGIS-NUR | Center for Geographic Information System and Remote Sensing of the National University of Rwanda |
| CHWs | Community Health Workers |
| GIS | Geographic Information Systems |
| ITNs | Insecticide Treated Nets |
| IWD | Inverse Weighting Distance |
| LLINs | Long Lasting Insecticide-treated Nets |
| IRS | Indoor Residual Spray |
| KIT | Royal Tropical Institute (Koninklijk Instituut voor de Tropen) |
| MINALOC | Ministry of Local Governance |
| MMI | Military Mutual Insurance |
| NISR | National Institute of Statistics of Rwanda |
| RAMA | La Rwandaise d' Assurance Maladie |
| RBC/MRC | Rwanda Biomedical Center/ Medical Research Center |
| REMA | Rwanda Environment Management Authority |
| RNRA/DLM | Rwanda Natural Resources Authority/Department of Land and Mapping |
| WHO | World Health Organization |

# 1. INTRODUCTION

Malaria is a disease that is threatening most tropical countries especially Sub-Saharan Africa. It is a result of many factors varying from environmental to socio-economic and therefore, its control requires a multilateral intervention. This chapter gives a brief background of malaria disease and defines some epidemiological concepts. It will also show the motivation of the research while defining objectives and hypotheses of the study.

## 1.1. Background

Malaria is a mosquito-borne infectious disease of humans and other animals caused by protists of the genus Plasmodium which are introduced into the circulatory system by the bite from an infected female Anopheles mosquito (Mwangangi et al., 2013). Human malaria is one of the most widespread vector borne diseases that is endemic throughout the tropical and subtropical regions of the World especially around the Equator (Ahmad et al., 2011).

Many studies about malaria were done in different parts of the World. Most of them tried to investigate the ecology of the vector (Nagasaki University, 2007; Smith et al., 2013), efficiency of control measures (Karema et al., 2012; Mackinnon, 2005) and the spatial disease modelling (Hay et al., 2006; Hay et al., 1998; Machault et al., 2011).

Most of researches described the interplay of biophysical, social and economic susceptibilities that define malaria risk (Stratton et al., 2008; WHO, 2008). Land-use patterns and agricultural practices combined with the climate and economic driven ecosystems change, are human actions that have favoured the breeding of malaria vectors, exposed populations to infection and facilitated the movement of malaria parasites (Packard, 2007). Indeed, the combination of different factors mentioned above has favoured the life cycle of the vector and its contact with the host (Stratton et al., 2008).

The next section defines key concepts used in this study.

### 1.1.1. Anopheles mosquito habitat

A species habitat is the location or environment where that organism is most likely to be naturally found (Plantinga et al., 2014). In fact, Anopheles mosquito suitable habitat is a location where biophysical conditions are adequate for its life cycle.

### 1.1.2. Malaria infection

Infection is defined by Autio et al. (2013) as "the invasion of a host organism's bodily tissues by disease causing organisms, their multiplication, and the reaction of host tissues to these organisms and the toxins they produce". Therefore, malaria infection is the invasion of the host organism by its causing pathogens (*Plasmodium sp.*) and their multiplication in blood cells.

### 1.1.3. Malaria prevalence

In disease modelling, the prevalence, or prevalence proportion is the proportion of a population found to have a disease or its condition. It is calculated by dividing the number of people found with disease or its condition with the total number of the study population, and is usually expressed as a fraction, as a percentage or as the number of cases per 1000, 10000 or 100000 people. Rothman (2012) suggests that malaria prevalence can be calculated at a specific point in time (Point prevalence) or over a given interval of time (period prevalence).

#### 1.1.4. Malaria vulnerability

Vulnerability is the susceptibility or weakness often associated with a particular situation such as illness, economy, age and gender (Stephenson et al., 2014). Malaria vulnerability is the susceptibility to be affected by its causal agent (*Plasmodium sp.*). According to Stratton et al. (2008), malaria vulnerability is influenced by demographic characteristics (population density, ages and sex) and socio-economic conditions (poverty and human behaviour). Malaria vulnerability is reduced by the advancement of control measures.

#### 1.1.5. Malaria Risk

In disease modelling, a risk is defined as the chance or likelihood that an undesirable event or effect will occur as a result of use or nonuse, incidence, or influence of a chemical, physical, or biologic agent, especially during a stated period of time; in other words, it is the probability of developing a given disease over a specified time period (The free dictionary, 2013). Malaria risk is the probability that an individual will be attacked by malaria in a given interval of time and in a known area. It increases with the number of people.

### 1.2. Research motivation

Malaria is a burden in Sub-Saharan Africa. Different measures for its control such as ITNs, IRS and antimalarial drugs are the most used and were proved relatively efficient (Karema et al., 2012; NISR, 2012a). However, some studies revealed that each of those measures were associated with behavioural and resistance challenges (Stratton et al., 2008).

To make control measures efficient requires an understanding of the disease spatial pattern, as successfully proven for the first time during the fifties. Snow (1855) was able to identify the source of cholera outbreak in Broad Street, London using GIS and the information was useful to the government that took the decision of blocking the contaminated water pump. From that time, many studies were carried all over the World in the domain of spatial modelling of diseases (Machault et al., 2011; Stevens & Pfeiffer, 2011).

The advancement of remote sensing allows to capture different biophysical variables (temperature, rainfall, and humidity, land cover, etc.) for different spatial and temporal scales. That was very important for many epidemiological studies (Anamzui-Ya, 2012; Machault et al., 2011; Munga et al., 2006). Satellite images are efficient tools for the detection of environmental factors associated with malaria risk. The integration of remotely and field acquired malaria related data (causing factors, control measures and incidence) through GIS and spatial statistics can provide decision makers in the domain of public health reliable and up to date information that is needed to detect and manage malaria disease.

It is essential to analyze malaria underlying factors to understand Malaria risk and elaborate relevant and efficient control measures. Along this line, several studies were able to model Malaria spatial patterns using GIS and Remote Sensing (Bizimana et al., 2009; Hassan et al., 2013; Konradsen et al., 1998; Machault et al., 2011). Few studies are available in the rural areas of Rwanda, which is prone to malaria.

## 1.3. Research justification

Environmental variables underlying anopheles mosquito habitat have been identified several decades ago (Bizimana et al., 2009; Jung, 2001; Lindsay & Martins, 1998; WHO, 1982; Wielgosz et al., 2012). In their review on malaria in the highlands of Africa, Wielgosz et al. (2012) classified irrigated agriculture among the factors that determine the habitat of malaria vector in Eastern Africa.

In Rwanda, studies about malaria targeted particularly malaria control policies and programmes. Thus, the focus was on the contribution of anti malarial drugs, mosquito nets and residual sprays (Gascon et al., 1986; Karema et al., 2012; Karema et al., 2006; NISR, 2012a; President's Malaria Initiative, 2013; Zeile et al., 2012), the prevalence of malaria in pregnant HIV positive women (Ivan et al., 2012) and the resistance of *Plasmodium falciparum* to antimalarial drugs (Gascon et al., 1986; Karema et al., 2012).

Aerial high-resolution photographs have been useful while identifying and delineating environmental variables such as land use in relation with a given disease. Thus, a detailed land use was important for epidemiological studies mainly targeting urbanized area (Anamzui-Ya, 2012; Bizimana et al., 2009; Hassan et al., 2013).

In Rwanda, the recent ortho-photographs produced for systematic land registration and Land Use Management Plan can allow further applications such as malaria risk modelling. Due to their high resolution (25cm), their use could be a fast and an accurate way to identify potential anopheles mosquito breeding sites and habitats.

In nature, diseases are not evenly distributed. Malaria is characterized by areas with high prevalence known as hot spots and areas of low prevalence known as cold spots. Each of the malaria distribution patterns is associated with a number of malaria causing factors. Getis and Ord hot spots analysis is a useful geostatistical technique that can help in malaria clusters identification. However, this approach is hardly used in epidemiological studies especially in malaria modelling in the East African region and worldwide.

The information gathered through the combination of GIS tools and spatial statistics analysis, could allow the spatial analysis of malaria underlying factors and to show their relationship with malaria prevalence at small scale within the rural administrative Sector of Ruhuha.

## 1.4. Research objectives

### 1.4.1. General objective

The aim of this study is to spatially model malaria risk factors in Ruhuha Sector. Factors that explain the occurrence of malaria in Ruhuha will be assessed using spatial modelling.

### 1.4.2. Specific objectives

1. To identify and map potential anopheles mosquito habitats in Ruhuha Sector.
   - ✓ Which anopheles mosquito habitats can be detected in the study area?
2. To determine and map malaria prevalence in Ruhuha Sector
   - ✓ Where are administrative areas with high malaria prevalence?
   - ✓ How is malaria prevalence clustered in the study area?
3. To identify factors underlying malaria risk in Ruhuha
   - ✓ What are the environmental, demographic, and economic variables underlying malaria risk in Ruhuha?
4. To determine malaria control measures in Ruhuha

- ✓ What are malaria control measures mostly applied in Ruhuha?

5. To determine the relationship between malaria prevalence and malaria risk factors
   - ✓ Is there any relationship between malaria prevalence and environmental, demographic, and economic variables and control measures?
   - ✓ What are causes underlying different malaria clusters?

### 1.5. Hypotheses

*Hypothesis 1:*

**$H_0$**: Malaria prevalence is uniformly distributed in the study area
**$H_1$**: Malaria prevalence is not uniformly distributed in the study area

*Hypothesis 2:*

**$H_0$**: Identified factors do not significantly explain malaria infection in Ruhuha
**$H_1$**: Identified factors significantly explain malaria infection in Ruhuha

*Hypothesis 3:*

**$H_0$**: Identified significant factors explain malaria infection with >50% correlation.
**$H_1$**: Identified significant factors explain malaria infection with <50% correlation.

*Hypothesis 4:*

**$H_0$**: Identified factors are not significantly different for all malaria clusters
**$H_1$**: Identified factors are significantly different for all malaria clusters

# 2. MALARIA BACKGROUND AND MODELLING

In this chapter, a great emphasis will be on malaria vector (anopheles mosquito) habitat suitability, malaria risk and its spatial modelling. These three components mentioned above will be put together to generate malaria conceptual framework on which this study will be based.

## 2.1. Anopheles mosquito habitat suitability

A good understanding of Anopheles mosquito requires understanding its ecology (Nagasaki University, 2007; Wielgosz et al., 2012; Zayeri et al., 2011). Depending on their ecological preferences, about 20 different *Anopheles* species are locally the most active *plasmodium sp.* vectors around the World and all of these species are operational during the night (WHO, 2013b). These anopheles species are vectors of the four parasite species that cause malaria in humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae* and *Plasmodium ovale*. *Plasmodium ovale* and *Plasmodium vivax* are the most common while *Plasmodium falciparum* is the most deadly (WHO, 2013b).

Anopheles mosquito breeds in water and each species has its own breeding preferences. For example, some prefer shallow collections of fresh water, such as puddles and rice fields (Munga et al., 2006; WHO, 2013b; Wielgosz et al., 2012). Irrigated farming increases nutrients and temperature which are favorable for the mosquito breeding and larvae survival (Wielgosz et al., 2012).

The biophysical (climatic and topographic) variables that can determine the regions with high endemicity have been object of different researches (Sipe & Dale, 2003; Zayeri et al., 2011). Anopheles mosquito proliferation depends on environmental factors like temperature, rainfall and humidity in association with vegetation cover and hydrology, especially water bodies (Sipe & Dale, 2003). Altitude is also an important factor and Anopheles mosquito prefers low altitude areas not only because they are characterized by high temperature and humidity especially in tropical regions but also because of their ability to retain water during and after rainy seasons (Fanello et al., 2007; Lindsay & Martins, 1998).

Certain man-induced environmental changes like deforestation, marshlands conversion and vegetation clearance for crop plantations favour ecological conditions that have a positive influence on the number and survival of Anopheles mosquito (Ahmad et al., 2011). These human activities have favoured anopheles mosquito either in creating breeding and resting sites or in favouring their contacts with humans (Verdonschot & Besse-Lototskaya, 2013). This has resulted in the predominance of three anopheles species (*Anopheles funestus* , *anopheles gambiae* and *anopheles arabiensis)* in the East African region where Rwanda is located (Mwangangi et al., 2013).

Anopheles mosquito proliferation requires the abundance of blood-meals and therefore, a travel distance is required from breeding sites to households where the vector-host contact becomes possible (Stoler et al., 2009). Verdonschot and Besse-Lototskaya (2013) reviewed articles about anopheles mosquito flight distance and found that the average flight distance for *anopheles sp.* was around 1000 m. However, the flight distance depends on the habits of the species and some species have a stronger dispersal capacity than others.

People near breeding sites (less than the average mosquito flight distance) are assumed to be at high risk of malaria while those beyond the average anopheles flight distance are less likely to be attacked by malaria (Liu et al., 2011). During their study in Accra, Stoler et al. (2009) found that malaria risk was higher within 1000 m from the vector breeding sites. However, the dispersal capacity of the mosquito also depends on the atmospheric conditions especially wind direction and the land use (Verdonschot & Besse-Lototskaya, 2013).

## 2.2. Malaria risk

Malaria is one of the most dangerous vector borne diseases in the World (Stratton et al., 2008). For example, an estimated 3.3 billion people were at risk of malaria in 2006 and the 1.2 billion at high risk (1 case per 1000 population) were living in the Sub-Saharan Africa (WHO, 2008). During the same year, an estimated 881 000 malaria deaths were reported, of which 91% were in Africa and 85% were of children under 5 years of age (WHO, 2008).

The rate of Malaria transmission is higher in areas where the mosquito lifespan is longer and where it prefers to feed on humans rather than other animals. The long lifespan and human blood meal preference the African anopheles species especially *anopheles arabiensis* explains the reason why more than 90% of the World's malaria deaths occur in Africa (WHO, 2013a).

In addition to naturally occurring ecological factors above mentioned, Donnelly et al. (2005) and Hassan et al. (2013) identified poverty, urban farming, deteriorating infrastructure and overcrowding in Sub-Saharan African urban areas as contributing factors to the development of conditions that modify anopheline mosquito habitats. The habitat style where people live near wetlands and water bodies, urban agriculture and the poor living conditions in these regions favour not only the breeding of the vector but also the vector-host contact (Smith et al., 2013).

Malaria risk becomes higher in rural areas of developing countries (Donnelly et al., 2005). A large number of malaria causing factors including the proximity to the vector breeding sites, the inadequate use of control measures, low income, illiteracy, land use and the house material play a big role (Stratton et al., 2008; Yamamoto et al., 2010). Stratton et al. (2008) mentioned the multiplicity of malaria causing factors in rural areas as the main cause of its persistence as they are difficult to control at the same time.

Malaria occurrence is higher in low income people. WHO (2012) suggested that these people are characterized by low access to health care facilities and lack of financial means to pay for vector control technologies such as ITNs and IRS and anti-malarial drugs. The same group of people is characterized by bad quality of house material that favours their contact with anopheles mosquito (Yamamoto et al., 2010).

Human immunity is another important factor. In Malaria endemic areas, partial immunity is developed over years of exposure, and even if it never provides complete protection, it reduces the risk that malaria infection will cause severe disease to adults. Therefore, the level of vulnerability is negatively associated with age and most malaria deaths in Africa occur in young children. Pregnant women are also highly affected, whereas in areas with less transmission and low immunity, all age groups are equally vulnerable (WHO, 2010). Malaria vulnerability and coping capacity will be discussed in the next sections on Rwanda.

## 2.3. Malaria in Rwanda

The environment and socio-economic context of Rwanda has favoured the proliferation of three Anopheles species (*Anopheles gambiae, Anopheles arabiensis* and *Anopheles funestus*) which are vectors for two Plasmodium species (*Plasmodium falciparum* and *Plasmodium vivax*)(Malaria Atlas Project, 2013). The topography, the land use, the climate, and the seasonality are factors that influence the malaria prevalence in Rwanda. Bizimana et al. (2009) identified land cover (rice plantations, sugar cane plantation, papyrus and wetlands), flooding, canalization and inundation as the main factors underlying malaria distribution in fringe zones of Kigali City. According to the same authors, the high risk is related to the combination of environmental conditions and high population density with poor living conditions.

In Rwanda, other factors that influence malaria in the country include high human concentration such as boarding schools in proximity to marshlands, population movement from low transmission to high transmission areas; irrigation patterns especially for rice crops which are mostly practiced in the Eastern and Southern parts of the country and cross-border movement of people especially in the Eastern and South-Eastern parts of the country (President's Malaria Initiative, 2013).

In recent years, a number of death cases were reported. For example, 670 and 380 Malaria deaths were reported in 2010 and 2011 respectively (WHO, 2013b). The Eastern and Southern Provinces are highly affected by Malaria. Particularly, children under five years and women are highly affected. During the Rwanda Demographic and Health Survey of 2010, the NISR (2012a) found that the rates of Children with Plasmodium parasites were 3.4% and 1.4% while the proportions of women with the parasite were 1.6% and 1% in the Eastern and the Southern Provinces respectively. In 2011, Malaria was the cause of 8% of mortalities under 5 years and 6% of all mortalities above 5 years.

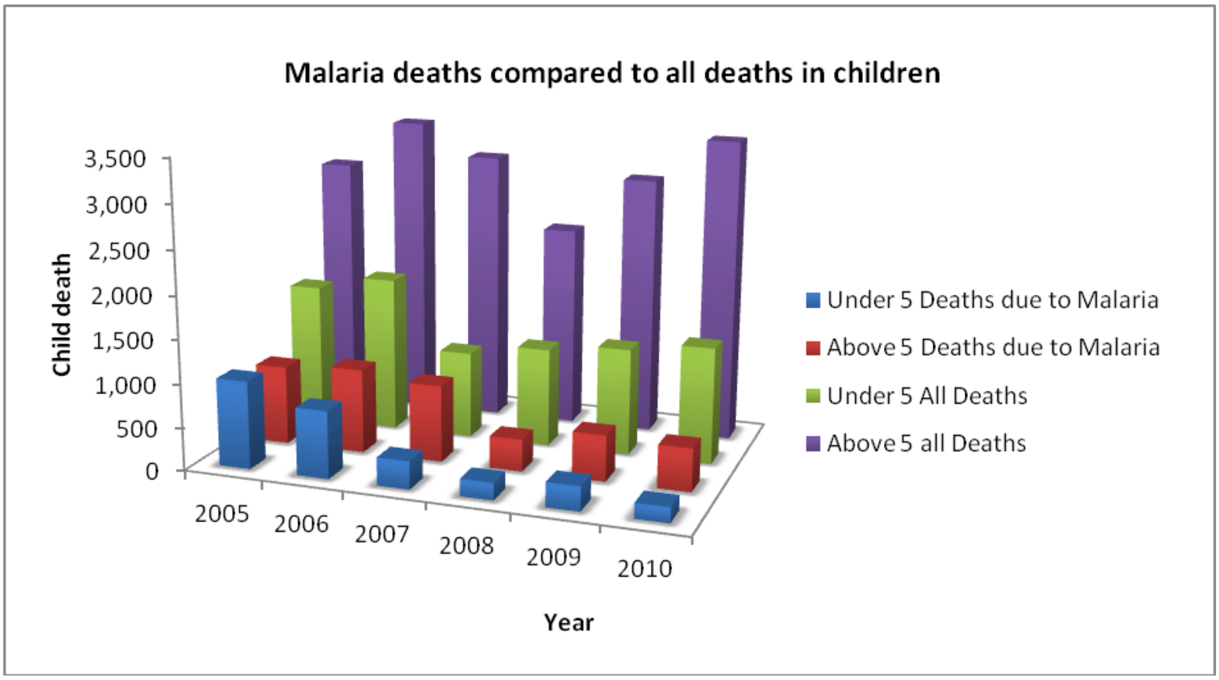



Figure 1. Malaria deaths compared to all deaths in children below 15 years (NISR, 2012b)

Figure 1 shows a decrease of malaria deaths in children from 2005-2010. This is attributed to the development of malaria control measures in the country (Karema et al., 2012). However, children above 5 years are still the most affected by malaria in Rwanda.

Malaria spread is also related to the level of wealth and to human behaviour, which differ from rural to urban regions. The NISR (2012a) suggested that the rate of Malaria is higher in rural than in urban areas.

This is related to the fact that, in addition to ecological conditions that limit the vector proliferation, people in cities have a higher revenue that can help them to pay treatment costs and access other malaria control technologies (WHO, 2012).

## 2.4. Malaria control measures in Rwanda

Malaria control has become a great concern for different governments and International Organizations and three main measures: ITNs, IRS and anti-malarial drugs have been object of the mobilization of billions of dollars (WHO, 2012). Nevertheless, each approach has revealed some weaknesses and they were criticized as being short time solutions (Stratton et al., 2008).

From the last two decades, the use of Insecticide-Treated Nets (ITNs) and Long-Lasting Insecticide-Treated Nets (LLINs) has been the most efficient malaria control measures in Rwanda. It has been incorporated in different visions and plans. The NISR (2012a) suggests that in 2010, 82% of households had at least one LLIN or ITN. The nets ownership was the highest in the Eastern Province (90% of households).

Residual sprays have also been used. However, they were qualified inefficient because anopheles mosquitoes have developed resistance to many types of insecticides and definite solutions to this resistance have not been found yet (Bigoga et al., 2012). The fact that this approach is used for the control of endophagic (eating indoors) and endophilic (resting indoors) mosquitoes has favoured exophagic (eating outdoors) and exophilic (resting outdoors) vector species.

As in many parts of the Sub-Saharan Africa, the parasite has developed resistance to different anti malarial drugs (Mackinnon, 2005; Zeile et al., 2012). For example, in 2001, the country changed anti-malarial treatment policy from chloroquine to amodiaquine and sulphadoxine-pyrimethamine (AQ+SP); five years later, the country shifted from AQ+SP to an artemisinin-based combination therapy (ACT), artemetherlumefantrine (Karema et al., 2012).

The combination of different Malaria control methods has been proved efficient (NISR, 2012a; President's Malaria Initiative, 2013). For example, between 2006 and 2012, these technologies contributed to the reduction of Malaria microscopically confirmed cases by 72% for all ages and by 82% for children below five years of age. Malaria death decreased by 47% for all age and 77% for Children below 5 years (Karema et al., 2012). However, the optimum has not been reached yet.

## 2.5. Malaria modelling

Modelling of malaria helps to describe the existing spatial patterns of the disease, to understand its causing factors especially the ecology of the vector and to predict for the future (Stevens & Pfeiffer, 2011). In their review, Stevens and Pfeiffer (2011) suggest that an adequate modelling of malaria must integrate the spatial and the classical statistics approaches.

The spatial modelling of malaria based on environmental variables such as temperature, rainfall, humidity and topographic variables especially altitude that determine anopheles mosquito habitat has been applied in different parts of the World (Gosoniu et al., 2009; Machault et al., 2011). Zayeri et al. (2011) were able to model malaria using environmental variables in Sistan and Baluchistan Province, Iran. The same authors used registered malaria data to compute malaria incidence rate.

Malaria prevalence is attributed to human activities especially agriculture and industrialization and behaviour such as immigration and the level of development of the disease control measures (Munga et al., 2006; Ngom & Siegmund, 2010; Smith et al., 2013). The interaction of the factors mentioned above contributes to the vector proliferation and its contact with humans as it was found by Stratton et al. (2008) in their review on the persistence of malaria.

In different parts of the World, many researches focusing on the process of malaria modelling were based on anopheles mosquito ecological and entomological data (Ahmad et al., 2011), while others were based on malaria incidence data (Bizimana et al., 2009; Grillet et al., 2010; Zeile et al., 2012). In whatever case, both studies were able to release scientifically useful information about malaria distribution in the areas of concern.

The process of malaria modelling combines classical and spatial statistics so that the integrated results could provide reliable and complete information that is useful for malaria control (Stevens & Pfeiffer, 2011). Different statistical models were used and proved efficient in the domain of epidemiological modelling. Those models are data-driven (presence-absence data or presence-only data) or knowledge-driven (multicriteria decision analysis). The following overview will focus on some of data driven models, which are mostly used in the domain of disease modelling.

#### 2.5.1. Regression models

Regression models are used when a dependent variable is explained by one or more predictors. They are divided into two categories:

a) *Linear regression*: It is useful when the relation between the event and its underlying factors is linear. In this case, only ratio data can be handled. Linear regression can be either univariate when the dependent variable is expressed by one explanatory variable or multivariate when the dependent variable is expressed by more than one independent variable (Crawley, 2007)

b) *Logistic regression model* proved useful when dealing with dichotomous data and has been used by different researchers, especially in the domain of disease modelling (Benndorf et al., 2011; Crawley, 2007; Yamamoto et al., 2010). This model is used when the disease occurrence is estimated using a set of malaria relevant causing factors, which can be ratio, nominal, or the combination of two.

The linear form of a regression model is

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \qquad 2.1$$

Which for logistic model, becomes

$$y = \log\left(\frac{\pi}{1-\pi}\right) \qquad 2.2$$

Where $\beta_0$ is the intercept, $\beta_1$, $\beta_2$, and $\beta_p$ are the coefficients of dependent variables and y is the independent variable.

If the coefficient of an independent variable is negative, it proves that this variable has a negative influence on the occurrence of the event; otherwise, it favours the event. For this property, regression models show not only the significance of each explanatory variable in the model, but also the direction of the relationship.

In a linear regression model, the coefficient of determination, $R^2$ helps to assess the estimates of the model. The higher the $R^2$ the better the model predicts the event while in logistic regression models; there are several ways of evaluating the estimated parameters. Below, three of them are discussed.

a. *Odds Ratio (OR):* It is the indicator of the probability resulting from the change in the predictor (see equation 2.3).

$$OR = \frac{\pi}{1-\pi} \qquad 2.3$$

Where $\pi$ is the probability of occurrence while $1-\pi$ is the probability of that the event could not occur. Odds ratio bigger than one (>1) indicates that as the value of the predictor increases, the probability of the event occurrence increases also (positive relationship). The odds ratio less than one (<1) indicates that as the predictor increases, the probability of occurrence decreases (negative relationship).

b. *Chi-square ($\chi^2$):*The chi-square is calculated

$$\chi^2 = \sum_{n} \frac{\left(O_i - E_i\right)^2}{E_i} \qquad 2.4$$

Where $O_i$ is the observed value, $E_i$ is the expected value for the $i^{th}$ observation and n is the number of cells.

To prove the significance of the independent variable in the model output, the probability value (p) is estimated. If it is less than the tolerance level ($p<0.05$ for most of epidemiological studies), there is a significant influence of the explanatory variable in the model output.

c. *Wald statistics*: This is useful when testing if a variable is a significant predictor in the model.

$$Wald(z) = \frac{\beta}{S_{E\beta}} \qquad 2.5$$

Where $\beta$ is the coefficient of the variable, and $S_{E\beta}$ is the standard error. If the coefficient is significantly different from 0, then, the explanatory variable contributes significantly to the model predictions.

Regression models have been applied in the domain of disease modelling in different parts of the Globe (Stevens & Pfeiffer, 2011; Stoler et al., 2009). In fact, logistic regression model has the advantage of dealing with dichotomous data, which cannot be handled by the linear regression. Both linear and logistic models rely on presence-absence data, are especially used in traditional statistical methods, and have been integrated with spatial models in the domain of disease modelling. Through this approach, Yamamoto et al. (2010) were able to assess household risk factors for clinical malaria in a semi-urban area of Burkina Faso while Stoler et al. (2009) were able to analyse distance threshold for the effect of urban agriculture on elevated self-reported malaria prevalence in Accra, Ghana.

### 2.5.2. Species Distribution Modelling (SDM)

Species Distribution Modelling is applied in the domain of spatial modelling of diseases and species modelling in general. This modelling approach is based on the collected presence data while taking the background as absence data (Stevens & Pfeiffer, 2011).

The SDM ,also known as Ecological Niche Factor Analysis (ENFA) or Diseases Distribution Modelling (DDM) in the domain of disease modelling, models the probability that a disease could appear in a given place (Stevens & Pfeiffer, 2011). This modelling approach is based on presence data and assumes that all the background is made of absence data. It uses the principle of Maximum entropy (MaxEnt) in which the expected value of each predictor under the estimated distribution matches its empirical average (Stevens & Pfeiffer, 2011).

MaxEnt is gaining credibility in epidemiological researches in different parts of the Globe (Stevens & Pfeiffer, 2011). Machado-Machado (2012) were able to map the suitability of dengue fever in Mexico. The same modelling approach was successively used by Chikerema et al. (2013) for the spatial analysis of *Bacillus anthracis* ecological niche in Zimbabwe.

Its advantage lies in the fact that it requires presence only data to make predictions from events (species or disease) observations and environmental layers that limit the species occurrence (Phillips et al., 2006). The second advantage is that it does not require a large set of point in representative data set for explanatory variables (Factors or constraints) to get accurate predictions (Chikerema et al., 2013). However, the assumption that the background is supposed to be absence data could put in doubt the reliability of the model output. This modelling approach is preferably used in spatial disease modelling when vector data are available.

#### 2.5.3. Hot spot analysis

Spatial statistics is a very useful tool in disease modelling but its application varies depending on the input data and expected output (Stevens & Pfeiffer, 2011). Many spatial models that have been developed were inspired by Tobler's law of spatial autocorrelation by which near things are more similar (Tobler, 1970). Based on this principle, models varying from simple hot spot analysis, simple interpolation and advanced simulations were developed.

Hot spots analysis is one of the most important objects of spatial modelling of diseases (Cromley & MCLafferty, 2012). It helps to detect zones of high risk in order to adapt appropriate intervention. However, their detection requires reliable approaches. Getis and Ord statistics (G*) is one of the approaches that are gaining credibility in the domain of hot spots analysis.

Getis and Ord is a local indicator of spatial autocorrelation, which uses the principle of spatial neighbouring to detect clusters of an event. The hot spot analyst tool calculates the Getis-Ord G* statistics for each feature in the input field. The resultant z-scores and p-values provide the information on whether features are spatially clustered or not.

To determine the statistical significance of spots, a feature will have a high value and be surrounded by other features with high values as well. A proportional comparison of the local sum for a feature and its neighbours is carried out and if there is a too high difference compared to the expected local value, a statistically significant z-score is assigned (Prasannakumar et al., 2011).

In accordance with Cromley and MCLafferty (2012), if **i** and **j** are point locations, $\mathbf{w_{ij}}$ is a spatial weight defining the nearness of point **j** to **i; $\mathbf{X_i}$** refers to the value of health indicator such as prevalence, the standardized $G^*_i$ statistics become

$$G_i^*(d) = \frac{\sum_j w_{ij}(d) X_j - w_i^* \bar{X}}{S\left\{\frac{\left[\left(nS_{ij}^*\right) - \left(w_i^*\right)^2\right]}{(n-1)}\right\}} \qquad 2.6$$

$$w_i^* = \sum_{ij}(d) \qquad 2.7$$

$$\bar{X} = \sum_j X_j \big/ n \qquad 2.8$$

$$s = \left\{ \sum \left( x_j - \bar{x} \right)^2 \right\}^{0.5} \qquad 2.9$$

For statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (hot spot) while for statistically significant negative z-scores, the smaller the z-score is, the more intense the clustering of low values (cold spot). The Gi_Bin is the easiest and quickest way for cluster identification as it shows different clusters at different confidence levels. In fact, the Gi_Bin varies between -3 and 3 where -3 represents the cold spot at 99% confidence, -2 represents the cold spot at 95% confidence, -1 represents the cold spot at 90% confidence, 0 is not significant while 1, 2 and 3 represent hot spot at 90%, 95% and 99% confidence levels respectively.

### 2.6. Malaria conceptual framework

The malaria conceptual framework used in this study is composed of five systems: (1) environmental variables, (2) anopheles mosquito (vector) life cycle, (3) parasite (*Plasmodium Sp.*) life cycle, (5) demographic and socioeconomic characteristics and (5) malaria control measures (see figure 2).

1. Environmental variables: These variables are grouped under three factors: climatic (temperature, rainfall, humidity) and topographic (altitude and slope) which are considered as the main determinants of anopheles mosquito habitat suitability at large (regional, country or global) scale. The third factor is ecology (hydrology, wetlands , vegetation cover and soil) which is mostly influenced by human activities and is an important determinant for anopheles mosquito breeding sites (Machault et al., 2011; Nagasaki University, 2007). It is the most important determinant of anopheles mosquito habitat at local level where there are small variations in climatic and topographic factors.

2. Vector life cycle: The vector life cycle is characterized by the breeding and the mortality, which determine its density. Control measures reduce the breeding and increases the mortality rate, which result in low densities.

3. The parasite cycle: The parasite cycle starts when humans are bitten by infected female anopheles mosquito. Incidence is detected when the parasite multiply quickly in non-immune people. If the parasite resists to anti-malarial drugs, it leads to death.

4. Demographic and socioeconomic characteristics: The level of vulnerability is associated with the demographic variables (household size, age and gender), socioeconomic variables (revenue, education, immigration) which has an impact on control measures resulting in the risk control. In whatever case, human behaviour has positive effect (creating ecological conditions that favour the vector) and negative effects (control measures) on malaria propagation.

5. Malaria control: It consists in vector control and in parasite control after human infection.
   The vector control consists of two methods: The natural control, which deals with environment (water ponds, wastes, and marshlands) management and land use planning (agriculture and settlement), and artificial control (use of ITNs and IRS). The effects of these technologies consist either in reducing the vector density or in breaking the vector-host contact (Killeen & Smith, 2007).
   The treatment with anti-malarial drugs is the only way to deal with the disease. However, not all treatments are successful because sometimes the parasite resists to drugs and still cause death.

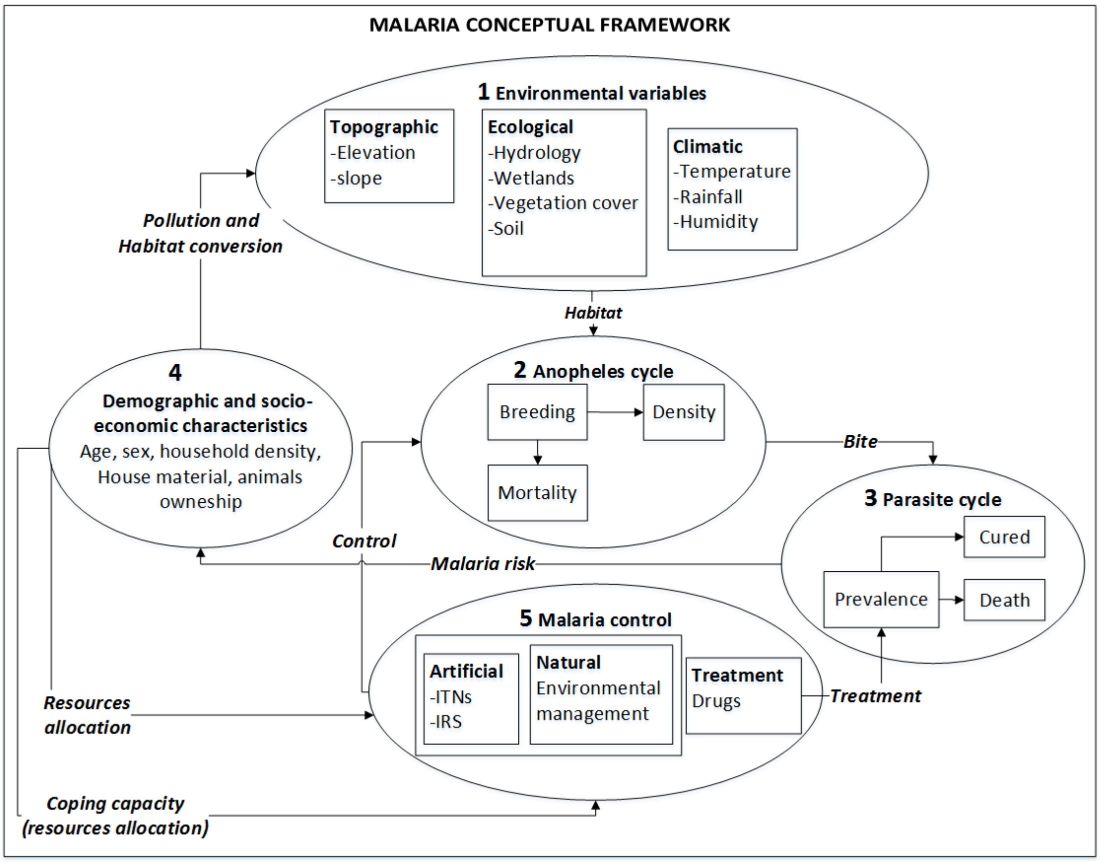



Figure 2. Malaria conceptual framework

The use of different malaria control methods does not provide definite solutions to the problem of malaria persistence. Artificial malaria control measures (ITNs and LLINs, IRS and anti-malarial drugs) which are respectively associated with behavioural, environmental and resistance challenges are the most used technologies in Rwanda. Natural vector control measures, which are environmentally healthy, are also successively used.

This study will focus on environmental, demographic, socioeconomic factors and control measures that are the main determinants of malaria prevalence at local scale of Ruhuha Sector.

## 2.7. Thesis outline

This thesis is subdivided into six chapters: introduction, literature review on malaria background and modelling, material and methods, results and discussion, and conclusions and recommendations.

*Chapter 1. Introduction*

In this first chapter, the research background is given, the key concepts are defined and objectives and research questions, formulated.

*Chapter 2. Malaria background and modelling*

This chapter gives the theoretical information about malaria causing factors, distribution, risk and spatial modelling. Malaria conceptual framework, which links the literature to the aim of the study, is produced.

*Chapter 3. Materials and methods*

This chapter describes the study area, focusing especially on the demography, climate, topography, socio-economy, and malaria control measures. Methods for data collection, data specifications, and data analysis are described. The softwares used are also mentioned.

*Chapter 4. Results and discussion*

The output of the analysis is available in this chapter in form of tables and graphs. The results about malaria vector habitat, malaria prevalence, malaria causing factors and control measures are presented and discussed. This chapter ends with the analysis of the relation between malaria infection and malaria causing factors.

*Chapter 5. Conclusion and recommendations*

In this part of the thesis, conclusions about the findings are formulated while recommendations are made not only to enforce the contribution of the research to the reduction of malaria in the study area but also for future researches.

# 3. METHODOLOGY

This section describes the study area with an emphasis on demography, economy, environment, malaria prevalence, and malaria control measures. It will also give an overview of the data collection, integration, and analysis. It ends by a generalized workflow summarizing all the steps that were followed for data analysis.

## 3.1. Study area

This study was carried in Ruhuha Sector. In fact, as shown in Figure 3, Ruhuha is one of the 13 administrative Sectors that make the Bugesera District in the Eastern Province of Rwanda. With an area of 43.4 $km^2$, Ruhuha Sector is divided in 5 administrative Cells: Ruhuha, Gatanga, Bihari, Kindama, and Gikundamvura. Due to data availability, the first four cells made by 28 villages were the object of this study.

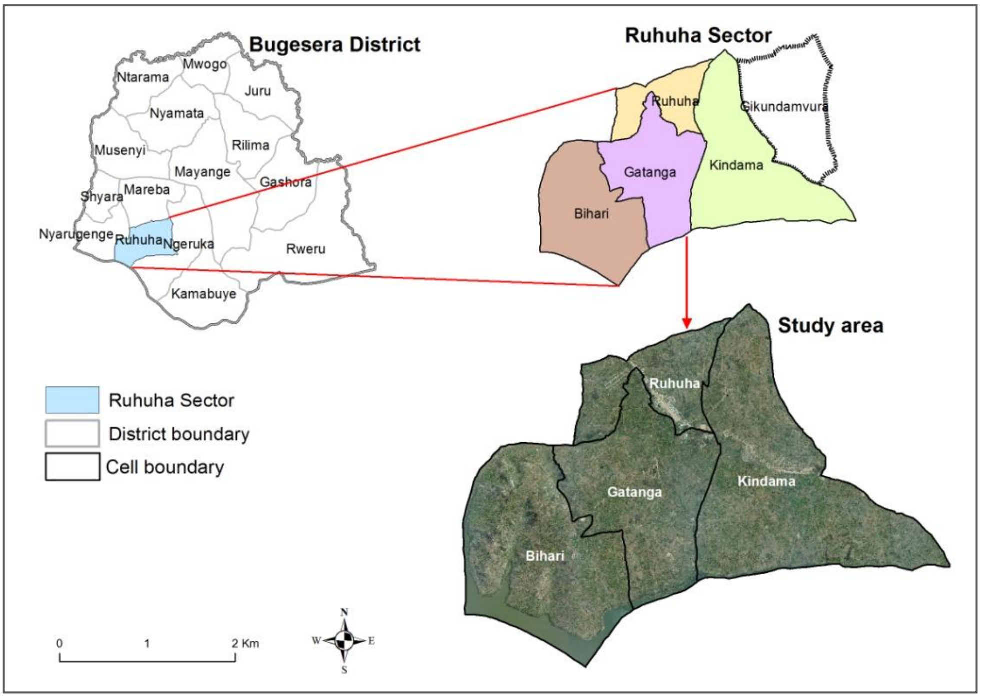



Figure 3. Study area

### 3.1.1. Demography

The study area was among the least populated regions of the country before 1994. Due to internal migration and to the return of Rwandese refugees to their homeland, it became very populated with a total population of 19504 inhabitants in 2010 as it is described by Figure 4 (NISR, 2010). The household density varies between 1 and 14 individuals.

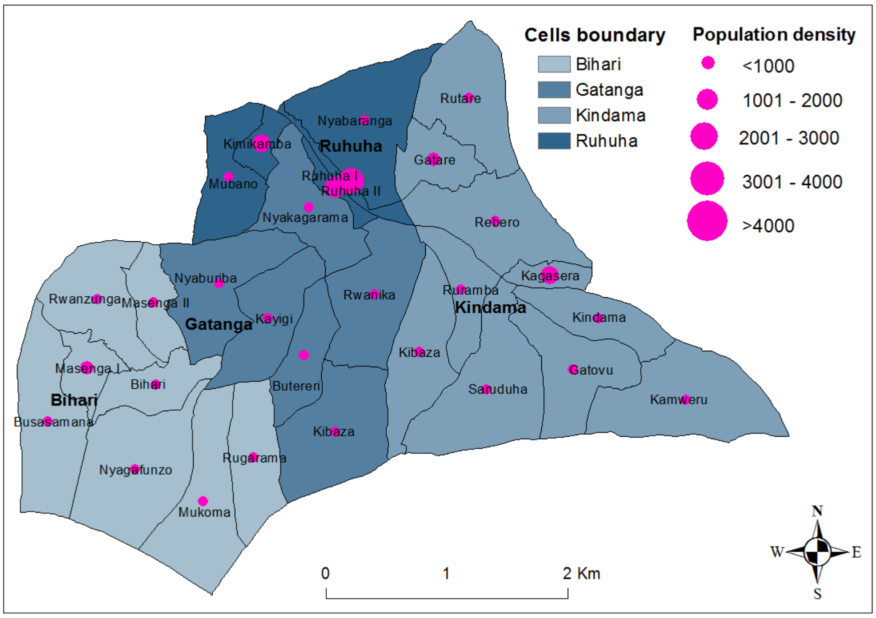



Figure 4. Population density

### 3.1.2. Economy

The population of the study area has agriculture as the main income generating activity. Domestic animals ownership is not developed and only about a half of surveyed households had domestic animals (cattle, poultry, sheep, goats, pigs and rabbits) while the other rural areas of the country count about 70% in term of domestic animals ownership (NISR, 2012a). Houses are predominantly made of un-burnt bricks (walls), iron sheets (roof) and clay on the floor (See figure 5).

### 3.1.3. Climate and topography

The study area is characterized by a low altitude varying between 1360-1520 m above sea level as described in figure 6. This low altitude compared to other regions of the country has a great influence on its monthly mean temperature varying between 20 and 30°C. The study area is characterized by the alternation of two rainy and two dry seasons.

Ruhuha Sector is characterized by a succession of low plateaux, hills, dry valleys, and swampy places. The latter are mostly exploited for agricultural activities or occupied by *papyrus sp.* and *Cyperus Sp.* (Especially near Cyohoha Lake). Those places have created a micro-environment that is favorable to vectors of diseases especially anopheles mosquito (Bugesera District, 2013).

Given the low difference of altitude (See Figure 6) which is consequently characterized by low differences in mean monthly rainfall and temperature, climatic variables that are important at large scale were assumed similar in the whole study area.

Figure 5. Predominant House materials

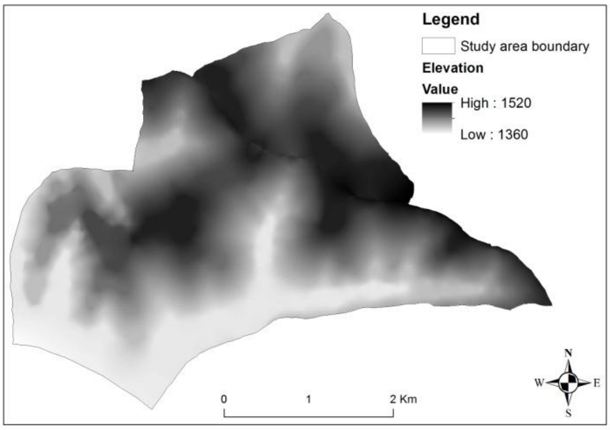



Figure 6. Elevation of the study area

#### 3.1.4. Malaria prevalence

Malaria is a serious danger in the study area and it is favoured by the environmental conditions in Ruhuha Sector and the Eastern Province of Rwanda in general (NISR, 2012a). It is the second cause of morbidity (23.3% of microscopically confirmed cases) after the sharp infections of the superior or lower respiratory ways which represent 27.3% (Bugesera District, 2013). Its effect could have become higher unless control measures had been taken.

#### 3.1.5. Malaria control measures.

The proliferation of anopheles mosquito in Bugesera District and particularly in Ruhuha Sector is a result of environmental, demographic, and economic factors. To face these malaria causing factors, three types of malaria control measures were adopted.

- Natural control measures which deal with the removal of potential anopheles breeding and resting sites near houses.
- Artificial control measures (mostly ITNs) consisting in breaking the vector-host contact
- Treatment, which consists in healing, infected people. In Ruhuha Sector, malaria infected people are treated at Ruhuha Health Center except children under five years that are regularly followed up by CHWs and treated at home when they are infected. If the infection persists, they are treated at the Health Center.

The next section explains the research design that includes data preparation, data collection and data analysis.

### 3.2. Data preparation

The potential land use of the study area was digitized from an orthophoto 25 cm resolution of August 2008 and 2009. The orthophoto was obtained through the stack of two layers, one produced in August 2008 and another produced in August 2009 for the national land management purpose. From the land use map, potential anopheles mosquito breeding sites and habitat were derived.

During the fieldwork, stratified random sampling was used. This approach consisted in collecting sample points within different land use types (strata) for the validation of the potential land use map and the identification of malaria vector breeding sites (see appendix 1). When leading to the locations of the vector breeding sites, sample points were located over different land use types

## 3.3. Data collection

This study was mainly based on secondary data consisting of malaria infection, demographic and economic data that were provided by the Rwanda Biomedical Center/Malaria Research program (RBC/MRC) and spatial data that were provided by the National Institute of Statistics of Rwanda (NISR). Primary data included identified anopheles mosquito breeding sites and land use. Table 1 shows all the collected data.

### 3.3.1. Secondary data

Secondary data were obtained from RBC/MRC in collaboration with the Royal Tropical Institute (KIT) through a research entitled "Empowering the Community towards Malaria Elimination". These data were collected from June to October 2013 by two teams: One made by socioeconomic surveyors who collected demographic and economic data and another made by laboratory technicians who were in charge of the malaria infection data collection. A code linking the results from the survey and from the laboratory for each household was created for data integration purposes. Another section of secondary data consisted of spatial information obtained from the NISR.

- *Demographic data*

Demographic data were collected per household and include household size, age and gender of all the members. The RBC surveyors visited each household used a questionnaire to fill in all demographic information that was provided by the head of the family. This questionnaire was supplied by RBC and can be presented on request.

- *Economic data*

The economic data that were collected consisted of animal ownership and household material. The economic data that were collected were also filled in the same questionnaire with the demographic data

- *Malaria control measures*

Information about malaria control measures consisting of artificial and natural approaches and treatment per household was also collected and filled in the same questionnaire.

The demographic, economic and malaria control measures data were directly recorded in tablets (Samsung Galaxy tabs). In addition, location (GPS coordinates) was recorded for each household. The collected data were then uploaded to the RBC / MRC server where they were combined from different villages and cleaned for further uses. 3232 households were surveyed.

- *Malaria infection data*

The blood samples of all individuals except children under 6 months in the surveyed households were collected on slides and malarial smear test was carried out at Ruhuha Health Center in accordance with malaria smear test guidelines (WHO, 2009). Individuals with parasites in the blood were qualified malaria positive while those without parasites were qualified malaria negative. The initials, age and the sex of all the tested individuals were recorded and marked in the laboratory result notebook and then in the RBC/MRC data register.

- *Spatial data*

Spatial data consisted of the shapefiles of the study area (sector, cells and village boundaries and water sources). They were all obtained from the National Institute of Statistics of Rwanda.

### 3.3.2. Primary data

The primary data collection was based on interviews with local key informers, visiting potential anopheles mosquito breeding sites, and validating the land use of the study area.

- *Interview with key informers*

Interviews were conducted with key informers especially the administrative cells Community Health Workers (CHWs) and staff from the Rwanda Biomedical Center (RBC) in charge of malaria control. Local people knowledge about anopheles mosquito breeding sites and habitat was gathered (see interview guidelines questions in appendix 2). Out of the interviews, irrigated farmlands and water reservoirs were identified as the main anopheles mosquito breeding sites in the study area.

- *Irrigated farmlands*

The study area has four marshlands (Nyaburiba, Nyabaranga, Nduhura, and Kibaza) where rice farming is developed. The households in the study area were close to marshlands and were separated by a distance varying between 200 and 2000 meters. Those kinds of irrigated farming (see figure 7) have also been confirmed by Nagasaki University (2007) as the main anopheles breeding sites in the western highlands of Kenya and Wielgosz et al. (2012) as the main anopheles mosquito breeding sites in the East Africa.

- *Water reservoirs*

The study area has been characterized by water shortage until the last two decades. Water reservoirs (see figure 8) were developed as a rainwater retention system that could help farmers to irrigate their crops the whole year. Most of those water reservoirs are close to households and were potentially favorable for anopheles mosquito life cycle as they were identified by Hassan et al. (2013) in Egypt and Ngom and Siegmund (2010) in Cameroun.

- *Land use*

The validation of the land use classes obtained from the orthophotograph was necessary to check whether the produced land use map was accurate enough for the further analyses. For the accuracy assessment, a set of 55 training points was collected during the fieldwork (see appendices 1 and 4).
With the Guidance of administrative cells CHWs, the coordinates and the elevation of each breeding site were marked using a hand-held Geographic Positioning System (GPS), GARMIN etrex 12 CHANNEL (Garmin Ltd, 2003) and the land use type was described (see appendix 5). The captured data were later imported via map source 6.16.3 software and processed in ArcGis 10.2 package.

Figure 7. Rice crop

Figure 8. Water reservoir

Table 1. Collected data

| *Data* | *Specification* | *Source* |
|---|---|---|
| Orthophoto | 25 cm resolution acquired in August 2008 and 2009<br>Projection: ITRF-2005 | RNRA/DLM |
| Study area shapefiles | Rwanda baseline maps 2010<br>Projection: ITRF-2005 | NISR |
| Household XY coordinates | GPS coordinates<br>Projection: WGS84 | RBC/MRC |
| Malaria infection data | Number of people with *Plasmodium* parasites per household | RBC/MRC |
| Population density | Household level | RBC/MRC |
| Population density | Cell and village levels | NISR |
| Age and sex of the population of study | Individual based | RBC/MRC |
| Animal ownership | Presence of animals in the household | RBC/MRC |
| House material | The material that makes the roof, walls and the floor | RBC/MRC |
| Use of control measures | Number of ITNs per Household, sprays, cutting bushes, clearing stagnant water and others | RBC/MRC |
| Anopheles micro-habitat in Ruhuha | GPS coordinates and elevation of Anopheles breeding sites<br>Projection: WGS84 | Field work |
| Land use of the study area | Training point for accuracy assessment<br>Projection: WGS84 | |

### 3.3.3. Data integration

The collected data were from three different sources: RBC/MRC, NISR and the fieldwork and therefore, before their analysis, they had to be integrated as summarized in Figure 9.

### 3.3.4. Data from RBC/MRC

The data from RBC consisted of laboratory results that were stored in laboratory notebook and in RBC/MRC general register and demographic, economic and malaria control measures that were stored on the online RBC/MRC server. A link code was created so that data from the three sources could be integrated.

- *Data from the RBC/MRC general register:* In this register, people to be tested for malaria infection were estimated at 10933 individuals. However, during the fieldwork it was realized that 2789 people were absent and 106 were children under 6 months (from whom blood samples were not collected for practical reasons). A total of 8144 people were recorded as tested for malaria infection, of which 287 were malaria positive and 7857 were malaria negative.

- *Data from the laboratory notebook:* From this notebook, 506 individuals among the tested for malaria infection were positive. However, only 487 individuals could be linked to their households for spatial analysis.
- *Households:* Locations of the surveyed households and their related demographic, economic and malaria control measures were downloaded from the RBC/MRC server.

The data from RBC general register were showing fewer malaria positive cases compared to the data from the laboratory notebook. This was because they were still entering the data and part of them was not entered yet and therefore, the laboratory notebook was more reliable. However, when the laboratory data were to be linked with household information, link codes were lacking for some of them. This was also attributed to the fact that some of the surveyed household information was not uploaded to the server yet due to technical problems. Recording errors could also result in false linking codes. In this study, only data that had a proper linking code were used for spatial analysis. Gender and sex were individual based and therefore, they were not included in spatial analysis.

#### 3.3.5. Data from NISR

These data consist of spatial information. They were linked with the data from RBC and the fieldwork through spatial join via the same link code created by RBC/MRC.

#### 3.3.6. Data from the field

Field data consisting of land use and anopheles mosquito breeding sites were linked to the household attributes from MRC and to administrative boundaries obtained from NISR

The spatial data were in different geographic projections (See Table 1). Before performing any data process, they were assigned the same local projection (TM_Rwanda) and the same coordinate system (GCS_ITRF_2005) through ArcGis 10.2 software.

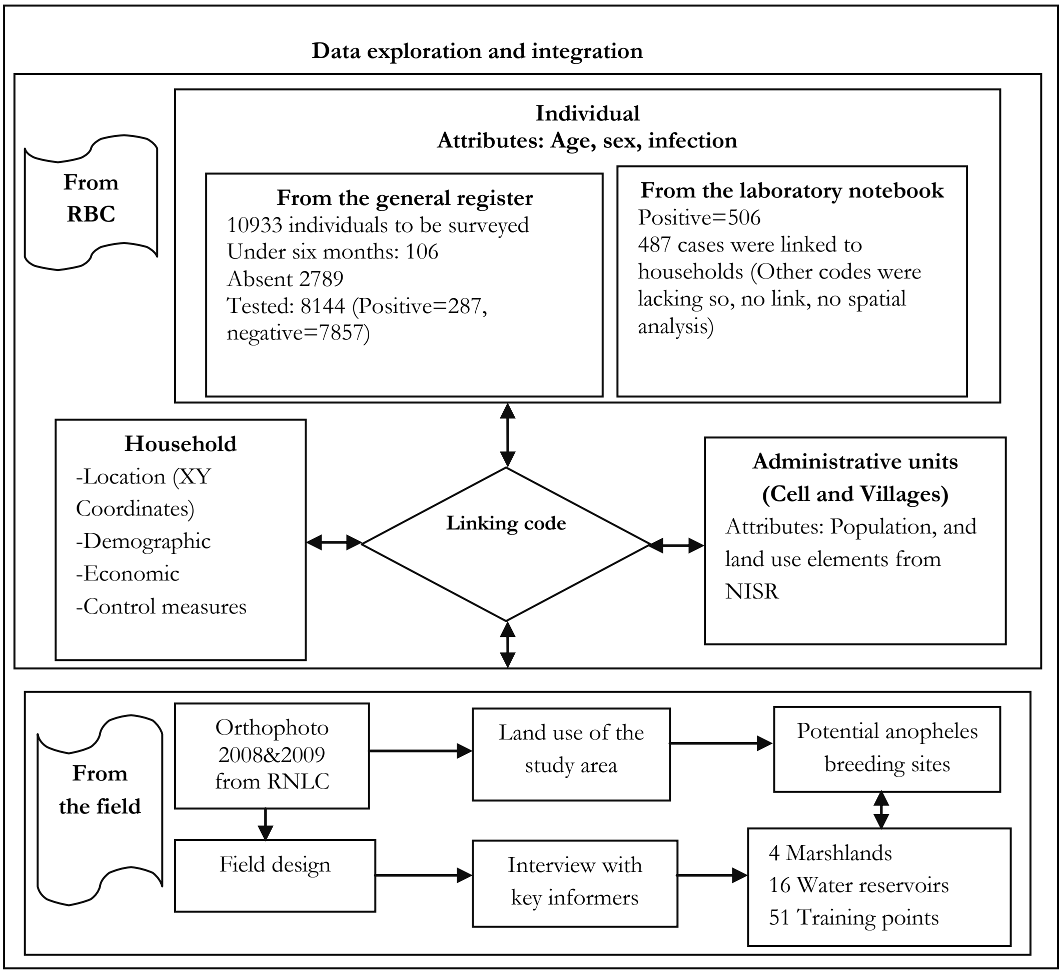



Figure 9. Data integration

## 3.4. Data analysis

In this section, the approaches for data analysis are presented in line with the objectives of this study. The way anopheles mosquito habitats, malaria prevalence, causing factors, control measures and their relationships were analyzed will be presented in more details in this section.

### 3.4.1. Potential anopheles mosquito habitat maps

The land use was delineated from an othophotograph of the study area. An accuracy assessment of the produced land use map was carried out to check its reliability for further analyses. Potential anopheles mosquito habitat map was created through the integration of the information extracted from the land use map and field observations (see figure 10).

- *Accuracy assessment*

The produced land use map was validated by a set of 55 training points collected from the field. The overall accuracy and Kappa statistics (K*) were calculated to measure the agreement between the land use types derived from the orthophotograph and the values observed from the field. Kappa statistics (K*) is an estimate to measure the agreement between the predicted value and the reference value. It varies between 0 (no agreement) and 1 (perfect agreement). A kappa Value above 0.80 represents a strong agreement, values between 0.40 and 0.80 represent a moderate agreement, while values below 0.40 represent a poor agreement (Phillips et al., 2006).

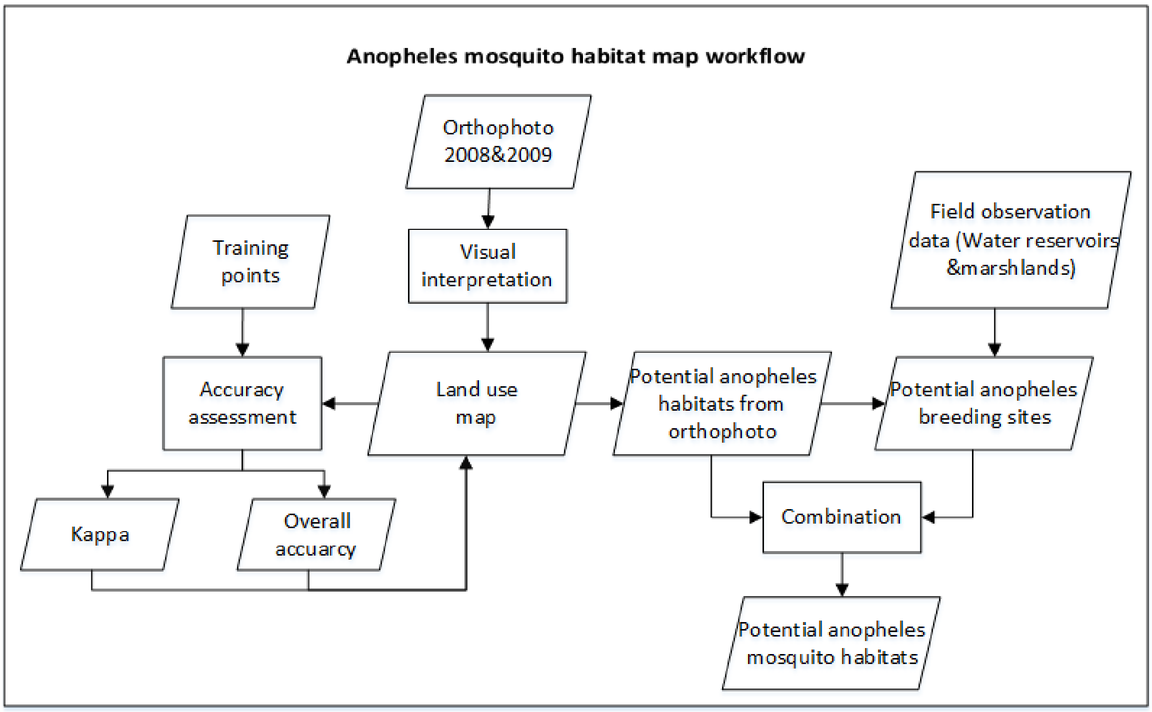



Figure 10. Flowchart for anopheles mosquito habitats map

### 3.4.2. Malaria prevalence and clustering

This section gives an overview about the way malaria prevalence and clustering was spatially analyzed. Malaria prevalence was analyzed at the cell and village levels. Getis and Ord (G*) index (Cromley & MCLafferty, 2012) was used to test the significance of malaria clusters within the surveyed households while IWD (Phillips et al., 2006) was used to interpolate malaria prevalence to the whole study area.

#### a. Malaria prevalence

Malaria prevalence was analyzed per household, per village and per cell level. It was considered as the ratio of the number of malaria positive cases over the household size (at the household level) or the total population (at village and cell levels) as it is shown in equation (3.1).

$$P = \frac{n_p}{n_t} \qquad 3.1$$

Where $P$ is malaria prevalence, $n_p$ is the number of malaria positive cases and $n_t$ is the total population.

#### b. Getis and Ord malaria clusters

Clustering malaria prevalence gives an idea about the spatial pattern of collected malaria prevalence data as malaria prevalence is assumed to be not homogeneous in all the surveyed areas (Cromley & MCLafferty, 2012). Hot Spot Analysis (Getis-Ord, Gi*) statistics with a threshold distance of 500 m and malaria prevalence per household as input field was used for malaria clusters detection (see figure 11). The threshold was chosen based on the mosquito flight distance. Stoler et al. (2009) found that malaria infection is higher between 200 m (minimum anopheles flight distance) and 1000 m (mean mosquito flight distance) having a high pick at 500 m. The way G* statistics works is described by the equation 2.6 in Section 2.5.

### c. IWD Malaria distribution map

Malaria clusters that were detected by G* statistics were point based. However, for planning purposes, it is important to know malaria distribution in the whole study area. The Inverse Distance Weighting (IWD) spatial analyst tool in ArcGis was used for this purpose with malaria prevalence as input field (see equation 3.2). IWD interpolates point values to surface raster (Watson & Philip, 1985) .

$$P_i = \frac{\sum_{j=1}^{G} \frac{P_j}{D_{ij}}}{\sum_{j=1}^{n} \frac{1}{D_j^n}} \qquad 3.2$$

Where $Pi$ is the property at location i; $P_j$ is the property at sampled location j; $Dij$ is the distance from i to j; $G$ is the number of sampled locations; and $n$ is the inverse-distance weighting power. The value of $n$, in effect, controls the region of influence of each of the sampled locations. As $n$ increases, the region of influence decreases until it becomes the point, which is closer to point $i$ than to any other. When $n$ is set equal to zero, the method is identical to simply averaging the sampled values (Watson & Philip, 1985).

The output is limited to the values of the interpolated field and its predictions cannot go beyond the maximum value. The output depends on whether the observations to interpolate are dense and sparse. The latter may lead to a weak simulation results.(Watson & Philip, 1985). This interpolation approach is appropriate for the type of data we had because households in the study area were close to each other.

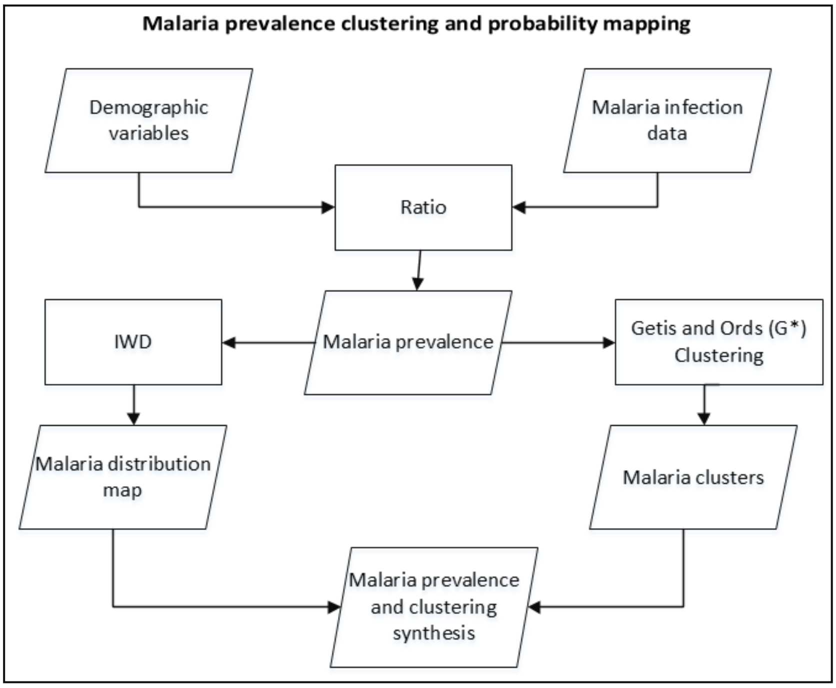



Figure 11. Flowchart for malaria prevalence and distribution mapping

### 3.4.3. Malaria causing factors

Malaria causing factors in the study area were determined from the literature and interviews with the local Community Health Works (CHWs) and the RBC/MRC staff in charge of malaria control. They were categorized in three groups: environmental, demographic, and economic. The identified malaria causing factors were either nominal or ratio (see table 10 in section 4.5.1). The latter consisted of the altitude and household size the distance to Anopheles breeding sites. To know the distance from each of the observed households to the nearest vector breeding, distance maps were created.

- *Distance map creation and normalization*

Many studies have revealed that areas that are close to anopheles mosquito breeding sites or to households with infected people are highly exposed to malaria (Gosoniu et al., 2009; Munga et al., 2006; Mutero et al., 2000). Accordingly, distance maps for water reservoirs, irrigated farmland, and households with infected people were created.

Data normalization enables to put all the continuous variables in the same range so that they can serve as inputs in models. A minimum-maximum linear transformation was used. By applying this algorithm, All the raster values were put in the same range between 0 and 1 (0-1) (Misaki et al., 2010). By using Arcmap, distance values were extracted and joined with households attributes so that they could be used in further analyses.

### 3.4.4. Malaria control measures in the study area

Malaria control measures were classified in natural, artificial and treatment. The latter was mostly represented by the distance to the Health Center. People that live far from health care facilities were identified the most vulnerable to malaria in most of Sub-Saharan African countries (Stratton et al., 2008). The distance map was created and normalized the same way as described in section 3.5.3.

### 3.4.5. Relationship between malaria infection and malaria causing factors

This section gives an overview of the analysis of the relationship between malaria prevalence and its causing factors. The potential relevant variables for malaria prediction were screened through the process of collinearity analysis (Correlation coefficients and VIFs calculation). The relationship between malaria and its causing factors is analyzed through a Chi-square ($\chi^2$) test for individual based variables (gender and age) and through a stepwise logistic regression for household based variables (see figure 12).

#### a. Collinearity analysis

Collinearity refers to the relationship of two variables which when included in the same regression model could cause information redundancy or multicollinearity. The first way of checking for collinearity is by running a correlation test between all pairs of independent variables, which will give a correlation matrix comparing all pairs of Pearson coefficients (r). In fact, if the independent variables are highly correlated (r close to -1 or +1), then collinearity becomes an issue. The correlated variables are likely to reflect the same reality, which will lead to biased results. The next step is the VIF analysis.

#### b. VIF analysis

During the process of collinearity analysis, correlation is complemented by Variance Inflation Factor (VIF) calculation (See equation 3.3).

$$\mathrm{VIF}_n = \frac{1}{1-R^2} \qquad 3.3$$

Where $R^2$ is the coefficient of determination for all the models, the last ($n^{th}$) explanatory variable excluded. The rule of thumb suggests that every variable with a VIF greater than 10 is excluded from the model (Lubetzky-Vilnai et al., 2013). However, based on the literature or expert knowledge a variable having a VIF higher than ten could not be excluded fro the model depending on its importance (Crawley, 2007).

**c. Regression model**

The logistic regression model as shown by the equation 2.2 in Section 2.5 was chosen for this study because it has the capacity predicting categorical variables by nominal and ratio variables at the same time. The stepwise logistic regression model will be used for testing the relationship between malaria infection (Presence of parasites in the blood=1, absence=0) and its causing factors and control measures. In fact, the stepwise regression helps to detect the model improvement if new predictors are added and is better for logistic regression with dummy variables (Field, 2009). Odds ratio (OR described by the equation 2.3 in section 2.5), Chi-square test ($\chi^2$ described by the equation 2.4 in section 2.5), Wald statistics (equation 2.5 in section 2.5) will be used to test the significance of the predicted coefficients (Lubetzky-Vilnai et al., 2013).

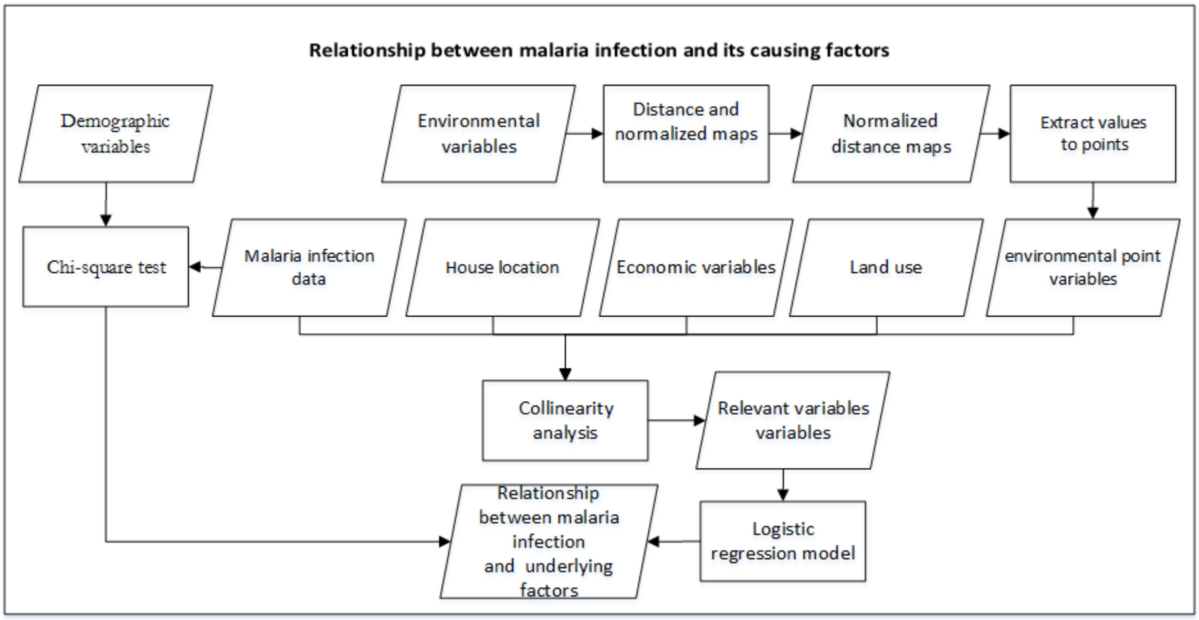



Figure 12. Analysis of the relationship between malaria infection and its underlying factors

### 3.4.6. Comparison between malaria clusters

Malaria clusters were compared to test their differences based on their causing factors and control measures. The hot spots were supposed to be at high risk compared to not significant and cold spots. One way ANOVA was used to test the mean difference among and between groups. When among group differences were detected, Tukey test was used to test between groups mean differences. Chi-square test (See equation 2.5 in section 2.5) was used for nominal variable differences testing among malaria clusters in the study area.

**a. ANOVA test**

Analysis of variance (ANOVA) is a collection of statistical models used to analyze the differences between group means and their associated variations (among and between groups)(Field, 2009). The ANOVA test uses the formula

$$F = \frac{MST}{MSE} \qquad 3.4$$

Where F is the statistics, testing the difference between means,

$$MST = \frac{SST}{n - k} \qquad 3.5$$

$$MSE = \frac{SSE}{k - 1} \qquad 3.6$$

Where

MST: Mean Square Total, SST: Sum Square Total, MSE: Mean Square Error; SSE: Sum Square Error, n: sample size, k: number of groups, n-k: degree of freedom of the Mean Total and k-1: degree of freedom of error.

When F is not statistically significant ($p>0.05$ for this study), there is a difference of means among groups. The Tukey test compares significantly different groups two by two.

**b. Chi-square test**

The Chi-square test is used to test differences between proportions. The chi-square test was used to test the difference in proportions of economic factors (house walls, house floors, and animal ownership) and malaria control measures among identified malaria clusters in the study area.

Figure 13 summarizes different steps for data analysis. It shows the ways malaria vector habitat was identified, steps for malaria prevalence analysis, the identification of malaria causing factors and control measures. It ends by the relation between malaria infection and its causing factors.

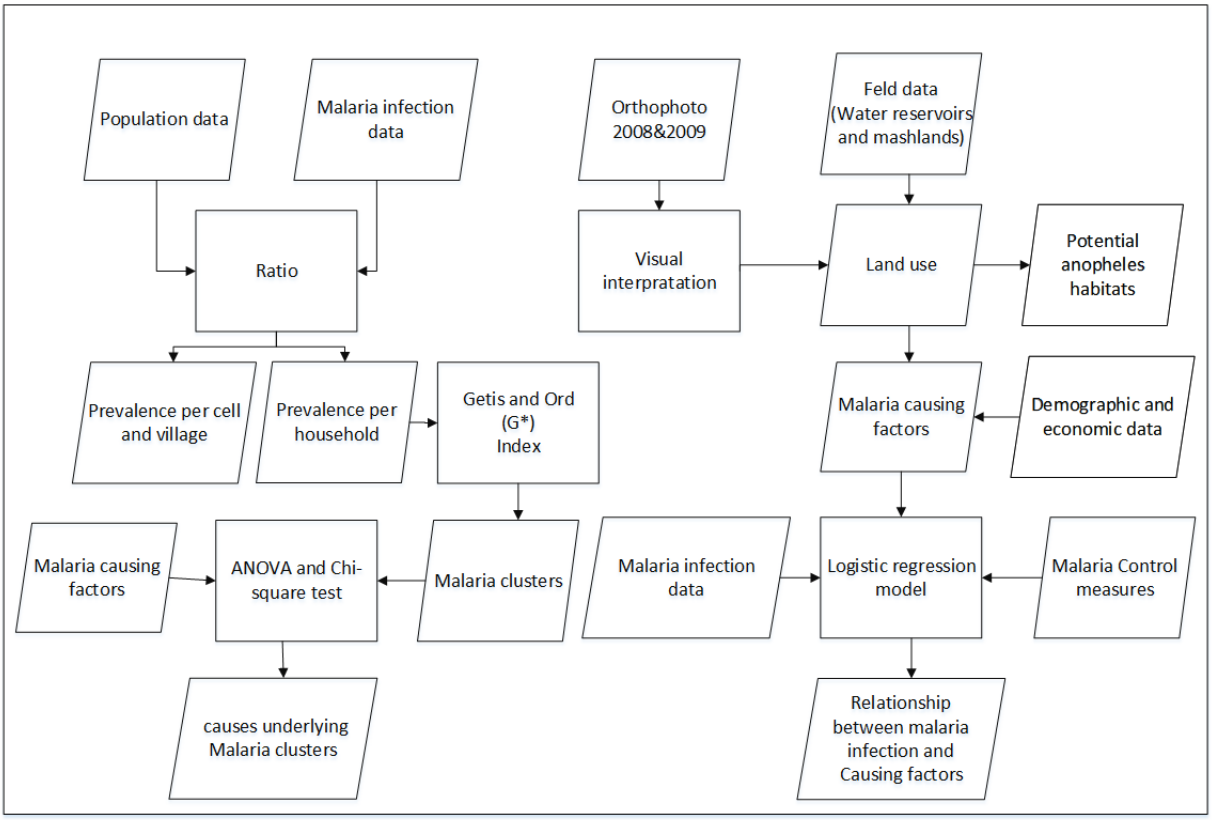



Figure 13. Generalized workflow

## 3.5. Used Software packages

The software packages used in this study were ArcGis 10.2, IBM SPSS Statistics 21, R-studio 0.97.551 and Microsoft Office 2010.

# 4. RESULTS AND DISCUSSION

In this chapter, the results are presented, interpreted, and discussed in line with the research objectives and questions, conceptual framework and the methodology of this study. Potential anopheles mosquito habitats are detected through the combination of the visual interpretation of an orthophotograph of the study area and data acquired from the field. Malaria prevalence is spatially analyzed within administrative boundaries (Cells and villages) and through spatial statistics by which clusters are detected. Malaria causing factors and control measures are identified. Finally, the relationship between malaria and its causing factors is analyzed through a logistic regression, one way ANOVA and chi-square tests.

## 4.1. Potential anopheles mosquito habitats in Ruhuha Sector

The interpretation of an orthophotograph (2008 and 2009) integrated with data acquired from the field resulted in the land use map of the study area. Potential anopheles mosquito habitats were derived from this land use.

### 4.1.1. Study area land use

Figure 14 shows that six land use types were detected in accordance with NISR (2010). However, as the study area was rural with buildings mixed with crops, the term "settlement" was used instead of built-up. Land use has a big influence on anopheles mosquito life cycle. For example, in their study in the western highlands of Kenya, Munga et al. (2006) found that water temperature in irrigated farmlands increases anopheles larvae survival rate and shortens the pupation period. The same authors suggest that due to their influence on the increase of temperature and water nutrient, irrigated farming has a positive effect on the larval survivorship and on the adults breeding.

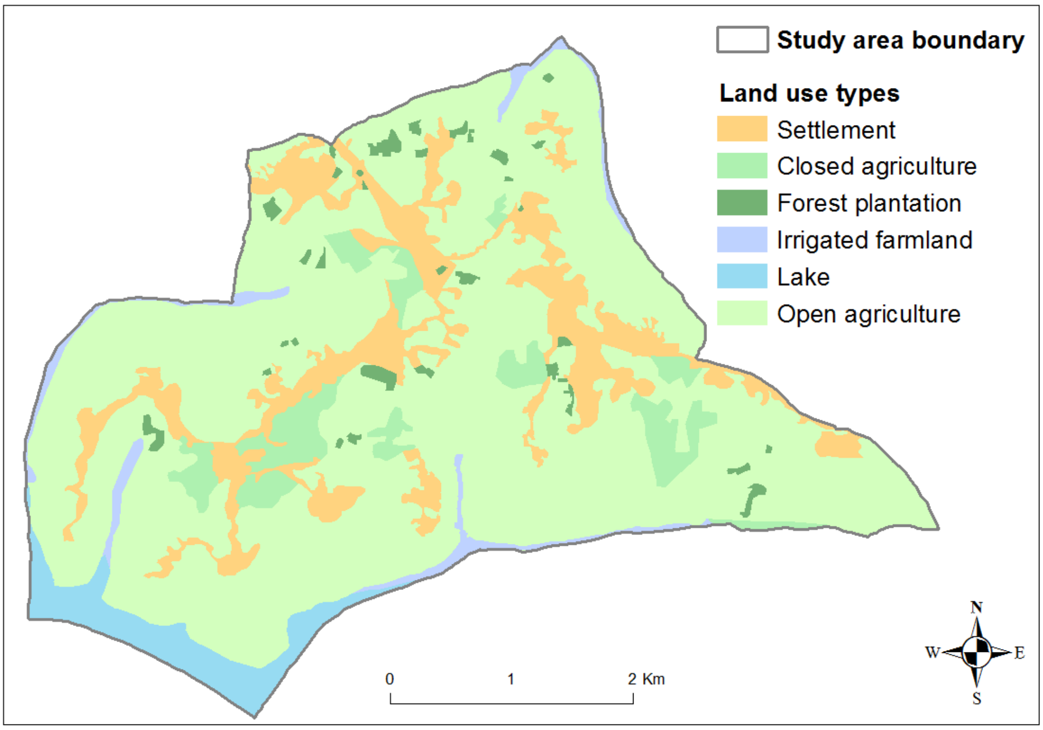



Figure 14. Land use of the study area

- *Accuracy assessment*

The overall accuracy and the kappa (K*) value were calculated (see Table 2). Fifty-five training points observed from the field were used to extract land use values derived from the orthophotograph and exported to IBM SPSS Statistics 21 for K* calculation while the overall accuracy was calculated in Arcmap 10.3 ©. In accordance with Phillips et al. (2006), the accuracy assessment proved a strong agreement ( K*=90%, overall accuracy=91%) and the produced land use map was therefore reliable for future analyses.

Table 2. Land use accuracy assessment table

| *Classes* | *Settlement* | *Forest plantation* | *Closed agriculture* | *Irrigation* | *Open agriculture* | *Accuracy (%)* |
|---|---|---|---|---|---|---|
| Settlement | 18 | 0 | 1 | 0 | 0 | 95 |
| Forest plantation | 0 | 8 | 1 | 1 | 0 | 80 |
| Closed agriculture | 0 | 0 | 10 | 0 | 1 | 91 |
| Irrigation | 0 | 0 | 0 | 4 | 0 | 100 |
| Open agriculture | 1 | 0 | 0 | 0 | 10 | 91 |
| Overall accuracy (%) | | | | | | 91 |
| Kappa (%) | | | | | | 90 |

#### 4.1.2. Potential anopheles mosquito habitats

From the land use map, potential anopheles mosquito habitats were derived. In accordance with the literature (Hassan et al., 2013; Munga et al., 2006; Stephen et al., 2006) and the results from the interview with local malaria experts, closed agriculture (especially made of banana and coffee plantations), forest plantations, irrigated farming (marshlands) and the lake (bushes of the shore) were considered as the potential anopheles habitats in the study area.

The mosquito requires particular conditions for its breeding. The first condition is the presence of stagnant water (Stoler et al., 2009; Verdonschot & Besse-Lototskaya, 2013; Yamamoto et al., 2010) which made irrigated farmlands and water reservoirs the potential anopheles mosquito breeding sites in the study area. Only irrigated farmlands were identifiable on the orthophoto for two reasons: their size was large enough and they were there the time the images were captured. However, their size have been narrowed in some places due to Rwanda National Wetlands Management Policy or extended in accordance with rice crop farming policy. So, those locations were updated based on field observations. Water reservoirs were only identifiable in the field, not only because of their small size, but also because they are recently established in the study area as solutions to irrigation water shortage for local farmers.

The identified potential anopheles breeding sites are relevant for malaria transmission in the study area, which is rural and semi-urban. Munga et al. (2006) proved the association of irrigated farmlands and malaria in rural areas of the Western Kenya. Yamamoto et al. (2010) found the association between malaria and water reservoirs in semi urban areas of Burkina Faso. The fact that the study area is characterized by rural and semi-urban places makes the relationship between malaria infection and the two types of anopheles mosquito breeding sites important.

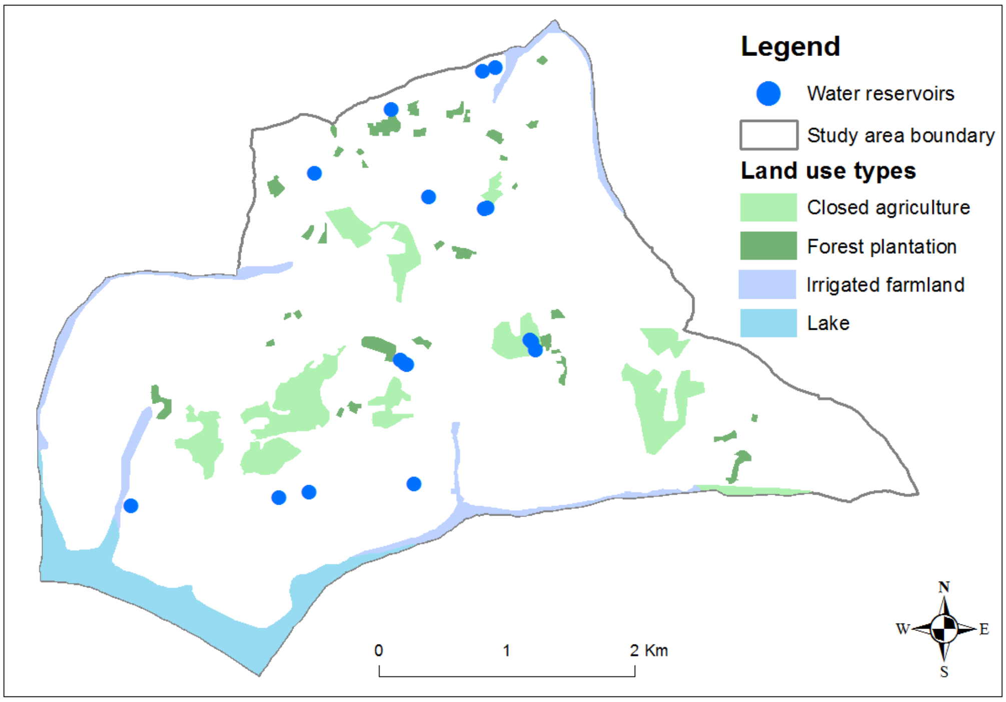



Figure 15. Integrated anopheles mosquito habitats

Potential anopheles mosquito habitats in the study area are made of irrigated farmlands, water reservoirs, closed agriculture, forest plantation, and the lake (See Figure15). Anopheles mosquito depends on blood meal and accordingly, its habitat combines the breeding and resting sites.

Irrigated farmlands and water reservoirs are places where ecological conditions are favorable for the adult mosquito breeding and the larva survival. This was confirmed by Wielgosz et al. (2012) in the East Africa and Yamamoto et al. (2010) in Burkina. In accordance with the findings of Stoler et al. (2009) in Accra and Munga et al. (2006) in the Western Kenya, irrigated farmlands and water reservoirs were considered as breeding sites and habitats at the same time because mosquitoes could stay there before their migration for blood meal search .

The intermediate habitats (forest plantations, closed agriculture, and vegetation on the lake shore) serve as resting sites during the mosquito flights. They can decelerate (if the canopy is dense) or accelerate (when the canopy is open) the flight depending on their types as found by Verdonschot and Besse-Lototskaya (2013) in their review on anopheles mosquito flight distance.

### 4.2. Malaria prevalence and clustering in the study area

Disease prevalence and distribution is the subject of most of epidemiological studies. Stevens and Pfeiffer (2011) suggested that it shows the level of problem and its spatio-temporal distribution which helps to determine the relevant intervention measures at the right place and time. Malaria prevalence will be determined and spatially analysed within administrative boundaries (cell and village) where it will be compared to the population density and beyond administrative boundaries using spatial statistics.

### 4.2.1. Malaria prevalence

Malaria prevalence was calculated as the ratio of the number of positive cases and the total number of the population. To avoid small numbers and decimals, malaria prevalence was multiplied by thousand to get the prevalence per thousand people in accordance with Rothman (2012). The same calculations were done from village to cell level. Malaria prevalence is higher (2.4%) in the study area compared to the country statistics (1.2%) as described by NISR (2012b). A large number of households were characterized by one infected individual.

Table 3. Malaria prevalence per village

| *Cell* | *Village* | *Area_sqkm* | *Number of households* | *Population* | *Population density* | *Infected people* | *Prevalence*1000* |
|---|---|---|---|---|---|---|---|
| Bihari | Bihari | 0.75 | 129 | 612 | 815 | 5 | 8 |
| | Busasamana | 1.72 | 73 | 303 | 176 | 6 | 20 |
| | Masenga 1 | 0.50 | 123 | 510 | 1014 | 19 | 37 |
| | Masenga 2 | 0.65 | 60 | 269 | 412 | 3 | 11 |
| | Mukoma | 1.68 | 108 | 565 | 337 | 11 | 19 |
| | Nyagafunzo | 2.09 | 162 | 837 | 401 | 12 | 14 |
| | Rugarama | 1.07 | 102 | 537 | 500 | 12 | 22 |
| | Rwanzunga | 1.39 | 114 | 527 | 379 | 22 | **42** |
| Total | | 9.86 | 871 | 4160 | 422 | 90 | 22 |
| Gatanga | Butereri | 0.82 | 123 | 662 | 808 | 18 | 27 |
| | Kayigi | 1.22 | 204 | 817 | 670 | 27 | 33 |
| | Kibaza | 1.74 | 151 | 656 | 378 | 29 | **44** |
| | Nyaburiba | 1.59 | 87 | 439 | 277 | 29 | **66** |
| | Nyakagarama | 1.24 | 143 | 760 | 612 | 9 | 12 |
| | Rwanika | 1.54 | 119 | 643 | 418 | 31 | **48** |
| Total | | 8.14 | 827 | 3977 | 489 | 143 | 36 |
| Kindama | Gatare | 0.70 | 163 | 862 | 1228 | 6 | 7 |
| | Gatovu | 1.33 | 74 | 347 | 261 | 10 | 29 |
| | Kagasera | 0.31 | 142 | 802 | 2566 | 6 | 7 |
| | Kamweru | 2.07 | 96 | 328 | 158 | 13 | **40** |
| | Kibaza | 1.35 | 93 | 656 | 275 | 29 | **44** |
| | Kindama | 0.89 | 138 | 607 | 681 | 21 | 35 |
| | Rebero | 1.67 | 206 | 920 | 552 | 15 | 16 |
| | Ruramba | 0.77 | 130 | 708 | 920 | 10 | 14 |
| | Rutare | 1.51 | 136 | 673 | 446 | 34 | **51** |
| | Saruduha | 2.03 | 113 | 609 | 300 | 21 | 34 |
| Total | | 12.63 | 1291 | 6512 | 516 | 165 | 25 |
| Ruhuha | Kimikamba | 0.32 | 167 | 892 | 2799 | 4 | 4 |
| | Mubano | 1.07 | 108 | 540 | 507 | 23 | **43** |
| | Nyabaranga | 1.94 | 172 | 858 | 443 | 27 | 31 |
| | Ruhuha 1 | 0.20 | 171 | 762 | 3838 | 4 | 5 |
| | Ruhuha 2 | 0.24 | 207 | 1035 | 4245 | 0 | 0 |
| Total | | 3.76 | 825 | 4087 | 1086 | 58 | 14 |
| Overall | | 34 | 3814 | 18736 | 545 | 456 | 24 |

Malaria is differently distributed in the study area with the highest prevalence in Gatanga Cell (See figure 16). According to local CHWs, the higher malaria prevalence in Gatanga cell could be justified by the fact that a large proportion of people in Gatanga are far from Ruhuha Health Center and have houses of low quality.

Malaria prevalence at village level is heterogeneously distributed as well (See Figure 17). Most of the villages in the study area are characterized by a high malaria prevalence compared to the national statistics. Nyaburiba, Rwanika and Kayigi of Gatanga Cell, Rutare, Kibaza and Kamweru of Kindama Cell, Rwanzunga of Bihari Cell and Mubano of Ruhuha Cell are characterized by higher malaria prevalence as shown in table 3. Compared to the WHO (2008) report suggesting that areas characterized by more than one malaria case per thousand people are at high risk, all the villages of the study area are at high risk of malaria.

High malaria prevalence areas were characterized by bad material houses, the proximity to anopheles mosquito breeding sites and a long distance to Ruhuha Health Center as they were found during the interview with local CHWs and RBC/MRC staff in charge of malaria control. According to the same source, lower level of education could be one of the causes underlying malaria prevalence though it was not assessed in this study.

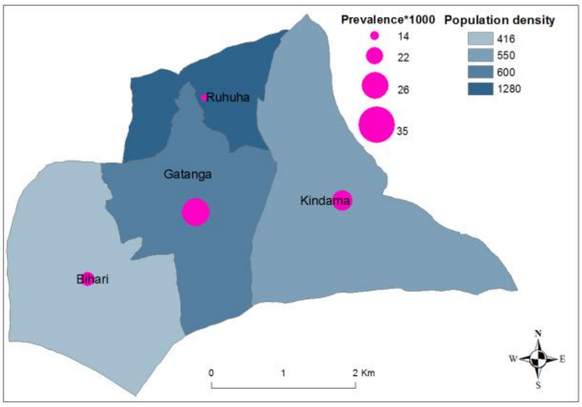



Figure 16. Malaria prevalence per Cell

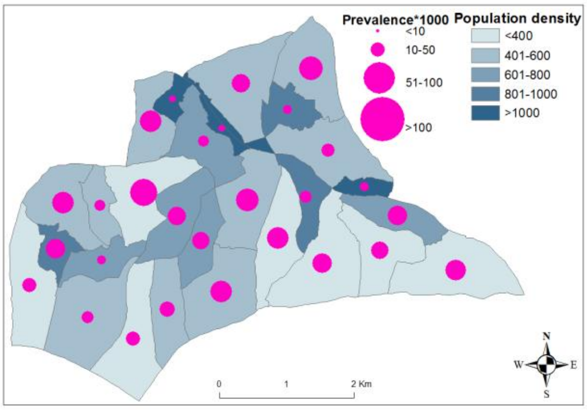



Figure 17. Malaria prevalence per Village

Malaria prevalence is higher in less populated areas (see figure17 and table 3). High malaria prevalence in less populated areas could be attributed to poor living conditions, education, and behaviour of people living in those remote areas.

#### 4.2.2. Malaria distribution

Getis-Ord (G*) statistics using malaria prevalence per household as input field and a threshold distance of 500 m was used for malaria hot spots detection (Cromley & MCLafferty, 2012). Inverse Weighting Distance (IWD) was used to interpolate the observed malaria prevalence to the whole study area.

- *Getis-Ord (G*) clusters*

Table 4 shows G* malaria clusters statistics (at 90-99% confidence level). Hot spots are characterized by positive significant (Gi_pvalues <0.10) GiZscore values, not significant spots are characterized by not significant (Gi_pvalue>0.10) GiZscore and Cold spots are characterized by significant negative GiZscore.

Table 4. Malaria clusters statistics

| *Cluster types* | *GiZScore* | *Gi_Pvalue* | *Gi_Bin* |
|---|---|---|---|
| Hot spot | [1.66,8.25] | [0.000,0.095] | [1,3] |
| Not significant | [-1.64,1.64] | [0.100,0.998] | 0 |
| Cold spot | [-4.71,1.65] | [0.000,0.098] | [-3,-1] |

The spatial pattern of malaria spots in the study area varied from one place to another and as in most of epidemiological cases, the number of surveyed households under hot spots is less than the number under cold and not significant malaria spots (see figure 18).

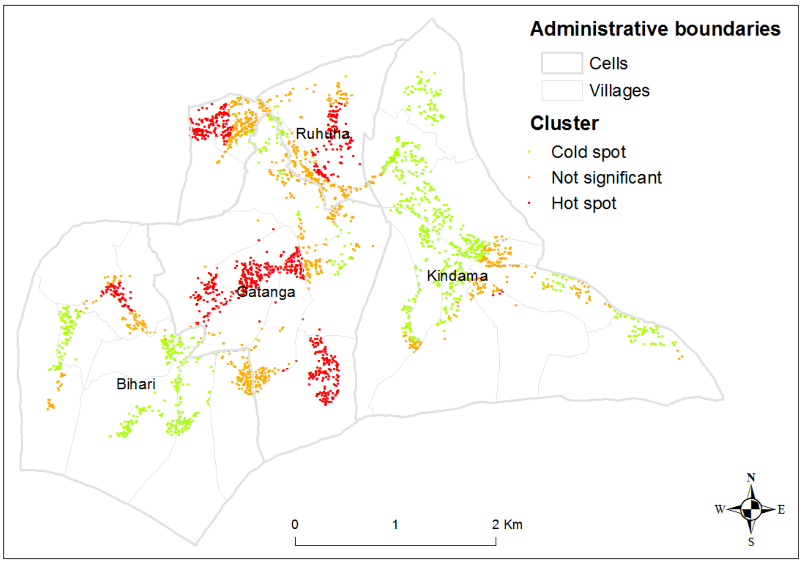



Figure 18. Malaria G* clusters point map

- *Inverse weighting Distance malaria distribution map*

In the study area, the heterogeneous distribution of malaria was proved by G* hot spot analysis. However, it concerned the observed points. For planning measures (building a new settlement, for example), it is important to interpolate the result of malaria prevalence in the whole study area. Figure 19 shows that Ruhuha sector is characterized by different malaria zones based on the inverse weighting distance (IWD) interpolation of the observed prevalence per household.

The probability of getting malaria is higher (Zone characterized by a high prevalence) near the place where the observed prevalence was higher too. This distribution map is realistic given the fact that humans are at the same time hosts and reservoirs of *Plasmodium Sp.* In accordance with Tobler (1970), malaria distribution in the study area is also justified by the principle of spatial autocorrelation by which neighbouring people share the same characteristics such as malaria causing factors and prevention measures and are therefore under the same risk, unless different intervention measures are relied upon. However, the fact that this interpolation approach does not take into consideration other malaria causing factors such as anopheles mosquito breeding sites could be a challenge for planning purposes.

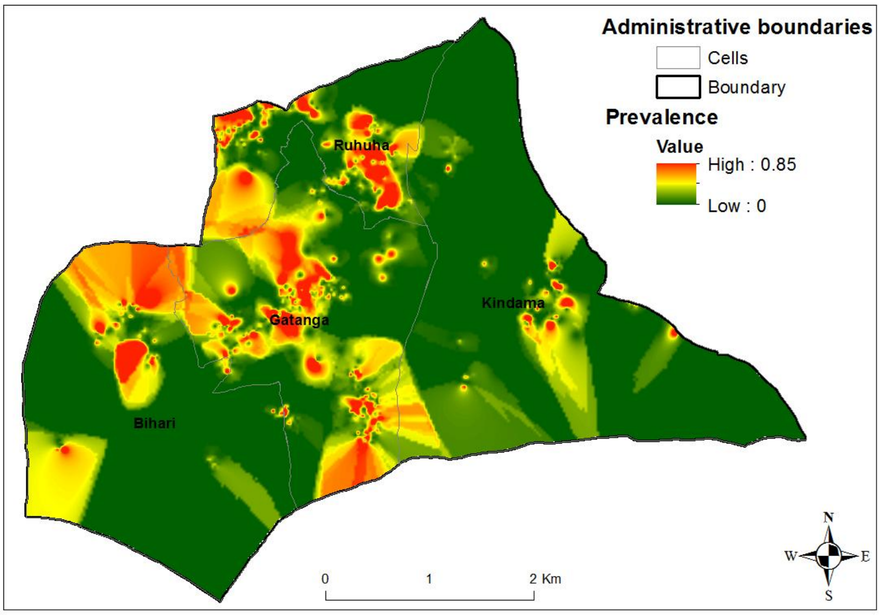



Figure 19. Malaria Prevalence IWD distribution map

### 4.3. Malaria causing factors in the study area

From the previous part, malaria was found to be higher in the study area compared to the national statistics. From the literature (Jung, 2001; Munga et al., 2006; Wielgosz et al., 2012; Yamamoto et al., 2010; Zayeri et al., 2011; Zeile et al., 2012) and interviews with local key informers and malaria experts from RBC/MRC, malaria causing factors were identified and grouped in three categories: environmental, demographic and economic (see table 5).

Table 5. Malaria causing factors

| | Factors | | |
|---|---|---|---|
| | *Environmental* | *Demographic* | *Economic* |
| **Indicators** | Altitude | Age | House quality (roof, wall and floor) |
| | Land use | Gender | Animal ownership |
| | Irrigated farmland | Household size | |
| | Water reservoirs | Density of houses | |

#### 4.3.1. Environmental variables

Environmental variables consist of the altitude, land use and breeding sites (irrigated farmlands and water reservoirs). The distance to the vector breeding sites plays an important role in malaria transmission given the mosquito flight distance as it has been suggested by Stoler et al. (2009) in their study in Accra, Ghana. For their integration, environmental variables were transformed by minimum-maximum linear transformation algorithm as shown by figure 21.

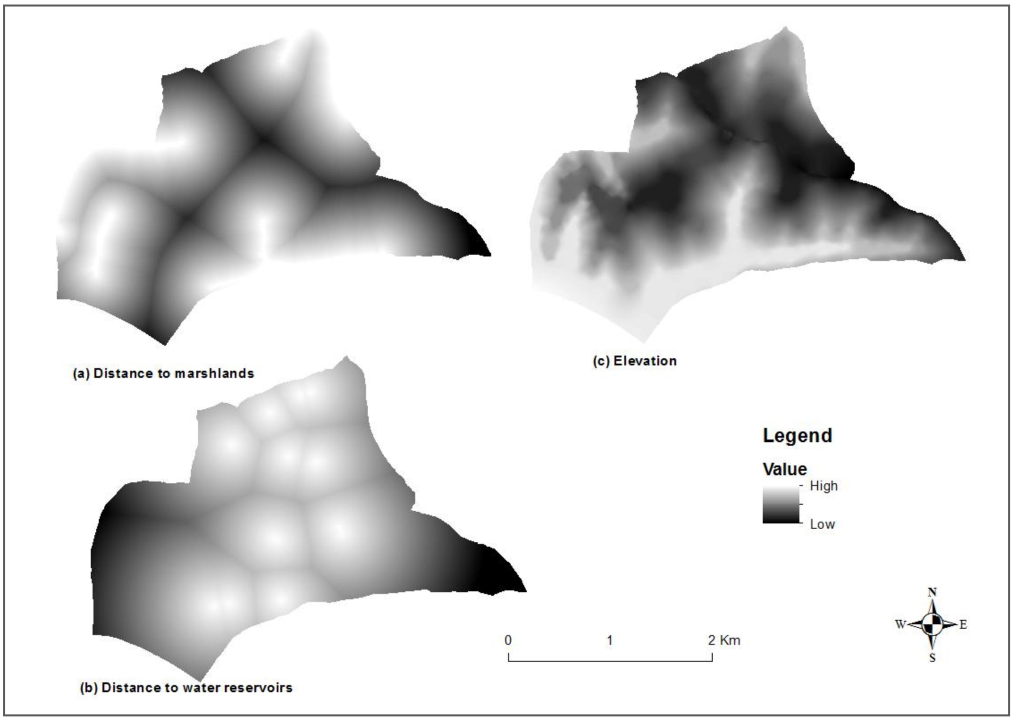


Figure 20. Normalized environmental variables distance maps

The surveyed households were located in settlements (>90%). However, small proportions of households were located in open and closed agriculture and in forest area (see figure 21). No household was located in marshlands, as it is not allowed by the Rwanda national habitat and wetland management policies.

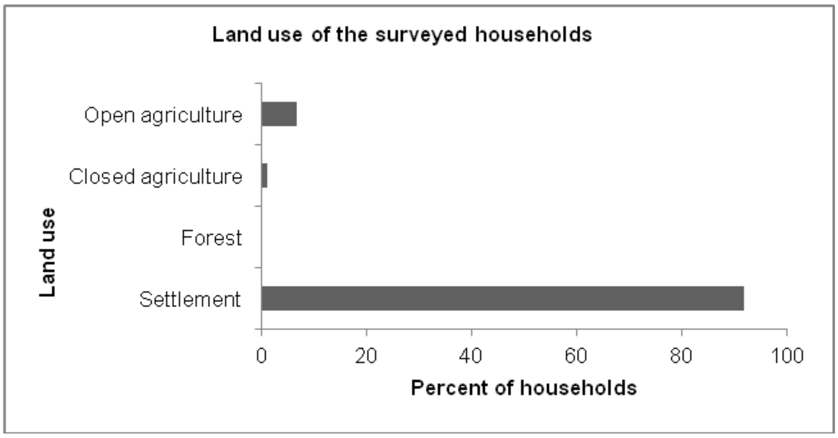



Figure 21. Land use of the surveyed household

#### 4.3.2. Demographic variables

Demographic variables consist of age, sex and household size that were proved to be related to malaria in Rwanda by NISR (2012a). It consisted also of density of houses, which defines the distance from one household to the nearest household with infected people.

In their review on malaria and urbanisation in the sub saharan Africa, Donnelly et al. (2005) suggested that the age and sex define the level of vulnerability while the household size defines the level of exposure due to malaria transmission from one household member to another. In the study area, the household size varied between 1 and 14 with a mean of four individuals. Females made a large proportion (53%) of the surveyed population compared to males who counted for 47%. This is realistic as at national level, the proportion of females is 52% of the total population (NISR, 2012a).

The surveyed population was categorized per age group in accordance with NISR (2012b). Four groups: small children (<5 years), children at school age (5-14 years), adults (15-65 years) and old people (>65 years) were made. Table 6 shows that the highest proportion of the surveyed population is made of adult people, followed by children at school age and children under five years.

Adult people are not vulnerable to malaria as it was found by WHO (2012) and are therefore not a big concern in malaria control process. In contrast, the same source found that children under 5 years are highly vulnerable to malaria. Children at school age are also vulnerable to malaria as mentioned by NISR (2012b).

This distance to household with infected people influences malaria transmission from house to house if anopheles mosquito flight distance is considered (see figure 22). Human beings are at the same time hosts and reservoirs of *Plasmodium sp.* and therefore if the distance that separates a household to the nearest with infected people increases malaria transmission decreases.

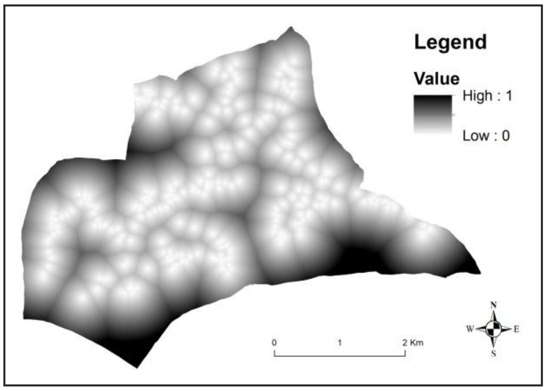



Figure 22. Normalized distance to household with infected people

Table 6. Surveyed population per age groups

| Age groups | Frequency | Percent |
|---|---|---|
| <5 years | 1732 | 21 |
| 5-14 years | 2427 | 29 |
| 15-65 years | 3984 | 47 |
| >65 years | 267 | 3 |
| Total | 8410 | 100 |

#### 4.3.3. Economic variables

House material and animal ownership were assessed as economic variables to explain malaria causes in the study area.

- *House material*

Houses were predominantly made of iron sheets (roof), un-burnt bricks (walls) and clay on the floor (see figure 23). Due to their similar quality, burnt bricks and cement blocks were grouped under "high quality walls"; unburnt bricks were considered as "medium quality walls" while mud and poles were considered as "low quality walls" (see figure 24.a, b and c). However, some burnt brick houses are not of high quality when they have holes (see figure 24.d). Cement, clay, and sand are the most material that make the house floors. Bricks, clay, and sand were both considered as "earth floors." Figure 23 shows proportions of house material in the study area

The house wall, roof and floor material have a negative or positive influence on malaria infection as it was hypothesised by (Yamamoto et al., 2010) during their study in Burkina Faso. This relationship will be assessed in section 4.5.2.

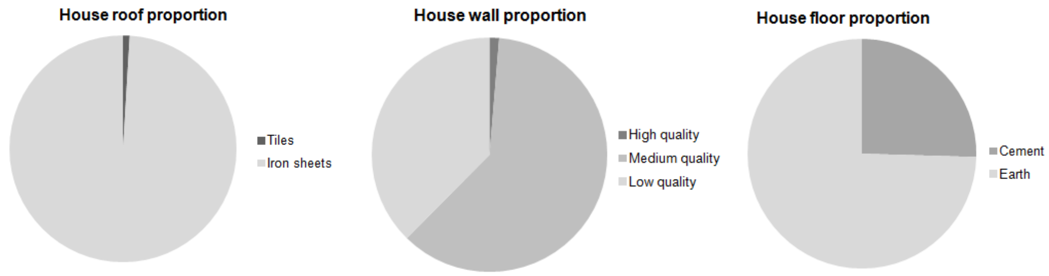



Figure 23. House material proportions

a. Burnt bricks houses (high quality)

b. Unburnt bricks houses (medium quality)

c. Mud houses (low quality)

d. Burnt bricks houses with open windows and holes in the walls

Figure 24. House quality in the study area

- *Animal ownership*

People live in the same plot or house with the domestic animals, which can serve as anopheles mosquito breeding sites via manure and urine. Within the study area, 52% of the surveyed households had domestic animals (see figure 25). This is not in line with other rural regions of Rwanda where 72% have domestic animals (NISR, 2012a).

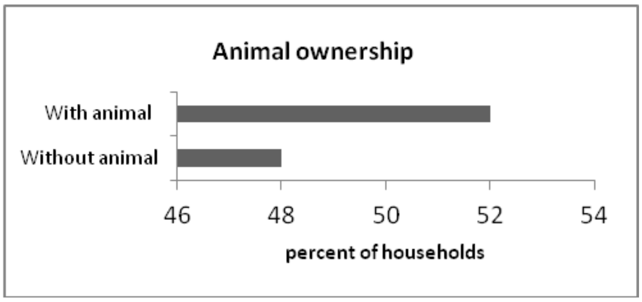



Figure 25. Animal ownership

### 4.4. Malaria control measures

Malaria control measures reduce the risk of malaria. For this study, malaria control measures were identified and grouped in three categories in accordance with the conceptual framework (See figure 2): natural, artificial and treatment.

#### 4.4.1. Natural malaria control measures:

Natural malaria control measures consist in removing bushes that can serve as anopheles resting sites and water ponds that are favorable for anopheles mosquito breeding (See table7).

People in the study area are used to cut bushes around houses because they are considered as anopheles mosquito resting sites and their increase favours the human-vector contact. Stagnant water is also favorable to anopheles mosquito breeding and its removal near houses is the best way to reduce the mosquito proliferation.

Table 7. Malaria natural control measures per administrative cell

| *Cutting bushes (%)* | | | | *Clearing stagnant water (%)* | | | |
|---|---|---|---|---|---|---|---|
| Cells | No | Yes | Total | Cells | No | Yes | Total |
| Bihari | 61 | 39 | 100 | Bihari | 63 | 37 | 100 |
| Gatanga | 65 | 35 | 100 | Gatanga | 66 | 34 | 100 |
| Kindama | 62 | 38 | 100 | Kindama | 66 | 34 | 100 |
| Ruhuha | 60 | 40 | 100 | Ruhuha | 62 | 38 | 100 |
| Total | 62 | 38 | 100 | Total | 64 | 36 | 100 |

The use of malaria natural control measures is still underestimated in the study area and only less than 40% of the surveyed households recognized one of them as efficient malaria control measure. This small proportion of the surveyed population that rely upon natural control measures is realistic because in the country, people are interested in the use of artificial control measures, which are in line with the country malaria reduction policy. In addition, the fact that natural malaria control measures are considered as basic hygiene and therefore, difficult to quantify could be another reason that they are still underestimated.

Though natural malaria control measures have not been the object of malaria researches in Rwanda, they were proved important by Munga et al. (2006) during their study in the Western Highlands of Kenya and by Zayeri et al. (2011) during their study in Sistan and Baluchistan Provinces of Iran. The same approaches were identified by Zeile et al. (2012) as the safest malaria control measures in Rwanda because they are not associated with pollution and resistance problems.

#### 4.4.2. Artificial malaria control measures:

Artificial malaria control measures consist in breaking the vector-host contact. They include the use of bed nets (ITNs), closed doors and windows, the use of repellents, IRS, and wearing long clothes during the evening and night when people are outdoors.

Different malaria artificial control measures are used in the study area as shown in Table 8. ITNs are used most and a large proportion of the surveyed households have at least one. Closing windows and doors is also another way of avoiding anopheles mosquito that could enter the house and bite humans. This approach is recognised by a small proportion of the surveyed households as important for malaria reduction. A very small proportion of the surveyed people use residual Sprays (IRS) and prophylaxis (see Table 8).

Table 8. Artificial malaria control measures per administrative cell

| *ITNs (%)* | | | | *Close windows (%)* | | | |
|---|---|---|---|---|---|---|---|
| Cells | No | Yes | Total | Cells | No | Yes | Total |
| Bihari | 11 | 89 | 100 | Bihari | 86 | 14 | 100 |
| Gatanga | 10 | 90 | 100 | Gatanga | 89 | 11 | 100 |
| Kindama | 9 | 91 | 100 | Kindama | 86 | 14 | 100 |
| Ruhuha | 11 | 89 | 100 | Ruhuha | 86 | 14 | 100 |
| Total | 10 | 90 | 100 | Total | 87 | 13 | 100 |
| *IRS (%)* | | | | *Prophylaxis (%)* | | | |
| Cells | No | Yes | Total | Cells | No | Yes | Total |
| Bihari | 97 | 3 | 100 | Bihari | 97 | 3 | 100 |
| Gatanga | 97 | 3 | 100 | Gatanga | 98 | 2 | 100 |
| Kindama | 99 | 1 | 100 | Kindama | 97 | 3 | 100 |
| Ruhuha | 98 | 2 | 100 | Ruhuha | 98 | 2 | 100 |
| Total | 98 | 2 | 100 | Total | 98 | 2 | 100 |

Different researchers in Rwanda assessed the importance of different malaria artificial control measures and they were in conformity with the results of this study. Karema et al. (2006), Zeile et al. (2012) and NISR (2012b) suggested a huge decrease of malaria prevalence from 2006-2012 due to the increase of the use of ITNs, IRS and treatment facilities. ITNs are distributed free of charge to local people in the study area as suggested by President's Malaria Initiative (2013) and therefore, they are used by a higher proportion of the population in Rwanda.

#### 4.4.3. Treatment:

Treatment consists of curing infected people who could serve as reservoirs for *Plasmodium malaria*. Malaria treatment is a good approach to avoid malaria transmission from household to household or among household members as humans are hosts and reservoirs of *plasmodium sp.* at the same time as mentioned by Zeile et al. (2012) during their study in the Southern highland of Rwanda.

Malaria treatment was the same in three of the four surveyed cells and a small proportion of households suggested treatment as malaria prevention measure (See Table 9). This underestimation of treatment as malaria control measure could be explained by the fact that most of the surveyed people do not consider treatment as control measure because prevention should come before the sickness. In the study area, malaria treatment is influenced by the distance to Health Center (see Figure 24). In fact, this distance was classified by WHO (2012) as one of the causes of malaria in the Sub Saharan African countries.

The efficiency of malaria control measures requires their integration. The combination of artificial (dominated by ITNs) natural approaches and treatment lowers malaria infection considerably. In line with NISR (2007), besides natural and artificial malaria control measures, the Community Health Insurance commonly known as "Mutuelle de Santé" and other types of health insurances especially RAMA and MMI, helped people in the study area to access and afford malaria treatment.

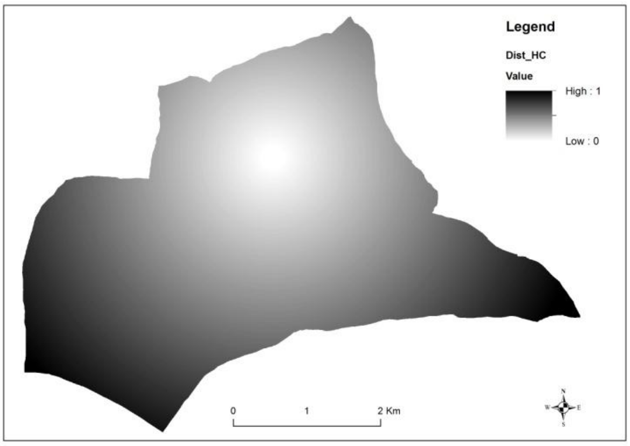



Figure 26. Normalized Health Center distance map

Table 9. Malaria treatment per cell

| | Treatment (%) | | |
|---|---|---|---|
| *Cells* | *No* | *Yes* | *Total* |
| Bihari | 98 | 2 | 100 |
| Gatanga | 98 | 2 | 100 |
| Kindama | 98 | 2 | 100 |
| Ruhuha | 99 | 1 | 100 |
| Total | 98 | 2 | 100 |

Malaria control measures are well known by the surveyed people through the sensitization by Community Health Workers (CHWs) as the interviewed local key informers reported them. However, their adequate use is still problematic as it is influenced by different factors such as the economic level, which determines the house quality and placement and the behaviour. Though it will not be assessed in this study, behaviour is important to choose and to use adequately relevant malaria control measures.

## 4.5. Relationship between malaria infection and underlying factors

Malaria infection is a result of its causing factors, which contribute to its increase and its control measures, which contribute to its reduction. In this section, malaria causing factors (environmental, demographic and economic) and control measures will be screened to determine the relevant ones that can be inputted in a stepwise logistic regression model to test their relationship with malaria infection. The household size is the only demographic variable that is household based and therefore can be included in the model while gender and age are individual based and will be analyzed separately.

### 4.5.1. Variables screening

Malaria causing factors were mentioned in table 5 and its control measures in tables 7, 8, 9 and figure 26. A reliable malaria prediction requires their integration as they come from different sources and are of different formats (Stevens & Pfeiffer, 2011). However, as in many epidemiological studies, before the integration of diseases causing factors and control measures, they must be pre-analyzed to evaluate their importance and pre-processed for their compatibility (Anamzui-Ya, 2012).

Not all the variables that have been explored in the previous part are equally important and it is necessary to decide about the ones, which could play a considerable role in malaria prediction.

*a. Environmental variables*

All environmental variables made especially of distances to irrigated farmland and to water reservoirs were taken into consideration. Altitude and the four land use classes that include the surveyed households were also taken into consideration (see figures 20 and 21).

*b. Demographic variables*

Household size, age, and gender were retained among malaria causing factors as they varied from one household to another and could explain the level of vulnerability towards malaria. However, the household size was inputted in the model as it was household based while age and sex are individual based and were analyzed differently. The distance to households with infected people was also taken into consideration.

*c. Economic variables*

Animal ownership was retained as one of the crucial economic factors. Houses are made of different materials. For house walls, high quality (combining burnt bricks and cement blocks), medium quality (unburnt bricks), and low quality (mud and poles) walls were taken into consideration. For house roof, iron sheet was almost the only material (See figure 23) and to consider it could not bring any benefit to the predictions. Cement and earth (bricks, clay, and sand) were the most important elements for house floor.

*d. Malaria Control measures*

Table 7 shows that natural malaria control measures (cutting bushes and removing stagnant water) were used by a considerable proportion of the surveyed households (>35%). Therefore, their inclusion in malaria prediction model is a benefit. Concerning malaria artificial measures, almost everybody used ITNs while small proportions of the surveyed population used IRS, prophylaxis, closing windows and doors and wearing long sleeves during the night. For this reason, they cannot be considered as good malaria predictors. The surveyed people underestimated treatment. However, as it was suggested by WHO (2012), it is influenced by the distance to the Health Center which will be taken into consideration in this section.

- **Variables coding**

The screened variables consisted of nominal and ratio. Before the input in the logistic regression model, nominal variables with more than two categories were coded in dummy variables while ratio variables remained the same.

When variables consist of more than one category which are exclusive, they must be coded before their input in the regression model through a process known as dummy coding (Field, 2009). Categorical variables (house walls, house floors and land use) with more than two classes, were coded in dummy variables in accordance with Field (2009).

Medium quality walls were taken as a control (reference) category while high quality and low quality walls were coded in presence (1) - absence (0). For the house floor, earth floor, which was the most predominant floor material, was taken as control category while cement floors were coded in presence-absence. Finally, Settlement that represented 90% of the land use where observations occurred was considered as reference category while forest, open agriculture, and closed agriculture were coded in presence-absence. Indeed, when dummy variables are used in a regression model, the control category is coded 0 for all the variables while other variables appear in the form of 1 and 0 and accordingly, the input dummy variables become n-1 where n stands for the number of categories (Field, 2009). Table 10 shows coded variables under description.

Table 10. Normalized and coded variables

| *Factor* | *Variable in LRM* | *Description* | *Nature* |
|---|---|---|---|
| Independent | y | 1-malaria infection, 0-no malaria infection | Dichotomous |
| Environmental | Alt | Altitude | Ratio |
| | Dist_marsh | Distance to marshland | Ratio |
| | Dist_reser | Distance to water reservoir | Ratio |
| | Settlement | 1-settlement, 0-other land use | Nominal |
| | Open_agr | 1-open agriculture, 0-other land use | Nominal |
| | Closed_agr | 1-closed agriculture, 0-other land use | Nominal |
| | Forest | 1-forest, 0-other land use | Nominal |
| Demographic | H_size | Number of individuals per household | Ratio |
| | Dist_HH | Distance to household with infected people | Ratio |
| Economic | Animown | 1-with animal, 0-without animal | Nominal |
| | High | 1- high quality walls, 0-other material | Nominal |
| | Medium | 1- medium quality walls, 0-other material | Nominal |
| | Low | 1-low quality walls, 0-other material | Nominal |
| | Cement | 1-cement floor, 0-other material | Nominal |
| | Earth | 1-Earth floor, 0-other material | Nominal |
| Control measures | Dist_HC | Distance to Health Center | Ratio |
| | Cut_bush | 1-Bush cut, 0-No | Nominal |
| | Remove_water | 1-Stagnant water removed, 0-No | Nominal |

**a. Collinearity test**

In statistical modelling, it is important to avoid the inclusion of related variables which can reflect the same reality leading to the issue of information duplicate or multicolinearity (Field, 2009). For this reason, collinearity test was performed. After viewing the correlation matrix, there were weak correlations between the altitude and the distance to marshlands (r=0.57), between the distance to water reservoir and the distance to Health Center (r=0.53) and cutting bush and removing stagnant water (r=0.68). So VIF was used as the next step to test collinearity.

All the variables had VIFs<10 (see Table 11) and therefore, collinearity was not a problem. The screened variables were then inputted in a stepwise logistic regression model. In fact, the stepwise regression model has the advantage of showing the model improvement as new predictors are added (Field, 2009).

Table 11. Independent variables VIF calculation

| *Variable* | *$R^2$* | *VIF* | *Variable* | *$R^2$* | *VIF* |
|---|---|---|---|---|---|
| Alt | 0.44 | 1.79 | High | 0.30 | 1.43 |
| Dist_marsh | 0.54 | 2.17 | Low | 0.27 | 1.37 |
| Dist_res | 0.61 | 2.56 | Cement | 0.51 | 2.04 |
| Dist_HH | 0.34 | 1.52 | Animown | 0.32 | 1.47 |
| Forest | 0.07 | 1.08 | Dist_HC | 0.66 | 2.94 |
| Closed_agr | 0.06 | 1.06 | Cut_bush | 0.68 | 3.13 |
| Open_agr | 0.13 | 1.15 | Remove_water | 0.69 | 3.16 |
| H_size | 0.29 | 1.41 | | | |

#### 4.5.2. Logistic regression

The relationship between malaria and its underlying factors was analyzed through a stepwise logistic regression where the Wald statistics and the OR proved their significance (See table 13). The logistic regression model, showed a low predictive power ($R^2$=13%). This is explained by the fact that the chosen variables are not the only malaria underlying factors in the study area.

Of environmental variables, altitude and distance to water reservoir were not good malaria predictors (p>0.05) and were not therefore included in the final model (See table 12). Of economic variables, only animal ownership was not a good predictor. All natural malaria control measures were also not good predictors (p>0.05) and were therefore excluded from the model during the initial step of the logistic regression (see table 12). The steps of the logistic regression model are shown in appendix 6.

Table 12. The initial step of logistic regression testing the significance of predictors

| | *Variable* | *Score* | *df* | *P-value* |
|---|---|---|---|---|
| Step 0 | H_size | 53.696 | 1 | **.000** |
| | Animown | 11.968 | 1 | .301 |
| | Cut_bush | 1.617 | 1 | .204 |
| | Remove_water | 2.819 | 1 | .093 |
| | Alt | .902 | 1 | .342 |
| | Dist_HH | 18.757 | 1 | **.000** |
| | Dist_marsh | 2.663 | 1 | **.013** |
| | Dist_reser | 5.127 | 1 | .064 |
| | Dist_HC | 20.366 | 1 | **.000** |
| | High | 11.861 | 1 | **.001** |
| | Low | 43.833 | 1 | **.000** |
| | Cement | 51.081 | 1 | **.000** |
| | Forest | .643 | 1 | .422 |
| | Closed_agr | 1.226 | 1 | .268 |
| | Open_agr | 1.187 | 1 | .276 |
| Overall Statistics | | 207.894 | 15 | .000 |

Table 13. Relationship between malaria infection and its causing factors

| *Variable* | *B* | *S.E.* | *Wald* | *df* | *P-value* | *OR* | *95% C. I. for OR* | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Lower* | *Upper* |
| Dist_marsh | -1.005 | .235 | 18.238 | 1 | .000 | .366 | .231 | .580 |
| Dist_HH | .463 | .110 | 17.825 | 1 | .000 | 1.589 | 1.282 | 1.970 |
| H_size | .253 | .029 | 77.368 | 1 | .000 | 1.288 | 1.217 | 1.363 |
| High | -1.487 | .618 | 5.788 | 1 | .016 | .226 | .067 | .759 |
| Low | .782 | .124 | 39.968 | 1 | .000 | 2.186 | 1.716 | 2.786 |
| Cement | -.942 | .200 | 22.243 | 1 | .000 | .390 | .264 | .577 |
| Dist_HC | -1.013 | .174 | 33.918 | 1 | .000 | .363 | .258 | .511 |
| Constant | -2.591 | .339 | 58.499 | 1 | .000 | .075 | | |

*$R^2$ =0.13 (Nagelkerke). Model $\chi 2$=207, p<0.001*

*b. Malaria infection towards environmental variables*

Distance to marshlands is negatively related to malaria infection (Wald=18.2 and OR<1). People living near irrigated farmlands are more exposed to the infection because mosquito can fly and reach them very easily and the higher the distance, the less the exposure. Similarly, in their study in the Western highlands of Kenya, Munga et al. (2006) found that people living near irrigated farmlands were more affected compared to those living far away. In their study in Accra, Stoler et al. (2009) suggested that people could live at least beyond 1000 m from irrigated farmlands and the further away the better.

*c. Malaria infection towards demographic variables*

Malaria infection increases with the household size (Wald=77.3 and OR>1). The increase of malaria with household size is realistic in the study area because the household size varied between 1 and 14 people with a mean of four individuals. In general, more people in the study area live in low income and large families with bad quality houses and are therefore prone to malaria.

In conformity with the results of this study, the high household size was also suggested among malaria underlying causes in the fringe zones of Kigali City by Bizimana et al. (2009). The association between malaria infection and the household size was also found by WHO (2008) in the annual malaria report while in their review on the persistence of malaria problem, (Stratton et al., 2008) mentioned poverty and increasing population as associated problems which are among the main causes of malaria in developing countries.

Malaria infection increased with the distance from household with infected people (Wald=17.8 and OR>1). This is not what was expected because malaria was supposed to decrease with this distance. This abnormal increase could be explained by the fact that the distance separating houses in the study area is shorter if the mosquito flight distance is considered. In addition, as the distance from one household to an other increases, the proximity to anopheles breeding sites could increase the exposure.

*d. Malaria infection towards economic variables*

House materials are the only economic malaria predictors (see table 13). Compared to houses with medium quality walls (reference category during the process of dummy coding), the infection is lower in houses made of high quality walls (Wald=5.7, OR<1) and higher in houses with low quality walls (Wald=39.9, OR>1). Malaria infection is lower in houses made of cement floor (Wald=31.3, OR<1) compared to those with earth floor (reference category).

Malaria infection was lower in houses with high quality walls (burnt bricks and cement blocks). Few houses of good quality made especially of burnt bricks or cement blocks for walls and tiles or iron sheets for roofs, have tight doors and windows so that mosquito could not enter easily except when doors or windows are open (See Figure 23.a). Conversely, most of houses in the study area are made of medium quality (unburnt bricks) walls which in most of cases have windows and doors that are not tight enough to block the mosquito entrance (See Figure 23.b) . In addition, a large proportion of houses with low quality (mud) walls (See Figure 23.c) were also proved positively associated with malaria infection. The quality of mud wall houses is the worst and they favour an easy contact between the mosquito and humans.

The relationship between poor housing quality and malaria infection that was found in this study is in confirmation with Stratton et al. (2008) in their review on the persistent problem of malaria. The same relationship is in contradiction with the results of Yamamoto et al. (2010) which suggested a lack of relationship between house material in the semi-urban areas of Burkina Faso.

*e. Malaria infection towards control measures*

Most of the people having difficulty in accessing healthcare services reported the distance to health center as the main cause. This assumption was contradicted by the results of the logistic model which showed that malaria infection is lower far from the Health Center (Wald=33.91 and OR<1). This result is also in contradiction with the association between malaria infection and the distance to Health Center as reported by WHO (2012). The same report suggested that malaria infection is higher far from health care facilities.

The distance to Health Center in the study area which varies between 0 and 5.3 km is less than the national standard according to which all the households should be within 90 minutes walking distance from a health care facility, that is approximately 6 km as suggested by NISR (2007) and therefore, it should not be a big problem.

#### 4.5.3. Relationship between malaria infection and age and gender

Malaria infection is a result of different demographic variables especially the age and gender that could justify the level of vulnerability and the household size and the distance to household with infected people that justify the level of exposure. The association between malaria infection, the household size (household based) and the distance to household with infected people has been shown in Table 13 and discussed previously. Age and gender are individual based and therefore were analyzed separately. Below, the relationship between malaria infection, gender, and age as described in Table 14 will be discussed.

Ages were grouped in four categories according to NISR (2012b) and malaria infection rate varied among different age categories ($\chi^2$= 141.5 , df=3, p<0.001). In fact, as shown by Table 14, children at school age (5-14 years) were characterized by a high malaria infection (11.5%) while children under five years come at the second place (6%). This is in conformity with NISR (2012a) suggesting that between 2006 and 2012 malaria microscopically confirmed cases decreased by 82% for children below five years due to the use of malaria control technologies. Indeed, as reported by CHWs, efficient malaria control measures such as local malaria treatment and proper use of ITNs were actually adopted for children under five years. Children at school age (5-14 years) are considered as adults while their levels of vulnerability and exposure are different. Malaria infection is lower in adults not only because their level of exposure is less but also because they have developed immunity against *plasmodium sp.* as it was confirmed also by WHO (2010) in the annually malaria report.

Table 14. Proportion of malaria infection per Gender and sex

| | | *Negative* | | *Positive* | | | |
|---|---|---|---|---|---|---|---|
| Age groups | Sex | Count | Percent | Count | Percent | Total | Overall total |
| <5 years | Female | 810 | 94 | 55 | 6 | 865 | 1732 |
| | Male | 824 | 94 | 54 | 6 | 877 | |
| 5-14 years | Female | 1166 | 92 | 112 | 9 | 1271 | 2427 |
| | Male | 1016 | 88 | 136 | 12 | 1156 | |
| 15-65 years | Female | 2567 | 98 | 62 | 2 | 2617 | 3984 |
| | Male | 1306 | 96 | 61 | 4 | 1367 | |
| >65 years | Female | 159 | 94 | 4 | 2 | 169 | 267 |
| | Male | 83 | 94 | 3 | 3 | 88 | |
| Grand Total | | 7923 | 94 | 487 | 6 | 8410 | 8410 |

A large proportion of males (52% of the surveyed population) were affected by malaria compared to females (48%) in the study area (See Table 14). This significant difference ($\chi^2$= 21.1, df=1,p<0.001) was in contraction with the results of NISR (2012a) in which women were more affected by malaria than men in Rwanda. The large proportion of males with malaria infection could be attributed to the reason that the majority of strong males were not tested because they were out of their households for daily activities so that the tested one were mostly children who are vulnerable to malaria and jobless adults who live in bad socio-economic conditions. The fact that most of the adult and school age males in the study area have the habit of staying outdoors during evening hours could expose them more than females who spend their evenings indoors taking care of children or busy with kitchen activities.

#### 4.5.4. Difference between malaria clusters

The results presented and discussed in section 4.2 show that malaria is differently distributed in the study area. Ruhuha Sector is characterized by malaria hot, not significant, and cold spots depending on the level of malaria prevalence. The detected heterogeneity is attributed to the difference in malaria causing factors or in the use of malaria control measures. For planning and intervention purposes, differences between malaria clusters will be analyzed in the next section.

The comparison of malaria clusters in the study area was done using one way ANOVA and Tukey tests for ratio variables and Chi-square test for nominal variables.

**a. ANOVA and Tukey tests**

ANOVA assumes that the data are normally distributed, then the test for normality (Shapiro-Wilk test and box plots) was carried out and proved that data were not normally distributed (Significant w at $p<0.05$) as shown in appendix 8. Therefore, log10 transformation was used for data normalization.

Table 15 proves a difference between the mean distance to marshland, water reservoir and household with infected people (Significant F at $p<0.05$). The mean household size, the mean altitude and the mean distance to Health Center are not significantly different ($p>0.05$). To show the differences between groups, the Tukey test was carried out for the significant variables.

Table 15. Comparison of continuous variables between different malaria clusters

| | | *Sum of Squares* | *df* | *Mean Square* | *F* | *P-value* |
|---|---|---|---|---|---|---|
| Dist_marsh | Between Groups | .533 | 2 | .267 | 5.673 | **.003** |
| | Within Groups | 150.498 | 3203 | .047 | | |
| | Total | 151.031 | 3205 | | | |
| Dist_reser | Between Groups | .803 | 2 | .402 | 4.133 | **.016** |
| | Within Groups | 311.193 | 3203 | .097 | | |
| | Total | 311.996 | 3205 | | | |
| H_size | Between Groups | .065 | 2 | .032 | 1.015 | .362 |
| | Within Groups | 102.140 | 3203 | .032 | | |
| | Total | 102.205 | 3205 | | | |
| Dist_HH | Between Groups | 26.299 | 2 | 13.149 | 27.685 | **.000** |
| | Within Groups | 1521.292 | 3203 | .475 | | |
| | Total | 1547.590 | 3205 | | | |
| Alt | Between Groups | .038 | 2 | .019 | 1.489 | .226 |
| | Within Groups | 40.392 | 3203 | .013 | | |
| | Total | 40.430 | 3205 | | | |
| Dist_HC | Between Groups | .185 | 2 | .092 | .907 | .404 |
| | Within Groups | 326.421 | 3203 | .102 | | |
| | Total | 326.606 | 3205 | | | |

The Tukey test (see table 16) proved a difference of mean distance to marshlands among the three detected malaria clusters. The mean difference of the distance to household with infected people was significant for hot spots compared to other clusters while the mean distance to water reservoir was different between insignificant and cold spots. Indeed, the hot spots were characterized by the lowest mean distances to marshlands and to household with infected people compared to other malaria clusters. Cold spots were characterized by the lowest mean distance to water reservoirs (see appendix 7).

Table 16. Multiple Comparisons (Tukey test) between malaria clusters

| *Dependent Variable* | *(I) Clusters* | *(J) Clusters* | *Mean Difference (I-J)* | *SE* | *P_value* | *95% CI Lower Bound* | *95% CI Upper Bound* |
|---|---|---|---|---|---|---|---|
| Dist_marsh | Hot spot | Cold spot | **-.03069*** | .00957 | .004 | -.0531 | -.0082 |
| | | Not significant | -.00910 | .00970 | .616 | -.0318 | .0136 |
| | Not significant | Cold spot | **-.02158*** | .00904 | .045 | -.0428 | -.0004 |
| | | Hot spot | .00910 | .00970 | .616 | -.0136 | .0318 |
| | Cold spot | Not significant | **.02158*** | .00904 | .045 | .0004 | .0428 |
| | | Hot spot | **.03069*** | .00957 | .004 | .0082 | .0531 |
| Dist_reser | Hot spot | Cold spot | .00230 | .01377 | .985 | -.0300 | .0346 |
| | | Not significant | -.03186 | .01395 | .058 | -.0646 | .0009 |
| | Not significant | Cold spot | **.03416*** | .01300 | .023 | .0037 | .0646 |
| | | Hot spot | .03186 | .01395 | .058 | -.0009 | .0646 |
| | Cold spot | Not significant | **-.03416*** | .01300 | .023 | -.0646 | -.0037 |
| | | Hot spot | -.00230 | .01377 | .985 | -.0346 | .0300 |
| Dist_HH | Hot spot | Cold spot | **-.22189*** | .03044 | .000 | -.2933 | -.1505 |
| | | Not significant | **-.16409*** | .03084 | .000 | -.2364 | -.0918 |
| | Not significant | Cold spot | -.05780 | .02873 | .110 | -.1252 | .0096 |
| | | Hot spot | **.16409*** | .03084 | .000 | .0918 | .2364 |
| | Cold spot | Not significant | .05780 | .02873 | .110 | -.0096 | .1252 |
| | | Hot spot | **.22189*** | .03044 | .000 | .1505 | .2933 |

*. The mean difference is significant at the 0.05 level.

Malaria hot spots are characterized by the proximity to irrigated farmlands. The high exposure of areas closer to anopheles breeding sites is attributed to the mosquito flight distance as it was proved by Stoler et al. (2009) during their study in Accra and Verdonschot and Besse-Lototskaya (2013) in their review about the mosquito flight distance. In their review on the relationship between malaria and agriculture in the East Africa, Wielgosz et al. (2012) found irrigated farmlands among the causes underlying malaria infection in the region, which was confirmed by the current study.

The proximity to household with infected people is also an important determinant of malaria hot spots, which are characterized by the shortest mean distance (see appendix 7). The significance of the mean distance to the household with infected people, hot spots having the lowest mean proves that malaria transmission from house to house is higher in hot spot areas. However, if anopheles mosquito flight distance is considered, the contamination from house to house could not present a big difference between the three clusters given the fact that the overall mean distance is short (<100m) as shown in appendix 7.

The mosquito flight distance contributes on the transmission of malaria from house to house given the fact that infected persons serve as *plasmodium sp.* reservoirs. This is a common determinant of malaria transmission in the study area and the lower the distance between one household and another with infected people, the higher the probability of being infected. The results of this study are similar to the findings of NISR (2012a) suggesting that transmission from house to house or between members of one household are among the causes of malaria in Rwanda

The distance to water reservoirs is not a big problem in the study area given the fact that it does not present any significant difference between hot and cold spots (see table 16). This is in confirmation with the results of the logistic regression (See table 12), which does not consider water reservoir among the relevant malaria predictors in the study area.

**b. Chi-square test**

The chi-square test proved a significant difference between houses made of high, medium and low quality walls ($\chi^2$=44.14, df=3, p<0.001). The same test proved that the three clusters had differences regarding house floors which were made of cement and earth ($\chi^2$=54.8, df=2, p<0.001). Land use, animal ownership and natural malaria control measures did not present any difference (p>0.05) for the three malaria clusters (See Table 17).

The difference between malaria clusters towards house material could not at its own explain the presence of malaria cold, insignificant and hot spots. The house material, which was proved related to malaria infection in the previous part, becomes controversial when malaria clusters are considered (see Table 17). The study area is rural and is characterized by a mixture of different quality houses so that it is not easy to separate them from one malaria cluster to another.

The mixture of houses of different wall and floor quality in each of the identified clusters makes the association between malaria infection and house material a common challenge in the whole study area. In all the clusters, there are houses made of high, medium, and low quality walls; while the floor was made by either cement or earth. The medium and low quality walls and earth floors were proved associated with the increase of malaria infection in the study area.

Summarized, if environmental variables are considered, malaria hot spots are determined by the proximity to irrigated farmlands. The proximity to water reservoirs, altitude and land use do not have any effect on malaria clustering in the study area.

Of demographic variables, malaria hot spots are determined by the distance to the nearest household with infected people and therefore, malaria transmission from house to house is higher in those areas. Household size has no effect on malaria clustering. This is because the whole study area is characterized by a mean household size of four individuals and, therefore, there is no mean variation between different malaria clusters.

If economic variables are considered, house materials (wall and floor) are among the determinants of malaria clusters. However, the fact that all the clusters are characterized by a mixture of houses made of different material, clustering is not efficient to assess the contribution of house material to malaria infection. Animal ownership does not have any effect on malaria clustering.

Malaria control measures (proximity to Health Center and natural control measures) do not play any significant role in malaria hot spot determination in the study area.

Table 17. Comparison of proportions of house material and malaria control measures per cluster

| *Clusters* | *Variables* | | | | *Total* | $\chi^2$ | *df* | *p-value* |
|---|---|---|---|---|---|---|---|---|
| House walls | | | | | | | | |
| | High | Medium | Low | - | | | | |
| Cold spot | 4 | 60 | 36 | - | 100 | | | |
| Hot spot | 3 | 73 | 24 | - | 100 | 44.14 | 4 | **0.00** |
| Not significant | 6 | 65 | 30 | - | 100 | | | |
| Total | 4 | 66 | 30 | - | 100 | | | |
| House floor | | | | | | | | |
| | Cement | Earth | - | - | | | | |
| Cold spot | 18 | 82 | - | - | 100 | | | |
| Hot spot | 31 | 69 | - | - | 100 | | | |
| Not significant | 29 | 71 | - | - | 100 | 54.82 | 2 | **0.00** |
| Total | 26 | 74 | - | - | 100 | | | |
| Total | 64 | 36 | - | - | 100 | | | |
| Cut bush | | | | | | | | |
| | No | Yes | - | - | | | | |
| Cold spot | 62 | 38 | - | - | 100 | | | |
| Hot spot | 63 | 37 | - | - | 100 | 0.77 | 2 | 0.67 |
| Not significant | 63 | 37 | - | - | 100 | | | |
| Total | 62 | 38 | - | - | 100 | | | |
| Remove_water | | | | | | | | |
| | No | Yes | | | | | | |
| Cold spot | 63 | 37 | | | 100 | 1.52 | | 0.46 |
| Hot spot | 65 | 34 | | | 100 | | | |
| Not significant | 64 | 36 | | | 100 | | | |
| Total | 64 | 36 | | | 100 | | | |
| Animal ownership | | | | | | | | |
| Cold spot | No | Yes | | - | - | 5.95 | 2 | 0.07 |
| Hot spot | 46 | 54 | | - | 100 | | | |
| Not significant | 47 | 53 | | - | 100 | | | |
| Total | 51 | 49 | | - | 100 | | | |
| | 48 | 52 | | - | 100 | | | |
| Land use | | | | | | | | |
| | Settlement | Forest | Closed_agr | Open_agr | - | | | |
| Cold spot | 91 | 0 | 1 | 8 | 100 | | | |
| Hot spot | 92 | 0 | 2 | 6 | 100 | 8.53 | 6 | 0.20 |
| Not significant | 93 | 0 | 1 | 6 | 100 | | | |
| Total | 92 | 0 | 1 | 7 | 100 | | | |

# 5. CONCLUSION AND RECOMMENDATIONS

Malaria prevalence in the study area is a result of both malaria causing factors and control measures. This section gives a synthesis about the output of this study in line with the objectives and with an emphasis on the verification of the hypotheses. Finally, recommendations are addressed to researchers and to local authorities.

## 5.1. General conclusion

Malaria infection is influenced by those land use types that serve as anopheles mosquito habitats (breeding and resting sites). Anopheles mosquito habitats can be accurately identified through the visual interpretation of a high-resolution image such as an orthophoto (25cm resolution) and the integration of ground survey data.

Malaria prevalence (number of infected people/the whole population) is higher in the study area compared to the country statistics and is differently distributed in the surveyed villages and cells. Malaria prevalence is higher in less populated areas.

Malaria transmission is a natural phenomenon and therefore, it does not consider administrative boundaries, which are arbitrary fixed. The need to detect malaria spatial pattern in the study area inspired the use of Getis and Ord statistics (G*) with malaria prevalence at household level as input field. The output of G* proved that malaria presented clusters in the study area which were qualified of hot (significant clusters of high values at $p<0.05$), not significant and cold (significant clusters of low values at $p<0.05$) spots. The G* results lead to the rejection of the first null hypothesis suggesting the uniform distribution of malaria in the study area.

The results of G* are point based and concern only the surveyed households. However, there is a need to interpolate the results to the whole study area for planning purposes (building new settlement for example). Inverse Weighting Distance (IWD) spatial analyst tool shows that the probability of getting malaria is higher in households neighbouring another household with high malaria prevalence and the influence decreases with the distance.

Malaria is caused by different factors, grouped into environmental (altitude, land use and distances to anopheles mosquito breeding sites), demographic (household size, age, gender and distance to household with infected people) and economic (house material and animal ownership).

Malaria vulnerability can be reduced by malaria control measures consisting either in reducing the vector density or in breaking the vector-host contact. Artificial malaria control measures (ITNs, IRS, closing windows and doors, use of repellents and wearing long clothes during evening hours) are used in the study area. They are dominated by the ITNs that are distributed by the Government of Rwanda free of charge to the local population and are therefore used by more than ninety percent of the surveyed households. A considerable proportion of the surveyed households also use natural malaria control measures consisting in removing potential anopheles mosquito breeding (stagnant water) and resting (bushes) sites near houses. Treatment is another malaria control measure, which limits malaria transmission from house to house or between household members. It is influenced by the distance to Ruhuha Health Center, which is the only public health care provider in the study area.

Malaria decreases with the distance to irrigated farmland. The other environmental variables (altitude, land use and distance to water reservoirs) are not associated with malaria infection if the whole study area is considered.

In the study area, malaria infection is also related to demographic factors. Malaria infection increases with the household size. It is higher in males than in females. Children at school age (5-14 years) are more exposed to malaria. Malaria infection increases with the proximity to household with infected people.

Of the economic factors, house floors and walls are related to malaria infection while roofs do not play any role. Animal husbandry was also not found to be related to malaria infection. Houses with earth floors are characterized by a higher malaria infection compared to those with cement floors. Compared to houses with medium quality (unburnt bricks) walls, malaria infection is higher in low quality (mud) walls and lower in high quality (burnt bricks and cement blocks) walls.

Of the malaria control measures, the distance to Ruhuha Health center was proved to be related to malaria infection. Natural malaria control measures do not play any role in malaria explanation.

The significant (Wald>1, $p<0.05$) contribution of environmental (irrigated farmland), demographic (household size and households with infected people) and economic (house material) malaria causing factors and control measures (proximity to Health Center) leads to the rejection of the second null hypothesis suggesting that malaria infection is not significantly explained by the identified factors.

The logistic regression model $R^2$ is 13%. This lead to the conclusion that the identified significant variables explain malaria infection with a 13% correlation and therefore, the third null hypothesis suggesting $R^2>50\%$ is rejected. The regression model shows a poor predictive power and therefore it is not worthwhile for practical use.

Malaria clusters that are detected using G* statistics are the result of the difference in underlying factors (causing factors and control measures). ANOVA and Tukey tests proved that the proximity to irrigated farmland and the distance to household with infected people determine malaria hot spots. This also confirms the results of the logistic regression.

Malaria control measures are similar among the three cluster types. The detected clusters are characterized by different proportions of house material as proved by a chi-square test. However, as the clusters are characterized by a mixture of houses varying from high to low quality, their accurate contribution in determining malaria clusters is difficult to identify by the clustering approach. Animal ownership and malaria control measures show no correlation with malaria hot spots.

The differences of mean distance to marshlands, to water reservoirs and to households with infected people as proved by ANOVA test and the differences of proportions of house material (wall and floor) as proved by chi-square test among malaria clusters lead to the rejection of the fourth null hypothesis suggesting a lack of differences. Therefore, signifantly different variables contribute to the determination of malaria clusters in the study area.

## 5.2. Recommendations

This study is a small but crucial contribution to the resolution of the malaria problem in Ruhuha Sector. It shows that malaria is a result of many factors varying from environmental, demographic, and economic to human behaviour. The following recommendations are proposed to researchers and to local authorities

### 5.2.1. To researchers

a. Malaria clusters were identified in this study and only differences among the three main clusters (cold, not significant and hot spots) were analyzed. There is a need to do a more in depth study analyzing causes underlying each of the detected hot spots in order to identify specific intervention measures.

b. This study relied on incidence data assuming the presence of the vector. As incidence data are only collected in settlement areas, for planning purposes, the interpolation of the results will not accurately reach the whole study area. There is a need to do further research, for example using Species Distribution Modelling, based on anopheles mosquito presence data, which can be collected for the whole study area.

c. The output of this study proves a lack of difference between malaria protected and unprotected people. There is a need to do a research, including the human education and behaviour in malaria underlying factors because they might influence the use of malaria control measures.

### 5.2.2. To local authorities

a. The output of this study shows that malaria infection is higher in children at school age. There is a need to take specific measures to prevent them from malaria as it was done successfully for children under 5 years.

b. In the study area, malaria varies with the quality of houses. The housing policy should be strengthened so that the number of houses with unburnt bricks, mud walls and earth floor, which are associated with the increase of malaria, should be lowered.

# LIST OF REFERENCES


Ahmad, R., Ali, W., Nor, Z., Ismail, Z., Hadi, A., Ibrahim, M., & Lim, L. (2011). Mapping of mosquito breeding sites in malaria endemic areas in Pos Lenjang, Kuala Lipis, Pahang, Malaysia. *Malaria Journal, 10*(1), 361.

Anamzui-Ya, J. A. (2012). *Spatial analysis and mapping of cholera causing factors in Kumasi, Ghana.* Enschede: MSc Thesis.University of Twente, Faculty of Geo-Information and Earth Observation (ITC).

Autio, A., Jalkanen, S., & Roivainen, A. (2013). Nuclear imaging of inflammation:homing-associated molecules as targets. *EJNMMI Res 3*, 1.

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.

Bigoga, J. D., Ndangoh, D. N., Awono-Ambene, P. H., Patchoke, S., Fondjo, E., & Leke, R. G. F. (2012). Pyrethroid resistance in Anopheles gambiae from the rubber cultivated area of Niete, South Region of Cameroon. *Acta Tropica, 124*(3), 210-214.

Bizimana, J. P., Amer, S., Iyikirenga, L., & Rutamu, D. R. (2009). *Geo - information for urban malaria risk analysis in fringe zones of Kigali City, Rwanda.*

Bugesera District. (2013). Bugesera District Retrieved 29/07, 2013, from http://www.bugesera.gov.rw/

Chikerema, S. M., Murwira, A., Matope, G., & Pfukenyi, D. M. (2013). Spatial modelling of Bacillus anthracis ecological niche in Zimbabwe. *Preventive Veterinary Medicine, 111*(1–2), 25-30.

Crawley, M. J. (2007). *The R Book*. London: John Wiley&Sons,Ltd.

Cromley, E. K., & MCLafferty, S. L. (2012). *GIS and Public Health, 2nd Edition*. Ney York, London: The Guilford Press.

Donnelly, M. J., PJ, M., Lengeler, C., Bates, I., D'Alessandro, U., Barnish, G., . . . Mutero, C. (2005). Malaria and urbanization in sub-Saharan Africa. *Malaria Journal*, 4:12.

Fanello, C. I., Karema, C., van Doren, W., Van Overmeir, C., Ngamije, D., & D'Alessandro, U. (2007). A randomised trial to assess the safety and efficacy of artemether–lumefantrine (Coartem®) for the treatment of uncomplicated Plasmodium falciparum malaria in Rwanda. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 101*(4), 344-350.

Field, A. (2009). *Discovering Statistics Using SPSS, Third Edition*. London-Thousand Oaks-New Delhi: SAGE Publications.

Garmin Ltd. (2003) Garmin etrex Personal navigator. Owner's manual. Taiwan: Garmin Ltd.

Gascon, J., Pluymaekakrs, J., & Bada, J. L. (1986). Chloroquine-resistant falciparum malaria from Rwanda. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 80*(2), 351.

Gosoniu, L., Vounatsou, P., Sogoba, N., Maire, N., & Smith, T. (2009). Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. *Comput. Stat. Data Anal., 53*(9), 3358-3371.

Grillet, M.-E., Jordan, G. J., & Fortin, M.-J. (2010). State transition detection in the spatio-temporal incidence of malaria. *Spatial and Spatio-temporal Epidemiology, 1*(4), 251-259.

Hassan, A. N., Nogoumy, N. E., & Kassem, H. A. (2013). Characterization of landscape features associated with mosquito breeding in urban Cairo using remote sensing. *The Egyptian Journal of Remote Sensing and Space Science*(0).

Hay, S. I., Graham, A. J., & Rogers, D. J. (2006). *Global Mapping of Infectious Diseases-Methods,Examples and Emerging Applications*. Oxford,UK: Academic Press.

Hay, S. I., Snow, R. W., & Rogers, D. J. (1998). From Predicting Mosquito Habitat to Malaria Seasons Using Remotely Sensed Data: Practice, Problems and Perspectives. *Parasitology Today, 14*(8), 306-313.

Ivan, E., Crowther, N. J., Rucogoza, A. T., Osuwat, L. O., Munyazesa, E., Mutimura, E., . . . Grobusch, M. P. (2012). Malaria and helminthic co-infection among HIV-positive pregnant women: Prevalence and effects of antiretroviral therapy. *Acta Tropica, 124*(3), 179-184.

Jung, R. K. (2001). *History of Malaria and Malaria Control in Nepal*. New Delhi: Aravali Printers & Publishers P.Ltd. New Delhi-20.

Karema, C., Aregawi, M., Rukundo, A., Kabayiza, A., Mulindahabi, M., Fall, I., . . . Binagwaho, A. (2012). Trends in malaria cases, hospital admissions and deaths following scale-up of anti-malarial interventions, 2000-2010, Rwanda. *Malaria Journal, 11*(1), 236.

Karema, C., Fanello, C. I., Van Overmeir, C., Van geertruyden, J.-P., van Doren, W., Ngamije, D., & D'Alessandro, U. (2006). Safety and efficacy of dihydroartemisinin/piperaquine (Artekin®) for

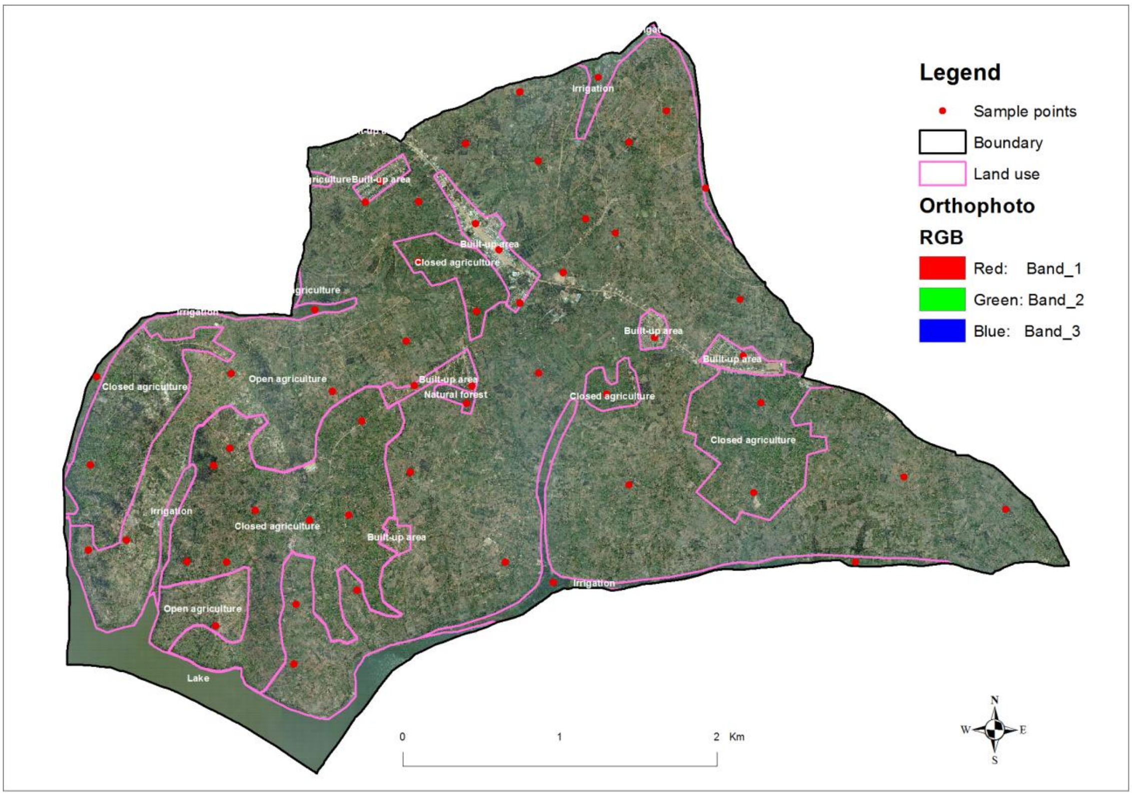
the treatment of uncomplicated Plasmodium falciparum malaria in Rwandan children. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 100*(12), 1105-1111.
Killeen, G. F., & Smith, T. A. (2007). Exploring the contributions of bed nets, cattle, insecticides and excitorepellency to malaria control: a deterministic model of mosquito host-seeking behaviour and mortality. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 101*(9), 867-880.
Konradsen, Matsuno, Y., Amerasinghe, F. P., Amerasinghe, P. H., & van der Hoek, W. (1998). Anopheles culicifacies breeding in Sri Lanka and options for control through water management. *Acta Tropica, 71*(2), 131-138.
Lindsay, S. W., & Martins, W. J. M. (1998). Malaria in the African Highlands: Past, Present, and Future. *Bulletin of the World Health Organization 76*(1), 33–45.
Liu, X.-B., Qi-Yong Liu, Yu-Hong Guo, Jing-Yi Jiang, Dong-Sheng Ren, Guang-Chao Zhou, . . . Gu, W.-D. (2011). The abundance and host-seeking behavior of culicine species (Diptera: Culicidae) and Anopheles sinensis in Yongcheng city, people's Republic of China. *Parasites & Vectors, 4*, 221.
Lubetzky-Vilnai, A., Ciol, M., & McCoy, S. W. (2013). Statistical Analysis of Clinical Prediction Rules for Rehabilitation Interventions: Current State of the Literature. *Archives of Physical Medicine and Rehabilitation*(0).
Machado-Machado, E. A. (2012). Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling. *Applied Geography, 33*(0), 82-93.
Machault, V., Vignolles, C., Borchi, F., Vounatsou, P., Pages, F., Briolant, S., . . . Rogier, C. (2011). The use of remotely sensed environmental data in the study of malaria. *Geospatial Health 5*(2), 151-168.
Mackinnon, M. J. (2005). Drug resistance models for malaria. *Acta Tropica, 94*(3), 207-217.
Malaria Atlas Project. (2013). Rwanda:Mosquito vectors and Malaria Risk Retrieved 30/06, 2013, from http://www.map.ox.ac.uk/explore/countries/rwa/#!mosquito-vectors
Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage, 53*(1), 103-118.
Munga, S., Noboru, M., Guofa, Z., Emmanue, l. M., Okeyo-owuor, J. B., Andrew K., G., & Yan, G. (2006). Association between Land Cover and Habitat Productivity of Malaria Vectors in Western Kenyan Highlands. *American Journal of Tropical Medicine and Hygiene, 74*(1), 69-75.
Mutero, C. M., Blank, H., Konradsen, F., & van der Hoek, W. (2000). Water management for controlling the breeding of Anopheles mosquitoes in rice irrigation schemes in Kenya. *Acta Tropica, 76*(3), 253-263.
Mwangangi, J. M., Ephantus J Muturi, Simon M Muriu, Joseph Nzovu, Midega, J. T., & Mbogo, a. C. (2013). The role of Anopheles arabiensis and Anopheles coustani in indoor and outdoor malaria transmission in Taveta District, Kenya. *Parasites & Vectors 6*, 114.
Nagasaki University. (2007). Spatial Relationship between Adult Malaria Vector Abundance and Environmental Factors in Western Kenya Highlands, from http://naosite.lb.nagasaki-u.ac.jp/dspace/handle/10069/19218
Ngom, R., & Siegmund, A. (2010). Urban malaria in Africa: an environmental and socio-economic modelling approach for Yaoundé, Cameroon. *Natural Hazards, 55*(3), 599-619.
NISR. (2007). *Service Provision Assessment Survey* Kigali: Government of Rwanda.
NISR. (2010). *Rwanda Baseline map*. Kigali: Government of Rwanda.
NISR. (2012a). Rwanda Demographic and Health Survey 2010 (pp. 411). Kigali, Rwanda.
NISR. (2012b). *Statistical Year Book 2012*. Kigali: Government of Rwanda.
Packard, R. M. (2007). *The making of a tropical disease: A short history of malaria.* Baltimore, MD: The Johns Hopkins University Press.
Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*(3–4), 231-259.
Plantinga, A. J., Helvoigt, T. L., & Walker, K. (2014). Critical habitat for threatened and endangered species: How should the economic costs be evaluated? *Journal of Environmental Management, 134*(0), 127-135.
Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment. *Procedia - Social and Behavioral Sciences, 21*(0), 317-325.
President's Malaria Initiative. (2013). Malaria Operational Plan,Rwanda FY (pp. 55). Kigali.
Rothman, K. J. (2012). *Epidemiology, An Introduction*. Oxford: Oxford University Press,Inc.
Sipe, N. G., & Dale, P. (2003). Challenges in using geographic information systems (GIS) to understand and control malaria in Indonesia. *Malaria Journal 2003* 2:36.

Smith, M. W., Macklin, M. G., & Thomas, C. J. (2013). Hydrological and geomorphological controls of malaria transmission. *Earth-Science Reviews, 116*(0), 109-127.
Snow, J. (1855). *On the mode of communication of cholera*. London: John Churchill.
Stephen, M., Noboru Minakawa, Guofa Zhou, Emmanuel Mushinzimana, Okeyo-Owuor J. Barrack, Andrew K. Githeko, & Yan, G. (2006). Association between land cover and habitat productivity of malaria vectors in western kenyan highlands. *Am. J. Trop. Med. Hyg., 74*(1), 69–75.
Stephenson, N., Davis, M., Flowers, P., MacGregor, C., & Waller, E. (2014). Mobilising "vulnerability" in the public health response to pandemic influenza. *Social Science & Medicine, 102*(0), 10-17.
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.
Stoler, J., Weeks, J. R., Getis, A., & Hill, A. G. (2009). Distance Threshold for the Effect of Urban Agriculture on Elevated Self-reported Malaria Prevalence in Accra, Ghana. *Am. J. Trop. Med. Hyg., 80*(4), 547–554.
Stratton, L., O'Neill, M. S., Kruk, M. E., & Bell, M. L. (2008). The persistent problem of malaria: Addressing the fundamental causes of a global killer. *Social Science & Medicine, 67*(5), 854-862.
The free dictionary. (2013). Medical dictionnary Retrieved 20/08, 2013, from http://medical-dictionary.thefreedictionary.com/risk).
Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography, 46*, 234–240.
Verdonschot, P. F. M., & Besse-Lototskaya, A. A. (2013). Flight distance of mosquitoes (Culicidae): A metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands. *Limnologica - Ecology and Management of Inland Waters*(0).
Watson, D. F., & Philip, G. M. (1985). A refinement of inverse distance weighted interpolation. *Geo-Processing, 2*, 15- 327.
WHO. (1982). *Manual on environmental management for mosquito control with special emphasis on malaria vectors*. Geneva: WHO.
WHO. (2008). World Malaria Report 2008 (pp. 183). Geneva: World Health Organisation.
WHO. (2009). *Malaria Microscopy Quality Assurance Manual ,Version 1*. Geneva.
WHO. (2010). World Malaria Report 2010 *WHO Global Malaria Programme* (pp. 137). Geneva: World Health Organisation

WHO. (2012). World Malaria Report 2012. Geneva: World Health Organisation.
WHO. (2013a). Global Health Observatory Data Repository. *Rwanda statistics summary (2002 - present)* Retrieved 30/06, 2013, from http://apps.who.int/gho/data/node.country.country-RWA?lang=en
WHO. (2013b). Malaria Fact sheet N°94 Retrieved 30/06, 2013, from http://www.who.int/mediacentre/factsheets/fs094/en/
Wielgosz, B., Mangheni, M., Tsegai, D., & Ringler, C. (2012). Malaria and Agriculture. A Global Review of the Literature With a Focus on the Application of Integrated Pest and Vector Management in East Africa and Uganda. *IFPRI Discussion Paper* (01232), 1-64.
Yamamoto, S., Louis, V. R., Sié, A., & Sauerborn, R. (2010). Household risk factors for clinical malaria in a semi-urban area of Burkina Faso: a case–control study. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 104*(1), 61-65.
Zayeri, F., Salehi, M., & Pirhosseini, H. (2011). Geographical mapping and Bayesian spatial modeling of malaria incidence in Sistan and Baluchistan province, Iran. *Asian Pacific Journal of Tropical Medicine, 4*(12), 985-992.
Zeile, I., Gahutu, J.-B., Shyirambere, C., Steininger, C., Musemakweri, A., Sebahungu, F., . . . Mockenhaupt, F. P. (2012). Molecular markers of Plasmodium falciparum drug resistance in southern highland Rwanda. *Acta Tropica, 121*(1), 50-54.

APPENDICES

**Appendix 1.** Identified land use during the pre-fieldwork stage



**Appendix 2.** Interview guideline questions

| Interview guideline questions<br>**Interviewer:**<br>**Date:**<br>**Cell:**<br>**Village:**<br>**Interviewee:**<br>**Position** | |
|---|---|
| What are the types of anopheles habitats in your area? | |
| How far from those habitats, are the cases you are receiving? | |
| What is the age group that is more infected by malaria? | |
| What is the gender that is more infected by malaria? | |
| What are control measures that are used in your area? | |
| What are the most used control measures? | |
| Others | |

**Appendix 3.**Field data collection sheet

| Field data collection sheet<br>**Observer:**<br>**Date:**<br>**Sector:** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Point N°** | **Cell** | **Village** | **X** | **Y** | **Elevation** | **Land use type** | **Observation** |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

**Appendix 4.** Sampled training points

| *Sector* | *cell* | *x* | *y* | *Elevation* | *Cover type* | *Land_use* | *Name* |
|---|---|---|---|---|---|---|---|
| Ruhuha | Bihari | 504027 | 9742790 | 1508 | Eucalyptus | Forest | |
| Ruhuha | Bihari | 503921 | 9742720 | 1512 | Coffee | Closed agriculture | |
| Ruhuha | Bihari | 503316 | 9742852 | 1501 | Built up | Settlement | |
| Ruhuha | Bihari | 503848 | 9741984 | 1481 | Crops | Open agriculture | |
| Ruhuha | Bihari | 503850 | 9741154 | 1433 | Crops | Open agriculture | |
| Ruhuha | Bihari | 503957 | 9742252 | 1465 | Banana | Closed agriculture | |
| Ruhuha | Bihari | 504614 | 9742294 | 1479 | Banana | Closed agriculture | |
| Ruhuha | Bihari | 504844 | 9742152 | 1471 | Eucalyptus | Forest | |
| Ruhuha | Bihari | 505812 | 9742074 | 1433 | Built up | Settlement | |
| Ruhuha | Bihari | 506184 | 9742272 | 1413 | Rice | Irrigation | Kibaza marshland |
| Ruhuha | Kindama | 507112 | 9742784 | 1424 | Built up | Settlement | |
| Ruhuha | Kindama | 506907 | 9745194 | 1461 | Built up | Settlement | |
| Ruhuha | Kindama | 506984 | 9745668 | 1453 | Crops | Open agriculture | |
| Ruhuha | Kindama | 507213 | 9745900 | 1449 | Eucalyptus | Forest | |
| Ruhuha | Kindama | 507866 | 9745958 | 1405 | Rice | Irrigation | Kijambari marshland |
| Ruhuha | Ruhuha | 506444 | 9744924 | 1459 | Crops | Open agriculture | |
| Ruhuha | Ruhuha | 506355 | 9745268 | 1452 | Forest | Forest | |
| Ruhuha | Ruhuha | 506240 | 9745802 | 1445 | Banana | Closed agriculture | |
| Ruhuha | Ruhuha | 506009 | 9745946 | 1449 | Coffee | Closed agriculture | |
| Ruhuha | Ruhuha | 505829 | 9745142 | 1486 | Built up | Settlement | |
| Ruhuha | Ruhuha | 505637 | 9745084 | 1500 | Built up | Settlement | |
| Ruhuha | Gatanga | 505750 | 9744388 | 1474 | Banana | Closed agriculture | |
| Ruhuha | Gatanga | 505565 | 9743830 | 1473 | Built up | Settlement | |
| Ruhuha | Gatanga | 506311 | 9743482 | 1484 | Forest | Forest | |
| Ruhuha | Gatanga | 505262 | 9743482 | 1474 | Built up | Settlement | |
| Ruhuha | Bihari | 501890 | 9743072 | 1364 | Forest | Forest | |
| Ruhuha | Bihari | 501763 | 9742946 | 1354 | Rice | Irrigation | Nyaburiba marshland |
| Ruhuha | Bihari | 501868 | 9742250 | 1352 | Crops | Open agriculture | |
| Ruhuha | Bihari | 503097 | 9741524 | 1424 | Crops | Open agriculture | |
| Ruhuha | Bihari | 503672 | 9741536 | 1435 | Built up | Settlement | |
| Ruhuha | Bihari | 504206 | 9741620 | 1429 | Banana | Closed agriculture | |
| Ruhuha | Bihari | 504225 | 9741796 | 1436 | Banana | Closed agriculture | |
| Ruhuha | Bihari | 503814 | 9742192 | 1485 | Coffee | Closed agriculture | |
| Ruhuha | Bihari | 503202 | 9742136 | 1457 | Crops | Open agriculture | |
| Ruhuha | Bihari | 502873 | 9748264 | 1409 | Crops | Open agriculture | |
| Ruhuha | Bihari | 502715 | 9742220 | 1386 | Crops | Open agriculture | |
| Ruhuha | Bihari | 502433 | 9742274 | 1406 | Forest | Forest | |
| Ruhuha | Bihari | 502260 | 9742194 | 1430 | Built up | Settlement | |
| Ruhuha | Gatanga | 506031 | 9744410 | 1499 | Built up | Settlement | |
| Ruhuha | Gatanga | 505563 | 9744354 | 1479 | Banana | Closed agriculture | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Ruhuha | Gatanga | 504932 | 9744262 | 1450 | Crops | Open agriculture | |
| Ruhuha | Gatanga | 504421 | 9744306 | 1404 | Rice | Irrigation | Nyaburiba marshland |
| Ruhuha | Gatanga | 504118 | 9743684 | 1445 | Crops | Open agriculture | |
| Ruhuha | Gatanga | 504815 | 9743314 | 1491 | Built up | Settlement | |
| Ruhuha | Gatanga | 505554 | 9743772 | 1486 | Built up | Settlement | Gatanga Center |
| Ruhuha | Gatanga | 505654 | 9743312 | 1460 | Forest | Forest | |
| Ruhuha | Gatanga | 505949 | 9742318 | 1426 | Built up | Settlement | Kibaza |
| Ruhuha | Ruhuha | 505944 | 9745766 | 1462 | Forest | Forest | |
| Ruhuha | Ruhuha | 506206 | 9745996 | 1451 | Built up | Settlement | |
| Ruhuha | Ruhuha | 506567 | 9746290 | 1438 | Crops | Open agriculture | |
| Ruhuha | Ruhuha | 505881 | 9746108 | 1440 | Forest | Forest | |
| Ruhuha | Ruhuha | 505119 | 9745840 | 1486 | Built up | Settlement | RuhuhaII |
| Ruhuha | Ruhuha | 504821 | 9745322 | 1462 | Banana | Closed agriculture | |
| Ruhuha | Ruhuha | 504484 | 9745560 | 1496 | Built up | Settlement | |
| Ruhuha | Ruhuha | 505645 | 9745192 | 1508 | Built up | Settlement | |
| Ruhuha | Ruhuha | 506048 | 9744602 | 1489 | Built up | Settlement | |

**Appendix 5**.Potential anopheles breeding sites

| Sector | cell | X | Y | Elevation | Cover Type | Name |
|---|---|---|---|---|---|---|
| Ruhuha | Bihari | 502007 | 9743488 | 1361 | Rice | Nyaburiba marshland |
| Ruhuha | Bihari | 501802 | 9742648 | 1357 | Rice | Nyaburiba marshland |
| Ruhuha | Bihari | 501726 | 9742648 | 1357 | Papyrus | Nyaburiba marshland |
| Ruhuha | Bihari | 502008 | 9741346 | 1351 | Trees | Cyohoha lake shore |
| Ruhuha | Bihari | 502567 | 9741622 | 1358 | Ibikangaga | Nyarufunzo marshland |
| Ruhuha | Bihari | 502607 | 9741720 | 1350 | Perennial crops | Marshland |
| Ruhuha | Bihari | 504377 | 9741786 | 1431 | Water reservoir | |
| Ruhuha | Bihari | 504705 | 9741846 | 1428 | Water reservoir | |
| Ruhuha | Bihari | 502715 | 9742220 | 1386 | Perennial crops | Nduhura marshland |
| Ruhuha | Gatanga | 504421 | 9744306 | 1404 | Rice | Nyaburiba marshland |
| Ruhuha | Gatanga | 503981 | 9744180 | 1387 | Rice | Nyaburiba marshland |
| Ruhuha | Gatanga | 505708 | 9743296 | 1467 | Water reservoir | |
| Ruhuha | Gatanga | 505764 | 9743258 | 1446 | Water reservoir | |
| Ruhuha | Gatanga | 505778 | 9743254 | 1451 | Water reservoir | |
| Ruhuha | Gatanga | 505862 | 9741928 | 1416 | Water reservoir | Butereri Primary School |
| Ruhuha | Gatanga | 506347 | 9741728 | 1376 | Rice | Kibaza marshland |
| Ruhuha | Gatanga | 506557 | 9741614 | 1358 | Rice | Kibaza marshland |
| Ruhuha | Ruhuha | 506025 | 9745102 | 1475 | Water reservoir | |
| Ruhuha | Ruhuha | 506619 | 9746482 | 1448 | Water reservoir | |
| Ruhuha | Ruhuha | 506756 | 9746526 | 1415 | Water reservoir | |
| Ruhuha | Ruhuha | 506889 | 9746680 | 1394 | Rice | Nyabaranga marshland |
| Ruhuha | Ruhuha | 505617 | 9746065 | 1468 | Water reservoir | |
| Ruhuha | Ruhuha | 504764 | 9745356 | 1465 | Water reservoir | |
| Ruhuha | Kindama | 507142 | 9743524 | 1488 | Water reservoir | |
| Ruhuha | Kindama | 507162 | 9743502 | 1478 | Water reservoir | |
| Ruhuha | Kindama | 507203 | 9743406 | 1474 | Water reservoir | |
| Ruhuha | Kindama | 506664 | 9744974 | 1462 | Water reservoir | |
| Ruhuha | Kindama | 506641 | 9744964 | 1458 | Water reservoir | |
| Ruhuha | Kindama | 507486 | 9747294 | 1378 | Water reservoir | |

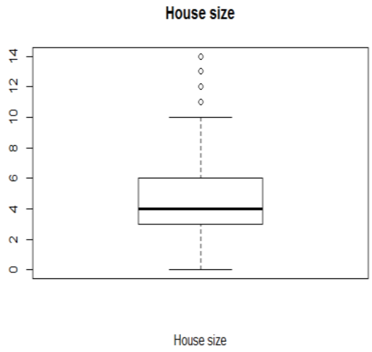
**Appendix 6.** Regression model steps

| | | B | S.E. | Wald | df | P-value | OR | 95% C.I.for OR | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Lower | Upper |
| Step 1 | H_size | .184 | .026 | 52.190 | 1 | .000 | 1.202 | 1.144 | 1.264 |
| | Constant | -2.906 | .138 | 443.605 | 1 | .000 | .055 | | |
| Step 2 | H_size | .230 | .027 | 70.428 | 1 | .000 | 1.259 | 1.193 | 1.328 |
| | Cement | -1.425 | .186 | 58.831 | 1 | .000 | .240 | .167 | .346 |
| | Constant | -2.875 | .145 | 395.871 | 1 | .000 | .056 | | |
| Step 3 | H_size | .227 | .028 | 67.862 | 1 | .000 | 1.254 | 1.188 | 1.324 |
| | Dist_HC | -.816 | .163 | 24.949 | 1 | .000 | .442 | .321 | .609 |
| | Cement | -1.515 | .189 | 64.585 | 1 | .000 | .220 | .152 | .318 |
| | Constant | -2.269 | .183 | 153.260 | 1 | .000 | .103 | | |
| Step 4 | H_size | .242 | .028 | 74.114 | 1 | .000 | 1.273 | 1.205 | 1.345 |
| | Dist_HC | -.868 | .158 | 30.218 | 1 | .000 | .420 | .308 | .572 |
| | Low | .662 | .121 | 29.974 | 1 | .000 | 1.939 | 1.530 | 2.457 |
| | Cement | -1.306 | .194 | 45.155 | 1 | .000 | .271 | .185 | .396 |
| | Constant | -2.582 | .195 | 176.058 | 1 | .000 | .076 | | |
| Step 5 | H_size | .245 | .028 | 75.862 | 1 | .000 | 1.278 | 1.209 | 1.350 |
| | Dist_marsh | -.745 | .224 | 11.016 | 1 | .001 | .475 | .306 | .737 |
| | Dist_HC | -1.079 | .174 | 38.354 | 1 | .000 | .340 | .242 | .478 |
| | Low | .699 | .122 | 32.803 | 1 | .000 | 2.011 | 1.583 | 2.554 |
| | Cement | -1.252 | .195 | 41.275 | 1 | .000 | .286 | .195 | .419 |
| | Constant | -1.863 | .290 | 41.384 | 1 | .000 | .155 | | |
| Step 6 | H_size | .247 | .028 | 75.966 | 1 | .000 | 1.280 | 1.211 | 1.353 |
| | Dist_HH | .441 | .110 | 16.143 | 1 | .000 | 1.554 | 1.253 | 1.927 |
| | Dist_marsh | -1.031 | .236 | 19.175 | 1 | .000 | .357 | .225 | .566 |
| | Dist_HC | -.981 | .174 | 31.764 | 1 | .000 | .375 | .267 | .527 |
| | Low | .791 | .124 | 40.671 | 1 | .000 | 2.205 | 1.729 | 2.811 |
| | Cement | -1.128 | .196 | 32.995 | 1 | .000 | .324 | .220 | .476 |
| | Constant | -2.528 | .337 | 56.420 | 1 | .000 | .080 | | |
| Step 7 | H_size | .253 | .029 | 77.368 | 1 | .000 | 1.288 | 1.217 | 1.363 |
| | Dist_HH | .463 | .110 | 17.825 | 1 | .000 | 1.589 | 1.282 | 1.970 |
| | Dist_marsh | -1.005 | .235 | 18.238 | 1 | .000 | .366 | .231 | .580 |
| | Dist_HC | -1.013 | .174 | 33.918 | 1 | .000 | .363 | .258 | .511 |
| | High | -1.487 | .618 | 5.788 | 1 | .016 | .226 | .067 | .759 |
| | Low | .782 | .124 | 39.968 | 1 | .000 | 2.186 | 1.716 | 2.786 |
| | Cement | -.942 | .200 | 22.243 | 1 | .000 | .390 | .264 | .577 |
| | Constant | -2.591 | .339 | 58.499 | 1 | .000 | .075 | | |

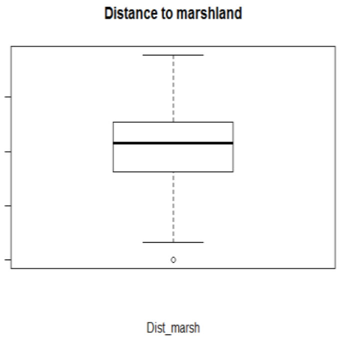
**Appendix 7.** Mean distances for different malaria clusters

| *Variable* | *Cluster* | | | | *95% CI* | | | |
|---|---|---|---|---|---|---|---|---|
| | | *N* | *Mean* | *Std. Deviation* | *Lower Bound* | *Upper Bound* | *Min* | *Max* |
| Alt | Cold spot | 1186 | 1468.64 | 36.05 | 1466.58 | 1470.69 | 809.00 | 1567.00 |
| | Not significant | 1117 | 1465.96 | 82.23 | 1461.14 | 1470.79 | 0.00 | 1575.00 |
| | Hot spot | 903 | 1465.07 | 58.99 | 1461.22 | 1468.92 | 0.00 | 1560.00 |
| | Overall | 3206 | 1466.70 | 61.78 | 1464.56 | 1468.84 | 0.00 | 1575.00 |
| Dist_marsh | Cold spot | 1186 | 1040.65 | 327.95 | 1021.97 | 1059.33 | 168.24 | 1891.04 |
| | Not significant | 1117 | 1027.03 | 371.13 | 1005.24 | 1048.81 | 0.00 | 1715.28 |
| | Hot spot | 903 | **995.01** | 356.23 | 971.75 | 1018.28 | 0.00 | 1762.03 |
| | Overall | 3206 | 1023.05 | 351.83 | 1010.87 | 1035.23 | 0.00 | 1891.04 |
| Dist_HH | Cold spot | 1186 | 96.20 | 85.41 | 91.34 | 101.07 | 0.00 | 609.07 |
| | Not significant | 1117 | 96.29 | 89.07 | 91.06 | 101.52 | 0.00 | 470.84 |
| | Hot spot | 903 | **84.18** | 87.03 | 78.50 | 89.87 | 0.00 | 450.03 |
| | Overall | 3206 | 92.85 | 87.30 | 89.82 | 95.87 | 0.00 | 609.07 |
| Dist_reser | Cold spot | 1186 | **691.29** | 539.65 | 660.55 | 722.04 | 25.50 | 3765.83 |
| | Not significant | 1117 | 758.44 | 546.40 | 726.36 | 790.52 | 0.00 | 3214.94 |
| | Hot spot | 903 | 702.13 | 490.52 | 670.10 | 734.17 | 0.00 | 3441.87 |
| | Overall | 3206 | 717.74 | 529.39 | 699.41 | 736.07 | 0.00 | 3765.83 |
| Dist_HC | Cold spot | 1186 | 1989.03 | 1069.90 | 1928.07 | 2049.98 | 24.35 | 5265.05 |
| | Not significant | 1117 | 2031.57 | 1124.50 | 1965.55 | 2097.59 | 0.00 | 4884.18 |
| | Hot spot | 903 | 2019.16 | 1041.76 | 1951.12 | 2087.19 | 0.00 | 4908.72 |
| | Overall | 3206 | 2012.34 | 1081.35 | 1974.89 | 2049.78 | 0.00 | 5265.05 |

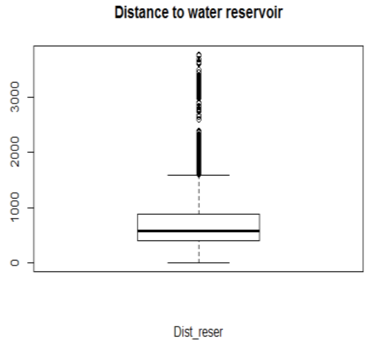
## **Appendix 8.** Test for normality



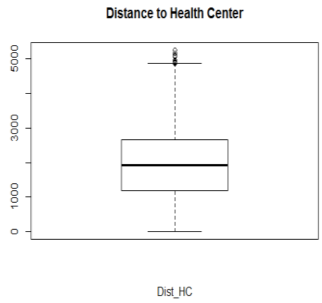

W = 0.9551, p-value < 2.2e-16



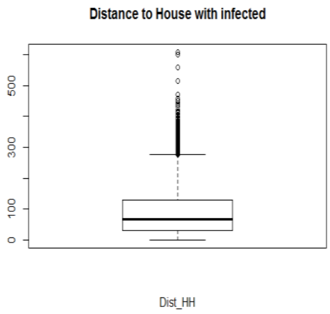

W = 0.9332, p-value < 2.2e-16



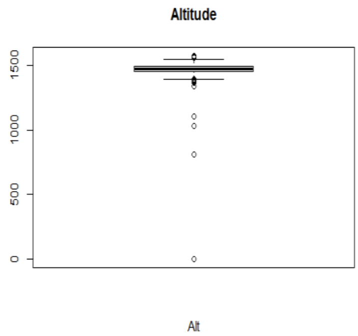

W = 0.8262, p-value < 2.2e-16




W = 0.7836, p-value < 2.2e-16




W = 0.8262, p-value < 2.2e-16




W = 0.3462, p-value < 2.2e-16