CLIMATIC VARIABLE SELECTION USING RANDOM FORESTS REGRESSION FOR MALARIA TRANSMISSION MODELLING IN MPUMALANGA PROVINCE, SOUTH AFRICA

Submitted in fulfillment of the academic requirements for the degree of Master of Science: Environmental Science

School of Agricultural, Earth and Environmental Sciences in the College of Agriculture, Engineering and Science of the University of KwaZulu-Natal

Thandi Kapwata

Student Number: 204502785

Supervisor:

Dr M. Gebreslasie

June 2015

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

DECLARATION 1 - PLAGIARISM

I, Thandi Kapwata, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written but the general information attributed to them has been referenced

b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed

……

CONTENTS

List of Figures iv
List of Tables v
List of appendices vi
ABSTRACT vii
Chapter 1 . INTRODUCTION 1
1.1 Background 1
1.2 Aim and Objectives of the study 2
1.2.1 Aim 2
1.2.2 Objectives 2
1.3 Limitations of the study 2
1.4 Structure of the thesis 3
Chapter 2 . LITERATURE REVIEW 4
2.1 BIOLOGY OF MALARIA 4
2.1.1 Vector Biology 4
2.1.2 Parasite Biology 6
2.1.3 Symptoms and effects of Malaria 7
2.2 SOCIAL AND ECONOMIC IMPACT OF MALARIA 8
2.3 MALARIA CONTROL AND TREATMENT 9
2.3.1 Vector Control 9
2.4 MALARIA TRANSMISSION 16
2.4.1 Intrinsic factors of malaria transmission 18
2.4.2 Extrinsic factors of malaria transmission 18
2.5 CLIMATIC RISK FACTORS OF MALARIA 19
2.5.1 Temperature 19
2.5.2 Rainfall 20
2.5.3 Land use and land cover 21
2.5.4 Normalized Difference Vegetation Index (NDVI) 22
2.5.5 Topography 22
2.5.6 Humidity 23
2.5.7 Climate change 23
2.6 APPLICATION OF GEOGRAPHICAL INFORMATION SYSTEMS, REMOTE SENSING AND SPATIAL STATISTICS IN MALARIA STUDIES 24
2.6.1 Geographical Information Systems (GIS) 25
2.6.2 Remote Sensing (RS) 26

2.6.3 Spatial statistics 27
2.7 CONCLUSION 30
Chapter 3 . STUDY AREA 32
3.1 Physical and Human Geography of Mpumalanga 32
3.2 Malaria Epidemiology Intervention Strategies 35
Chapter 4 . METHODOLGY 38
4.1 Malaria Case Data 38
4.2 Climatic data 39
4.2.1 Land surface temperature (LST) 39
4.2.2 Rainfall 39
4.2.3 Humidity 40
4.2.4 Natural vegetation index (NDVI) 40
4.2.5 Altitude 40
4.2.6 Lag rainfall and lag temperature 40
4.2.7 Extraction of climatic data 40
4.3 Statistical method 40
4.3.1 Random Forest for Regression 41
4.3.2 Model validation 43
Chapter 5 . RESULTS AND DISCUSSION 43
Chapter 6 . CONCLUSIONS AND RECOMMENDATIONS 69
6.1 Conclusions 69
6.2 Limitations and recommendations 70
Chapter 7. References 71
Chapter 8. Appendices 83

List of Figures

Figure 3-1: Map showing location of Mpumalanga Province 35
Figure 5-1. The weekly number of reported malaria cases in Ehlanzeni District for 2006 44
Figure 5-2. The monthly number of reported malaria cases in Ehlanzeni District for 2006 44
Figure 5-3. Variable importance index of predictor variables averaged from 50 runs of RF 47
Figure 5-4. Variable importance index of random forest algorithm using top 3 statistically significant variables 50
Figure 5-5: Association between altitude and malaria cases 52
Figure 5-6. Trend of lag rainfall and malaria cases 53
Figure 5-7. Trend of NDVI and malaria cases 54
Figure 5-8. Trend of temperature and malaria cases 55
Figure 5-9. Trend of rainfall and malaria cases 55
Figure 5-10. Trend of lag temperature and malaria cases 56
Figure 5-11. Trend of relative humidity and malaria cases 57
Figure 5-12. Observed malaria cases vs predicted malaria cases 59
Figure 5-13. Maps showing predicted cases of malaria by sub-place 62
Figure 5-14. Error estimates of monthly predictive models 68

List of Tables

Table 2- 1. Insecticides recommended by WHO for IRS 10
Table 2- 2. Insecticides used in ITN impregnation 12

Table 5-1. Variables selected as top predictor variables for each month 51

List of appendices

Appendix 1: Graphs of total malaria cases for all sub places by month……92

ABSTRACT

Malaria is one of the wold's most prevalent vector borne diseases with sub-Saharan Africa bearing the highest burden of reported cases. Climate is one of the major determinant factors of malaria transmission as it influences the spatial and temporal pattern of transmission. It is therefore important to be able to understand the relationship between climatic variables and malaria transmission because an understanding of the interactions between them at a local level is an important part in potential outbreaks, targeting vector control strategies, and developing malaria early warning systems.

This study covered the Ehlanzeni district of Mpumalanga province in South Africa. It was aimed at determining the climatic variable importance of temperature, lag temperature, rainfall, lag rainfall, humidity, altitude and NDVI in relation to malaria transmission. The random forest algorithm was used to relate the climatic variables extracted from remote sensing imagery and malaria case data collected from health facilities in order to establish individual measures of variable importance and to develop a spatial and temporal prediction models.

In this study altitude appeared to be the most responsible variable for malaria transmission because it was most frequently selected as one of the top variables with the highest variable importance followed by NDVI and temperature. The combination of climatic variables that produced the highest coefficient of determination values was altitude, NDVI, and temperature. This suggests that these 3 variables have high predictive capabilities and as a result they should be selected for spatial and temporal modelling of malaria.

Furthermore, it was expected that the predictive models generated by the random forest algorithm could be used as an operational malaria early warning system using forecast climatic variable identified in this study in order to assist in containing any potential reoccurrence of malaria after elimination,

CHAPTER 1 . INTRODUCTION

1.1 Background

The African continent bears the highest burden of malaria globally with 85% of cases being reported with sub-Saharan Africa accounting for 90% of malaria-related deaths (RBM, 2013). This is due to a combination of reasons including the climate of much of sub-Saharan Africa which is suitable for malaria transmission, economic constrain to fund national treatment and control measures, and resistance to antimalarial drugs and insecticides (Blumberg and Frean, 2002). South Africa is situated in a low malaria transmission zone characterized by lower incidence of confirmed cases (Sharp et al., 2007). Malaria is endemic in three provinces of South Africa; transmission is seasonal and predominantly occurs when temperatures are favourable for vector survival (Gerritsen et al., 2008). However, transmission in the country has decreased over the years, and is now limited to the low-altitude, north-eastern parts of Limpopo, Mpumalanga and KwaZulu-Natal Provinces. *Anopheles arabiensis* is the only mosquito vector that transmits malaria. *Anopheles merus* is found in abundance in some areas but has not been implicated in malaria transmission despite the fact that it plays a significant role in other countries where it occurs. *Anopheles funestus* used to be a very important vector before it was eliminated through years of Indoor Residual Spray (IRS) with Dichlorodiphenyltrichloroethane (DDT) (Hargreaves et al., 2000). In South Africa, approximately 95 % of malaria infections are due to the parasite, *Plasmodium falciparum* (Grimwade et al., 2004). About 5 million people, (10% of overall population) are at risk of malaria infection.

Mpumalanga province contributes 44% of the country's notified malaria cases; malaria in the province is seasonal due to the alternate dry and rainy seasons. It follows the pattern of starting with the first rains in October, peaking in January and remaining high until May (Silal et al., 2013, Ngomane and De Jager, 2012). Mpumalanga has maintained a successful control program encompassing rapid detection and treatment of confirmed malaria cases at primary health care facilities and vector control through IRS with insecticides and focal larviciding (Govere et al., 2000). As a result malaria cases are greatly reduced in the province to a point, where the province achieved the minimum elimination incidence level according to the RBM frame work strategy. Maharaj et al. (2012) suggested that Mpumalanga intervention efforts must focus on control and pre-elimination phase that require targeting malaria hot spot areas.

Climatic variables have a great impact on the life cycle of the mosquito vector and the malaria parasite (Craig et al., 1999) and studies have shown that malaria risk and transmission intensity exhibit significant spatial and temporal variability related to variations in climate, altitude, topography, and human settlement pattern (Gosoniu et al., 2006). It is important to be able to identify which climatic conditions are conducive to malaria transmission because this allows for the prediction of malaria transmission and the effective targeting of

malaria control as well as the identification of high risk areas (Stresman, 2010). Furthermore being able to characterize spatial and temporal patterns of malaria would provide us with a better understanding of the climatic and anthropogenic drivers of the disease (Clements et al., 2009). An interpretation of the relationship and interactions between environmental variables and malaria at a more detailed level is also an important part in developing malaria early warning systems, identifying potential outbreaks, and targeting vector control strategies (Ngomane and De Jager, 2012).

This study, therefore, aimed to rank explanatory climatic variables according to their individual scores of variable importance so as to allow for the elimination of variables with little importance as well as to establish what combinations of variables were highly related to malaria by evaluating their prediction performance. The analysis makes use of a recent advancement in statistical learning methods – the random forest approach. This study was conducted at small scale spatial level to allow for more accurate selection of important predictor variables and more detailed prediction of malaria because climatic conditions vary over small distances. Further, this study will use malaria case data that cover the entire study area without any missing data. In this case we identified malaria case data from period of 1 December 2005 to 31 December 2006.

1.2 Aim and Objectives of the study

1.2.1 Aim

Transmission of malaria is influenced by many factors, variability in climatic factors being amongst them. This study aimed to assess the relationship between climatic variables and malaria transmission by determining the variable importance of each climatic variable to quantify its influence on malaria as well as to evaluate what combination of climatic factors was most associated to malaria. The study site is the Ehlanzeni district of the Mpumalanga province in South Africa.

1.2.2 Objectives

In order to achieve the main aim of the study the following objectives are identified:

1. Perform random forest algorithm to relate monthly extracted climatic data and monthly malaria case data to establish a measure of variable importance for each variable.
2. To develop spatio-temporal malaria transmission models using the selected climatic variables

1.3 Limitations of the study

1. Malaria is a notifiable disease however, there is a possibility that not all malaria cases in the study area were diagnosed and subsequently recorded by the National Department of Health malaria control program. Also cases diagnosed and treated in private facilities could have gone unreported.

1.4 Structure of the thesis

Chapter 1 consists of the introduction which outlines the background of the study as well as the aims and objectives to be achieved and the expected limitations. Chapter 2 provides an overview of the epidemiology of malaria as well as the current use of geographical information systems (GIS) and remote sensing (RS) techniques and statistics in malaria studies. Chapter 3 offers a description of the study area (physical and human geography) and malaria intervention measures currently in place. Chapter 4 is the methodology section. It outlines the methods employed to acquire and process remotely sensed climatic data and malaria case data and it describes the random forest algorithm that was used to make an association between climatic variables and malaria. The results of the study are presented in Chapter 5 along with a detailed discussion of the findings. Chapter 6 is the conclusion section and it provides concluding statements drawn from the results of the study.

CHAPTER 2 . LITERATURE REVIEW

2.1 BIOLOGY OF MALARIA

2.1.1 Vector Biology

Malaria is transmitted by female mosquitoes of the genus *Anopheles*, about 70 species of *Anopheles* transmit malaria but only about 30 of these are of importance as vectors (Hsiang et al., 2009). In any given area, just a few *Anopheles* species are responsible for the transmission of malaria (Williams and Bloland, 2002). Individual species differ in their breeding and biting behaviour and the main characteristics that determine whether a mosquito is a major vector of malaria is its blood feeding preferences (predominantly animal or human) and longevity (Hemingwayd, 2009). In Sub-Saharan Africa the primary vectors of malaria are belong to *Anopheles gambiae or Anopheles funestus* groups of species (Godfray, 2013).

The *Anopheles gambiae* complex was initially regarded as one species with ecological salt-water variants; it has now been split into seven distinct species including two of the most efficient human vectors worldwide, *Anopheles gambiae* and *Anopheles arabiensis*. These vectors coexist widely over much of their range however *Anopheles gambiae* is mostly found in humid environments while *Anopheles arabiensis* is mostly found in drier areas (Fontenille and Simard, 2004). They have also been found to be highly dependent on humans for their feeding, resting and to a certain extent breeding habits (Fontenille and Simard, 2004).

The *Anopheles funestus* group comprises of nine species that are morphologically very similar in the adult stage (Spillings et al., 2009). Four of those species namely *Anopheles funestus*, *Anopheles vaneedeni*, *Anopheles parensis* and *Anopheles aruni* have identical morphology at all life stages. *Anopheles leesoni* is the most distict at both egg and larval stage, *Anopheles confuses* is easily identifiable by larval characteristics. *Anopheles rivulorum* and *Anopheles brucei* also have distinctive larvae although it is almost impossible to differentiate between the two. The ninth species *Anopheles fuscivenous* is known only from the adult stage and by their unique chromosomal banding arrangements (Coetzee and Fontenille, 2004). The species *Anopheles* funestus and *Anopheles Rivulorum* are widely distributed throughout sub-Saharan Africa and *Anopheles Parensis* is the most common species in South Africa, Swaziland and eastern Africa however the extent of the distribution of the other members in the group is largely unknown or is more localized (Mulamba et al., 2014, Choi et al., 2012).

During its growth and metamorphosis, the mosquito passes through four distinct phases – egg, larva, pupa and adult (CDC, 2014). The immature stages (the first three) are aquatic meaning they depend on free standing water for their survival and development and lasts for between 5 – 14 days. The adult stage of the mosquito is when the female adult is a malaria a vector and they can live up to a month (Williams and Bloland, 2002).

Adult mosquitoes usually mate within 1-2 days after their emergence, they normally mate in the evenings and in many species this is preceded by the formation of swarms of males (Warrell and Gilles, 2002). During copulation, a male passes spermatozoa into the female which in turn passes them to her spermatheca. Most female anophelines are anautogenous (female must obtain a blood meal to provide the proteins and amino acids required for the maturation of the eggs. However females of a few species are autogenous [can mature and lay their first batch of eggs without a blood meal but blood is necessary for subsequent ovipositions] (Baton and Ranford-Cartwright, 2005). After feeding the swollen abdomen of the female mosquitoes appears bright red and as digestion occurs, the abdomen darkens and the ovaries enlarge and appear whitish through the abdomen. At the half-way stage when much of the blood has been digested and ovarian development is half completed, the mosquito is said to be semi-gravid (Beier, 1998).

Beier (1998) further elaborated that once all the blood has been digested and the eggs have matured, much of the dilated abdomen appears whitish and the mosquito is said to be gravid. After the female lays her eggs, the abdomen appears empty and the female is classified as unfed. This cycle from unfed to blood-fed to half-gravid to gravid to unfed again is called the gonotrophic cycle and is repeated several times (four or five) until the female dies. The importance of this cycle is that if its duration is known, the number of gonotrophic cycles can be used as a means to determine the physiological age of the mosquito which is considered an important determinant in malaria epidemiology(Mattingly, 1969).

Females of some species including *Anopheles gambiae*, may require two blood meals before the first batch of eggs can develop, some species may require three to four blood meals but this is not common. The quantity of blood ingested at a single feed is dependent on the size of the mosquito. Plant sugars are a major source of energy for both male and female mosquitoes; they feed on sugary exudates from fruit, honey dew and even damaged or intact plant tissues (Mattingly, 1969).

The females of most species of *Anopheles* feed on warm –blooded animals, usually mammals. They are attracted to their hosts by a range of factors, once it locates a host, the female mosquito makes a minute incision in the skin and penetrates a capillary vessel and blood feeding commences and is usually completed within 1 minute. In anophelines, feeding mostly takes place between dusk and dawn but some species feed during the day in densely shaded woods and forests. Some malaria vectors predominantly feed outside (exophagic) while others may feed inside houses (endophagic). Once the female mosquito has fed on blood, it seeks shelter to rest, digest and develop her eggs. Some anophelines rest inside houses (endophilic) and those that rest in various outdoor sites are termed exophilic (Warrell and Gilles, 2002).

Several studies have shown that *Anopheles Funestus* has late-night feeding patterns, being most active between midnight and the early hours of the morning (Oyewole and Awolola, 2006, Robert et al., 2006), they are also the most endophilic (rest indoors) and anthropophilic (feed on humans) members of the group (Dabire et al., 2007, Awolola et al., 2005, Antonio-Nkondjio et al., 2006). It is important to take the resting

behaviour of adult mosquitoes into account when planning control measures because vector control is determined largely by feeding and resting patterns of mosquitoes. For example endophilic mosquitoes can be controlled by indoor residual spraying whereas exophilic vectors are best controlled by destroying breeding sites (Warrell and Gilles, 2002).

2.1.2 Parasite Biology

There are four species of *Plasmodium* that infect humans; these are *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*. *Plasmodium falciparum* is the most virulent and common of these human malaria parasites in Sub-Saharan Africa and it accounts for almost all the malaria mortality. According to reports, Sub-Saharan Africa bears over 90 percent of the global *Plasmodium falciparum* burden (Robert et al., 2003). *Plasmodium falciparum* is also distinguished by its ability to bind to endothelium (thin layer of cells that lines the interior surface of blood vessels and lymphatic vessels) during the blood stage of infection and to sequester in organs, including the brain (Greenwood et al., 2008). *Plasmodium vivax* is less deadly but highly disabling, also the ability of *Plasmodium* vivax and *Plasmodium ovale* to remain dormant for months as hypnozoites in the liver makes infection with these parasites difficult to eradicate. *Plasmodium malariae* does not form hypnozoites but it can persist for decades as an asymptomatic blood stage infection (Greenwood et al., 2008), it differs from the other species by its morphological characteristics and its slow development in both the human and insect host (Warrell and Gilles, 2002).

Bannister and Mitchell (2003) discuss that malaria infection and illness start when a single–celled parasite of the genus *Plasmodium* invades the human blood stream. Infected female mosquitoes inject motile parasites known as sporozoites into the victims' bloodstream while taking a blood meal, within minutes parasites invade liver cells and start to reproduce. In one to two weeks, infected liver cells rupture releasing thousands of new parasites known as merozoites which then invade red blood cells and undergo further cycles of asexual reproduction during the course of which many erythrocytes will be erupted. A few merozoites transform into male and female (sexual) stages capable of infecting new mosquitoes, these stages are called gametocytes. Once ingested by a new mosquito during a blood meal, male and female gametes are formed and fuse within the insect's gut ultimately spawning forms that invade its salivary glands from which they enter the next human hosts (Hsiang et al., 2009). Gametocytes are the sexual stages of the malaria parasite and are primarily responsible for its transmission to the mosquito vector (Drakeley et al., 2006).

Each of the developmental stages of the life cycle of the *Plasmodium* parasite that are discussed above represents a potential target at which the life cycle can be interrupted to prevent transmission of the parasite between the mosquito vector and humans (Greenwood et al., 2008).

2.1.3 Symptoms and effects of Malaria

The malaria pathogen enters the human body through the bites of infectious mosquitoes and it immediately colonises the liver. In the liver, it divides multiple times, eventually rapturing the cell to produce a new form of pathogen that infects red blood cells (Godfray, 2013).

Uncomplicated malaria presents with fever and nonspecific symptoms such as vomiting, diarrhoea, fatigue, abdominal discomfort, muscle and joint aches followed by fever, chills, perspiration, anorexia and worsening malaise (WHO, 2006a) . Severe malaria caused by *Plasmodium falciparum* and is characterized by multi organ failure including renal failure in adults whereas in children it presents with prostration, respiratory distress, severe anaemia and/ or cerebral malaria also children with severe malaria rarely present with the classical features of circulatory shock. Retinal changes occur in many patients with severe malaria and a specific pathology that may aid diagnosis was recently described (Greenwood et al., 2008).

Severe malaria affects several tissues and organs although the most evident manifestations appear to involve a single organ such as the brain. Metabolic acidosis has proved to be the most significant pathophysiological feature that dominates classical clinical syndromes of cerebral malaria and severe malarial anaemia (Miller et al., 2002, Marsh et al., 1995). Taylor et al. (1993) found that metabolic acidosis leads directly to respiratory distress which in most cases is a lactic acidosis. The causes of lactic acidosis in children with severe malaria range from increased production of lactic acid by parasites to decreased clearance by the liver (English et al., 1996). Marsh et al. (1995) report that respiratory distress could potentially result from several underlying processes acting alone or in combination. In non-immune adults with severe malaria, respiratory distress often signals the development of pulmonary edema which may lead to adult respiratory distress syndrome. In addition, chronic malaria is known to be a causal factor in anaemia and is associated with hyper-reactive malarial splenomegaly (rare chronic complication of malaria), chronic renal damage, nephrotic syndrome (a non-specific kidney disorder) and Burkitts lymphoma (Sachs and Malaney, 2002)

Slutsker and Marston (2007) explored the interactions between HIV and malaria and they reported that malaria transiently increases HIV viral load and consequently increases the likelihood of HIV transmission, acute and chronic malaria infections can alter the immune system and the body's response to vaccines and increase vulnerability to other infections.

The two population groups most at risk of being infected with malaria are children under the age of 5 and pregnant women. Under conditions of high transmission, the burden of disease rests disproportionately on very young children. However, under conditions of lower transmission, the burden of disease is distributed throughout childhood and this results in an increase in the proportion of children who develop cerebral malaria - which is characterized by impaired consciousness, seizures and comas (O'Meara et al., 2008).

Pregnant women are a high risk group due to pregnancy associated immune suppression and the affinity for *Plasmodium falciparum* for the placenta (O'Meara et al., 2008). Erythrocytes infected with *Plasmodium falciparum* congregate in the maternal placental vascular space, where the parasites replicate (Guyatt and Snow, 2004) . The detrimental consequences of malaria in pregnant women include miscarriages, still-birth, severe anaemia in the mother and low birth weights in infants which results in a substantial increase in the risk of infant mortality (O'Meara et al., 2008).

2.2 SOCIAL AND ECONOMIC IMPACT OF MALARIA

Previous studies have suggested that malaria costs African countries $12 billion annually and impedes economic development considerably (Breman et al., 2004). A typical African family may spend up to 25% of their income on malaria prevention and control (Breman et al., 2004). Spielman (2003) reported that malaria has slowed economic growth in African countries by 1.3% per year, the compounded effects of which are a gross domestic product level now up to 32% lower than it would have been had malaria been eradicated from Africa in 1960. Nur (1993) found that malaria affected economic productivity through its effect on work capacity (time off work due to illness and repeated malaria attacks cause disability), land use (extent of cultivated land and choice of crop) and labour quality (malaria can affect cognitive development). It is also reported that mortality in malaria endemic areas discourages economic investment in those areas (Teklehaimanot, 2005).

According to a report by the Centres for Disease Control and Prevention (CDC, 2014), the cost of illness due to malaria to individuals and their families include the purchase of drugs for treating malaria at home, the costs of travelling to and from health facilities for treatment, absence from work due to illness or to take care of sick family members, expenses for preventative measures and expenses for funerals in the case of deaths. Further, the CDC (2014) stated that the costs to government include maintenance of health facilities, purchase of drugs and supplies, public health interventions against malaria (insecticide spraying, distribution of insecticide-treated bed nets .

A study by Castillo-Riquelme et al. (2008) aim to evaluate the household burden of malaria in South Africa and Mozambique found that around 50% of people stopped work and school related activities while sick with malaria. Malaria episodes typically last between 4.4 and 7 days. The same study also found that a care giver was needed in most cases to look after people with malaria, on average the time caregivers sacrificed from their own activities ranged from 1.1 to 2.7 days. Russell (2004), noted that labour time is lost as a result of illness caused by malaria and this means that household capacity to earn income is reduced. Malaria also hampers children's schooling and social development through absenteeism and permanent neurological and other illnesses associated with severe episodes of the disease (RBM, 2004).

Mmbando et al. (2011) found that households of a lower to moderate socio-economic status were at a more than 60% higher risk of malaria when compared to households with a higher socio-economic status. This

result is supported by a study by Coleman et al. (2010) that showed houses built of mud walls were at a 6 fold risk of malaria compared to houses that were built of brick walls.

In rural African communities, the economy is reliant on agriculture and the malaria transmission season generally coincides with planting or harvesting seasons, resulting in reduced agricultural productivity and output. This is because illness resulting from malaria might result in a delay in planting crops or harvesting (Teklehaimanot, 2005). Some evidence suggests that families living in malaria endemic areas tend to favour less labour-demanding crops over higher value crops that are more labour-intensive, possibly in anticipation of sickness due to malaria (Abdullateef and Oluwatoyin, 2011).

2.3 MALARIA CONTROL AND TREATMENT

The 5 key strategies identified by WHO (World Health Organisation) for malaria control are surveillance, vector control, health promotion, case management and cross border initiatives (RBM, 2005). In keeping with these strategies, South Africa has been implementing evidence based malaria control interventions. These are listed by Moonasar et al. (2012) as being surveillance, vector control, health promotion, case management and cross border malaria initiatives.

2.3.1 Vector Control

The primary aim of malaria vector control is to reduce the vectorial capacity of local vector populations below the critical threshold needed to achieve a low malaria reproduction rate (malERA, 2011). As a result of the long extrinsic incubation time of *Plasmodium* in its *Anopheles* vectors, the most effective vector control strategies currently in place rely on insecticide interventions like indoor residual insecticide sprays (IRS) and long-lasting insecticide-treated nets (LLINs) that reduce vector daily survival rates (Enayati and Hemingway, 2010). The vector species within the *Anopheles gambiae* and *Anopheles funestus* groups of species (primary vectors of malaria in Sub-Saharan Africa) feed and rest indoors at night therefore ITNs and IRS are effective against them (Hemingwayd, 2009).

2.3.1.1 Indoor Residual Spraying (IRS)

IRS has been used as a method of vector control since the 1930s (Moonasar et al., 2012). It repels mosquitoes from entering houses and kills female mosquitoes that rest inside houses after taking a blood meal (Pluess et al., 2010). Mosquitoes typically have a blood meal every 2-3 days and IRS ensures that few survive the approximately 12 days that are required for malaria parasites to complete part of their life cycle in the vector mosquito (WHO, 2006b). Griffin et al. (2010), stated that the killing effect of IRS relies primarily on the indoor resting (endophilic) nature of the mosquito species as well as its human blood index (HBI).

The effectiveness of house spraying for malaria control is dependent on adherence to the criteria specified for the insecticide and the application procedure, public acceptance of spraying, the availability of well-

maintained equipment, adequately trained spraying personnel, efficient supervision and strong financial support (WHO, 2006b).

A report by RBM (2005) states that IRS is a highly effective method for malaria control and is responsible for rapid reduction in malaria transmission during epidemics. Out of 63 countries providing information on their use of insecticides, in 2013, 53 reported using pyrethroids. Carbamates were used by 12 countries and 13 countries reported using an organophosphate (WHO, 2014) . WHO recommends that countries should select an insecticide based on the local situational analysis, dichlorodiphenyltrichloroethane (DDT) is one of the 12 insecticides that can be used for IRS (RBM, 2005).

Table 2-1 provides a description of the 12 insecticides that are recommended by WHO for use in IRS (Sadasivaiah et al., 2007).

Table 2- 1. Insecticides recommended by WHO for IRS

Insecticide	Class	Recommended dosage of active ingredient (g/m2)	Duration of effective action (months)
DDT	Organochlorine	1-2	>6
Fenitrothion	Organophosphate	2	3-6
Malathion	Organophosphate	2	2-3
Pirimiphos-methyl	Organophosphate	1-2	2-3
Propoxur	Carbamate	1-2	3-6
Bendiocarb	Carbamate	0.1-0.4	2-6
Alpha-cypermethrin	Pyrethroid	0.02-0.03	4-6
Cyfluthrin	Pyrethroid	0.02-0.05	3-6
Deltamethrin	Pyrethroid	0.02-0.025	3-6
Etofenprox	Pyrethroid	0.1-0.3	3-6
Lambda-cyhalothrin	Pyrethroid	0.02-0.03	3-6
Bifenthrin	Pyrethroid	0.025-0.05	3-6

DDT, deltamethrin and carbamates remain the preferred chemicals for use for IRS. The insecticidal properties of DDT were discovered in 1939, further tests confirmed the practical value of DDT in malaria vector control and large scale use of DDT for vector control was introduced in 1946 (Russell, 1946, Hemingway and Ranson, 2000). According to WHO (2006b), DDT has a duration of effective action of 6 months and its effectiveness against indoor resting mosquitoes led to the adoption of the Global Eradication Programme of Malaria in

1955, coordinated and supported by WHO. This campaign was based on the periodic use of IRS with DDT for 3 – 5 years to interrupt malaria transmission, this should then be followed by active case detection and surveillance (Sadasivaiah et al., 2007).

The first trial testing of indoor residual spraying of insecticides in South Africa was carried out in KwaZulu - Natal province in 1931. In 1932, a wide spread residual house-spraying programme was undertaken using pyrethrum but it was replaced by DDT in 1946. Pyrethroid deltamethrin replaced DDT in 1996 in accordance with international trends (Maharaj et al., 2012). This was because the environmental protection agency (EPA) banned DDT in 1972 due to ecological concerns (Raghavendra et al., 2011).

However pyrethroid resistant *Anopheles funestus* which was previously eradicated re-emerged in Northern KwaZulu Natal (Hargreaves et al., 2003) and scientists discovered that South Africa had a pyrethroid resistant strain of mosquito that had originated from Mozambique (Biscoe et al., 2004) . Consequently there was an increase in malaria cases from 1996 to 2000 culminating in epidemics. The national Department of Health then decided to revert to using DDT for malaria vector control (Mabaso et al., 2004). In September 2006, WHO lifted the ban on the use of DDT to combat malaria in Africa and other countries where the vectors were still susceptible to DDT (WHO, 2006a). In the past five years, all malaria endemic provinces have reported spray coverage rates greater than 80% which is the level that is recommended by the World Health Organization to achieve maximum impact (Maharaj et al., 2012).

2.3.1.2 Insecticide Treated Bednets (ITNs) and Long Lasting Insecticide Treated Nets (LLIN)

RBM (2005), reported that in areas of malaria transmission where sustained vector control is required, ITNs are the main strategy employed for malaria prevention. It further states that all countries south of the Sahara, the majority of Asian malaria endemic countries and some American countries have adopted ITNs as a key part of their malaria control strategies.

Most of the 97 countries with on-going malaria transmission distribute ITNs free of charge and 83 countries distribute ITNs or LLINs to all age groups (WHO, 2014). Evidence shows that over the past 10 years, the proportion of people with access to an ITN has increased considerably in sub-Saharan Africa however not all households have enough nets to protect all members of the household (WHO, 2014).

A paper by Griffin et al. (2010) provides four effects of ITNs, they kill mosquitoes that land on them, they have a repellency which results in a longer gonotrophic cycle and possible diversion to a human host, they provide a direct protective effective for individuals sleeping under them by forming a physical barrier between the infected mosquito and the individual and finally, they result in a reduction in transmission from infected individuals to susceptible mosquitoes.

WHO (2014) report that over the past 10 years, the proportion of people with access to an ITN has increased markedly in sub-Saharan Africa however not all households have enough nets to protect all members of the households.

ITNs have been used as a method of vector control in many malaria endemic countries and where infrastructure restricts the use of IRS (Raghavendra et al., 2011). Traditional ITNs are treated with an insecticide that can last for a period of up to one year and are retreated at least every year after that to remain effective. (Sexton, 2011). A study by D'alessandro et al. (1995) claims that insecticide treated nets are twice as effective as untreated nets and provide greater than 70% protection compared with having no net. They have also proved to be a cost effective prevention method against malaria. Also most African *Anopheles* bites occur at night and therefore the provision of bed nets reduces risk of being bitten and consequently results in a reduction in disease (Godfray, 2013).

Table 2- 2. Insecticides used in ITN impregnation

Insecticide	Dosage (mg a.i/m^2)
Alpha-cypermethrin	20-40
Bifenthrin	50
Cyfluthrin	50
Deltamethrin	15-25
Etofenprox	200
Lambdacyhalothrin	10-20
Permethrin	500

Long lasting insecticide treated nets (LLIN) are impregnated with insecticide that can last for a period of up to three to five years (Fullman et al., 2013). LLINs are becoming more widely available because they maintain high coverage levels for longer durations Pyrethroids on treated nets work in three ways, they act as a killing agent when the insect makes contact with the insecticide by landing on the net; they have an irritating effect and therefore insects can only rest briefly on the treated fabric and the formulation in which the pyrethroid is presented contains volatiles that deter mosquitoes (Imbahale et al., 2011).

In some countries where resources are available, malaria control programs use both IRS and ITNs. The benefits of this combination are that it reduces transmission and therefore the malaria burden of disease more efficiently than one method would have, it increases the overall coverage of vector control protection and it delays the emergence of insecticide(Kleinschmidt et al., 2009) resistance by using different classes of insecticide for IRS and LLINs (Kleinschmidt et al., 2009).

In South Africa, IRS has been the backbone of the national malaria control program because it is more cost effective method of control in comparison to treated bed nets. An investigation by NDoH (2009) provided evidence that the use of insecticide treated bed nets (ITBNs) would cost the government an additional R111 per additional case averted. The investigation also showed that time costs are higher with treated nets than IRS because ITBNs requires people to participate in a detailed sleeping patterns survey, initial distribution of nets as well as retreatment of nets also required substantial amount of time compared to IRS.

2.3.2 Health Promotion

Health promotion is the process of enabling people to increase control over and to improve their health. It comprises of 3 approaches , firstly advocacy for health, secondly enabling all people to achieve their full potential and thirdly, mediating between the different interests in society in the pursuit of health (WHO, 1986).

Health promotion in South Africa involves educating communities about malaria control to ensure that they comply with instructions from spray operators during IRS campaigns and that they take personal, protective measures against being bitten by malaria infected mosquitoes (Moonasar et al., 2012).

The paper by Moonasar et al. (2012) further explains that the health promotion strategy employed by the South African Malaria Control Programme has two components such as preventative and curative. The preventative component comprises of ensuring that communities comply with instructions from spray operators during IRS campaigns and take personal, protective measures against being bitten by malaria infected mosquitoes. The curative component comprises of educating communities to recognize signs and symptoms of malaria so as to seek early treatment. Information, education and communication (IEC) about malaria is delivered through several channels, including health facilities, community events, radio, posters and pamphlets.

A study by Hlongwana et al. (2011) to investigate the knowledge and practises (KAP) of residents of Bushbuckridge in the Mpumalanga Province of South Africa towards malaria found that the main sources of malaria information were health facilities and radio. The study further concluded that residents had a good understanding of the reasons for spraying. However, knowledge of malaria signs and symptoms was found to be inadequate.

A similar study conducted by Hlongwana et al. (2009) in Swaziland found that health facilities were again the most important source of malaria information. They also found that there was little information originating from traditional community district meetings, community health workers and rural health motivators.

2.3.3 Surveillance

Malaria surveillance allows for early detection and immediate treatment of malaria cases (RBM, 2005). It also ensures that outbreaks of malaria are tracked and interventions can be implemented to avoid epidemics.

Proper surveillance of malaria cases requires reliable notification of cases and efficient capturing and storage of case information. Malaria has been a notifiable disease in South Africa since 1956, therefore by law, health authorities in the country should be notified of all malaria cases. In addition, effective malaria information systems (MIS) exist in the three malaria endemic provinces of South Africa which are Limpopo, Mpumalanga and KwaZulu Natal.

Malaria case data is routinely collected by malaria control staff in the 3 endemic provinces and entered into the MIS of each province. Also active case detection of index cases and screening of populations in close proximity to confirmed malaria cases are conducted in Mpumalanga and KwaZulu Natal provinces. However this activity is limited in the Limpopo Province.

2.3.4 Cross boarder malaria initiatives

Several cross border malaria initiatives such as the Trans-Limpopo Malaria Initiative (TLMI) and the Lubombo Spatial Development Initiative (LSDI) were undertaken to reduce the transmission of malaria in the border areas of South Africa (NDoH, 2009).

The aim of the TLMI was to reduce malaria transmission between and within the Matebeleland South Province of Zimbabwe and the Limpopo Province of South Africa. Its main strategies were case management, vector control, surveillance and health promotion (Moonasar et al., 2012).

The LSDI was a joint programme between the governments of Mozambique, Swaziland, southern Mozambique (Maputo province) and north eastern KwaZulu Natal into a globally competitive economic zone. However these areas are malaria endemic and therefore malaria control was a paramount precursor to economic development. Therefore a tri-national malaria program was launched between the 3 countries (Sharp et al., 2007).

2.3.5 Case Management

Malaria case management involves confirming malaria cases and treatment of individuals who test positive. Most malaria endemic countries (89 out of 99 countries with on-going malaria transmission) have adopted WHO's policy to test all patients that are suspected of having malaria. Evidence of infection with malaria can be obtained using light microscopy or rapid diagnostic tests (RDT) (WHO, 2006a).

For efficient case management South Africa uses definitive diagnosis for malaria case confirmation and treatment, malaria diagnosis and treatment are free of charge in public health care facilities (Moonasar et al., 2012). RDTs detect parasite-specific antigens or enzymes in a small amount of blood and some can detect more than one species. Microscopy is the microscopic examination of Giemsa stained blood smears (Coleman, 2009). The choice of using RDTs or microscopy to diagnose malaria depends on local circumstances such as the skills available, patient case load, epidemiology of malaria and the possible use of microscopy for the diagnosis of other diseases (WHO, 2006a).

RDTs are easy to use and interpret and they allow for immediate findings and treatments (Endeshaw et al., 2008, Rolland et al., 2006). A study by Björkman and Mårtensson (2010) found that RDT sensitivity is dependent on parasite density and well performed lab studies showed that RDTs had a sensitivity greater than 95% for high parasite densities and mostly greater than 75% for low parasite densities. Studies under field conditions showed sensitivity greater than 90% again dependant on parasite density. Field trials showing that RDTs had a high degree of sensitivity and specificity and were user friendly led to the rapid acceptance of RDTs in malaria endemic regions (Munga et al., 2006).

Although microscopy is the gold standard in malaria diagnosis, it is sometimes a less appealing option because it is labour intensive and requires trained staff and quality equipment which are a scarce attributes in countries that lack resources (Batwala et al., 2010). However microscopy is advantageous because it allows for differentiation between species, quantification of the parasite density and the ability to distinguish clinically important asexual parasite stages from gametocytes which may persist without causing problems (Bloland and Organization, 2001).

The South African National Department of Health was the first health ministry in Africa to implement a definitive malaria diagnosis policy using RDTs at all public health facilities in 1996. This action was due to the delayed and substandard malaria microscopic results (Munga et al., 2006).

Historically, chloroquine was the drug of choice for treatment of non-severe or uncomplicated malaria and for chemoprophylaxis however drug resistance has dramatically reduced its truthfulness (Bloland and Organization, 2001)

WHO (2006a) state that resistance to *Plasmodium falciparum* has been observed in all currently used antimalarials (amodiaquine, choloroquine,mefloquine, quine and sulphadoxine-pyrimethamine [SP])and more recently in artemisinin derivatives. *Plasmodium. vivax* has developed resistance rapidly to SP in many areas while resistance to chloroquine is confined largely to Indonesia, Papua New Guinea and parts of Oceania.

In view of the documented resistance to anti-malarial drugs, WHO (2006a) guidelines recommend that uncomplicated *Plasmodium falciparum* malaria should be treated with an artemisinin-based combination therapy (ACT). In areas where chloroquine is still effective, it should be used to treat *Plasmodium vivax* malaria however in areas where resistance to chloroquine is documented, it should be treated with an appropriate ACT. In order to prevent relapses, both chloroquine and ACT should be administered with a 14 day course of primaquine (WHO, 2014).

South Africa treated uncomplicated malaria with chloroquine and quinine was used to treat complicated malaria, drug resistance to chloroquine was first reported in KwaZulu Natal in 1987, consequently from 1988 SP was used to treat malaria in KwaZulu Natal. Chloroquine resistance was reported in Mpumalanga and

Limpopo provinces in 1997 prompting a change to SP for malaria treatment in these provinces as well (Moonasar et al., 2012).

However SP resistance began to increase steadily in the mid-1990s and reached 80% by 2000 therefore drug policy in KwaZulu Natal changed to Coartem in 2001 followed by Limpopo in 2004. Mpumalanga province used SP-artesunate from 2001 to 2005 and changed to artemether-lumefantrine (AL) which trades as Coartem in January 2006 (Moonasar et al., 2012).

Coartem is an ACT and treatment of malaria with ACT has the advantage of reducing malaria transmission by decreasing gametocyte carriage while curing the disease (Barnes et al., 2009). Artemisinins cause a rapid and substantial reduction in the parasite biomass, irrespective of the parasites resistance to other antimalarials, the remaining parasites are then killed off by the high concentrations of the companion drug. There is a sound scientific reasoning for using ACTs, in general, combining different drugs with independent modes of action will increase the chances of killing parasites and greatly decrease the probability that an infected person will develop resistance to both drugs(Olliaro, 2005).

Currently in all endemic and non-endemic provinces in South Africa, AL is the recommended first line antimalarial treatment of uncomplicated *Plasmodium falciparum* malaria. Primaquine is used to treat *Plasmodium vivax* and P*lasmodium ovale* infections (NDoH, 2010)

Prompt and proper treatment of patients infected with malaria parasites as well as efficient malaria control interventions that are effectively executed contribute to a reduction in parasite prevalence through reduced transmission.

2.4 MALARIA TRANSMISSION

Malaria transmission refers to a vector mosquitoes actively transmitting malarial infections in human populations at particular locations (Carter et al., 2000) and is a function of the three-way interaction between humans, Plasmodium parasites and the mosquito vectors that transmit them between hosts (Lyimo and Ferguson, 2009, Dobson, 1999). Malaria transmission is dependent on human-vector contact rate, vector and parasite survival, parasite-development rate inside vectors and human and vector population size (Lyimo and Ferguson, 2009). The intensity of transmission is associated with the frequency with which a person is exposed to the bite of an anopheline mosquito infected with malaria sporozoites and therefore to the possibility of becoming infected with malaria parasites (Carter et al., 2000).

The intensity of malaria transmission is usually discussed in terms of the malaria sporozoite inoculation rate, which is also known as the entomological inoculation rate [EIR](Carter et al., 2000). EIR is important because it provides an estimate of the passage of malaria parasites from infective anopheline species to human populations and is calculated as the product of the human biting rate (an estimation of the density of mosquitoes per person) and sporozoite index (an estimation of the proportion of vectors with sporozoites in

their salivary glands). EIR is expressed as the number of infective bites per person per year (Robert et al., 2003).

EIR is a useful indicator of malaria transmission and as such it is an important indicator of the impact of vector control measures. It can be used to quantify the impact of insecticide treated nets (ITNs), indoor residual spraying (IRS) and source reduction on malaria transmission(Shaukat et al., 2010, Kelly-Hope and McKenzie, 2009). EIR is a more direct measure of transmission intensity however it has a few disadvantages such as it is difficult to measure, it is not standardized, it has no standard protocols and there is variability in methodologies to calculate it (Hay et al., 2005). Another notable disadvantage expressed by Robert et al (2003) is that it cannot be considered an exact measure of transmission because not all bites from infected anopheline species succeed in infecting humans.

Another indicator of malaria transmission potential is the standard Ross-MacDonald mathematical model that predicts malaria transmission in terms of the reproductive rate R_0, which is defined as the number of new cases generated by one infected person in a population of susceptible people(Lyimo and Ferguson, 2009).

A report by WHO (2006a) states that in locations where malaria transmission is stable (populations are continuously exposed to a fairly constant rate of malarial inoculations) and if the inoculation rate is high (EIR > 10/year), partial immunity to clinical malaria and its severe manifestations is acquired early in childhood. The above scenario is common in Sub-Saharan Africa and in such cases; acute clinical malaria is almost always confined to young children who suffer high parasite densities. In stable and high-transmission areas, adolescents and adults are partially immune and rarely suffer clinical disease. The report went on to state that in areas of unstable malaria transmission, rates of inoculation fluctuate greatly over seasons and years (EIRs are usually <5/year and often <1/year). This reduces the acquisition of immunity and results in people of all ages suffering acute clinical malaria, epidemics are also most likely to occur in areas of unstable malaria transmission when inoculation rates increase rapidly

A study conducted by Mouchet et al. (1998) revealed that in stable malaria regions, transmission occurs at high levels every year and can either be perennial or seasonal and the human population acquires a strong immunity during the first few years of life. They also found that transmission is lower in regions of unstable malaria and the population acquires less immunity and epidemics occur in all age classes.

Malaria transmission is influenced by intrinsic and extrinsic factors. Breman (2001), suggests that the most important intrinsic factors include host (human) immunity, parasite species and the species of anopheline mosquito and its longevity. He further lists the most significant extrinsic factors as climate, social and economic conditions, political commitment and effectiveness of control and prevention efforts.

It is important to understand the relationship between extrinsic and intrinsic drivers of malaria to be able to understand the causes and patterns of malaria transmission (Childs and Boots, 2010).

2.4.1 Intrinsic factors of malaria transmission

The first to be discussed is host immunity. When exposed to malaria infection, human populations have varying susceptibility to infection and severity of sickness (Breman, 2001). A study by Allison (1954) found that sickle cell and other traits that alter red blood cell (RBC) structure limit parasite multiplication within the RBCs. Additionally Miller et al. (1976) reported that *Plasmodium vivax* requires the presence of the Duffy blood factor on the surface of RBCs in order to enter them; since over 90% of sub-Saharan Africans lack the Duffy factor, *Plasmodium vivax* is not found in countries in sub-Saharan Africa. The immunity status of the host is crucial in determining the clinical response to infection and transmission. A study by Baird and Snow (2007) suggests that populations in malaria endemic areas acquire immunity over time because they are continually exposed to malaria parasites. Although these populations do not obtain full immunity and low level parasite infections may still occur, they are generally protected against severe malaria.

The second important intrinsic factor in malaria transmission is the parasite, *Plasmodium falciparum* is the most virulent of the 4 species if plasmodia that affect humans and causes severe illness (Gupta et al., 1994). *Plasmodium vivax* and *Plasmodium ovale* cause relapses months after infection and *Plasmodium malariae* has the mildest clinical manifestations of the 4 and causes fevers that recur at 3 day intervals. The third factor is the anopheline mosquito and its longevity, there are about 400 species but only 60 transmit malaria and only 30 are of major importance (Bruce-Chwatt, 1980). Several studies have reached the conclusion that the

Anopheles gambiae complex and *Anopheles funestus* are the most efficient vectors of malaria transmission in Africa (White, 1974, Coluzzi, 1992). Also Malaria transmission is only possible if the longevity of the vector is sufficient to complete sporogony (Coosemans et al., 1992).

2.4.2 Extrinsic factors of malaria transmission

The development of the mosquito larvae and survival and behaviour of adult mosquitos are dependent on various environmental factors. Temperature, rainfall and humidity have long been associated with the dynamics of malaria vector population and consequently the transmission of malaria (Yé et al., 2007). Warm temperatures, heavy rainfall and high humidity are conducive to mosquito breeding and longevity and parasite sporogony (Breman, 2001). Studies by Kiang et al. (2006) and Tanser et al. (2003) provide additional evidence that sustained malaria transmission is dependent on suitable environmental conditions for both vector and parasite; rainfall however appears to have the greatest influence of all environmental factors. Research by Small et al. (2003) concluded that temperature also has a considerable influence on mosquito vector feeding intervals, population density and longevity.

The effect of land use and land cover on malaria transmission is that they affect the availability and suitability of larval habitats and topography is important because a close relationship was observed between malaria risk and elevation (Tuno et al., 2005, Adimi et al., 2010).

Climate change influences malaria transmission because it may directly influence the behaviour and geographical distribution of mosquitoes and the life cycle of the parasite resulting in changes in patterns and incidence of transmission (Martens et al., 1995, Tanser et al., 2003).

Social and economic conditions of a community also have an influence on malaria transmission. Improved education and increased economic activity led to an increase in the ability to manufacture purchase and effectively use insecticides which has resulted in a great reduction in malaria transmission. Poverty alleviation has also been associated with a decrease in malaria endemicity due to improved access to health facilities and quality housing (Andrews et al., 1950, Bruce-Chwatt and De Zulueta, 1980). Human population movement influences malaria transmission because infected people move from areas where malaria is endemic to areas where the disease has been eradicated leading to a resurgence of malaria (Martens and Hall, 2000).

2.5 CLIMATIC RISK FACTORS OF MALARIA

Climate is an important determinant of malaria transmission dynamics because it determines the spatial and temporal pattern of malaria transmission dynamics (Brooker, 2007). It affects various facts of malaria epidemiology such as mosquito reproduction, sporogony and mosquito biting behaviour (Karthe, 2010).

2.5.1 Temperature

A study by Paaijmans et al. (2010) examined the effects that daily temperature dynamics have on what they identified as three key mosquito life history parameters. They examined the effects of temperature fluctuations on immature mosquito development and survival; survival of adult mosquitoes and on the length of the gonotrophic cycle. The results provide empirical evidence that parasite infection, growth and development; immature mosquito development and survival; length of gonotrophic cycle and adult survival are all sensitive to daily variation in temperature. And since all the above mosquito-related traits combine to determine malaria transmission, this study highlights the direct influence of temperature on malaria transmission.

According to (Lindsay and Martens, 1998), the minimum temperature for *Plasmodium falciparum* development is between 16^0C and 18^0C, therefore a drop in temperature results in a decline in the risk of infection because parasite development is restricted (Musa et al., 2012). The sporogenic cycle takes about 9 to 10 days to complete at temperatures of 28^0C (Craig et al., 1999). Studies have shown that the optimum temperature range for parasite development in the female *Anopheles* is between 25^0C and 30^0C and development ceases below 16^0C (Snow and Omumbo, 2006). At temperatures above 40^0C, the daily vector

survival is zero because thermal death for mosquitoes occurs between 40 – 42^{0}C (Craig et al., 1999). Another effect of increased temperature on malaria transmission is that it causes mosquitoes to develop faster and feed at shorter intervals because blood meals are more rapidly digested and the whole genotrophic cycle (process of blood-feeding, egg maturation and oviposition) is accelerated (Grover-Kopec et al., 2006, Hay et al., 2000, Musa et al., 2012).

Alemu et al. (2011) discuss studies that suggest that minimum temperature is the most significant factor for malaria transmission in comparison to other meteorological factors, this means that a rise in minimum temperature would encourage the survival of *Plasmodium* and *Anopheles* during different seasons, accelerating the transmission dynamics of malaria. One such study was conducted in Shuchen County, China, by (Bi et al., 2003) and their results showed a strong correlation between monthly mean minimum temperature and monthly mean incidence of malaria.

2.5.2 Rainfall

Rainfall provides breeding sites for mosquitoes by providing surface water in which female *Anopheles* can lay eggs and it prolongs the vector life span by increasing water availability (Casman and Dowlatabadi, 2002). *Anopheles gambiae s.l.* breed more effectively in temporary and turbid water bodies such as those formed by rain, in contrast, *Anopheles funestus* breed more effectively in more permanent water bodies. However, both temporary and permanent water bodies are dependent on rain (Coleman, 2009)**.** Rain is also related to humidity and saturation deficit, both of which influence mosquito survival, because rainfall increases relative humidity and thus the longevity of the adult mosquito (Kiang et al., 2006, Craig et al., 1999). Mosquitoes generally do not live long enough to complete their transmission cycle when relative humidity is below 60% (Grover-Kopec et al., 2006).

Mabaso et al. (2006) conducted a study to illustrate the role of rainfall in malaria transmission in Zimbabwe. Their findings were that high annual malaria incidence coincide with high rainfall and relatively warm conditions while low incidence years coincide only with low rainfall (Mabaso et al., 2006). These findings were supported by a later study by Himeidan et al. (2007) to investigate the role of climatic variables on malaria transmission in eastern Sudan which found a significant positive association between high rainfall and malaria slide positive rate (SPR).

However, an excess of rainfall destroys breeding sites and flushes mosquito larvae out of small ponds and therefore reduces transmission of malaria. The relationship between rainfall and the development of breeding sites is also dependant on the topography, run off and soil type of the region (Hay et al., 2000) Also the lack of rainfall does not necessarily result in a reduction of larval populations. This is because the lack of rainfall may create new habitats such as pools and puddles and thus lead to an increase in larval populations (Kiang et al., 2006). A study by Danuor et al. (2010) corroborated the above findings by concluding that

excessive rainfall in an area in Ghana was followed by a decrease in malaria cases and low rainfall was followed by a rise in malaria cases.

2.5.3 Land use and land cover

A study by (Munga et al., 2009) investigated the impact of land use and land cover on the availability and suitability of anopheline larval habitats found that *Anoheles Gambiae* and *Anoheles.funestus* larvae occurred frequently in open and unlit habitats in farmlands and pastures. Land cover type was also found to influence the suitability and availability of anopheline larval habitats through its effects on temperature and food conditions, Munga et al. (2006) found that land cover type affects water temperature and available nutrients in aquatic habitats.

Furthermore Afrane et al. (2006) concluded that land use and land cover have a profound impact on vectorial capacity of anopheline mosquitoes by influencing survivorship and biting frequency of *Anoheles Gambiae*. This was proved by observed significant increases in the net reproductive rate of mosquitoes in deforested areas in Kenya which suggested that deforestation enhances mosquito reproductive fitness thus increasing mosquito growth potential (Munga et al., 2009).

Changes in land use can influence malaria transmission in several ways, such as, increased cattle grazing creates more open temporary habitats that can provide mosquito breeding habitats; it may change the physical and chemical properties of mosquito larval habitats and it may also change the microclimate of mosquito larval habitats (Minakawa et al., 2005). The clearance of forests to grow crops or create pastures leads to open landscapes which when puddled provide suitable breeding sites (Lindsay and Martens, 1998), it also shortens the sporogenic development time of *Plasmodium falciparum* in *Anoheles Gambiae* (Munga et al., 2009).

In a study conducted by Lindblade et al. (2000) to investigate the effect of land use change on malaria transmission in a highland area in Uganda, it was established that mosquito density during the wet season was consistently higher in cultivated swamps than in natural swamps. The same study also indicated that changes in vegetation due to change of land over altered evapotranspiration systems and modified local climates, this was proved by minimum and maximum temperatures being uniformly higher near cultivated swamps. These findings were echoed in a study in Kenya that concluded that cultivated swamps receive more exposure to sunlight than natural swamps therefore the ambient air temperature in cultivated swamps was significantly higher than in natural swamps. Also, mosquito larval predators may be more prevalent in natural swamps than in cultivated swamps (Minakawa et al., 2005). Another possible reason why clearing natural swamps increases malaria transmission is that the papyrus found in swamps produces oils which form a thin layer on the water surface and this could prevent mosquito larvae from breathing thus the clearing of swamps provides surface water that is ideal for the breeding of mosquitoes (Lindsay and Martens, 1998).

Vegetation type and growth affect the abundance of the malaria vector because vegetation provides a resting place, protection from climatic conditions and sugar feeding supplies for adult mosquitoes. It also influences the presence of animal or human hosts and consequently affects the availability of blood meals (Ceccato et al., 2005). There is also evidence that habitats covered by grass has more anopheline mosquitoes that those with other types of vegetation and open land, this could be due to the grass protecting mosquito larvae from being flushed away by running water, protecting larvae from predation and offering newly emerged adult mosquitoes a shaded resting place (Imbahale et al., 2011).

Urban environments also have an effect on malaria transmission, replacing vegetation with asphalt and concrete may reduce the number of larval habitats. However, in urban areas, where vegetation remains, combined with urban farming provides sufficient aquatic habitats for mosquitoes (Robert et al., 2003).

2.5.4 Normalized Difference Vegetation Index (NDVI)

NDVI is unitless with a theoretical range of -1 to +1, however most values fall between 0 and 0.7 (Britch et al., 2008). it is an index that is calculated by dividing the difference between reflectance in the red and near infrared spectral regions by the sum of the reflectance in the same two bands. It correates to a number of biophysical parameters including leaaf area index, biomass and fraction of absorbed photosynthetically active radiation (Wayant et al., 2010). It is dependant on both land cover and atmospheric conditions (Machault et al., 2011)

Positive values near 0 are indicative of bare soil or little or no vegetation and values closer to 0.7 are indicative of dense vegetation (Britch et al., 2008).The higher the NDVI value, the denser or healthier the vegetation is regarded as being. The growth of vegetation cover results in a cool shaded environment that is conducive to the development of aquatic stages of the mosquitoe life cycle (Okogun et al., 2003). Additionally, an abundance of vegetation cover provides adult mosquitoes with shade for resting poitions and breeding sites (Gilliet, 1971). This is important because it has been proven that mosquitoes prefer cool, shaded areas for biting and breeding activities (Okogun et al., 2003).

2.5.5 Topography

Many of the factors that are important to mosquito development and survival such as meteorological conditions, vegetation, water body characteristics and land use appear to be related to topography – mainly landform and elevation (Ndoen et al., 2010). It was shown that for a study conducted in Tanzania that the intensity of transmission is directly related to altitude and that parasite prevalence is a good indication of this (Drakeley et al., 2005). Further studies in the Kenyan Highlands support these findings, a linear relationship was found between *Plasmodium falciparum* prevalence in the study group and altitude (R^2 = 0.98) with a 15.9% reduction in prevalence for every 50 meter increase in altitude along the transect which originated from the bottom of Yala River valley and terminating at Sigalaga village, 4km uphill (Githeko et al., 2006).

An investigation into the spatial distribution of Anopheline larval habitats in Western Kenya showed that topography affected the formation of aquatic habitats for mosquito larvae (Minakawa et al., 2005). Stagnant aquatic habitats are more prevalent in valley bottoms than on hills because it is more difficult for water to accumulate on hill slopes due to surface run off, these bottom stagnant aquatic habitats are formed as a result of run off from uphill, and seepage from springs and groundwater. Also, groundwater levels are lowered during the dry season and therefore the distribution of stagnant aquatic habitats becomes more confined towards valley bottoms (Minakawa et al., 2005). The spatial distribution of larval habitats is often constrained by topography and water drainage (Munga et al., 2009).

2.5.6 Humidity

Humidity is another important climatic variable related to malaria transmission because adequate humidity is required to ensure mosquito survival (Jawara et al., 2008). A study by Huang et al. (2011) that performed a temporal correlation analysis to analyse the relationship between rainfall, temperature, humidity and malaria incidence found that monthly relative humidity was most closely correlated to monthly malaria incidence.

Adult mosquitoes are dependent on specific moisture content in the air and will desiccate if the climate is too dry (Stresman, 2010). Vittor et al. (2009), Afrane et al. (2008) and Tuno et al. (2005) found that decreasing humidity had an impact on mosquito fitness and parasite development. For a mosquito to transmit the parasite, it must survive long enough to bite an infected person, surpass the extrinsic incubation period of the parasite and then bite an uninfected person (Yamana and Eltahir, 2013). Relative humidity below 60% shortens the lifespan of mosquitoes (Yé et al., 2007) and therefore humidity has a significant effect on malaria transmission (Yamana and Eltahir, 2013).

Studies have further shown that relative humidity directly influences mosquito activities such as biting and breeding rates (Bi et al., 2003). De Casas and Carcavallo (1995) report that high relative humidity stimulates metabolic processes of mosquitoes and low relative humidity results in mosquitoes feeding on blood more frequently to compensate the dehydration.

2.5.7 Climate change

Climate change is expected to affect malaria indirectly by changing ecological relationships that are important to the organisms involved in malaria transmission (vector, parasite and host) and directly by causing changes in temperature, rainfall and relative humidity (Alemu et al., 2011). This will result in the modification of the behaviour and distribution of malaria vectors and a change in the length of the life cycle of the parasite (Alemu et al., 2011) and hence a change in the distribution and incidence of malaria.

The second assessment report by the Intergovernmental Panel on Climate change (IPCC) published in 1996 concluded that the earth's mean temperature will increase by $1 – 3^0$ C over the coming century (Danuor et

al., 2010). However, larger local increases could occur over a shorter time period due to deforestation, this could result in increases of up to 3 – 4^0 C (Lindsay and Martens, 1998). These findings are significant because malaria is considered one of the major vector – borne diseases most sensitive to changing environmental conditions resulting from climate change (Parham and Michael, 2010).

Parham and Michael (2010) modelled the effects of climate change on malaria and discovered that changes in rainfall strongly govern malaria endemicity, invasion and extinction by influencing vector abundance. Furthermore, they revealed that temperature effects have a complex relationship with malaria transmission and a stronger influence on the rate of the disease by affecting multiple parts of the pathogen life cycle but only when sufficient rainfall is available to sustain vector development and survival.

A model created by Martens et al. (1995) to assess the potential impact of climate global change on malaria incidence showed that simulation experiments (using the climate changes predicted by the IPCC) indicate a wide spread increase in transmission potential of the malaria mosquito population and an expansion of the areas that are conducive to malaria. The model also showed that in the highly endemic malarial areas of tropical Africa, the incidence of malaria and consequently the number of years of life lost due to malarial disease may increase however in malarial areas of lower endemicity, the incidence of infection is far more sensitive to climate changes.

2.6 APPLICATION OF GEOGRAPHICAL INFORMATION SYSTEMS, REMOTE SENSING AND SPATIAL STATISTICS IN MALARIA STUDIES

Geographical Information Systems (GIS) and Remote Sensing (RS) and spatial statistics play a pivotal role in malaria studies because they allow for the acquisition of detailed, accurate and continuous environmental data, they are a resource that enables the processing this data as well as the integration of environmental and epidemiologic data into models and they enable an understanding of the spatial and temporal relationships between malaria cases and environmental variables (Thomson et al., 1999). Therefore GIS, RS and spatial statistics are tools for enhancing planning and implementation of efficient and effective large scale malaria control programs (Himeidan et al., 2007).

The interactions between GIS, RS and spatial statistics in malaria research are emphasized by project MALAREO which is a mixed European-African consortium focused on providing suitable remotely sensed data for two applications. The first application is in epidemiological modelling using mostly Bayesian statistics to predict malaria incidence, prevalence and risk. The second is for use in producing a range of thematic maps to assist in the daily work of national malaria control programs to support management of integrated vector control which includes planning of IRS and distribution of ITNs or larviciding (Bauwens et al., 2012).

The application of GIS and RS in malaria control is primarily for the classification and mapping of the distribution of sources of malaria transmission and malaria risk down to the household level where possible

so that control efforts in endemic situations and intervention strategies in epidemic situations may be directed efficiently (Hay et al., 2000, Nájera et al., 1998, Carter et al., 2000). These systems provide a structure that is capable of collecting data linked to geographical location from different sources and storing it in a form which permits subsequent analysis and synthetic presentation in map form (Carter et al., 2000). GIS and RS make it possible to monitor and analyse environmental factors that influence malaria transmission, in this way associations between environmental variables and the distribution of the different species responsible for malaria transmission can be understood and this knowledge can be used to fight the disease (Ceccato et al., 2005).

2.6.1 Geographical Information Systems (GIS)

The Malaria Atlas Project (MAP) is a case study of the efficient use of GIS in malaria research. As has been discussed earlier, malaria is a vector borne disease and the anopheline mosquitoes responsible for its transmission are very sensitive to climate. Hay et al. (2009) elaborate that the MAP team collected climatic data and community based parasite prevalence data to develop global maps of malaria risk. This was done primarily to predict the endemicity of *Plasmodium falciparum* by developing global maps of malaria risk.

Booman et al. (2000) described how GIS can be used to plan a malarial control programme in Mpumalanga province, South Africa by producing risk maps at town and village level. A database containing malaria case data as well as the geographical area (exact position of villages and towns) was developed, this allowed for GIS to be used to display malaria cases and their specific location in Mpumalanga spatially. Malaria incidence was also calculated (total number of new cases occurring in each age cohort of females and males divided by the total person-years and then multiplied by 1000), this allowed for the creation of a thematic map displaying malaria risk. Spatially defining and depicting the incidence of malaria allows for proper planning to enable efficient use of available resources, focusing on districts with the highest risk of malaria.

A similar study was later conducted by Kelly et al. (2013); it also used GIS for malaria control in Solomon Islands and Vanuatu. GIS tools were used to automatically locate and map the distribution of reported confirmed malaria cases at household level, and this could be done because positive cases were geo-referenced using household location. This study allowed for the detection of priority geographic areas (where there is a high number of reported malaria cases) so that follow up activities could be conducted and it assisted with appropriate measures being implemented timeously.

A paper by Booman et al. (2003) discussed how GIS can also be used for monitoring the progress of indoor spraying activities as was done in Mozambique as part of the multilateral Lubombo Spatial Development Initiative. The smallest administrative unit boundary at which data was captured (localities) were drawn onto maps and digitally captured into GIS, this then enabled the production of spraying progress maps because weekly spray activities (number of structures sprayed, volume and type of insecticide per structure) were captured into a database for each locality. Maps depicting spray coverage were also produced from this data

and these were useful for tracking the progress of spray teams, areas that showed a lower coverage percentage than expected were then investigated to determine the cause of this and the necessary measures to correct the problems instituted. Insecticide application rates can also be monitored from the information entered into the database, it is important to keep track of application rates because under application reduces the overall effectiveness of insecticides and over-application is wasteful as available resources should be used effectively.

Shirayama et al. (2009) used GIS to assist in the monitoring of intervention coverage and health outcomes in Laos province in Asia by visually displaying the uneven distribution of ITNs and health outcomes. They were also able to highlight areas where malaria cases were concentrated as well as areas with low ITN coverage and poor adherence to interventions. This data and subsequent maps were greatly beneficial to decision makers and local health staff.

2.6.2 Remote Sensing (RS)

RS can be defined as the collection of data by instruments measuring physical and biological characteristics of some objects without direct contact, sensors on board satellites record electromagnetic radiation reflected or emitted by the earth's surface. Earth observing satellites are equipped with instruments that provide high spatial resolution images that have a low temporal resolution (revisit time can be several days or weeks) or low resolution platforms that provide images once or several times a day (Machault et al., 2011).

Climate, seasonality, rainfall patterns, temperature, humidity and the presence of vegetation and surface water are all directly related to malaria transmission. This data can be linked to many malariometric indices such as breeding sites, vector density (larvae densities and adult mosquitoes are closely related), entomological inoculation rate(EIR) and parasite prevalence, morbidity and mortality (environmental and climatic indicators have been used in models explain malaria morbidity and mortality), detection of changes (image processing allows for detection of environmental changes over time that can be associated with new, increased or decreased malaria risk) and urban malaria mapping [urbanization has been shown to have an effect on malaria transmission] (Ceccato et al., 2005). Furthermore, RS images that provide information about environmental factors can serve as predictors of vector distribution patterns and of average malaria transmission levels (Curran et al., 2000, Rogers et al., 2002). Therefore remote sensing in the form of satellites for earth observation is important because it allows for the monitoring of these environmental factors which influence the reduction or re-emergence of malaria (Ceccato et al., 2005).

This is shown in a study by Kiang et al. (2006) that was carried out to examine the environmental dependency of malaria transmission in Thailand. Environmental data (temperature, precipitation, relative humidity and Normalised Difference Vegetation Indices (NDVI)) was obtained through RS. Neural network methods (statistical learning algorithms) were then used to model malaria transmission based on these variables. Out of a total of 19 provinces, malaria transmission was correctly predicted in 11 provinces and over predicted in

only 8. Also an investigation by Gaudart et al., 2009 coupled environmental remotely sensed data with field study data in order to create a malaria transmission model for a locality in the Sudanese savannah area. The model showed that the seasonal pattern of *Plasmodium falciparum* incidence was significantly explained by NDVI. A later study by Midekisa et al. (2012) also developed time series models for malaria early warning in Ethiopia by quantifying the relationship between malaria cases and remotely sensed environmental variables. They were also able to display the temporal patterns of malaria risk.

In addition to providing environmental data associated with malaria transmission, research by Mushinzimana et al. (2006) showed that remote sensing can also be used to identify geographic features that are associated with larval habitats of malaria vectors and ultimately to predict the spatial distribution of aquatic habitats. The importance of this is that if larval habitats are identified then it is possible to control anopheline larva through environmental management or larvicides to assist in combating malaria transmission.

RS has proved to be a useful source of environmental data that influences malaria transmission and GIS have emerged as a powerful tool to link and display information from many different sources such as environmental and disease data in a spatial context. However the mapping capabilities of GIS and RS are limited because they are unable to quantify the relation between environmental factors and malaria risk and to produce model based predictions, this is where the importance of spatio-temporal modelling lies (Gosoniu et al., 2006).

2.6.3 Spatial statistics

Spatial statistics are similar to traditional statistics except that they have the added advantage of integrating spatial relationships into the calculations. Data can be described as continuous or discrete and can be spatial aggregations or individual observations at points in space. The spatial location of data can be regular or irregular and from a spatial continuum or a discrete set (Cressie, 1993). Data that is used in spatial statistics is categorized as geostatistical, lattice or point pattern data. Geostatistical data is data that has co-ordinates and values, lattice data is spatial data that is indexed over a lattice of points and point pattern data pertain to the location of 'events' of interest, the main interest is in the locations of all occurrences of some event (Cressie, 1993).

Data that are close together in space and time are often more alike than those that are far apart, this phenomenon is referred to as spatial auto correlation. Spatial autocorrelation shows the association or relationship between the same variable in 'near-by' areas and it is important that spatial statistical models take it into account as failure to allow for spatial autocorrelation leads to the significance of covariates being overstated (Boyd et al., 2005{Thomson, 1999 #296)}. Also spatial correlation may arise because of omitted or unobserved covariates, and incorporating the spatial random effect in the model further allows these to be accounted for (Ver Hoef et al., 2001).

2.6.3.1 Spatio-temporal modelling

The advantage of spatially dependant models is that they have a more parsimonious description than classical trend-surface models, they have more stable spatial extrapolation properties and they yield more efficient estimators of explanatory-variable effects (Cressie, 1993). Statistical methods including linear and logistic regression are often used in disease mapping which is a term used to define the modelling of spatial variation of risk of disease over an area of interest (White, 2012). An understanding of the spatio-temporal variation in malaria provides a basis for effective disease control planning and monitoring. This is because characterizing spatial and temporal patterns of clinical malaria provides insights into the important drivers of the disease such as climatic variables (rainfall and temperature) that influence seasonal patterns and human factors that influence long term trends (Clements et al., 2009).

Moreover spatio-temporal modelling of malaria in relation to various environmental factors requires precise disease and environmental data both in space and time for the models to be able to describe patterns in detail for efficient targeting of disease prevention or treatment in space and time (Ostfeld et al., 2005).

When Bayesian geostatistical models are applied to malaria risk data, they can be used to quantify environment-disease relations, identify significant environmental predictors of malaria transmission and they provide model based predictions of malaria risk (Gosoniu et al., 2006). The Bayesian approach is advantageous because it allows for the modelling of hierarchal datasets and it incorporates spatiotemporal autocorrelation (Clements et al., 2009).

There are several studies that utilise the Bayesian approach in spatial modelling for malaria mapping, an example of such a study is one by Gosoniu et al. (2006) that used climatic and environmental data (NDVI, rainfall, temperature, season length and water bodies) and malaria prevalence data in Mali to develop stationary and non-stationary Bayesian models to assess the malaria-environmental relationship to predict malaria risk and to develop malaria risk maps using predicted malaria risk data.

A similar analysis was carried out to predict and map malaria risk in Malawi to identify areas where the greatest malaria control and intervention strategies should be focused (Kazembe et al., 2006). However, for this analysis, only point referenced prevalence data for children aged 1 – 10 years was used and in addition to temperature and rainfall, elevation and potential evapotranspiration were also added in the Bayesian model for risk assessment and improved prediction.

Bayesian geostatistical techniques can also be used to predict malaria risk at locations where malaria data is not available by using a kriging technique. Kriging techniques provide a framework for predicting (interpolating) values of a variable of interest at unobserved locations given a set of spatially distributed data, incorporating spatial autocorrelation and computing uncertainty measures around model predictions (Noor et al., 2008). The advantages of Bayesian geostatistical modelling are that they incorporate spatial correlation

in data and they allow for errors of estimation to be quantified making it possible to assess the precision of the map and significance of covariates.

Some spatio-temporal models use the Poisson distribution which is a discrete probability distribution for the counts of events that occur randomly in a given interval of time or space. Kleinschmidt et al. (2000) sought to test the hypothesis that the spatial distribution of climatic conditions in winter is responsible for the variation in intensity of malaria transmission in a district in KwaZulu Natal. To achieve this, they performed a spatial statistical analysis of malaria incidence rates in a district in KwaZulu Natal. This was done using generalized linear mixed models (GLMM) with a Poisson distribution which took into account the effect of climatic variables on malaria cases. The results from the model were used to produce maps of predicted malaria incidence.

Mabaso et al. (2005)also used a Poisson model within a Bayesian framework to describe the relationship between seasonality in malaria and environmental covariates in Zimbabwe. The Poisson model was used to analyse the relationship between environmental factors and the number of malaria incident cases, all covariates (temperature, rainfall, NDVI, vapour pressure) showed significant associations with malaria cases and were therefore included in the spatio-temporal analysis.

Negative binomial regression is another type of data analysis that is incorporated into spatio-temporal modelling; it can be used for over-dispersed count data, that is when the conditional variance exceeds the conditional mean. Negative binomial regression can be considered as a generalization of Poisson regression since it has the same mean structure as Poisson regression and it has an extra parameter to model the over-dispersion (Hilbe, 2011). Mabaso et al (2006) used Bayesian negative binomial models to perform a spatio-temporal analysis of the relationship between malaria incidence and climatic covariates in Zimbabwe. Preliminary negative binomial regression analysis was carried out to assess the relationship between annual malaria incidence and annual values of each climatic covariate. Bayesian negative binomial models were used to examine the association between inter-annual variation in malaria incidence and a combination of climatic covariates that were selected from the preliminary analysis.

Kleinschmidt et al. (2001) used linear models to determine the climatic and environmental variables such as temperature, rainfall and distance to water bodies that are associated with small scale malaria incidence rates, Alemu et al. (2011) also conducted linear regression to assess the relationship between observed malaria cases and meteorological variables such as rainfall, temperature and relative humidity. Regression modelling was used in a study by Riedel et al. (2010) to determine the association between malaria parasitaemia risk and environmental variables including temperature, rainfall, NDVI, altitude and land use. Regression modelling was also the basis of a study by Gosoniu et al. (2009) to understand the influence of environmental predictors on the number of observed malaria cases.

All the above spatio-temporal modelling studies focused on showing the temporal and spatial variations of malaria as influenced by various climatic factors. However studies have shown that clinical malaria case data are more closely related to seasonality (Thomson et al., 1999, Hay et al., 1998) and therefore to fully understand malaria transmission at a more detailed level, the relationship between malaria cases and environmental variables should be analysed on a monthly basis for any given year. It would also further improve efforts for malaria control if we are able to identify the climatic variables that play the most influential role in malaria transmission monthly throughout the year.

Variable selection is a procedure that is important in the process of creating parsimonious and well identifiable models. It has received little attention within a geostatistical modelling framework and is usually performed as part of an explanatory analysis; it is carried out separately to the geostatistical model fit (Chammartin et al., 2013). However, over the past few years, several approaches towards variable selection have emerged. Some examples are the stochastic search variable selection (SSVS) developed by George and McCulloch (1993), the variable selection sampler used by Kuo and Mallick (1998), the Gibbs variable selection by Dellaportas et al. (2002) and the parameter expanded normal mixture of inverse-gama (peNMIG) geostatistical variable selection that was proposed by Ishwaran and Rao (2005).

Taking the above into consideration, random forests can be proposed as a simple non-parametric modelling technique that can be employed to identify which variables are most related to malaria cases for each month of the year. This is because random forests have been proven to be capable of creating accurate predictive models and they can identify the most relevant and informative predictor variables from a set of candidates and provide a measure of variable importance within the predictive model.

Random forests are an ensemble method that uses many decision tree models for classification and regression problems (Breiman, 2001). A subset of the training data, with replacement, is used to train each tree and the remaining data are used to estimate error and variable importance (Horning, 2013).

Furlanello et al. (2003) used random forests to predict the presence of tick borne diseases using environmental variables as predictors. The result was risk maps showing probability of tick presence under various environmental conditions. They were also able to evaluate the importance of each explanatory variable (predictor) variable using the random forest variable importance algorithm.

2.7 CONCLUSION

Studies have shown that the distribution of malaria can be estimated successfully based only on climatic conditions (Craig, 2007). Climatic variables determine both the distribution of endemic malaria and the distribution and frequency of epidemic malaria in Africa (Cox et al., 1999, Mabaso et al., 2007).

The relationships identified and applied in the preceding sections on climate and malaria transmission highlight the possibility of assertively relating malaria transmission both spatially and temporally to climatic

variables such as temperature, rainfall, humidity. However none of them are able to provide a measure of variable importance to quantify the individual importance of these variables in relation to malaria as well as provide an indication of what interactions between these climatic factors were most related to malaria.

Furthermore, the described methods have limitations such as the restricted functional form of the association pattern (linear models are the most common and most restrictive), measures of variable importance are only available for a small range of methods and ordinal scaled variables are often treated as if they were measured on an interval or ratio scale (Strobl et al., 2009). Furthermore the use of linear and logistic regression modelling to predict infection risk is limited by the inability of these methods to consider spatial correlation of infection and environmental variables. This leads to underestimation of standard errors of the covariate coefficient resulting in erroneous inference and justifies the need for assessment of uncertainties inherent in data and modelling techniques (Brooker, 2007). Another limitation of standard parametric models is that interaction effects of high order usually cannot be included (Strobl et al., 2009).

Another shortcoming of predictive modelling is variable selection, this is due to analytical problems caused by over-fitting, confounding and non-independence in data (Craig et al., 2007). Babyak (2004)suggested that both manual and automated stepwise selection procedures are not recommended because of frequent over-fitting and the resulting phantom degrees of freedom.

This study therefore will use the random forest algorithm developed by Leo Breiman at the University of California, Berkeley to identify relevant and informative climatic variables that are closely related to malaria cases for each month of the year of 2006 and to produce monthly predictive models.

Random forest is a non-parametric modelling technique which is widely used in prediction and classification problems (Garge et al., 2013). It is a statistical method that has been shown to be able to deal with a large number of predictor variables even in the presence of complex interactions and highly correlated variables (Shih, 2011). In addition, random forests allow predictor variables that would otherwise have been outplayed by a strong competitor, to enter the ensemble. Therefore interaction effects that would otherwise have been unnoticed are revealed (Strobl et al., 2009). Finally, random forest can identify relevant predictor variables by means of variable importance measures (Strobl et al., 2008). It is a powerful new approach to data exploration, data analysis and predictive modelling which was developed by Leo Breiman at the University of California, Berkeley. Rossi et al. (2005) also found that random Forest variable importance ranking proved more stable than stepwise variable selection approaches that are available for logistic regression that are known to be affected by order effects. Also, a variable with a high variable importance in random Forest that is not included in stepwise regression may indicate that the variable works in interactions that are too complex to be captured by parametric regression models (Strobl et al., 2009). Further to this, testing and rejecting many variables increases the probability of finding a significant predictor purely by chance and

standard errors in predictive models are underestimated because this sifting process remains undeclared (Babyak, 2004, Harrell, 2001).

Random Forests are basically an ensemble method using many decision tree models that can be used for classification and/or regression (Breiman, 2001). A subset of the training data, with replacement, is used to train each tree and the remaining data are used to estimate error and variable importance(Horning, 2013).

The environmental parameters that will be used as input data are NDVI (normalized difference vegetation index), land surface temperature, lag land surface temperature, rainfall, lag rainfall, altitude and humidity. These variables were chosen because they were shown to have the greatest impact on malaria transmission in numerous studies (Kiang et al., 2006, Tanser et al., 2003, Small et al., 2003, Parham and Michael, 2010, Alemu et al., 2011, Imbahale et al., 2011, Zhou et al., 2004, Yé et al., 2007, Amek et al., 2012a)These will be obtained using remote sensing and GIS techniques and analysis, most of which were covered in the review. Malaria case data is available at locality level for the entire Mpumalanga province. Because the intention is to establish which variable influences malaria cases the most for each month, these values will input simultaneously into the random forest. It is expected that important interactions or interrelationships will be detected between all the input variables and the outcome of predicted malaria cases and this will be used help explain the climatic factors driving malaria transmission and further explain the seasonality of malaria Mpumalanga province in South Africa.

CHAPTER 3 . STUDY AREA

3.1 Physical and Human Geography of Mpumalanga

Mpumalanga Province is situated in the north-east of South Africa. It shares international borders with Mozambique and Swaziland in the east and local borders with KwaZulu Natal and Free State in the south, Gauteng in the west and Limpopo in the north. Mpumalanga has a population of 4 128 000 people which constitutes 7.8% of the national population (NAFCOC, 2014). It has the third highest unemployment rate among the country's nine provinces and its poverty rate of 39.4% is higher than the national rate (Mpumalanga Province Department of Finance, 2013). The majority of Mpumalanga's population, including the majority of the poor, is located in areas of low economic activity. The areas with the highest concentrations of poverty are Broader KwaMhlanga, Siyabuswa, Bamokgoko which are located in Nkangala District Municipality; Bushbuckridge, KaNyamazane; Nkomazi in Ehlanzeni District Municipality and

Elukwatini in Gert Sibande (Mpumalanga Provincial Government, 2009).

Mining, manufacturing and services are the main economic activities in the province. Mpumalanga is rich in coal reserves and the South Africa's major power stations are located in this province. In addition to coal, Mpumalanga also produces steel and vanadium (Mpumalanga Province Department of Finance, 2013). Extensive forestry plantations, timber processing and paper mills, chrome alloy and steel manufacturing all form part of the manufacturing sector (Mpumalanga Provincial Government, 2013). Agriculture in Mpumalanga comprises of sugar production and crops such as tropical and sub-tropical fruit, maize, wheat, sunflowers, potatoes, nuts as well as livestock farming (NAFCOC, 2014). Tourism also contributes significantly to the province's GDP with Mpumalanga being the third most visited province by foreign visitors (Mpumalanga Province Department of Finance, 2013).

The vegetation of Mpumalanga Province can be divided into three biomes, namely, grassland, savannah and forest. The grassland biome comprises 60 % of the province covering the Highveld and the escarpment hills. Grasslands are important because they play a vital role in water conservation and providing water. Savannas comprise 39% of the province and they cover the lower, warmer regions and are commonly referred to as "bushveld". The vegetation varies from open to dense, thicket-like bushveld and is dominated by woody plants which are mostly deciduous and resistant to fire, drought and browsing. The forest biome accounts for 0.5% of Mpumalanga and occur in areas that receive more than 750mm of rainfall per year, the closed canopies that they form provide moist, shady growing conditions (SANBI, 2007, Schmidt et al., 2002).

About 60% of Mpumalanga Province comprises of the grassland biome and 39% comprises of savanna. Wayant et al. (2010) suggest that grasslands and savanna are more sensitive to climatic fluctuations (especially precipitation) compared to forests and thus NDVI reacts more quickly.

Located between latitudes 22^{o} - 34^{o} S, South Africa's climate is strongly influenced by its position in relation to the major circulation features of the southern hemisphere (Tyson, 1986, Benhin, 2006) Therefore the country is often under the influence of high–pressure systems of the subtropical high-pressure belt. Climates of countries in subtropical regions are characterized by a high degree of intra-annual and inter-annual variability, South Africa being no exception with rainfall in particular being erratic in both time and spatial distribution (Benhin, 2006, Hulme et al., 2001)

Other factors that influence the climate of South Africa are the topography and surrounding oceans. The western and eastern escarpments lead to a high plateau of about 1250m above sea level. The plateau experiences hot summers and cold winters, however, the climate of coastal plains is moderated by oceans and therefore they experience milder winters. The east coast is characterized by a warm and humid climate

due to the Algulhas current, while the west coast is influenced by the cold Benguela current and upwelling (oceanographic phenomenon of the wind-driven rise of deep, dense and cold water to the ocean surface) resulting in this region having an arid climate with lower temperatures (Lutjeharms et al., 2001, Archer et al. (2010)).

There are several factors that affect the distribution of temperature over South Africa, namely latitude, altitude, continentality, temperature region, topographic index and longitude. All else being equal, temperatures decrease with increasing latitude southwards as well as with an increase in altitude. The importance of continentality (position of a site with respect to the source of moisture) is that oceans have a moderating influence on temperature. Temperature regions influence temperature because South Africa comprises of a series of regions in which different dominant factors determine temperature regimes in different seasons of the year. Topographic index is important because of night-time cold air drainage into valley bottoms; this is particularly significant in hilly inland areas. Longitude is mainly important in regions with an east-west alignment (Schulze et al., 1993).

The distribution of rainfall is influenced by altitude, continentality, aspect and rainfall type. It has been proven that rainfall increases with rising altitude, even small terrain features have been shown to play an important role in enhancing rainfall. Continentality affects rainfall because the further inland a moisture laden air mass must travel; the more likely it is that the precipitable water will be reduced due to the orographic effect of previous upliftings and this reduces the chances of rainfall. Aspect is important because it affects the direction of rain-bearing wind. Lastly rainfall type in southern Africa is determined by predominantly frontal systems occurring in winter and convective storms in summer (Schulze and McGee, 1978).

Mpumalanga province has two major regions, the Highveld escarpment and the sub-tropical Lowveld plains, the Northern Drakensberg Mountains are the boundary between these two regions. The Lowveld is sub-tropical because of its proximity to the warm Indian Ocean; in contrast, the Highveld is cooler because of its altitude of 1700m to 2300m above sea level. The Lowveld is very hot in summer and warm in winter and the Highveld is warm to hot in summer but cold at night in winter (Cadman, 2007). Summer temperatures range between 20^{o}C and 38^{o}c with the highest summer maximum temperature (48^{o}C) being recorded in the Mpumalanga and Northern Cape Provinces. Winter temperatures range between 6^{o}C and 20^{o}C (Benhin, 2006).

Southern Africa is described as a predominantly semi-arid region with high intra-seasonal and inter-annual variability with extreme events such as droughts and floods occurring frequently. The average annual rainfall for South Africa is 450mm and rainfall is highly variable in space and a west-east gradient in rainfall totals is evident (rainfall decreases from east to west). Only 10% of the country including the Mpumalanga lowveld,

receives an annual precipitation of more than 750mm (Benhin, 2006). The majority of rainfall occurs in the summer half of the year (October to March) and it peaks between December and February when most of southern Africa receives 80% of its rainfall (Davis, 2011).

In southern Africa, humidity reaches a minimum in winter and a maximum in summer. Mpumalanga province experiences relative humidity in the province ranging from 26% (dry) to 97% (very humid) throughout the course of the year (Davis, 2011). The Highveld often experiences severe frost while the Lowveld is mostly frost free. On occasion, a cold front intrudes into the interior of South Africa resulting in what is referred to as a cold snap. In extreme cases they can cause snowfall in Free State, the Highveld regions and Mpumalanga (Archer et al., 2010).

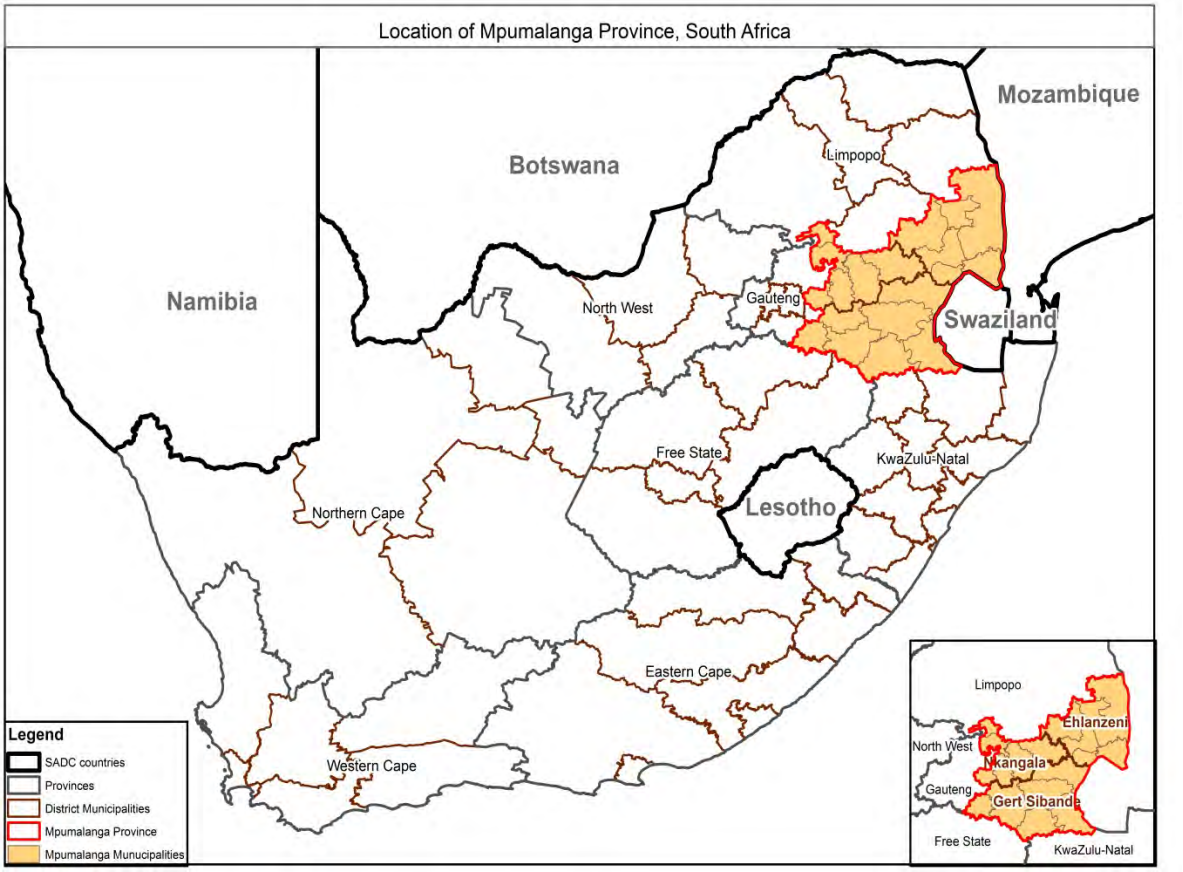


Figure 3-1: Map showing location of Mpumalanga Province

3.2 Malaria Epidemiology Intervention Strategies

In South Africa, malaria is currently restricted to low altitude (below 1000m above sea level) border regions, namely, the provinces of Limpopo, Mpumalanga and KwaZulu Natal (Maharaj et al., 2012). About 5 million people, (10% of overall population of South Africa) are at risk of malaria infection (RBM, 2013). Malaria is

seasonal and predominantly occurs when temperatures are favourable for vector survival, which is generally from September to May with a peak in the rainy months of December and January. Malaria risk areas are characterized by relatively low transmission so the population at risk does not necessarily develop immunity and consequently the population is at risk of severe malaria (RBM, 2013).

Malaria transmission refers to *anopheline* mosquitoes actively transmitting malarial infections in human populations at particular locations (Carter et al., 2000) and is a function of the three-way interaction between humans, Plasmodium parasites and the *anopheline* mosquito vectors that transmit them between hosts (Lyimo and Ferguson, 2009). Malaria is transmitted by female mosquitoes of the genus *Anopheles*, about 70 species of *Anopheles* transmit malaria but only about 30 of these are of importance as vectors (Feachem et al., 2009). The predominant malaria vectors in South Africa belong to *Anopheles gambiae* or *Anopheles funestus* complexes (Maharaj et al., 2013).

There are four species of *Plasmodium* that infect humans; these are *Plasmodium Falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*. *Plasmodium falciparum* is the most virulent and common of these human malaria parasites in Sub-Saharan Africa and it accounts for almost all the malaria mortality. According to reports, Sub-Saharan Africa bears over 90% of the global *Plasmodium falciparum* burden (Robert et al., 2003).

Malaria infection and illness start when a single–celled parasite of the genus Plasmodium invades the human blood stream. Infected female mosquitoes inject motile parasites known as sporozoites into the victims' bloodstream while taking a blood meal. Parasites then invade liver cells and start to reproduce. In one to two weeks, infected liver cells rupture releasing thousands of new parasites known as merozoites which then invade red blood cells and undergo further cycles of asexual reproduction during the course of which many erythrocytes will be erupted. A few merozoites transform into male and female (sexual) stages capable of infecting new mosquitoes, these stages are called gametocytes. Once ingested by a new mosquito during a blood meal, male and female gametes are formed and fuse within the insect's gut ultimately spawning forms that invade its salivary glands from which they enter the next human hosts (Feachem et al., 2009).

Each of the developmental stages of the life cycle of the *Plasmodium* parasite that are discussed above represents a potential target at which the life cycle can be interrupted to prevent transmission of the parasite between the mosquito vector and humans (Greenwood et al., 2008).

In South Africa, sustained malaria control over many decades has succeeded in stopping transmission throughout most of the country (O'Meara et al., 2010). Policy development on all malaria-control interventions takes place at a national level, however the implementation of these policies remains the

responsibility of the provincial departments of health. The three malaria endemic provinces have provincial malaria control programmes which operate as a vertical programme consisting of IRS (indoor residual spraying), surveillance and health promotion (RBM, 2013).

IRS is based on a mosaic approach in which pyrethroids are used for cement-brick structures and DDT is used for traditional mud-walled structures carbamates are also used in some instances (Brooke et al., 2013). IRS spray activities are carried out by malaria spray teams before the main transmission season, usually the end of December and follow up spraying continues until March in areas where it is needed. DDT and WHO Pesticide Evaluation Scheme accredited pyrethroids are used during these spray operations (Govere et al., 2000)

Proper surveillance of malaria cases requires reliable notification of cases and efficient capturing and storage of case information. Malaria has been a notifiable disease in South Africa since 1956, therefore by law, health authorities in the country should be notified of all malaria cases. In addition effective malaria information systems exist in the three malaria endemic provinces (Moonasar et al., 2012).

Health promotion can be defined as enabling people to increase control over and to improve their health. The National Department of Health lists key components of health promotion as (*i*) to advocate at political, healthcare worker and community level, (*ii*) to develop and distribute material on environmental and vector control, chemoprophylaxis, personal protection, signs and symptoms of malaria and (*iii*) to ensure adherence to treatment and to collaborate with partners (Groepe et al., 2013).

Historically, South Africa treated uncomplicated malaria with chloroquine and quinine was used to treat complicated malaria, drug resistance to chloroquine was first reported in KwaZulu Natal in 1987, consequently from 1988 sulphadoxine-pyrimethamine (SP) was used to treat malaria in KwaZulu Natal. Chloroquine resistance was reported in Mpumalanga and Limpopo provinces in 1997 prompting a change to SP for malaria treatment in these provinces as well (Moonasar et al., 2012).

However SP resistance began to increase steadily in the mid-1990s and reached 80% by 2000 therefore drug policy in KwaZulu Natal changed to Coartem in 2001 followed by Limpopo in 2004. Mpumalanga province used SP-artesunate from 2001 to 2005 and changed to Coartem in January 2006 (Moonasar et al., 2012).

Several cross border malaria initiatives were undertaken to reduce the transmission of malaria in the border areas of South Africa (Moonasar et al., 2012). These include the Trans-Limpopo Malaria Initiative (targeting Zimbabwe and Limpopo), the Lubombo Spatial Development Initiative (targeting northern KwaZulu Natal,

Mpumalanga, southern Mozambique and Swaziland) and the MOZIZA Initiative (targeting northern Mozambique and southern Zimbabwe), (RBM.2013).

There are four phases within the malaria elimination continuum, namely, prevention, elimination, pre-elimination and control. KwaZulu Natal is the only province in which all the districts have achieved the minimum elimination incidence level, therefore in KwaZulu Natal, all efforts can now focus on malaria elimination. However, both Limpopo and Mpumalanga have districts at each of the four categories on the elimination continuum therefore in these provinces intervention efforts must focus on control and pre-elimination. (Maharaj et al., 2012).

CHAPTER 4 . METHODOLGY

This chapter outlines the methods used to acquire and process remotely sensed images for use in geographical information systems (GIS) software and the method used to extract climatic data from the images, the methods used to obtain malaria case data as well as the statistical method employed to analyse the relationship between the various climatic variables and malaria cases in the Ehlanzeni district of Mpumalanga Province in South Africa. Some of the methods discussed were derived from the research studies discussed in the literature review in Chapter Three.

4.1 Malaria Case Data

For this study, the individual malaria case data was aggregated to sub place level for a total of 396 sub places within the study area. Sub place boundaries are the smallest administrative boundaries that are assigned a community name and represent a local social boundary equivalent to a split suburb or merged suburb in urban formal areas, a locality in the informal areas and a village in the traditional areas.

The malaria case data was provided by the Office of Malaria Research of the South African Medical Research Council (SAMRC). They obtained records of cases of malaria from the provincial integrated malaria information system (IMIS) which is regulated by the Mpumalanga Malaria Control Program of the Department of Health. This system was developed by the SAMRC, a national research organisation in South Africa, using Microsoft Access for data entry and validation. Malaria morbidity and mortality data consisting of both passive and active cases based on definitive diagnosis reported from December 2005 to December 2006 were provided from the IMIS for the purposes of this study. Only malaria cases that were reported in the Ehlanzeni district were selected for analysis as it is a malaria prone district in Mpumalanga. The district borders Swaziland and Mozambique and therefore there are many imported cases that are recorded, for this analysis, only local cases (people who contracted malaria in South Africa) were extracted from the database.

Data consisted of the following variables - date of diagnosis, gender, age, type of mosquito species that infected the individual and their residential locality, health facility name (where individual was screened), municipality, province and source country.

The malaria case data used in this study was for the year 2006, this was because it was the most comprehensive dataset that was made available.

4.2 Climatic data

Climatic data for Mpumalanga Province was obtained from remotely sensed images. Data for all climatic variables was obtained for the period of 1 December 2005 to 31 December 2006.

The main reason for using remotely sensed climatic information is that meteorological stations are often located far apart and have a poor network density. However satellite based data provide a high spatial data density and therefore provide more accurate climatic information because they provide continuous data over larger areas. Satellite sensor observations also have another advantage in that data can be gathered at the actual location of interest. Detailed and accurate climatic data is important because climate variability can affect the spatial and seasonal distribution of malaria transmission.

4.2.1 Land surface temperature (LST)

The MODIS/Aqua LST and Emissivity product was obtained from the moderate resolution imaging spectroradiometer (MODIS) sensor that is an instrument which records information from two satellites, one in the morning (Terra) and one in the afternoon (Aqua). Daily LST data was obtained from NASA's EODIS Reverb tool website http://reverb.echo.nasa.gov/reverb/ at a resolution of 1km x 1km in HDF-EOS (hierarchal data format – earth observation system) format. The HDF images were then converted to GeoTIFF format using the Modis Reprojection Tool (MRT) version 4.1 and the projection selected was geographic. The images contained temperature values in Kelvin (K); the raster calculator tool in ArcGIS version 10 was used to apply a conversion factor as well as to convert temperature values into degrees Celsius (^{0}C).

4.2.2 Rainfall

Rainfall estimate data (RFE) was retrieved from the http://www.esrl.noaa.gov/ site. RFE was obtained from RFE version 2.0 that was implemented by the National Oceanic and Atmospheric Administration's (NOAA) Climate Prediction Centre. RFE v2.0 obtains the final daily rainfall estimation by combining all satellite data using the maximum likelihood estimation method, thereafter, Global Telecommunication System (GTS) station data are used to remove bias. The daily data are then summed up to produce decadal (10 day) totals in mm for each month at a resolution of 8km x 8km in BIL file format. The images were then projected into a geographic projection and converted into GeoTIFF format using ArcGIS v10.

4.2.3 Humidity

Remotely sensed data for monthly humidity was obtained from the NOAA's site http://www.esrl.noaa.gov/ in netCDF file format at a resolution of 0.25^0 x 0.25^0 in percentages. A tool in ArcGIS v10 called Make NetCDF Raster Layer was used to convert the images into GeoTIFF format.

4.2.4 Natural vegetation index (NDVI)

Daily NDVI data was obtained from MODIS through NASA's EODIS Reverb tool website http://reverb.echo.nasa.gov/reverb/ at a resolution of 1km x 1km in HDF-EOS format. The HDF images were then converted to GeoTIFF format using the Modis Reprojection Tool (MRT) v 4.1 and the projection selected was geographic. The raster calculator tool was used to apply a conversion factor to obtain the correct units of reporting NDVI. NDVI is a proxy measure of vegetation cover and moisture that ranges in value from 1 to -1. When the values are positive, it indicates the presence of vegetation and negative values or values close to zero represent barren soil or water surfaces.

4.2.5 Altitude

Altitude was obtained by creating a triangulated irregular network (TIN) in ArcGIS v10 using contour lines at 5m intervals. Contours are a common source of digital elevation data, all the vertices of the contour lines are used as mass points for triangulation. TIN is a digital data structure used in geographic information systems for the representation of the physical land surface made up of irregularly distributed nodes with three-dimensional coordinates that are arranged in a network of non-overlapping triangles. Contour data was obtained from http://www.planetgis.co.za.

4.2.6 Lag rainfall and lag temperature

A lag time is a period between two related events; lagged effects with a lag of 1 month for LST and rainfall were added to climatic variables to account for possible delay of the effect of these predictor environmental variables on the number of malaria cases.

4.2.7 Extraction of climatic data

The zonal statistics tool in ArcGIS v10 was used to extract averages of the climatic variables obtained through remote sensing. The output was a CSV file that could be imported into RStudio to use when performing the randomForest algorithm. Each CSV file contained monthly values for all the discussed climatic variables for each administrative boundary (sub place) in the Ehlanzeni district.

4.3 Statistical method

The relationship between climatic variables and the occurrence of malaria cases was analysed using the random forests (RF henceforth) algorithm.

The RF algorithm is based on model aggregation ideas for regression and classification problems that was developed by Breiman (2001). For the statistical analysis in this study, the RF package (available at

http://cran.r-project.org/web/packages/randomForest/index.html) was downloaded into RStudio which is an integrated development environment (IDE) for R. R is a programming language and software environment for statistical computing and graphics.

There two common methods in ensemble learning, boosting (used bySchapire et al. (1998) and bagging used by Breiman (1996). Boosting uses successive trees to give extra weight to points that were incorrectly predicted by earlier predictors and in the end, a weighted vote is taken for prediction. In bagging, however, successive trees do not depend on earlier trees as each is independently constructed using a boot-strap sample of the data set. In the end, a simple majority vote is taken for prediction(Liaw and Wiener, 2002).

Random forests is the latest addition to ensemble learning, it adds an additional layer of randomness to bagging. Each tree is constructed using a different bootstrap sample of the data, and each node is split using the best split among a subset of predictors randomly chosen at that node (Breiman, 2001). This seemingly counterintuitive strategy has proved to perform very well compared to many other classifiers such as discriminant analysis, support vector machines and neural networks and it is also robust against overfitting (Breiman, 2001).

4.3.1 Random Forest for Regression

The RF algorithm as explained by Liaw and Wiener (2002) is as follows:

1. Draw n_{tree} bootstrap samples from the original data
2. For each of the bootstrap samples, grow an *unpruned* classification or regression tree, with the following modification: at each node, instead of choosing the best split among all predictors, randomly sample m_{try} of the predictors and choose the best split from among those variables
3. Predict new data by aggregating the predictions of the n_{tree} trees (i.e. majority votes for classification, average for regression)

An estimate of the error rate can be obtained based on the training data, by the following:

1. At each bootstrap iteration, predict the data not in the bootstrap sample (what Breiman calls “out-of-bag” or OOB data) using the tree grown with the bootstrap sample. The OOB sample is the set of observations which are not used for building the current tree, it is used to estimate the prediction error and then to evaluate variable importance.
2. Aggregate the OOB predictions. On average each data point would be out-of-bag around 36% of the times, so aggregate these predictions. Calculate the error rate and call it the OOB estimate of error rate.

As stated previously, the RF algorithm produces an important piece of additional information which is a measure of the importance of predictor variables that is referred to as variable importance. Variable

importance is the increasing in mean of the error of a tree (mean square error (MSE) for regression and misclassification rate for classification) in the forest when the observed values of this variable are randomly permuted in the OOB (out-of-bag) samples (Genuer et al., 2010). There are two main reasons for quantifying variable importance, the first is to rank variables used in a model in order of importance so as to identify relevant predictor variables (true predictors) and the second is to interpret data and to understand the impact of predictor variables in prediction of outcomes or their causal effect (Strobl et al., 2008).

RF variable importance can be defined as, for each tree *t*:

- Consider the associated OOB_t sample
- Denote by $errOOB_t$ the error of a single tree t on this OOB_t sample
- Randomly permute the values of X^j in OOB_t to get a perturbed sample denoted by OOB_t^j and compute $errOOB_t^j$, the error of predictor t on the perturbed sample.

$$VI(X^j) = \frac{1}{ntree}\sum_t \left(errOOB_t\dot{j} - errOOB_t\right)$$

RF for regression was used to create a model to predict cases of malaria for each sub place for each month of the year 2006. The monthly *x* input variables (independent predictor variables) were rainfall, temperature, altitude, humidity, NDVI and lag temperature and rainfall extracted for each sub place. The monthly *y* input variable was the sum of malaria cases for each sub place in the study area. Variable importance index was used to select the top three predictor variables in each model.

Averages over 50 runs of the random forest algorithm were used on each dataset because studies by Geurts et al. (2006), Genuer et al. (2010) and Abdulsalam et al. (2011) show that 50 runs provided accurate, reliable results. A total of 50 runs of RF with n_{tree} = 2000, default m_{try} and n = 396 observations were performed for each month. The variable importance index of each run was recorded and the climatic variables were ranked in descending order of importance averaged from the 50 runs. These results are shown in the results section.

A default value for m_{try} (number of input variables randomly chosen at each spilt) was used because choice of m_{try} makes little difference to the overall result because RF is not particularly sensitive to this value.

The value of n_{tree} (number of trees to grow) = 2000 was used because it resulted in a low error rate and it ensured that each every input gets predicted a number of times to produce reliable results.

The call to function randomForest in the RF package is:

randomForest (x=x, y=y, ntree=2000, importance =TRUE)

Type of random forest : regression

Number of tress: 2000

Number of variables tried at each split : 1

4.3.2 Model validation

When performing the RF algorithm, there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error. This is because it is estimated internally as described previously (error rate). However, assessing model fit is an important step in data analysis and the co-efficient of determination also referred to as R^2 is a very simple tool to assess the quality of the fit of a linear regression model by proving an indication of the suitability of the chosen explanatory variables in predicting the response (Renaud and Victoria-Feser, 2010). Therefore model validation was performed by calculating the co-efficient of determination for each monthly predictive model generated by the RF algorithm.

CHAPTER 5 . RESULTS AND DISCUSSION

Figures 5-1 and 5-2 shows a time series of the weekly and monthly counts of malaria cases respectively for the study site from January 2006 to December 2006, there is a pronounced seasonal trend that is evident in the data. It is clear that malaria cases fluctuated throughout the course of the year and that transmission was seasonal peaking between October and May. There was a decline in malaria cases between June and September followed by an increase in transmission. The climatic factors responsible for this pattern of seasonality will be explored in the ensuing discussion. The seasons that South Africa experiences are defined by (Kruger and Shongwe, 2004) as being autumn (March to May), winter (June to August), spring (September to November) and summer (December to February).

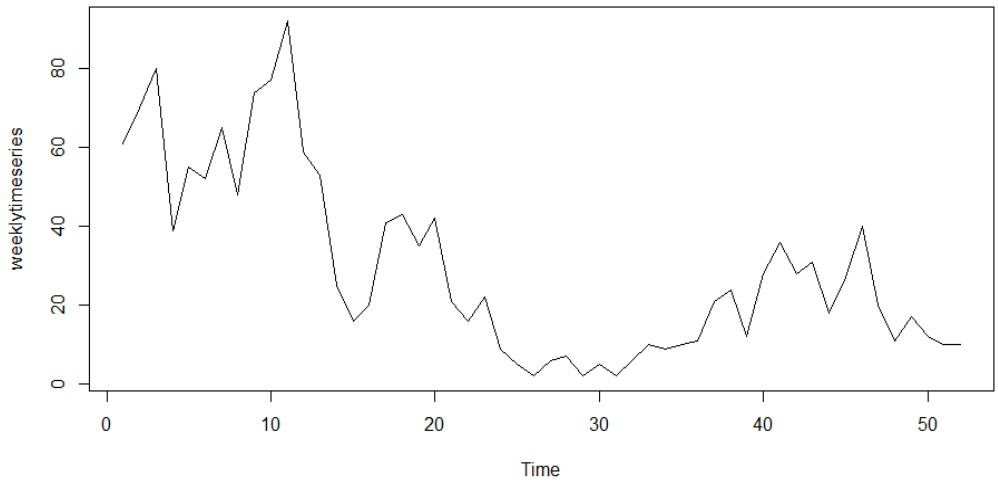


Figure 5-1. The weekly number of reported malaria cases in Ehlanzeni District for 2006

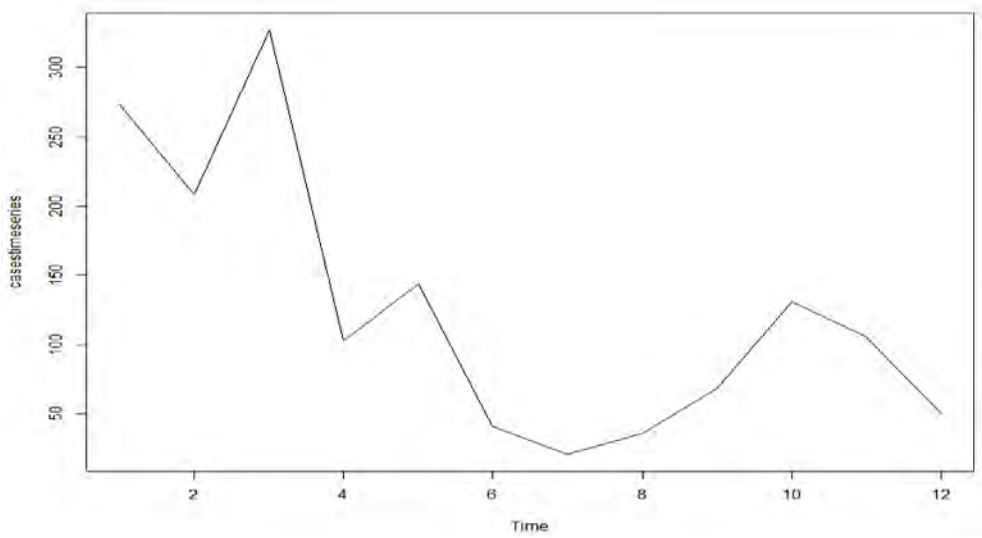


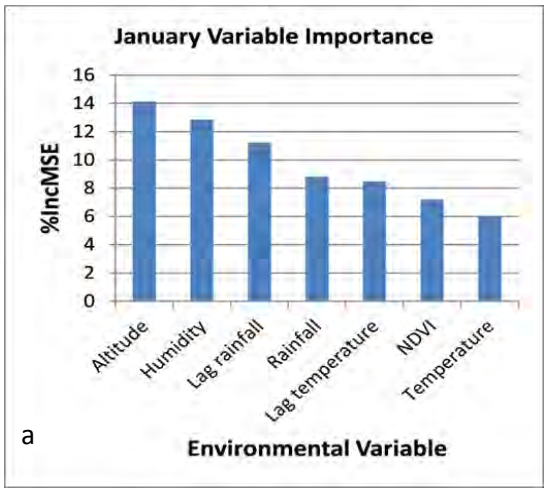
Figure 5-2. The monthly number of reported malaria cases in Ehlanzeni District for 2006

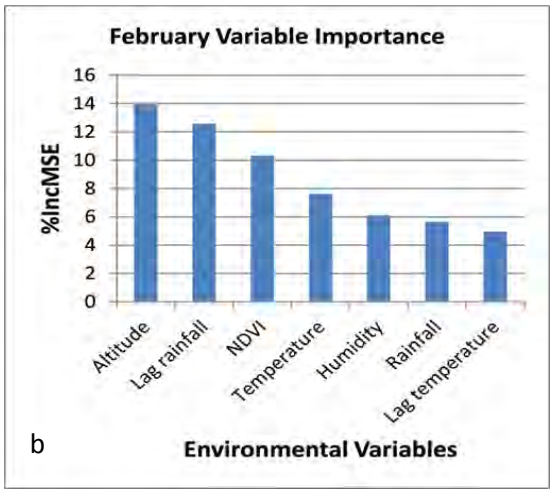
Appendix 1 shows graphs that represent the distribution of malaria cases in the Ehlanzeni District across all the sub places on a monthly basis for detailed scrutiny. As all the histograms show, there is a presence of a high number of zeroes in the data. These zeroes are true zeroes because they represent the absence of malaria cases in those sub places, for this reason, they cannot be excluded from the random forest algorithm. Cameron and Trivedi (2013) state that real life data frequently displays over dispersion through excess zeroes and a zero value has special appeal in many situations because it partitions the population into subpopulations in a meaningful way. If we apply this statement to this study, the importance of zeroes in the dataset of malaria cases is that the zero values separate sub places that have actual observed malaria cases and those that do not have any cases. However the inclusion of this large number of zeroes in the predictive

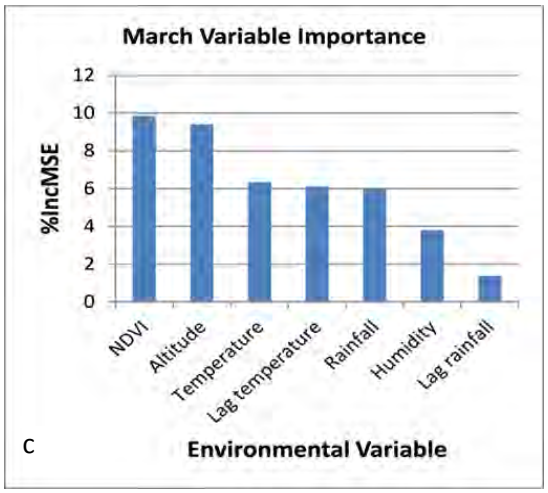
models resulted in them adding additional mass at the zero values resulting in a higher probability of this value.

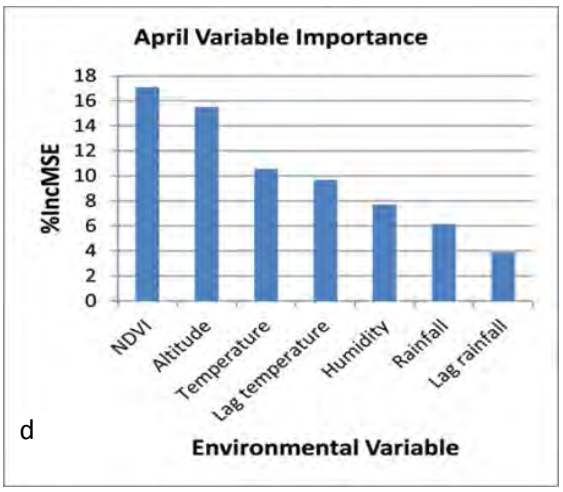
Average monthly mean temperature, rainfall, humidity, NDVI, lag temperature, lag rainfall and altitude were extracted for each sub place in the Ehlanzeni district. To observe the correlation between meteorological variables and malaria cases, monthly malaria cases were regarded as the dependant variable while meteorological variables were regarded as independent variables. The random forest (RF) algorithm was used to examine the strength of the relationship between meteorological variables and malaria cases by providing a measure of variable importance.

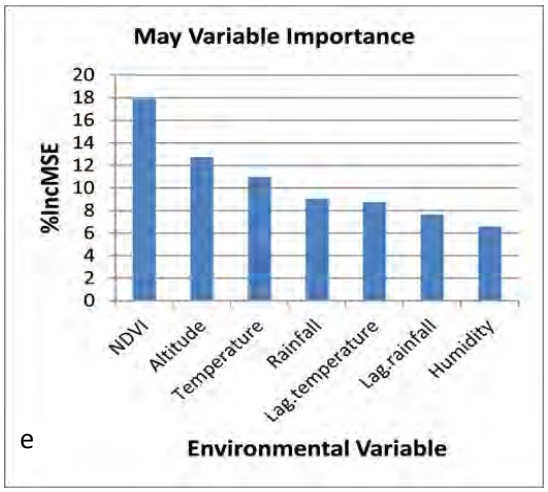
A common question that needs to be addressed in modelling is what predictor variables are more important to a model and to what degree they are important. Figure 5-3 below shows the variable importance ranking of climatic variables averaged from 50 runs of the RF algorithm using all 7 predictor variables for each month of the year, the variables are sorted in decreasing order of RF scores of importance.

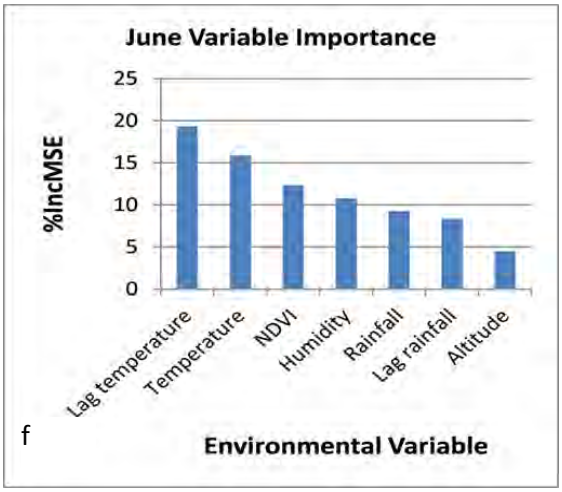

January Variable Importance
%IncMSE
16
14
12
10
8
6
4
2
0
Altitude
Humidity
Lag rainfall
Rainfall
Lag temperature
NDVI
Temperature
Environmental Variable
a


February Variable Importance
%IncMSE
16
14
12
10
8
6
4
2
0
Altitude
Lag rainfall
NDVI
Temperature
Humidity
Rainfall
Lag temperature
Environmental Variables
b


March Variable Importance
%IncMSE
12
10
8
6
4
2
0
NDVI
Altitude
Temperature
Lag temperature
Rainfall
Humidity
Lag rainfall
Environmental Variable
c


April Variable Importance
%IncMSE
18
16
14
12
10
8
6
4
2
0
NDVI
Altitude
Temperature
Lag temperature
Humidity
Rainfall
Lag rainfall
Environmental Variable
d


May Variable Importance
%IncMSE
20
18
16
14
12
10
8
6
4
2
0
NDVI
Altitude
Temperature
Rainfall
Lag temperature
Lag rainfall
Humidity
Environmental Variable
e


June Variable Importance
%IncMSE
25
20
15
10
5
0
Lag temperature
Temperature
NDVI
Humidity
Rainfall
Lag rainfall
Altitude
Environmental Variable
f

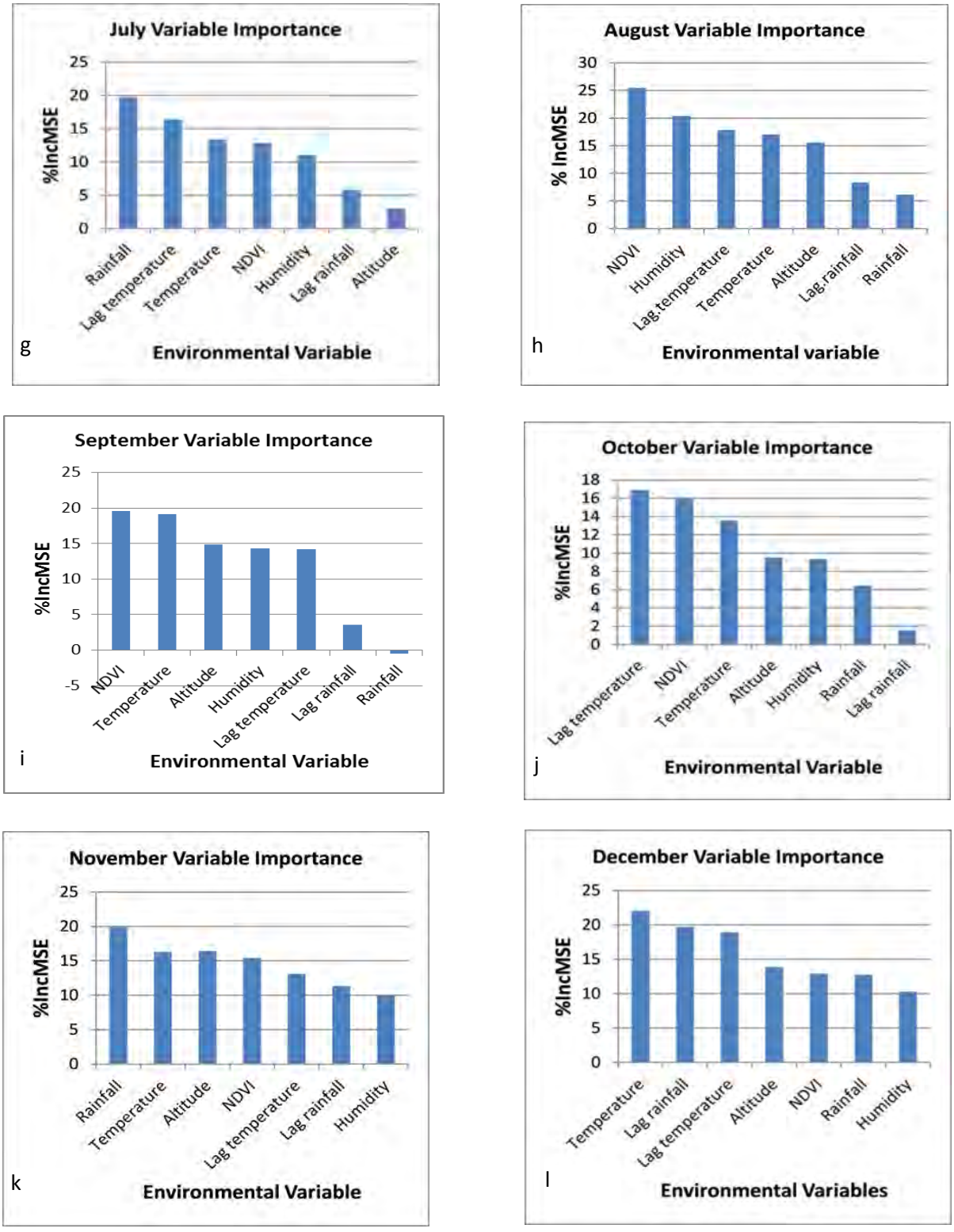
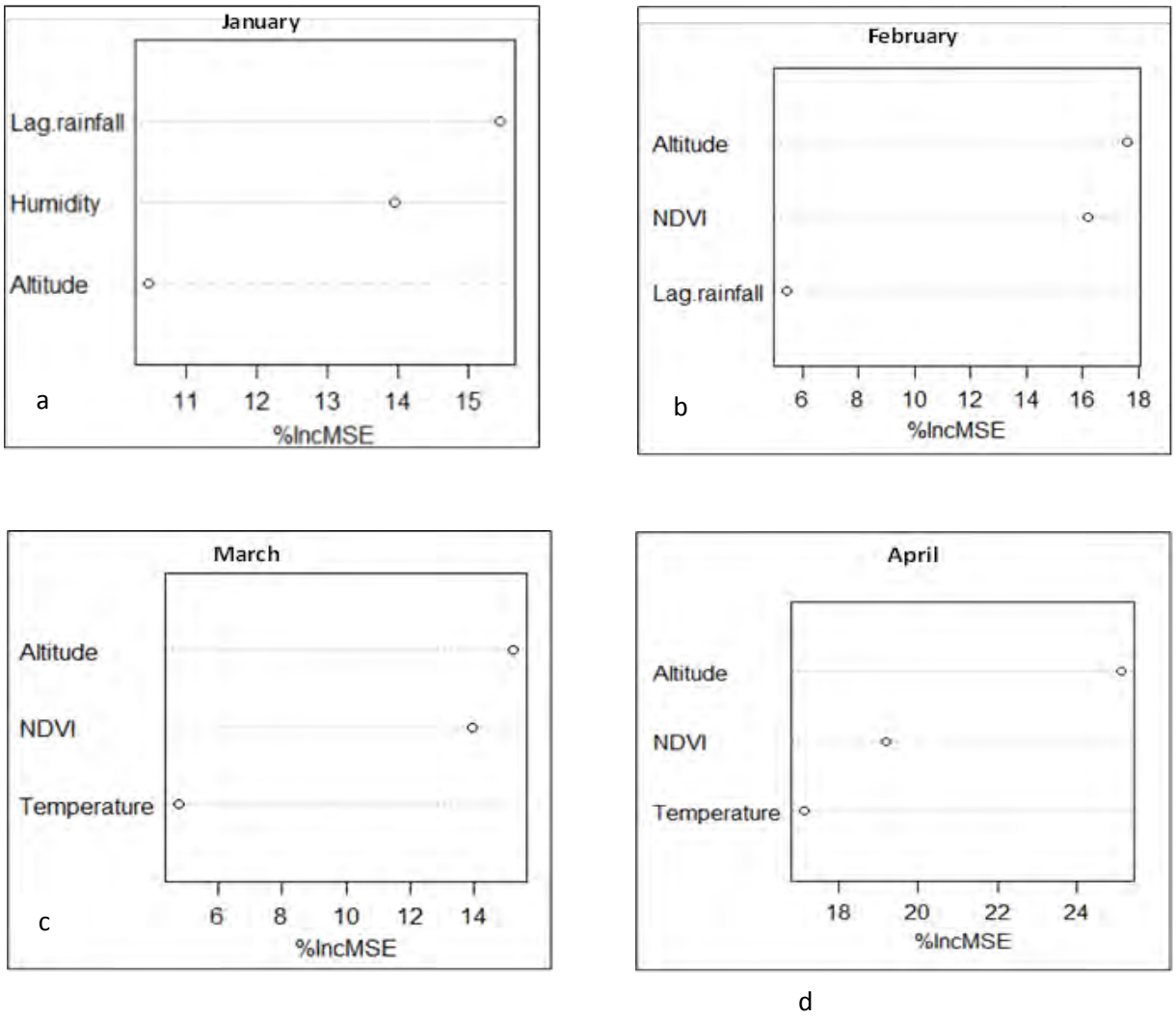


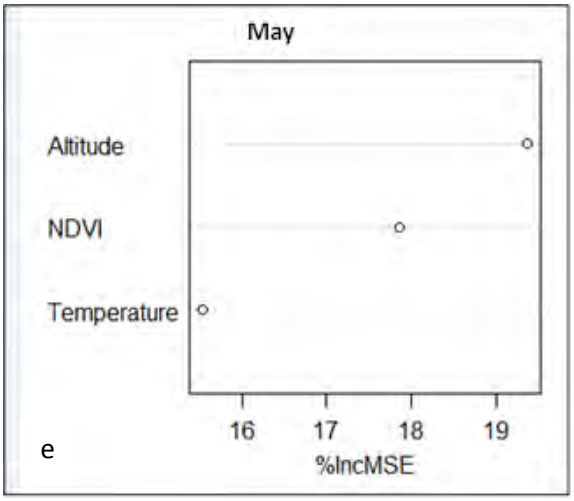
Figure 5-3. Variable importance index of predictor variables averaged from 50 runs of RF

This step was performed as part of a variable selection procedure to select the top 3 statistically significant variables that are highly correlated to the response variable. The higher the percent increase in mean squared error (%IncMSE) value of a predictor is, the higher the importance of that predictor in predicting the outcome (malaria cases). A low importance value indicates a poor relationship between the predictor variable and the

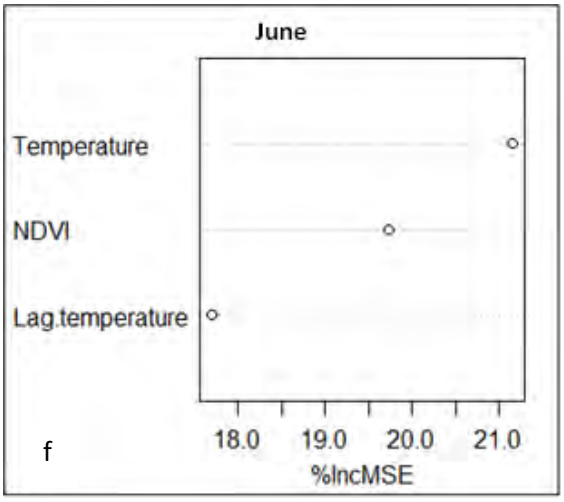
outcome. The importance of preliminary elimination and ranking is to cancel the variables of small importance for further RF analysis.

The RF algorithm was then performed on the data for each month using the top 3 important variables that were identified as being highly related to the response variable in the previous step. Figure 5-4 shows the ranking patterns of the climatic variables. There was no single variable that was consistently dominant throughout all the months.

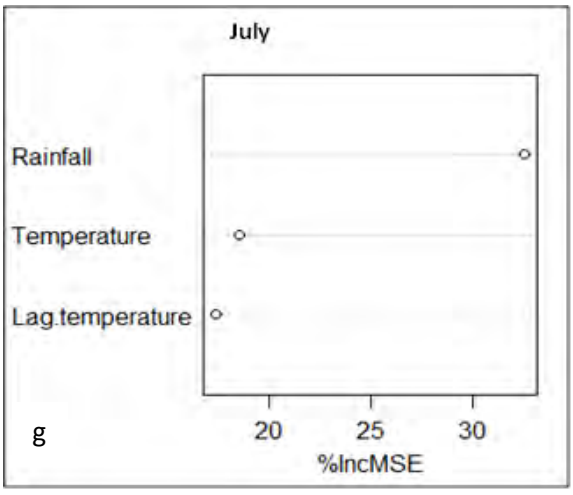



May
Altitude
NDVI
Temperature
16
17
18
19
%IncMSE

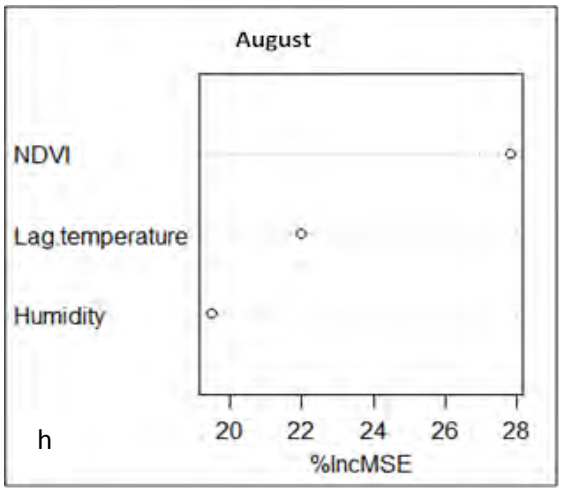
e


June
Temperature
NDVI
Lag.temperature
18.0
19.0
20.0
21.0
%IncMSE

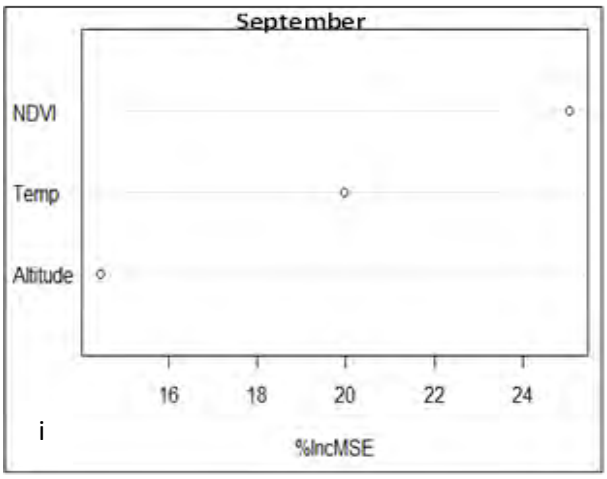
f


July
Rainfall
Temperature
Lag.temperature
20
25
30
%IncMSE

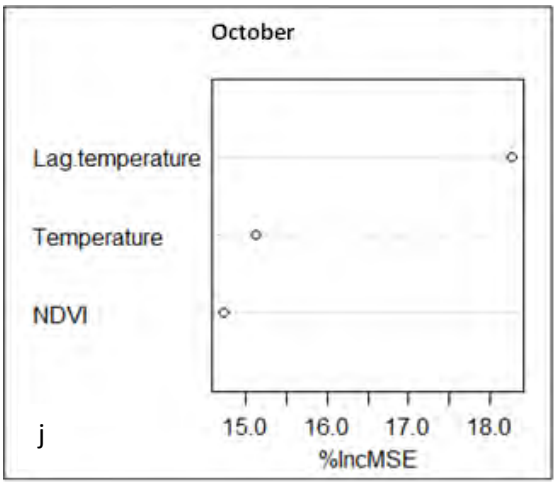
g


August
NDVI
Lag.temperature
Humidity
20
22
24
26
28
%IncMSE

h


September
NDVI
Temp
Altitude
16
18
20
22
24
%IncMSE

i


October
Lag.temperature
Temperature
NDVI
15.0
16.0
17.0
18.0
%IncMSE

j

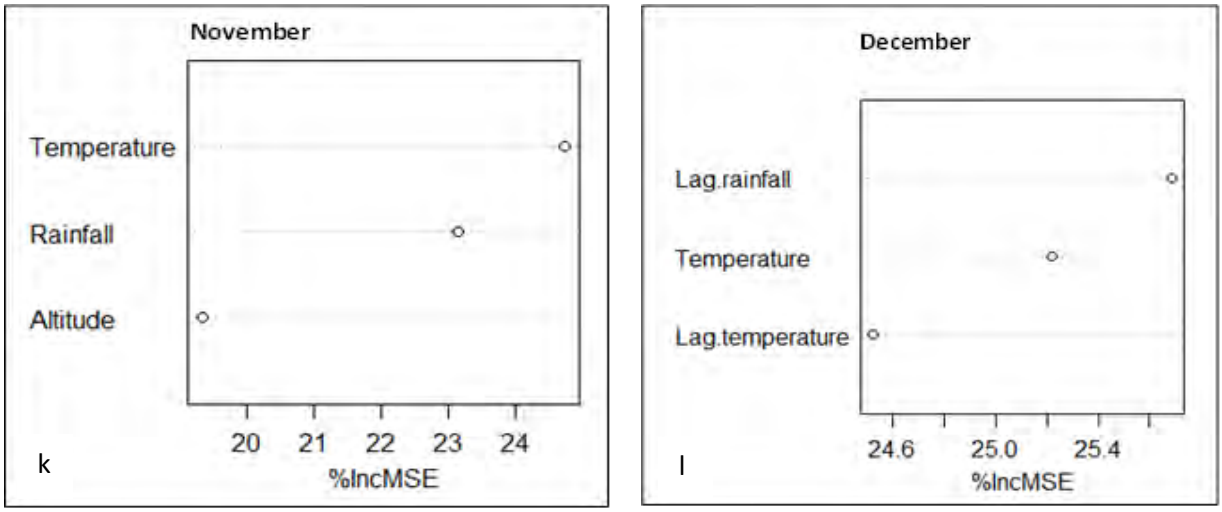


Figure 5-4. Variable importance index of random forest algorithm using top 3 statistically significant variables

All the climatic variables in Figure 5-4 are relevant predictor variables because they have significant values for %IncMSE whereas irrelevant variables have values that are negative or closer to zero. The significant variables identified above can then be used in more sophisticated modelling techniques and further analysis of the influence of climate on malaria transmission.

Altitude, NDVI and temperature were much more likely to be selected as the most important variables for predicting malaria cases in each month; all 3 were selected 8 times. These appear to be the most reliable predictors because they are associated consistently with the response, variables that explain the most observations are selected most frequently. Lag temperature and lag rainfall appeared 5 and 3 times respectively and humidity and rainfall both appeared 2 times.

A study by Craig et al. (2007) to develop a malaria risk map using a systematic variable selection process found that out of 50 potential explanatory variables from 8 environmental data themes, 3 were identified as being independently and significantly associated to malaria prevalence, these were rainfall, temperature and altitude. The results in this study are similar to this finding because temperature and altitude were 2 out of 3 climatic variables that were identified as being consistently associated with malaria throughout the year.

When comparing the top predictor variables for each month, Figure 5-4g shows that rainfall had the highest measure of importance because it had the highest %IncMSE of about 34%, therefore it was the climatic variable that made the greatest contribution to the prediction of malaria in the monthly models. This is an important finding because rainfall and temperature have been proven to be two of the major environmental variables triggering malaria epidemics in warm semi-arid and high altitude areas because epidemics occur in these regions after excessive rain or increases in temperature (Ceccato et al., 2012). Rainfall was followed by NDVI (Figure 5-4h) and lag rainfall (Figure 5-4l) with %IncMSE values of 28 and 26 % respectively.

Table 5-1 shows that of all 7 variables, altitude was selected as the top predictor variable for 4 months of the year (February, March, April and May). Lag rainfall was the top predictor for the months of January and December, NDVI was the top predictor for August and September, temperature was the top predictor variable for June and November, rainfall was the top predictor for only one month (July) and lag temperature was also the top predictor for only one month of the year (October).

Table 5-1. Variables selected as top predictor variables for each month

	Altitude	Humidity	Lag rainfall	Rainfall	Lag temperature	NDVI	Temperature
January			✓				
February	✓						
March	✓						
April	✓						
May	✓						
June							✓
July				✓			
August						✓	
September						✓	
October					✓		
November							✓
December			✓				
Count	**4**	**0**	**2**	**1**	**1**	**2**	**2**

Altitude seems to be the most robust predictor variable because it was selected as a top predictor variable for more months than the other climatic variables. Ngomane et al. (2012) found that malaria incidence was more pronounced in the low altitude region of Ehlanzeni district in comparison to the high altitude regions of Nkangala and Gert Sibande districts. This result suggests that altitude has an effect on malaria transmission and it echoes an investigation by Kazembe et al. (2007) that documented a higher rate of malaria incidence at elevations below 1500m when compared to higher altitudes. Further studies show that altitude is a key determinant of malaria risk because it restricts mosquito habitats. This is due to the findings that with every 1000 m increase in elevation, temperature decreased by 6^0C. (Patz et al., 2008). Craig et al. (2007) also found a strong positive association with malaria prevelance, they report an increase in logit(p) of 1 every 160m. A drop in temperature results in a decline in the risk of infection because parasite development is restricted; the minimum temperature for *Plasmodium falciparum* development is said to be between 16^0C and 18^0C (Lindsay and Martens, 1998).

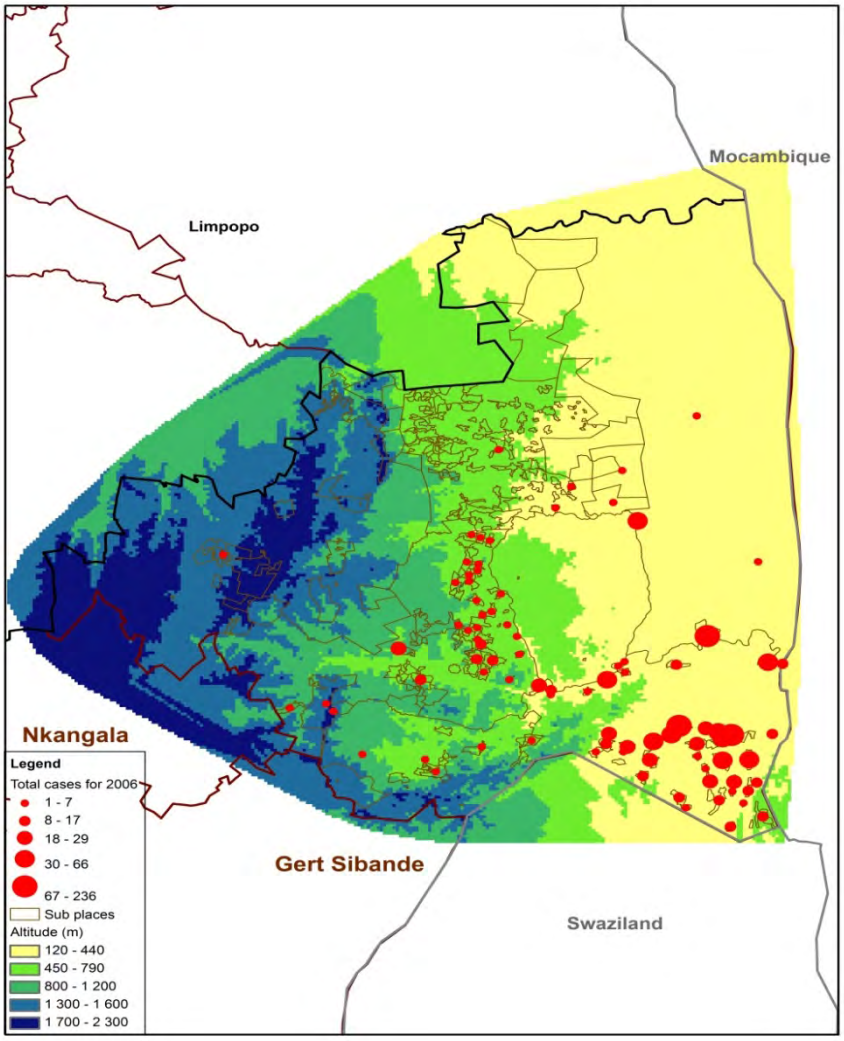


Figure 5-5: Association between altitude and malaria cases

Figure 5-5 depicts the relationship between altitude and malaria; it is evident that the occurrence of malaria cases decreases from east to west as altitude increases. The highest number of cases occurs at altitudes less that 600m and decreases towards the west at altitudes greater than 600m.

Lag rainfall, NDVI and temperature were the second most common predictors, each having been selected as a top predictor for 2 months.

Rainfall was lagged by one month because the *Anopheles vector* takes two weeks to complete their life cycle and a further two weeks to generate parasites in the new host (Kumar et al., 2014). Davis (2011), states that rainfall in South Africa reaches a peak between December and February. Figure 5-6 shows that the rainfall in February had the greatest influence on the number of malaria cases in March of 2006. This finding corroborates the study by Ngomane et al. (2012) that analysed the relationship between rainfall and malaria and found that depending on the amount of rainfall, upsurges in malaria transmission were observed with a

time lag of 1-2 months. This is because epidemiologically, rainfall has a delayed impact on malaria incidence due to the incubation and latent phases of the parasite in the vector and the host(Silal et al., 2013). Similarly Loevinsohn (1994) also suggest that rainfall impacts on malaria through lag periods of 1 – 2 months not only through its influence on parasite development but also because there is a delay in runoff and in seepage collecting in low-lying breeding sites. The general trend observed in the data analysis of monthly totals of lag rainfall throughout the year and monthly malaria cases was an increase in lag rainfall corresponds to an increase in malaria cases. This finding is further supported by Alemu et al. (2011) who concluded that total monthly rainfall was associated with occurrence of malaria with a month lag effect in Jimma town in South West Ethiopia.

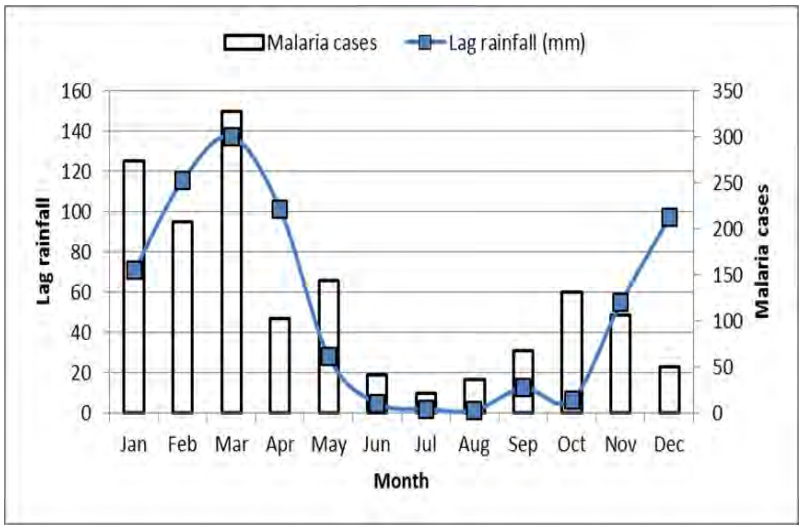


Figure 5-6. Trend of lag rainfall and malaria cases

NDVI can be defined as a measure of vegetation conditions and values range between +1.00 to -1.00 (Haque et al., 2010), the higher the NDVI value, the denser or healthier the vegetation is regarded as being. NDVI values above 0.2 usually represent areas covered by vegetation, and negative values represent water or buildings and asphalt (Machault et al., 2011). Figure 5-7 shows that the monthly NDVI values in the study area were all well above 0.2 throughout the year, this shows that the area is well vegetated thus increasing the chances of survival for adult mosquitoes. it is also evident that NDVI values drop during the months with lower rainfall (May, June, July, August, September). This decline in NDVI coincides with a decline in malaria cases for those same months. These results are similar to those obtained by Gaudart et al. (2009) that show there was a high mortality rate of *Anopheles* at the lowest values of NDVI (dry seasons).Conversely, the months with higher rainfall and hence healthier vegetation have higher cases of malaria.

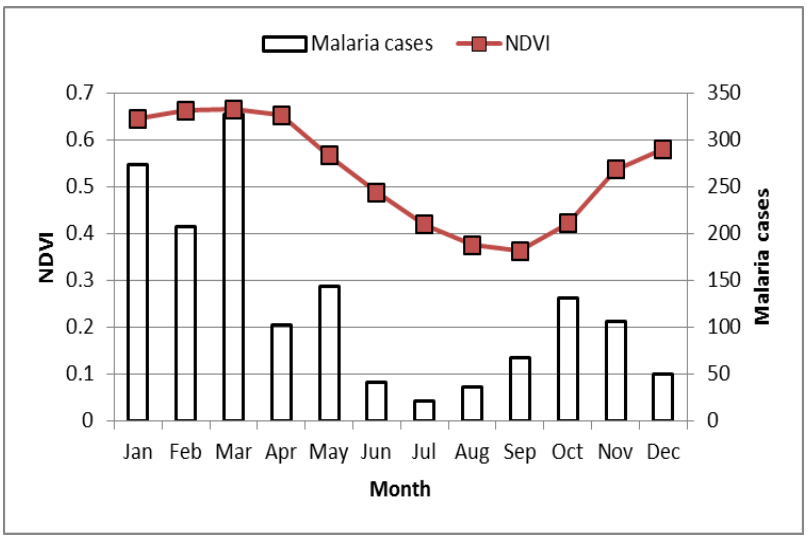


Figure 5-7. Trend of NDVI and malaria cases

A study by Eisele et al. (2003) found that the number of potential anopheline larval sites increased as NDVI increased. This finding is substaiated by a study by Amek et al. (2012b) in which a Bayesian model was fitted to predictict mosquito density using rainfall, temperature and NDVI as predictive variables. The results showed that NDVI was positively associated with mosquito density. These studies indicate that the higher the value of NDVI, the higher the malaria risk. An explanation for this is that high NDVI values represent healthy vegetation and since vegetation improves adult mosquito survival by providing resting sites, protection from climatic conditions and sugar feeding supplies for adult mosquitoes, there is a strong likelihood that high NDVI values represent an abundance of mosquitoes and therefore increased malaria transmission and consequently an increase in malaria risk. This study did not use land cover as a predictor variable because NDVI is a measure of the amount of photosynthetically active vegetation, and is thus a proxy for land cover (Hay et al., 2002).

Snow and Gilles (2002) concluded that at 25^{o}C, *Plasmodium falciparum* requires only 12 days to undergo parasite development but at 20^{0}C, it requires over 30 days to undergo development and render a mosquito infectious. The more time the parasite needs to mature, the probability that a mosquito will live long enough for the parasite to spread the infection decreases (Ikemoto, 2008). Figure 5-8 below supports the above findings because it illustrates that the highest malaria incidence was reported in March when the temperature was 25.56^{0}C, in line with the conclusion that the parasite develops faster at an optimum temperature of 25^{o}C resulting in increased transmission of malaria and the lowest incidence was reported in June when the average temperature was 20.96^{0}C. It is also thought that temperature could affect the gender of mosquitoes because Yang et al. (2009) found that female mosquitoes survived more than males in the optimum temperature range of nearly 25^{0}C, this finding is important because malaria is transmitted by female mosquitoes of the genus *Anopheles*.

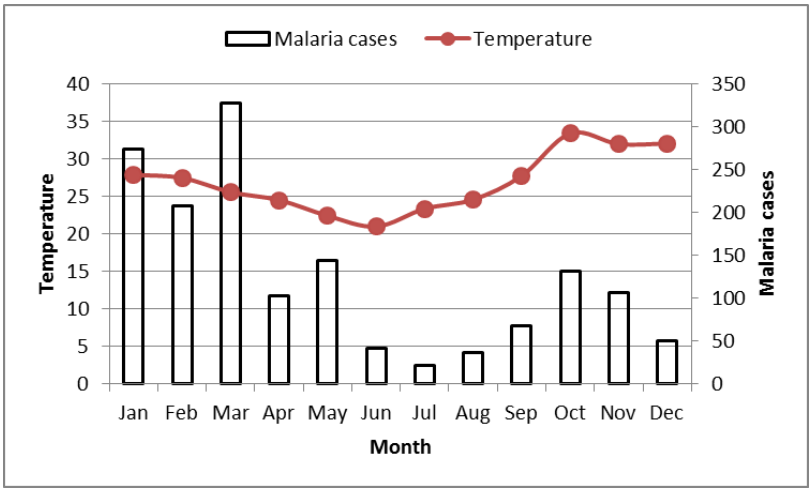


Figure 5-8. Trend of temperature and malaria cases

Rainfall and lag temperature were both variables that were least likely to be selected because they were the top predictors for only one month each.

The general pattern observed in Figure 5-9 is that heavy rainfall coincides with a high number of malaria cases and in this study, the heaviest rainfall occurred in March. Rainfall plays a significant role in the mosquito life cycle because the immature stages of *Anopheles* mosquitoes are aquatic meaning they depend on free standing water for their survival and development (Warrell and Gilles, 2002) and this is why rainfall is important in malaria transmission because water provides a habitat for mosquitoes to lay their eggs and for the development of *Anopheles* larvae and pupae (Stresman, 2010). Studies have shown that *Anopheles* can breed in sites where water has been present for 10-14 days, depending on the time required for the mosquito life cycle to take place (Stresman, 2010).

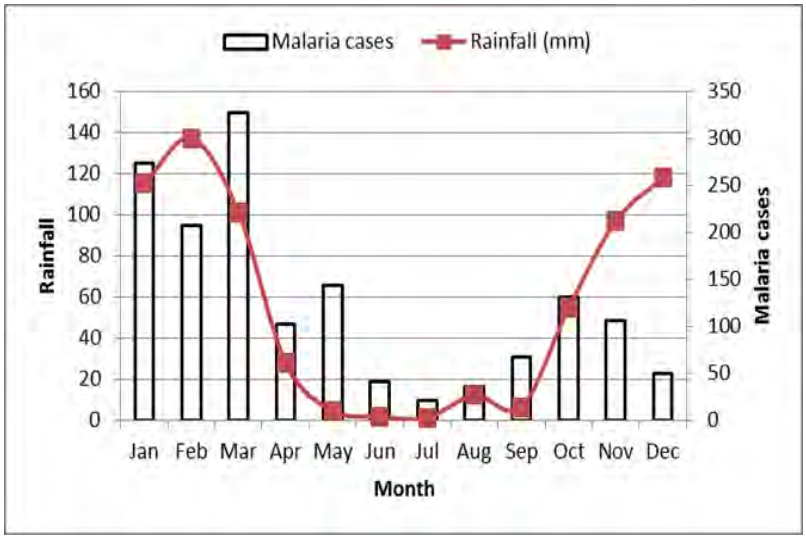


Figure 5-9. Trend of rainfall and malaria cases

The study conducted by Fournet et al. (2010) found that the parasitological values differed between the rainy and dry seasons and this showed that the malaria transmission dynamic increased during the rainy season where the density of vectors also increased. This dynamic is clearly shown in figure 5-9 above where it is evident that malaria transmission increased in the rainy months and decreased during the drier, winter months. Govere et al. (2001) state that the decline in mosquito populations and consequently transmission at the end of the rainy season is due to breeding sites drying up.

Figure 5-10 illustrates that the lag of the lowest temperature (20.96^{0}C for July) resulted in the lowest number of malaria cases (21 cases) reported for the year.

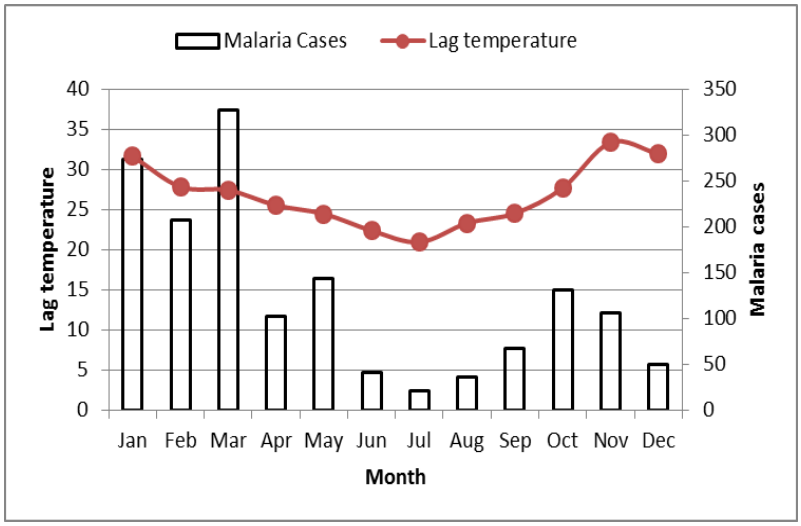


Figure 5-10. Trend of lag temperature and malaria cases

Correlation and regression analysis on monthly climatic variables and monthly malaria incidence in Shuchen County, China, suggest that temperature and rainfall act on malaria with a lag of one month (Bi et al., 2003). Therefore it was significant that both lag temperature and lag rainfall were selected as top predictor variables with the RF algorithm. Furthermore it has been shown that there is a 1 – 2 month lag between peak anomalous temperature and rainfall and malaria epidemics (Githeko et al., 2012). Reid et al. (2010) also report that a study conducted in the east African Highlands shows that a 1^{0}C increase in minimum temperature with a lag time of 1 – 2 months.

Although humidity was not selected as a top predictor variable for any of the months, it was selected as the second and third important variable for the months of January and August respectively. Humidity plays an important role in malaria transmission because adult mosquitoes are dependent on specific moisture content in the air and they desiccate if the air is too dry. The high surface area to volume ratio of mosquitoes makes them especially sensitive to desiccation at low humidity levels (Yamana and Eltahir, 2013).

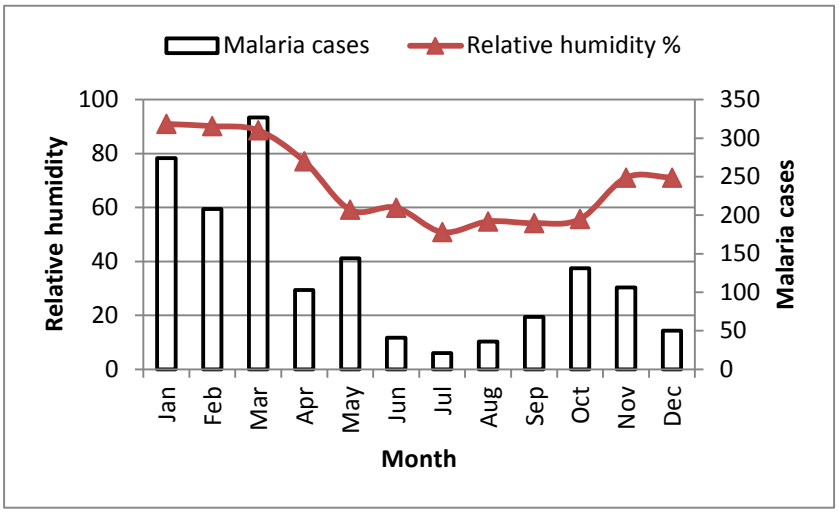
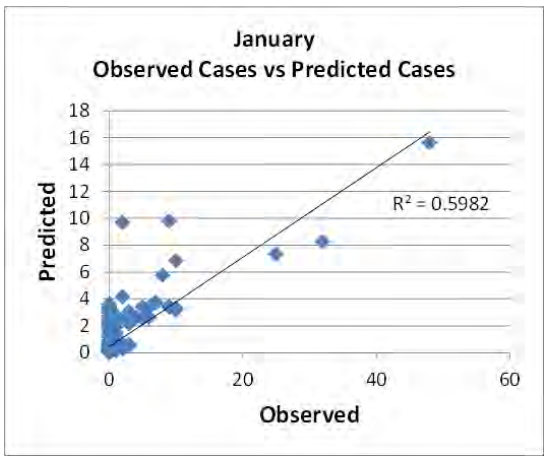
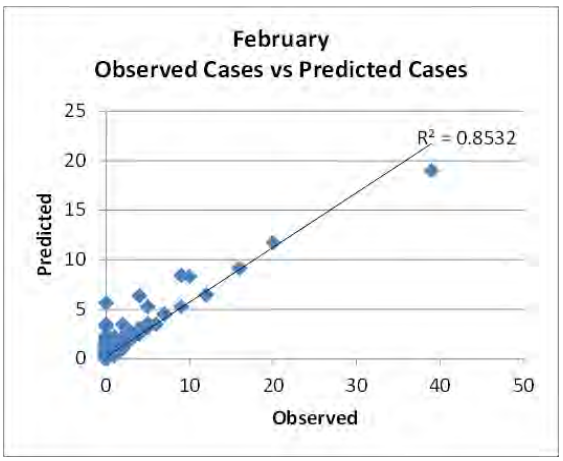


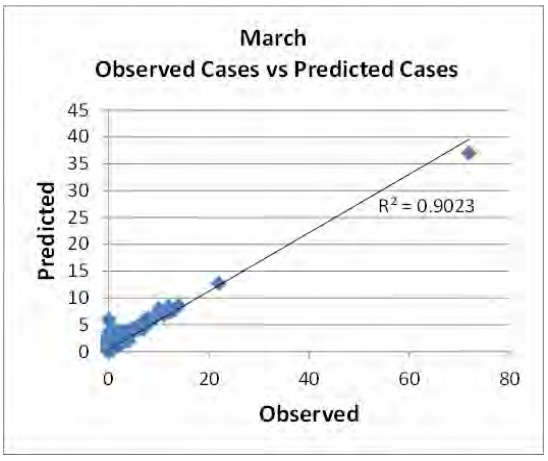
Figure 5-11. Trend of relative humidity and malaria cases

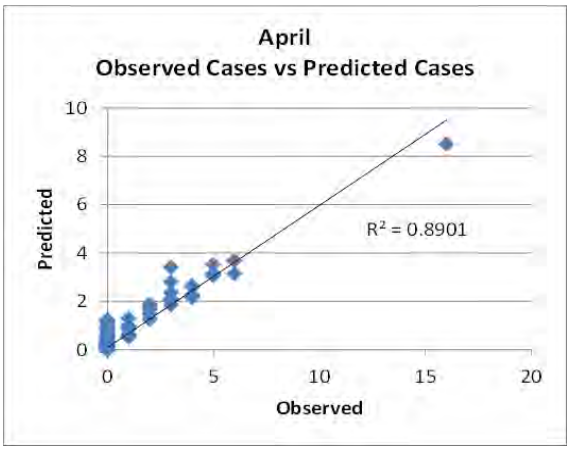
Figure 5-11 shows that the most humid month in the Ehlanzeni district in 2006 was January with a value of 90.97% which is relatively high. A study by Jawara et al. (2008) showed that numbers of both male and female mosquitoes collected at 4 study sites in the Gambia increased towards the end of the dry season as humidity began to increase. The importance of the influence that humidity has on malaria transmission is augmented by an investigation by Yé et al. (2007) that concluded that the risk of clinical malaria in children increased exponentially when relative humidity exceeded 60%. Humidity in South Africa reaches a maximum in summer and a minimum and winter, as seen in the data analysis; malaria transmission reaches a peak in the summer months and declines in the winter months due to the lower survival rate of mosquitoes in the drier winter months.

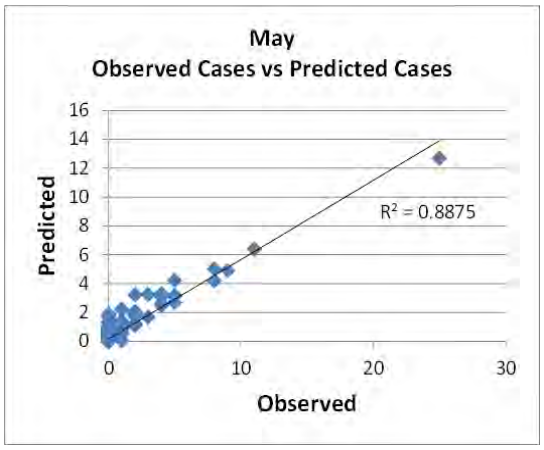
R^2 is an important output of regression analysis; it is interpreted as the proportion of the variance in the dependant variable that is predictable from the independent variable (Draper and Smith, 2014). It is calculated as the square of the correlation between predicted y scores and actual y scores and ranges from 0 to 1 (Hahn, 1973). The larger the R^2 value is, the more variability is explained by the linear regression model. It was stated earlier that in addition to providing measures of variable importance, the RF algorithm also produces predictive models. Figure 5-12 shows the prediction performance of the climatic variables. March, April and May had the highest R^2 values of 0.9023, 0.8901 and 0.8875 respectively. This indicates that the combination of climatic variables for the models for each of these 3 months yielded the highest R^2 compared with other combinations and they were able to explain a high percentage of the total variation in observed malaria cases resulting in high model accuracy.

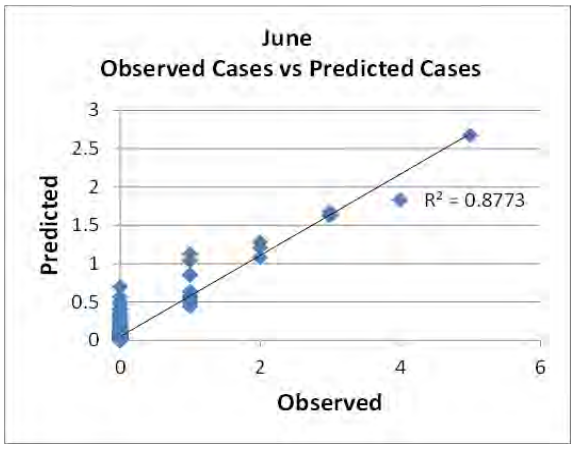

January
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.5982
0
2
4
6
8
10
12
14
16
18
0
20
40
60

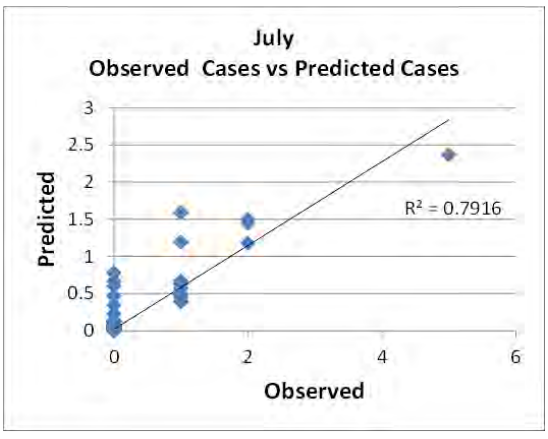

February
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.8532
0
5
10
15
20
25
0
10
20
30
40
50

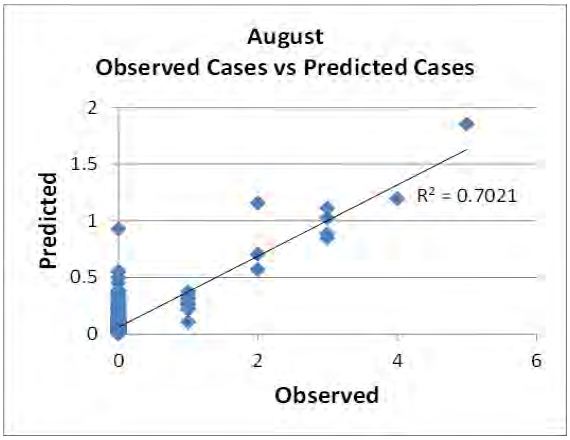

March
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.9023
0
5
10
15
20
25
30
35
40
45
0
20
40
60
80


April
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.8901
0
2
4
6
8
10
0
5
10
15
20


May
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.8875
0
2
4
6
8
10
12
14
16
0
10
20
30


June
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.8773
0
0.5
1
1.5
2
2.5
3
0
2
4
6


July
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.7916
0
0.5
1
1.5
2
2.5
3
0
2
4
6


August
Observed Cases vs Predicted Cases
Predicted
Observed
R² = 0.7021
0
0.5
1
1.5
2
0
2
4
6

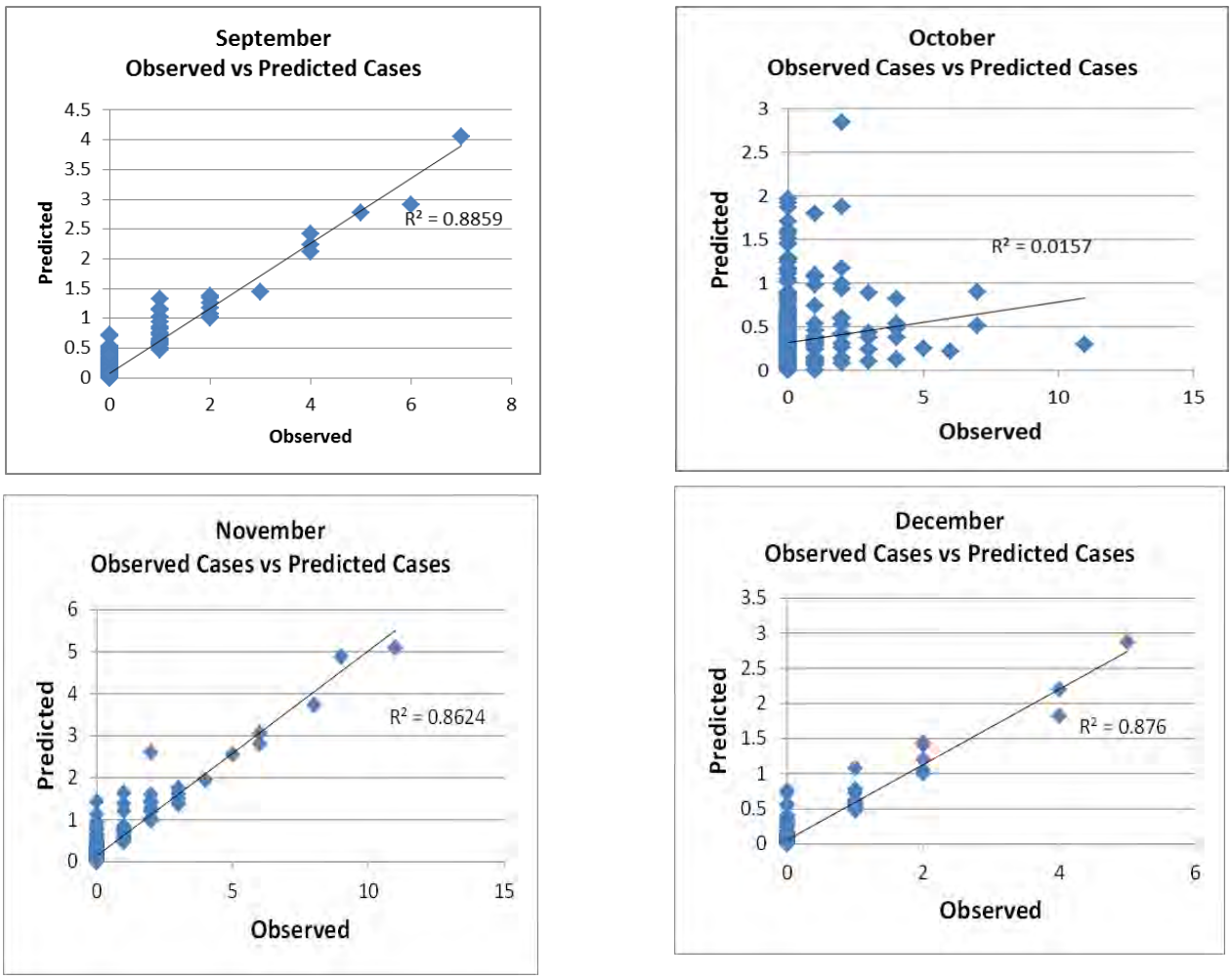
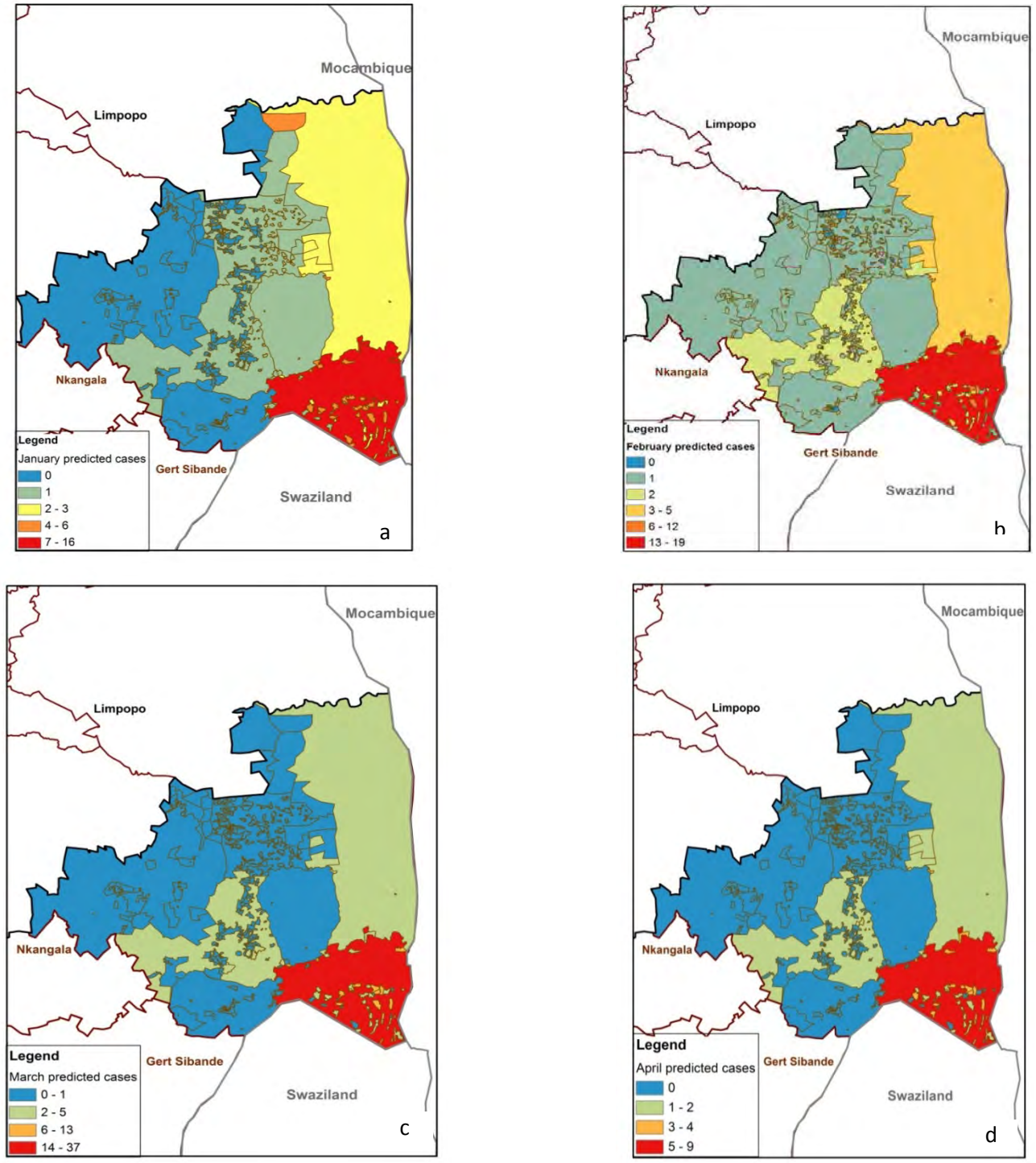
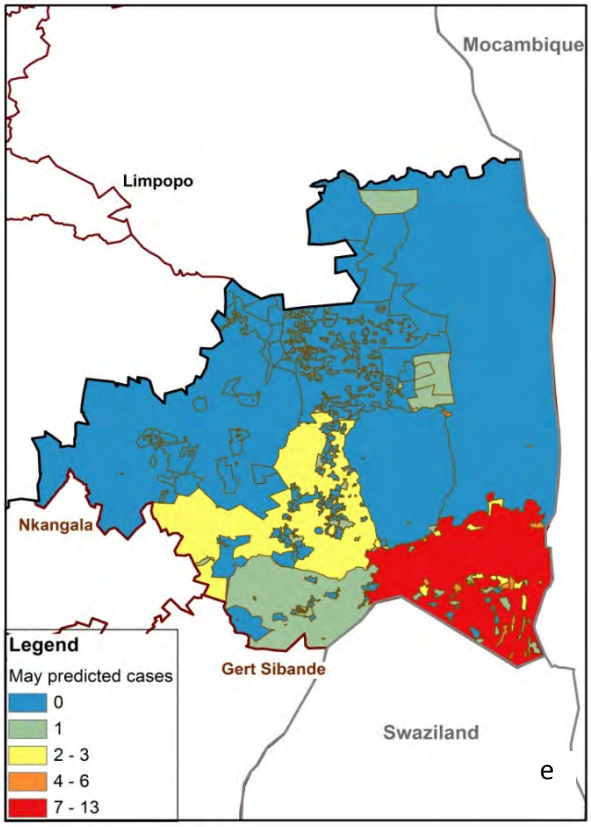


Figure 5-12. Observed malaria cases vs predicted malaria cases

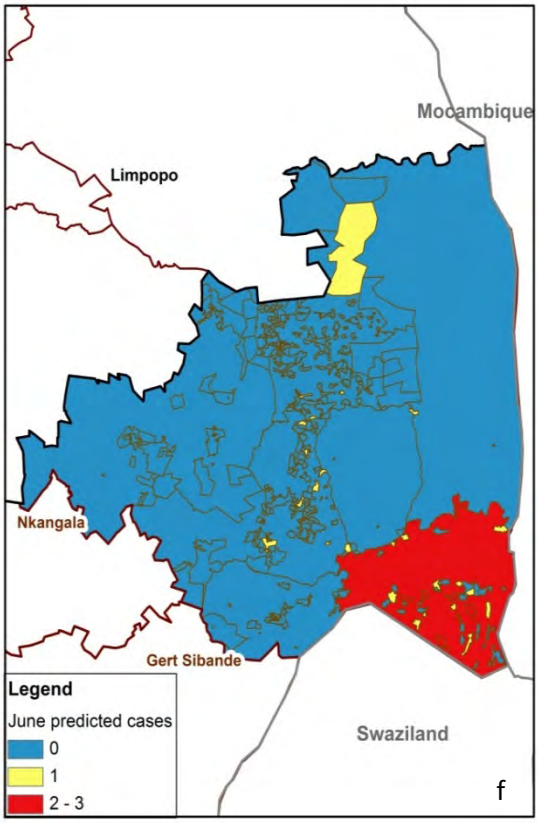
Regression tree methods such as RF are relatively easy to understand and implement however the results are sensitive to small changes in the data, especially outliers (Faraway, 2005) furthermore the RF algorithm does not predict extreme values accurately (Horning, 2013) and it usually considers values greater than 10 as outliers (Breiman, 2002). This is shown in Figure 5-12 because the predictive models identified observed cases of malaria above 10 as outliers and this resulted in decreased model accuracy because these values were poorly prediced. Although excluding these values wold have increased model accuracy, they could be excluded because they were actual recoded values of malaria cases in the sub places of the Ehlanzeni district.

Figure 5-13 depicts maps showing the monthly spatial distribution of predicted cases of malaria in the Ehlanzeni district obtained from the predictive models. These maps are considered a starting point in the process of understanding and describing the distribution of malaria cases as predicted for each month using the top 3 statistically significant climatic variables.

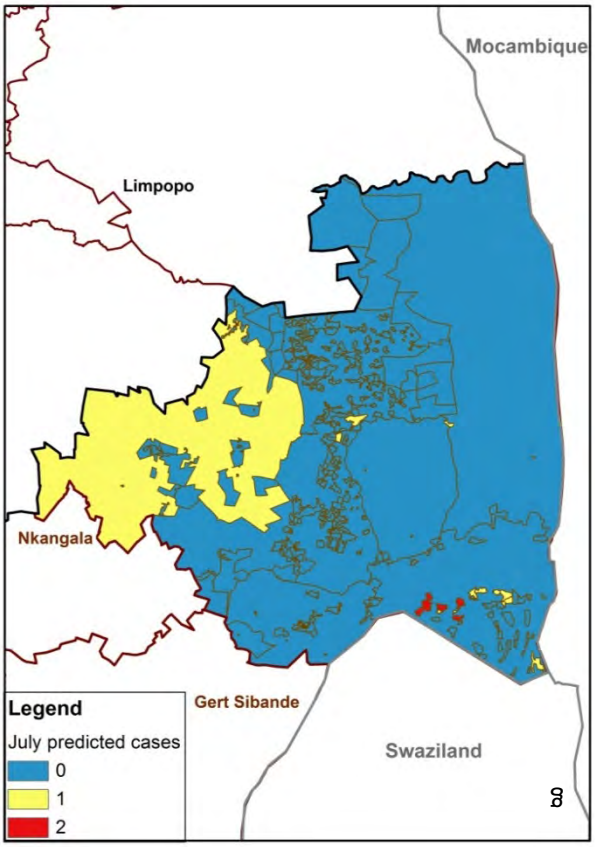



Mocambique
Limpopo
Nkangala
Gert Sibande
Swaziland
Legend
May predicted cases
0
1
2 - 3
4 - 6
7 - 13

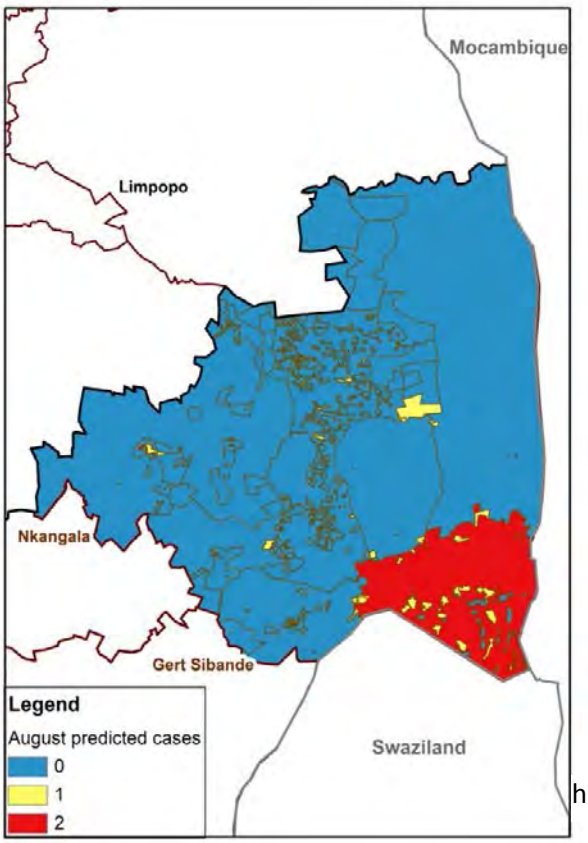
e


Mocambique
Limpopo
Nkangala
Gert Sibande
Swaziland
Legend
June predicted cases
0
1
2 - 3

f


Mocambique
Limpopo
Nkangala
Gert Sibande
Swaziland
Legend
July predicted cases
0
1
2

g


Mocambique
Limpopo
Nkangala
Gert Sibande
Swaziland
Legend
August predicted cases
0
1
2

h

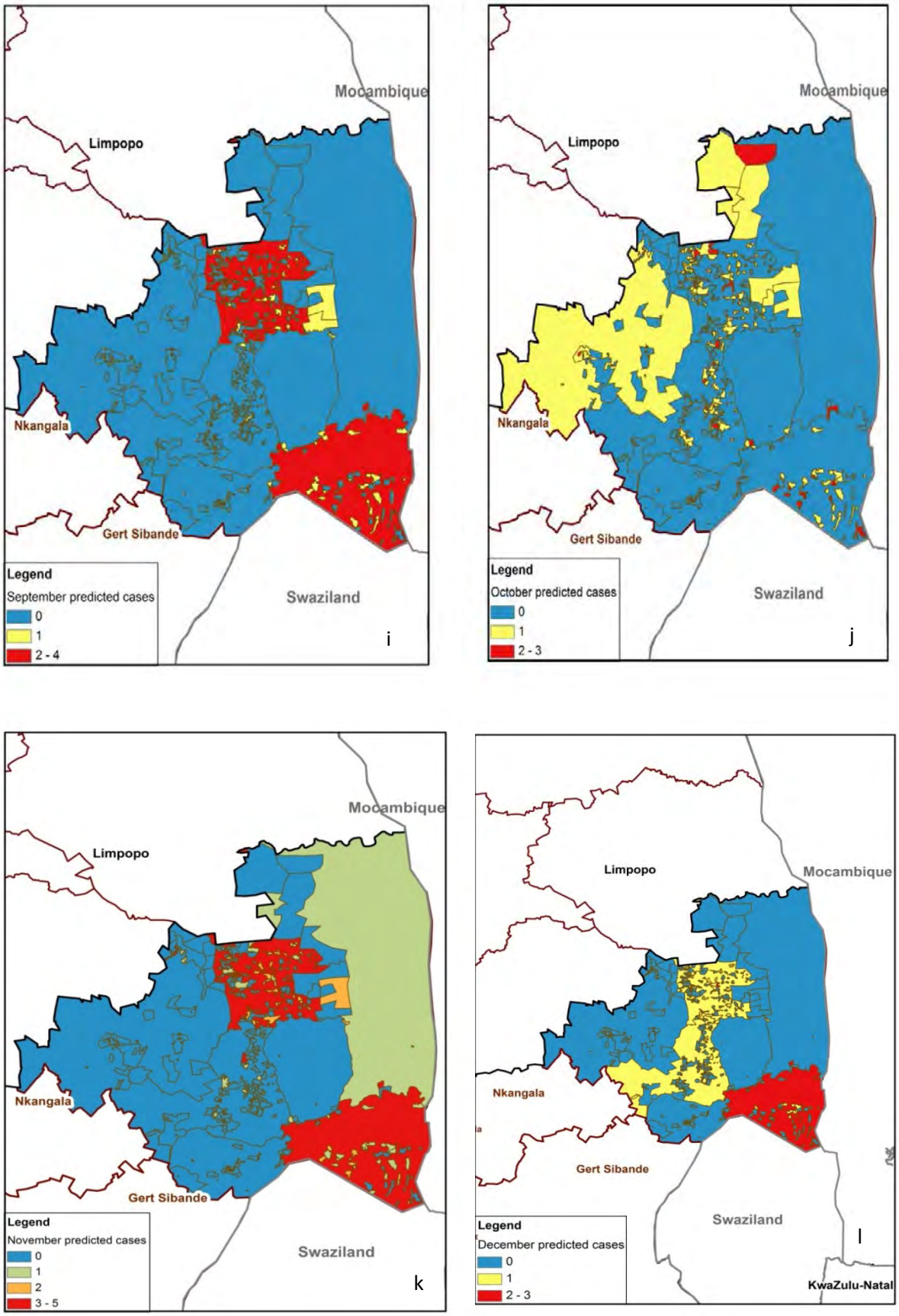
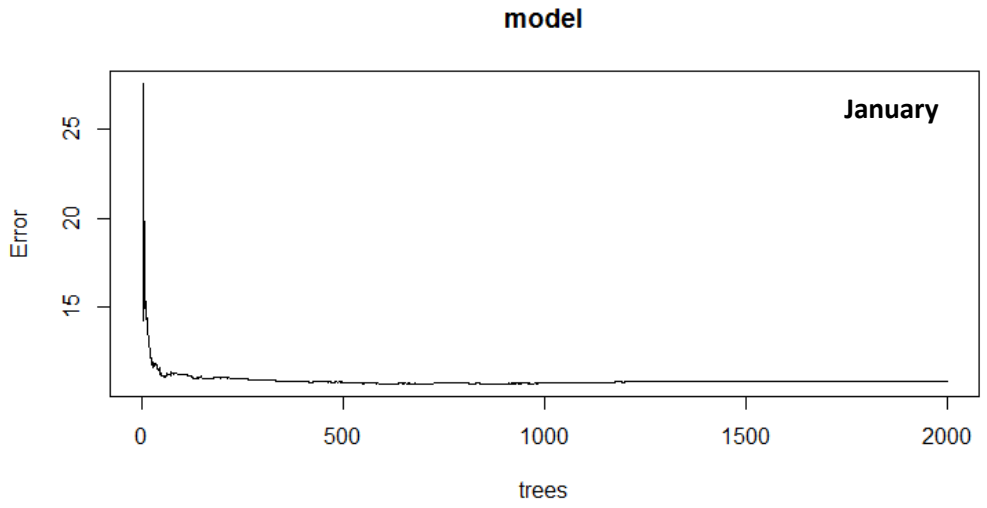
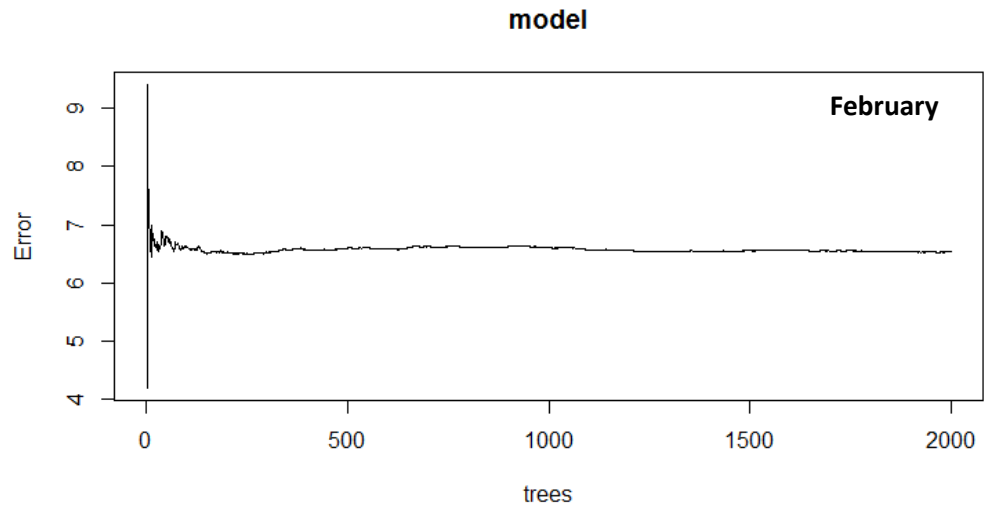


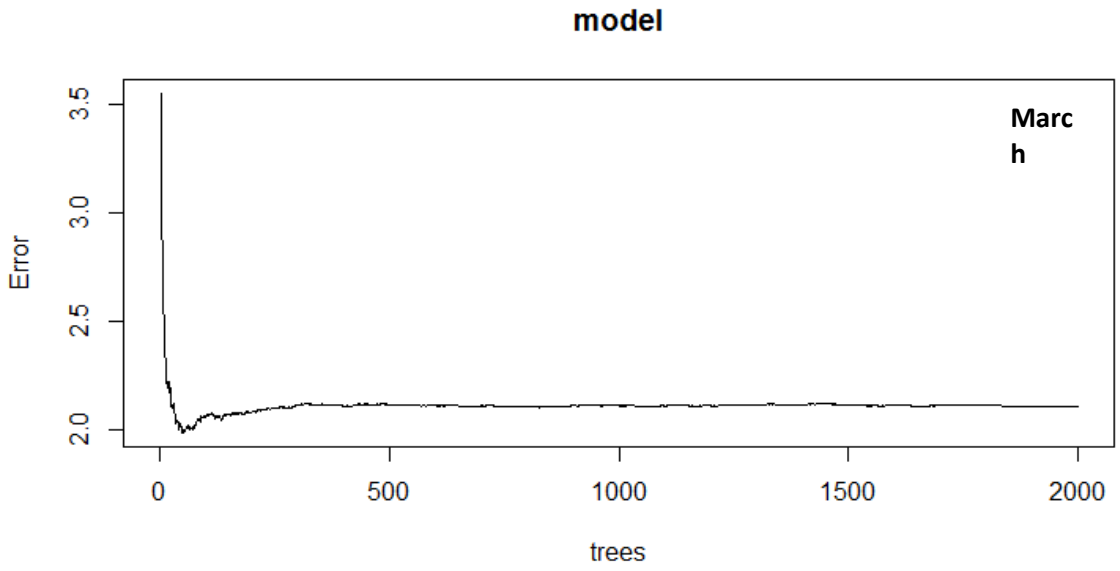
Figure 5-13. Maps showing predicted cases of malaria by sub-place

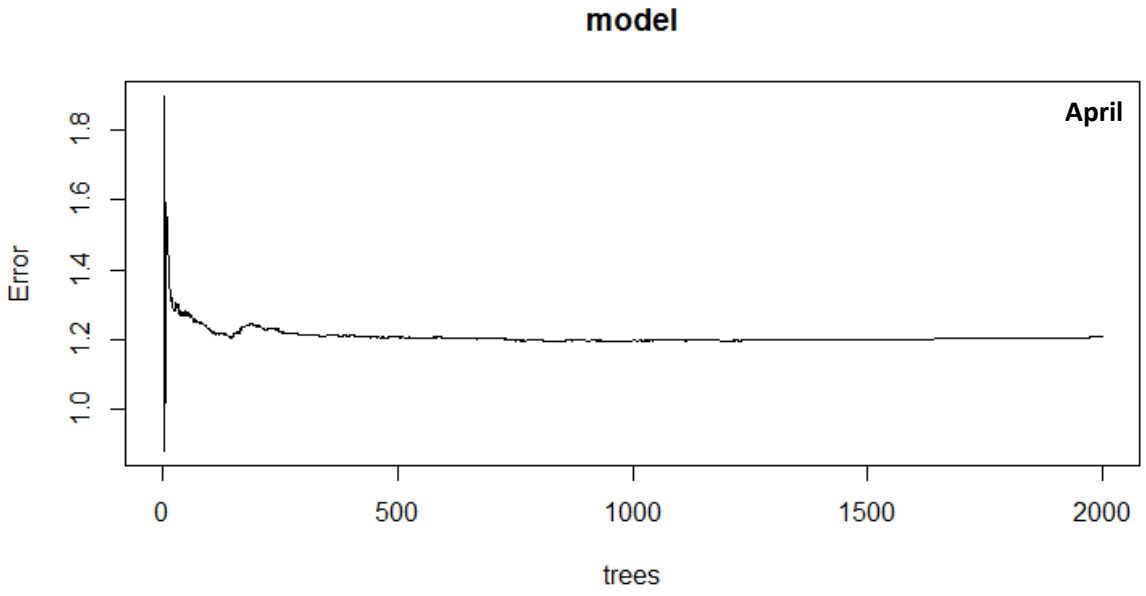
The use of maps to interpret the results of the random forest prediction makes it easy to display the spatial effect of climatic variables on malaria and it shows areas where there are a large number of cases. This plays an important role in monitoring the distribution of malaria and the predictive maps in Figure 5-13 could be used as a tool to guide resource allocation and identify areas for further investigation.

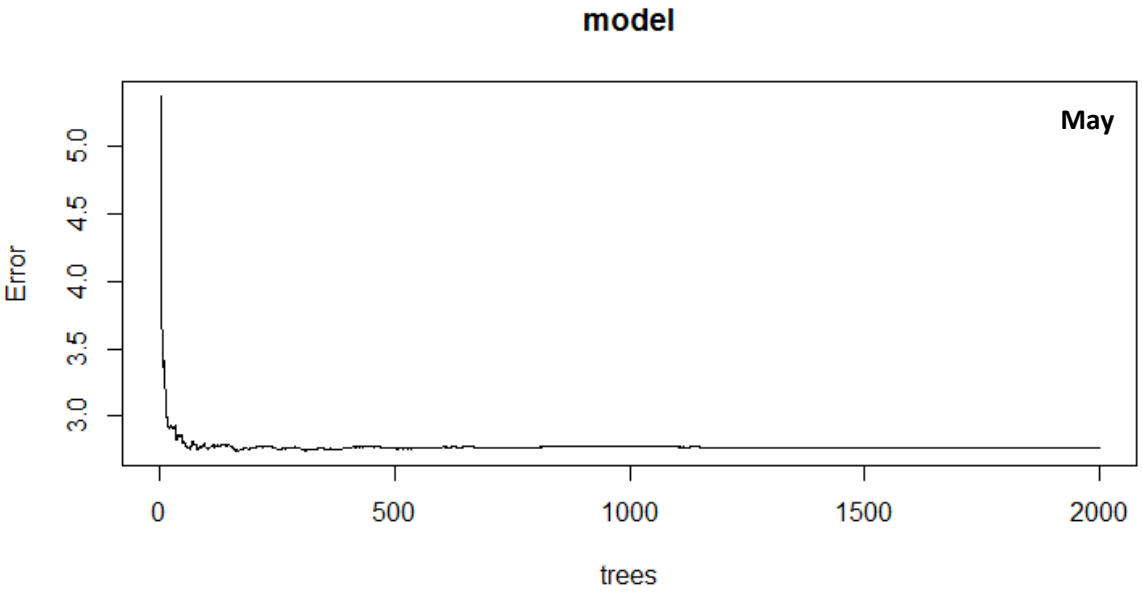
Another useful by product of the RF algorithm is the error estimate. When performing the RF algorithm, after each tree of the random forest is built, the forest makes predictions on each individual raw training dataset observation relying on the previously described technique of bagging. After the first tree, there are many observations in which the forest does not predict because for these observations, a bootstrapped observation was used (Brence and Brown, 2004). Because there are only a few points used in prediction, the general trend observed in Figure 5-14 is that the error rate of the predictive models starts off relatively high then drops sharply ass the number of observations increases and later stabilizers.

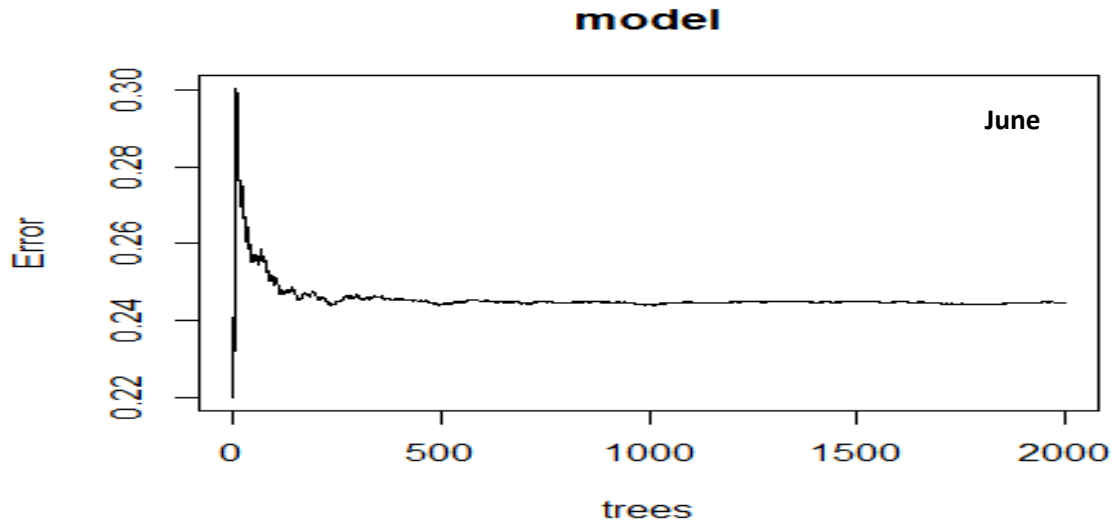


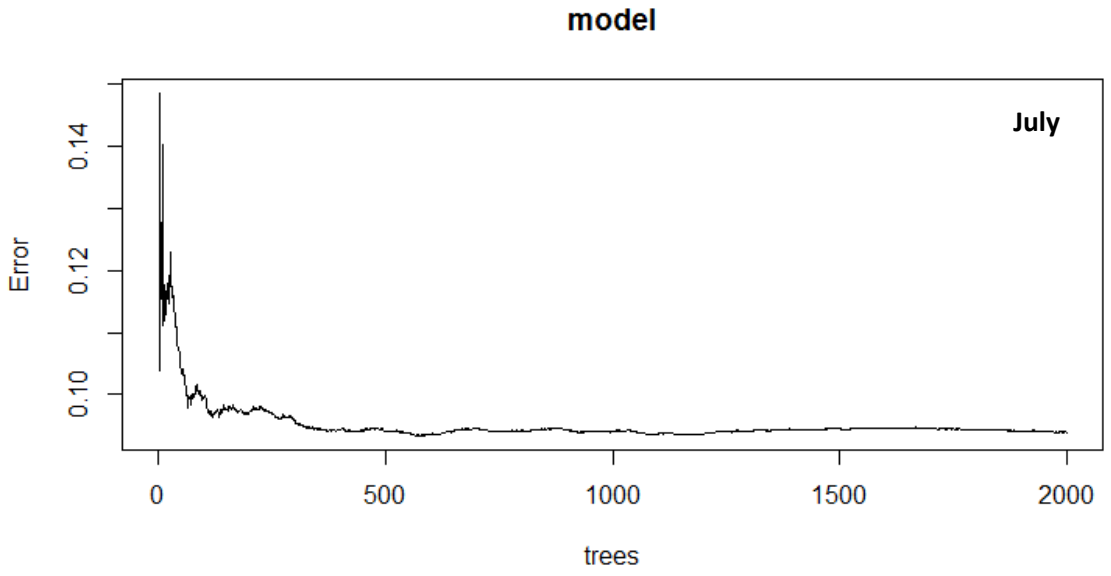


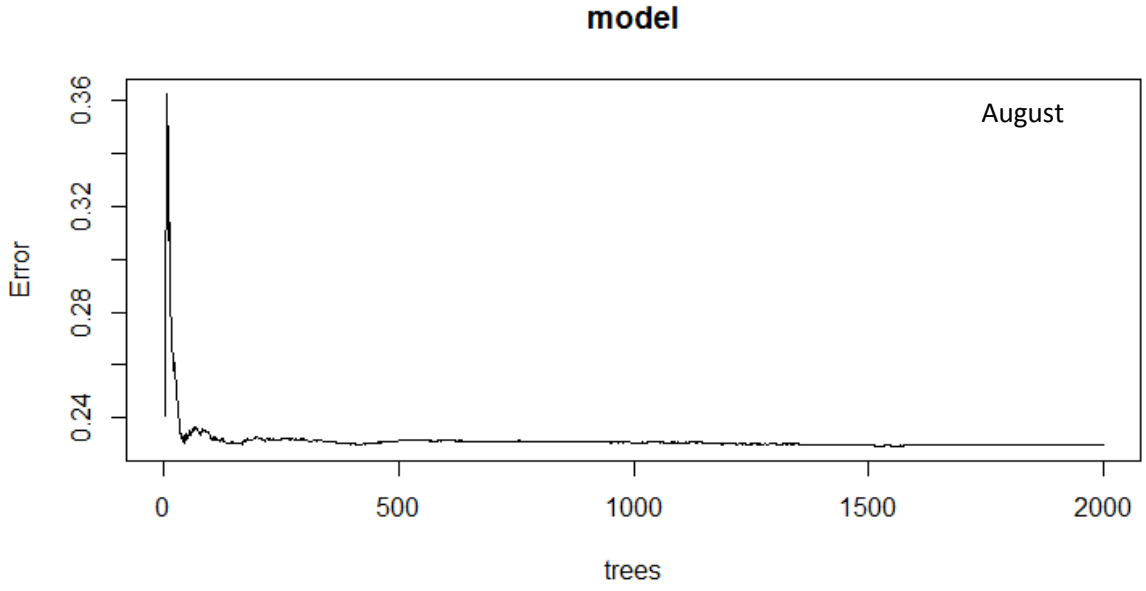

model
March
Error
2.0
2.5
3.0
3.5
0
500
1000
1500
2000
trees

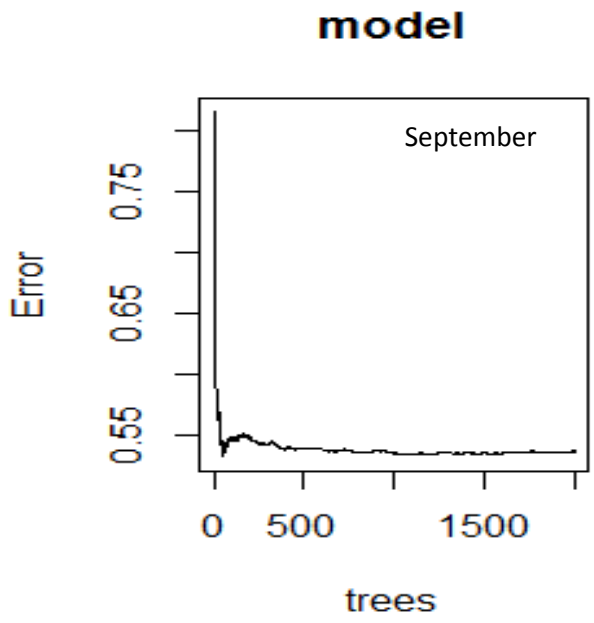

model
April
Error
1.0
1.2
1.4
1.6
1.8
0
500
1000
1500
2000
trees

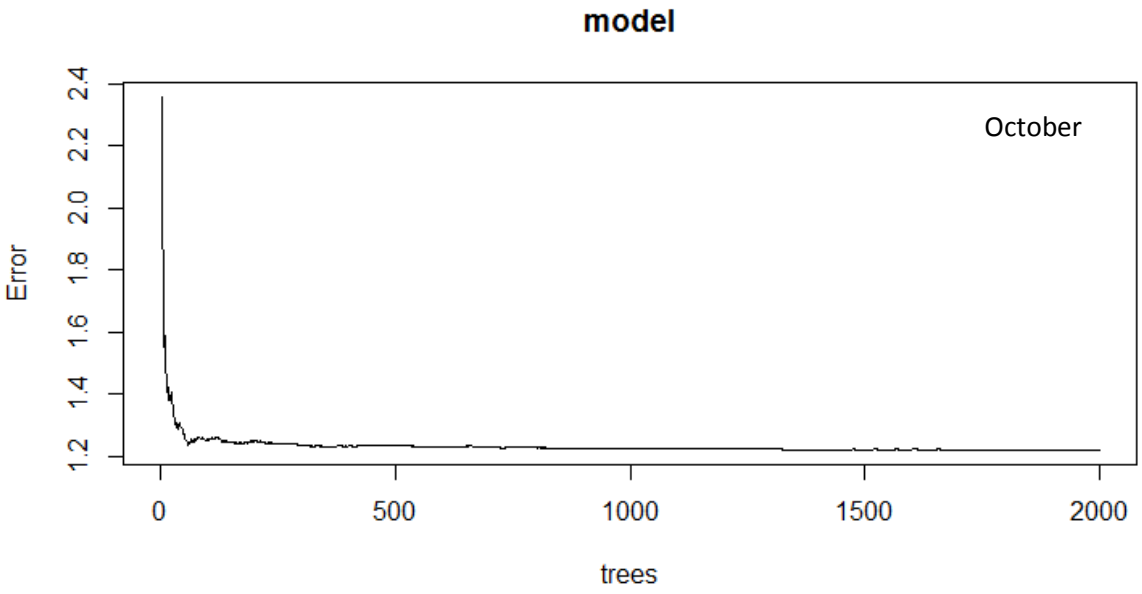

model
May
Error
3.0
3.5
4.0
4.5
5.0
0
500
1000
1500
2000
trees


model
June
Error
0.22
0.24
0.26
0.28
0.30
0
500
1000
1500
2000
trees


model
July
Error
0.10
0.12
0.14
0
500
1000
1500
2000
trees


model
August
Error
0.24
0.28
0.32
0.36
0
500
1000
1500
2000
trees


model
September
Error
0.55
0.65
0.75
0
500
1500
trees


model
October
Error
1.2
1.4
1.6
1.8
2.0
2.2
2.4
0
500
1000
1500
2000
trees

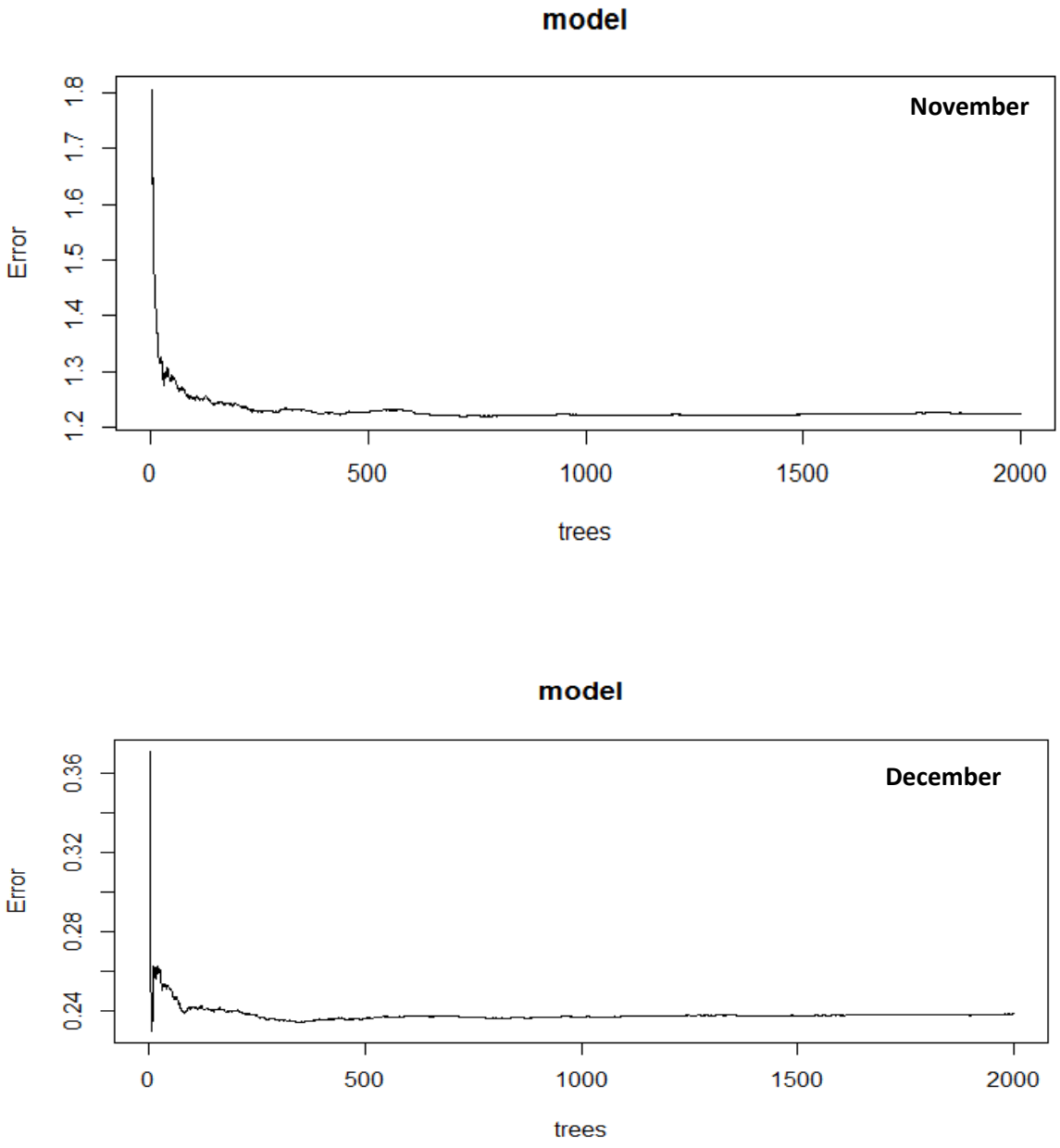


Figure 5-14. Error estimates of monthly predictive models

CHAPTER 6 . CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

6.1 Conclusions

Literature has shown that there are relationships between climate and malaria incidence globally, therefore in order to be able to control and prevent malaria cases, an understanding of the influence of climate on malaria is essential. This will allow for better mathematical modelling of transmission which might lead to improved allocation of limited resources for malaria vector control strategies and this will in turn have a greater impact on malaria control.

This study used the random forest (RF) algorithm to determine which climatic variable as well as what combination of multiple variables had the greatest influence on malaria transmission on a monthly basis for the year of 2006 by providing a measure of variable importance for each predictor and producing predictive models. Climatic variables, namely, NDVI, temperature, lag temperature, rainfall, lag rainfall, altitude and humidity were obtained using remote sensing and GIS techniques and analysis. Malaria case data was obtained from the malaria information system database.

Altitude was most frequently selected as the top predictor variable, it was selected as the dominant predictor for 4 months of the year and it was the most commonly selected predictor throughout the year compared to the other 6 variables. It was therefore the most robust climatic predictor influencing malaria followed by lag rainfall, NDVI and temperature (each selected as the dominant variable for 2 months). Rainfall and lag temperature were each selected as dominant predictor variables for one month each.

The RF algorithm is capable of handling highly correlated variables and because it can provide a measure of importance for each predictor variable, it can be used as a tool to filter out irrelevant predictor variables and focus on those that have the greatest impact on malaria. This can be viewed as a time saving exercise for data collection and experimental run time because for future studies, efforts can be concentrated towards only collecting data for significant climatic variables.

The RF algorithm not only produces a measure of variable importance, it also generates predictive models. An important finding in this study was that the predictive models for the months of March, April and May had the highest model accuracy due to the fact that the combination of climatic factors produced the highest R^2 values. All 3 months had the same top 3 statistically significant climatic variables that were ranked in the same order of importance, altitude was first followed by NDVI and temperature was last. The only difference was that the values for variable importance for each variable differed for each month.

There is also potential for the predictive models generated by the RF algorithm to be used as an operational malaria early warning system by determining areas that have the potential to have a high number of malaria cases. This is because forecasted climatic data can be used as input into the models to predict malaria cases.

The RF algorithm was fast and easy to execute and the results of the study support the findings of a study by Dasgupta et al. (2011) that the algorithm has been proven to provide low bias, low variance predictions because of the low bias nature of the component trees and the averaging across independent bootstrap samples, respectively.

6.2 Limitations and recommendations

The RF algorithm is capable of processing thousands of records. Therefore, a follow up to this study would be to obtain malaria case data and climatic data over a longer period of time, such as the past decade, and obtain measures of variable importance for all the climatic variables for each month over each of the years to evaluate the pattern of statistically significant climatic variables most associated with cases of malaria.

The influence of altitude on malaria has been presented in this study as well as others(Woyessa et al., 2013, Bødker et al., 2003) therefore for future malaria studies for the Ehlanzeni district, it would be more useful to classify malaria cases according to varying altitudes (low level, mid-level, high) to be able to evaluate and describe the effect of climatic variables on malaria at various altitudes.

The predictive models in this study obtained from the RF algorithm did not take malaria intervention methods into account. However, in order to be able to create accurate early warning systems, climate, control methods and malaria data should all be incorporated into predictive models. This is because both climate and IRS (the primary intervention method in Mpumalanga) and larviciding will influence the transmission of malaria across the various seasons. Spray coverage and type of insecticide should also be factored into the model at a later stage because they both influence transmission as well.

Performing the RF algorithm over a period of a decade would also be beneficial because RF cannot predict beyond the range of data in the training data therefore the more data is available, the greater and more reliable the prediction ability of predictive models will become.

Studies by Strobl et al. (2008) and Nicodemus et al. (2010) suggest that randomForest variable importance measures tend to be biased towards correlated predictor variables however this can be overcome by considering different values for the RF package fine tuning parameter mtry and by selecting a sufficient number of trees to ensure that results produced with different runs of the algorithm do not vary systematically.

CHAPTER 7. REFERENCES

ABDULLATEEF, U. & OLUWATOYIN, A. M. 2011. Socio-Economic Impact Of Malaria Epidemics On Households In Nigeria: Microevidence From Kwara State. *International Journal of Asian Social Science,* 1**,** 188-196.

ABDULSALAM, H., SKILLICORN, D. B. & MARTIN, P. 2011. Classification using streaming random forests. *Knowledge and Data Engineering, IEEE Transactions on,* 23**,** 22-36.

ADIMI, F., SOEBIYANTO, R. P., SAFI, N. & KIANG, R. 2010. Research Towards malaria risk prediction in Afghanistan using remote sensing. *Malaria Journal,* 9**,** 125.

AFRANE, Y. A., LITTLE, T. J., LAWSON, B. W., GITHEKO, A. K. & YAN, G. 2008. Deforestation and vectorial capacity of Anopheles gambiae Giles mosquitoes in malaria transmission, Kenya. *Emerging Infectious Diseases,* 14**,** 1533.

AFRANE, Y. A., ZHOU, G., LAWSON, B. W., GITHEKO, A. K. & YAN, G. 2006. Effects of microclimatic changes caused by deforestation on the survivorship and reproductive fitness of Anopheles gambiae in western Kenya highlands. *The American journal of tropical medicine and hygiene,* 74**,** 772-778.

ALEMU, A., ABEBE, G., TSEGAYE, W. & GOLASSA, L. 2011. Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia. *Parasit Vectors,* 4**,** 30.

ALLISON, A. C. 1954. Protection afforded by sickle-cell trait against subtertian malarial infection. *British medical journal,* 1**,** 290.

AMEK, N., BAYOH, N., HAMEL, M., LINDBLADE, K. A., GIMNIG, J., ODHIAMBO, F., LASERSON, K. F., SLUTSKER, L., SMITH, T. & VOUNATSOU, P. 2012a. Spatial and temporal dynamics of malaria transmission in rural Western Kenya. *Parasit Vectors,* 5**,** 86.

AMEK, N., BAYOH, N., HAMEL, M., LINDBLADE, K. A., GIMNIG, J. E., ODHIAMBO, F., LASERSON, K. F., SLUTSKER, L., SMITH, T. & VOUNATSOU, P. 2012b. Spatial and temporal dynamics of malaria transmission in rural Western Kenya. *Parasit Vectors,* 5**,** 86.

ANDREWS, J. M., QUINBY, G. E. & LANGMUIR, A. D. 1950. Malaria Eradication in the United States*. *American Journal of Public Health and the Nations Health,* 40**,** 1405-1411.

ANTONIO-NKONDJIO, C., KERAH, C. H., SIMARD, F., AWONO-AMBENE, P., CHOUAIBOU, M., TCHUINKAM, T. & FONTENILLE, D. 2006. Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. *Journal of medical entomology,* 43**,** 1215-1221.

ARCHER, E., ENGELBRECHT, F., LANDMAN, W., LE ROUX, A., VAN HUYSSTEEN, E., FATTI, C., VOGEL, C., AKOON, I., MASERUMULE, R. & COLVIN, C. 2010. *South African risk and vulnerability atlas*, Department of Science and Technology.

AWOLOLA, T., OYEWOLE, I., KOEKEMOER, L. & COETZEE, M. 2005. Identification of three members of the Anopheles funestus (Diptera: Culicidae) group and their role in malaria transmission in two ecological zones in Nigeria. *Transactions of the Royal Society of tropical medicine and Hygiene,* 99**,** 525-531.

BABYAK, M. A. 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine,* 66**,** 411-421.

BAIRD, J. K. & SNOW, R. W. 2007. Acquired immunity in a holoendemic setting of Plasmodium falciparum and P. vivax malaria. *The American journal of tropical medicine and hygiene,* 76**,** 995-996.

BANNISTER, L. & MITCHELL, G. 2003. The ins, outs and roundabouts of malaria. *Trends in parasitology,* 19**,** 209-213.

BARNES, K., CHANDA, P. & AB BARNABAS, G. 2009. Impact of the large-scale deployment of artemether/lumefantrine on the malaria disease burden in Africa: case studies of South Africa, Zambia and Ethiopia. *Malaria Journal,* 8**,** S8.

BATON, L. A. & RANFORD-CARTWRIGHT, L. C. 2005. Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito. *Trends in parasitology,* 21**,** 573-580.

BATWALA, V., MAGNUSSEN, P. & NUWAHA, F. 2010. Are rapid diagnostic tests more accurate in diagnosis of Plasmodium falciparum malaria compared to microscopy at rural health centres. *Malar J,* 9**,** 10.1186.

BAUWENS, I., JONAS, F., GEBRESLASIE, M., DLAMINI, S., AHMED, F. & PENELOPE, V. 2012. MALAREO–Earth Observation in Malaria Vector Control and Management.

BEIER, J. C. 1998. Malaria parasite development in mosquitoes. *Annual review of entomology,* 43**,** 519-543.

BENHIN, J. K. 2006. Climate change and South African agriculture: Impacts and adaptation options. CEEPA Discussion paper.

BI, P., TONG, S., DONALD, K., PARTON, K. A. & NI, J. 2003. Climatic variables and transmission of malaria: a 12-year data analysis in Shuchen County, China. *Public health reports,* 118**,** 65.

BISCOE, M. L., MUTERO, C. M. & KRAMER, R. A. 2004. *Current policy and status of DDT use for malaria control in Ethiopia, Uganda, Kenya and South Africa*, IWMI.

BJÖRKMAN, A. & MÅRTENSSON, A. 2010. Risks and benefits of targeted malaria treatment based on rapid diagnostic test results. *Clinical infectious diseases,* 51**,** 512-514.

BLOLAND, P. B. & ORGANIZATION, W. H. 2001. *Drug resistance in malaria*, World Health Organization Geneva.

BLUMBERG, L. & FREAN, J. 2002. Malaria control in South Africa-challenges and successes. *South African Medical Journal,* 92.

BØDKER, R., AKIDA, J., SHAYO, D., KISINZA, W., MSANGENI, H., PEDERSEN, E. & LINDSAY, S. 2003. Relationship between altitude and intensity of malaria transmission in the Usambara Mountains, Tanzania. *Journal of Medical Entomology,* 40**,** 706-717.

BOOMAN, M., DURRHEIM, D. N., LA GRANGE, K., MARTIN, C., MABUZA, A. M., ZITHA, A., MBOKAZI, F. M., FRASER, C. & SHARP, B. L. 2000. Using a geographical information system to plan a malaria control programme in South Africa. *Bulletin of the World Health Organization,* 78**,** 1438-1444.

BOOMAN, M., SHARP, B. L., MARTIN, C. L., MANJATE, B., LA GRANGE, J. J. & DURRHEIM, D. N. 2003. Enhancing malaria control using a computerised management system in southern Africa. *Malaria Journal,* 2**,** 13.

BOYD, H. A., FLANDERS, W. D., ADDISS, D. G. & WALLER, L. A. 2005. Residual spatial correlation between geographically referenced observations: a Bayesian hierarchical modeling approach. *Epidemiology,* 16**,** 532-541.

BREIMAN, L. 1996. Bagging predictors. *Machine learning,* 24**,** 123-140.

BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.

BREIMAN, L. 2002. Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA*.

BREMAN, J. G. 2001. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *The American journal of tropical medicine and hygiene,* 64**,** 1-11.

BREMAN, J. G., ALILIO, M. S. & MILLS, A. 2004. Conquering the intolerable burden of malaria: what's new, what's needed: a summary.

BRENCE, J. R. & BROWN, D. E. 2004. *Analysis of Robust Measures for Random Forest Regression.* University of Virginia.

BRITCH, S. C., LINTHICUM, K. J., ANYAMBA, A., TUCKER, C. J., PAK, E. W., MALONEY JR, F. A., COBB, K., STANWIX, E., HUMPHRIES, J. & SPRING, A. 2008. Satellite vegetation index data as a tool to forecast population dynamics of medically important mosquitoes at military installations in the continental United States. *Military medicine,* 173**,** 677-683.

BROOKE, B., KOEKEMOER, L., KRUGER, P., URBACH, J., MISIANI, E. & COETZEE, M. 2013. Malaria vector control in South Africa. *SAMJ: South African Medical Journal,* 103**,** 784-788.

BROOKER, S. 2007. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 101**,** 1-8.

BRUCE-CHWATT, L. J. 1980. *Essential malariology*, William Heinemann Medical Books Ltd.

BRUCE-CHWATT, L. J. & DE ZULUETA, J. 1980. *The rise and fall of malaria in Europe: a historico-epidemiological study*, Published for the Regional Office for Europe of the World Health Organization by Oxford University Press, Walton Street, Oxford OX2 6DP.

CADMAN, M. 2007. *Exploring our Provinces: Mpumalanga*, Jacana.

CAMERON, A. C. & TRIVEDI, P. K. 2013. *Regression analysis of count data*, Cambridge university press.

CARTER, R., MENDIS, K. N. & ROBERTS, D. 2000. Spatial targeting of interventions against malaria. *Bulletin of the World Health Organization,* 78**,** 1401-1411.

CASMAN, E. A. & DOWLATABADI, H. 2002. *The contextual determinants of malaria*, Resources for the Future.

CASTILLO‐RIQUELME, M., MCINTYRE, D. & BARNES, K. 2008. Household burden of malaria in South Africa and Mozambique: is there a catastrophic impact? *Tropical Medicine & International Health,* 13**,** 108-122.

CDC 2014. Impact of Malaria. *http://www.cdc.gov/malaria/malaria_worldwide/impact.html*.

CECCATO, P., CONNOR, S., JEANNE, I. & THOMSON, M. 2005. Application of Geographical Information Systems and Remote Sensing technologies for assessing and monitoring malaria risk. *Parassitologia,* 47**,** 81-96.

CECCATO, P., VANCUTSEM, C., KLAVER, R., ROWLAND, J. & CONNOR, S. J. 2012. A Vectorial Capacity Product to Monitor Changing Malaria Transmission Potential in Epidemic Regions of Africa. *Journal of Tropical Medicine,* 2012**,** 6.

CHAMMARTIN, F., HÜRLIMANN, E., RASO, G., N'GORAN, E. K., UTZINGER, J. & VOUNATSOU, P. 2013. Statistical methodological issues in mapping historical schistosomiasis survey data. *Acta Tropica,* 128**,** 345-352.

CHILDS, D. Z. & BOOTS, M. 2010. The interaction of seasonal forcing and immunity and the resonance dynamics of malaria. *Journal of The Royal Society Interface,* 7**,** 309-319.

CHOI, K. S., KOEKEMOER, L. L. & COETZEE, M. 2012. Population genetic structure of the major malaria vector Anopheles funestus ss and allied species in southern Africa. *Parasit Vectors,* 5**,** 283.

CLEMENTS, A., BARNETT, A. G., CHENG, Z. W., SNOW, R. W. & ZHOU, H. N. 2009. Space-time variation of malaria incidence in Yunnan province, China. *Malar J,* 8**,** 180.

COETZEE, M. & FONTENILLE, D. 2004. Advances in the study of Anopheles funestus, a major vector of malaria in Africa. *Insect Biochemistry and Molecular Biology,* 34**,** 599-605.

COLEMAN, M., COLEMAN, M., MABASO, M. L., MABUZA, A. M., KOK, G., COETZEE, M. & DURRHEIM, D. N. 2010. Household and microeconomic factors associated with malaria in Mpumalanga, South Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 104**,** 143-147.

COLEMAN, S. 2009. *Studies of Entomological Parameters and Perception of Malaria Transmission on the Kwame Nkrumah University of Science and Technology campus, in the Ashanti Region of Ghana.*

COLUZZI, M. 1992. Malaria vector analysis and control. *Parasitology Today,* 8**,** 113-118.

COOSEMANS, M., WERY, M., MOUCHET, J. & CARNEVALE, P. 1992. Transmission factors in malaria epidemiology and control in Africa. *Memórias do Instituto Oswaldo Cruz,* 87**,** 385-391.

COX, J., CRAIG, M., LE SUEUR, D. & SHARP, B. 1999. Mapping malaria risk in the highlands of Africa. *MARA/HIMAL Technical Report***,** 114.

CRAIG, M. 2007. The Temporal and Spatial Distribution of Malaria in Africa, with Emphasis on Southern Africa. *University of Basel*.

CRAIG, M. H., SHARP, B. L., MABASO, M. L. & KLEINSCHMIDT, I. 2007. Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure. *International Journal of Health Geographics,* 6**,** 44.

CRAIG, M. H., SNOW, R. W. & LE SUEUR, D. 1999. A Climate-based Distribution Model of Malaria Transmission in Sub-Saharan Africa. *Parasitology Today,* 15**,** 105-111.

CRESSIE, N. A. C. 1993. *Statistics for spatial data*, J. Wiley.

CURRAN, P. J., ATKINSON, P. M., FOODY, G. M. & MILTON, E. J. 2000. Linking remote sensing, land cover and disease. *Advances in Parasitology,* 47**,** 37-80.

D'ALESSANDRO, U., OLALEYE, B., MCGUIRE, W., THOMSON, M., LANGEROCK, P., BENNETT, S. & GREENWOOD, B. 1995. A comparison of the efficacy of insecticide-treated and untreated bed nets in preventing malaria in Gambian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 89**,** 596-598.

DABIRE, K., BALDET, T., DIABATE, A., DIA, I., COSTANTINI, C., COHUET, A., GUIGUEMDE, T. & FONTENILLE, D. 2007. Anopheles funestus (Diptera: Culicidae) in a humid savannah area of western Burkina Faso: bionomics, insecticide resistance status, and role in malaria transmission. *Journal of medical entomology,* 44**,** 990-997.

DANUOR, S., TAY, S., ANNOR, T., FORKUO, E., BOSOMPEM, K. & ANTWI, V. The impact of climate var-iability on malaria incidence and prevalence in the forest zone of Ghana-A case study at two (2) hospitals located within the Ku-masi Metropolitan area of the Ashanti Region of Ghana. 2010. 2nd International Conference: Climate, Sustainability and Develop ment in Semi-arid Regions, Fortaleza-eara, Brazil.

DASGUPTA, A., SUN, Y. V., KÖNIG, I. R., BAILEY-WILSON, J. E. & MALLEY, J. D. 2011. Brief Review of Regression-Based and Machine Learning Methods in Genetic Epidemiology: The Genetic Analysis Workshop 17 Experience. *Genetic Epidemiology,* 35**,** S5-11.

DAVIS, C. 2011. Climate risk and vulnerability: a handbook for Southern Africa. *Council for Scientific and Industrial Research, Pretoria, South Africa,* 25.

DE CASAS, S. I. C. & CARCAVALLO, R. U. 1995. Climate change and vector-borne diseases distribution. *Social Science & Medicine,* 40**,** 1437-1440.

DELLAPORTAS, P., FORSTER, J. J. & NTZOUFRAS, I. 2002. On Bayesian model and variable selection using MCMC. *Statistics and Computing,* 12**,** 27-36.

DOBSON, M. 1999. The malariology centenary. *Parassitologia,* 41**,** 21-32.

DRAKELEY, C., SUTHERLAND, C., BOUSEMA, J. T., SAUERWEIN, R. W. & TARGETT, G. A. 2006. The epidemiology of Plasmodium falciparum gametocytes: weapons of mass dispersion. *Trends in parasitology,* 22**,** 424-430.

DRAKELEY, C. J., CARNEIRO, I., REYBURN, H., MALIMA, R., LUSINGU, J. P. A., COX, J., THEANDER, T. G., NKYA, W. M. M. M., LEMNGE, M. M. & RILEY, E. M. 2005. Altitude-Dependent and -Independent Variations in Plasmodium falciparum Prevalence in Northeastern Tanzania. *Journal of Infectious Diseases,* 191**,** 1589-1598.

DRAPER, N. R. & SMITH, H. 2014. *Applied regression analysis*, John Wiley & Sons.

EISELE, T. P., KEATING, J., SWALM, C., MBOGO, C. M., GITHEKO, A. K., REGENS, J. L., GITHURE, J. I., ANDREWS, L. & BEIER, J. C. 2003. Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya. *Malaria Journal,* 2**,** 44.

ENAYATI, A. & HEMINGWAY, J. 2010. Malaria management: past, present, and future. *Annual review of entomology,* 55**,** 569-591.

ENDESHAW, T., GEBRE, T., NGONDI, J., GRAVES, P. M., SHARGIE, E. B., EJIGSEMAHU, Y., AYELE, B., YOHANNES, G., TEFERI, T. & MESSELE, A. 2008. Evaluation of light microscopy and rapid diagnostic test for the detection of malaria under operational field conditions: a household survey in Ethiopia. *Malaria Journal,* 7**,** 118.

ENGLISH, M., WARUIRU, C., AMUKOYE, E., MURPHY, S., CRAWLEY, J., MWANGI, I., PESHU, N. & MARSH, K. 1996. Deep breathing in children with severe malaria: indicator of metabolic acidosis and poor outcome. *The American journal of tropical medicine and hygiene,* 55**,** 521-524.

FARAWAY, J. J. 2005. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*, CRC press.

FEACHEM, R. G. A., PHILLIPS, A. A., HWANG, J., COTTER, C., WIELGOSZ, B., GREENWOOD, B. M., SABOT, O., RODRIGUEZ, M. H., ABEYASINGHE, R. R., GHEBREYESUS, T. A. & SNOW, R. W. 2009. Shrinking the malaria map: progress and prospects. *The Lancet,* 376**,** 1566-1578.

FINANCE, M. D. O. 2013. *Socio-economic review and outlook of Mpumalanga* [Online]. Available: http://finance.mpu.gov.za/documents/ea.SERO.June.2013.pdf.

FONTENILLE, D. & SIMARD, F. 2004. Unravelling complexities in human malaria transmission dynamics in Africa through a comprehensive knowledge of vector populations. *Comparative Immunology, Microbiology and Infectious Diseases,* 27**,** 357-375.

FOURNET, F., CUSSAC, M., OUARI, A., MEYER, P.-E., TOÉ, H. K., GOUAGNA, L.-C. & DABIRÉ, R. K. 2010. Diversity in anopheline larval habitats and adult composition during the dry and wet seasons in Ouagadougou (Burkina Faso). *Malar J,* 9**,** 78.

FULLMAN, N., BURSTEIN, R., LIM, S., MEDLIN, C. & GAKIDOU, E. 2013. Nets, spray or both? The effectiveness of insecticide-treated nets and indoor residual spraying in reducing malaria morbidity and child mortality in sub-Saharan Africa. *Malaria Journal,* 12**,** 62.

FURLANELLO, C., NETELER, M., MERLER, S., MENEGON, S., FONTANARI, S., DONINI, D., RIZZOLI, A. & CHEMINI, C. GIS and the random forest predictor: Integration in R for tick-borne disease risk assessment. Proceedings of DSC, 2003. Citeseer, 2.

GARGE, N. R., BOBASHEV, G. & EGGLESTON, B. 2013. Random forest methodology for model-based recursive partitioning: the mobForest package for R. *BMC bioinformatics,* 14**,** 125.

GAUDART, J., TOURÉ, O., DESSAY, N., DICKO, A. L., RANQUE, S., FOREST, L., DEMONGEOT, J. & DOUMBO, O. K. 2009. Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali. *Malar J,* 8**,** 61.

GENUER, R., POGGI, J.-M. & TULEAU-MALOT, C. 2010. Variable selection using random forests. *Pattern Recognition Letters,* 31**,** 2225-2236.

GEORGE, E. I. & MCCULLOCH, R. E. 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association,* 88**,** 881-889.

GERRITSEN, A., KRUGER, P., VAN DER LOEFF, M. & GROBUSCH, M. P. 2008. Malaria incidence in Limpopo Province, South Africa, 1998–2007. *Malar J,* 7**,** 162.

GEURTS, P., ERNST, D. & WEHENKEL, L. 2006. Extremely randomized trees. *Machine learning,* 63**,** 3-42.

GILLIET, J. 1971. Mosquitoes: The World Naturalist. Richardo Clay (The Chaucer Press) ltd.

GITHEKO, A., OTOTO, E. & GUIYUN, Y. 2012. Progress towards understanding the ecology and epidemiology of malaria in the western Kenya highlands: opportunities and challenges for control under climate change risk. *Acta Tropica,* 121**,** 19-25.

GITHEKO, A. K., AYISI, J. M., ODADA, P. K., ATIELI, F. K., NDENGA, B. A., GITHURE, J. I. & YAN, G. 2006. Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control. *Malaria Journal,* 5**,** 107.

GODFRAY, H. C. J. 2013. Mosquito ecology and control of malaria. *Journal of Animal Ecology,* 82**,** 15-25.

GOSONIU, L., VOUNATSOU, P., SOGOBA, N., MAIRE, N. & SMITH, T. 2009. Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. *Computational Statistics & Data Analysis,* 53**,** 3358-3371.

GOSONIU, L., VOUNATSOU, P., SOGOBA, N. & SMITH, T. 2006. Bayesian modelling of geostatistical malaria risk data. *Geospatial Health,* 1**,** 127-139.

GOVERE, J., DURRHEIM, D., COETZEE, M. & HUNT, R. 2001. Malaria in Mpumalanga Province, South Africa, with special reference to the period 1987-1999: research letter. *South African Journal of Science,* 97**,** p. 55-58.

GOVERE, J., DURRHEIM, D., COETZEE, M., HUNT, R. & LA GRANGE, J. 2000. Captures of mosquitoes of the Anopheles gambiae complex (Diptera: Culicidae) in the Lowveld region of Mpumalanga Province, South Africa. *African entomology,* 8**,** 91-99.

GREENWOOD, B. M., FIDOCK, D. A., KYLE, D. E., KAPPE, S. H., ALONSO, P. L., COLLINS, F. H. & DUFFY, P. E. 2008. Malaria: progress, perils, and prospects for eradication. *The Journal of clinical investigation,* 118**,** 1266.

GRIFFIN, J. T., HOLLINGSWORTH, T. D., OKELL, L. C., CHURCHER, T. S., WHITE, M., HINSLEY, W., BOUSEMA, T., DRAKELEY, C. J., FERGUSON, N. M. & BASÁÑEZ, M.-G. 2010. Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies. *PLoS medicine,* 7**,** e1000324.

GRIMWADE, K., FRENCH, N., MBATHA, D. D., ZUNGU, D. D., DEDICOAT, M. & GILKS, C. F. 2004. HIV infection as a cofactor for severe falciparum malaria in adults living in a region of unstable malaria transmission in South Africa. *Aids,* 18**,** 547-554.

GROEPE, M., URBACH, J., JOOSTE, H., HLONGWANA, K., BAKER, L., MISIANI, E. & MAYET, N. 2013. Health promotion: From malaria control to elimination. *SAMJ: South African Medical Journal,* 103**,** 799-800.

GROVER-KOPEC, E., BLUMENTHAL, M. B., CECCATO, P., DINKU, T., OMUMBO, J. & CONNOR, S. 2006. Web-based climate information resources for malaria control in Africa. *Malaria Journal,* 5**,** 38.

GUPTA, S., HILL, A., KWIATKOWSKI, D., GREENWOOD, A. M., GREENWOOD, B. M. & DAY, K. P. 1994. Parasite virulence and disease patterns in Plasmodium falciparum malaria. *Proceedings of the National Academy of Sciences,* 91**,** 3715-3719.

GUYATT, H. L. & SNOW, R. W. 2004. Impact of malaria during pregnancy on low birth weight in sub-Saharan Africa. *Clinical Microbiology Reviews,* 17**,** 760-769.

HAHN, G. J. 1973. The coefficient of determination exposed! *Chemtech,* 3**,** 609-612.

HAQUE, U., HASHIZUME, M., GLASS, G. E., DEWAN, A. M., OVERGAARD, H. J. & YAMAMOTO, T. 2010. The role of climate variability in the spread of malaria in Bangladeshi highlands. *PLoS ONE,* 5**,** e14341.

HARGREAVES, K., HUNT, R. H., BROOKE, B. D., MTHEMBU, J., WEETO, M. M., AWOLOLA, T. S. & COETZEE, M. 2003. Anopheles arabiensis and An. quadriannulatus resistance to DDT in South Africa. *Medical and Veterinary Entomology,* 17**,** 417-422.

HARGREAVES, K., KOEKEMOER, L., BROOKE, B., HUNT, R., MTHEMBU, J. & COETZEE, M. 2000. Anopheles funestus resistant to pyrethroid insecticides in South Africa. *Medical and veterinary entomology,* 14**,** 181-189.

HARRELL, F. E. 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer.

HAY, S., SNOW, R. & ROGERS, D. 1998. From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives. *Parasitology Today,* 14**,** 306-313.

HAY, S. I., GUERRA, C. A., GETHING, P. W., PATIL, A. P., TATEM, A. J., NOOR, A. M., KABARIA, C. W., MANH, B. H., ELYAZAR, I. R. & BROOKER, S. 2009. A world malaria map: Plasmodium falciparum endemicity in 2007. *PLoS medicine,* 6**,** e1000048.

HAY, S. I., GUERRA, C. A., TATEM, A. J., ATKINSON, P. M. & SNOW, R. W. 2005. Urbanization, malaria transmission and disease burden in Africa. *Nature reviews. Microbiology,* 3**,** 81-90.

HAY, S. I., OMUMBO, J. A., CRAIG, M. H. & SNOW, R. W. 2000. Earth observation, geographic information systems and Plasmodium falciparum malaria in sub-Saharan Africa. *In:* S.I. HAY, S. E. R. D. J. R. (ed.) *Advances in Parasitology.* Academic Press.

HAY, S. I., ROGERS, D. J., RANDOLPH, S. E., STERN, D. I., COX, J., SHANKS, G. D. & SNOW, R. W. 2002. Hot topic or hot air? Climate change and malaria resurgence in East African highlands. *Trends in parasitology,* 18**,** 530-534.

HEMINGWAY, J. & RANSON, H. 2000. Insecticide resistance in insect vectors of human disease. *Annual review of entomology,* 45**,** 371-391.

HEMINGWAYD, J. 2009. 9| SuppreSSing the vector. *Shrinking the Malaria Map***,** 140.

HILBE, J. M. 2011. *Negative Binomial Regression*, Cambridge University Press.

HIMEIDAN, Y. E., HAMID, E., THALIB, L., EL BASHIR, M. I. & ADAM, I. 2007. Climatic variables and transmission of falciparum malaria in New Halfa, eastern Sudan.

HLONGWANA, K. W., MABASO, M., KUNENE, S., GOVENDER, D. & MAHARAJ, R. 2009. Community knowledge, attitudes and practices (KAP) on malaria in Swaziland: a country earmarked for malaria elimination. *Malar J,* 8**,** 1-8.

HLONGWANA, K. W., ZITHA, A., MABUZA, A. M. & MAHARAJ, R. 2011. Knowledge and practices towards malaria amongst residents of Bushbuckridge, Mpumalanga, South Africa. *African Journal of Primary Health Care & Family Medicine,* 3**,** 9 pages.

HORNING, N. 2013. Introduction to decision trees and random forests. *American Museum of Natural History's*.

HSIANG, M. S., PANOSIAN, C. & DORSEYC, G. 2009. 5| underStanding Malaria. *Shrinking the Malaria Map***,** 81.

HUANG, F., ZHOU, S., ZHANG, S., WANG, H. & TANG, L. 2011. Temporal correlation analysis between malaria and meteorological factors in Motuo County, Tibet. *Malar J,* 10**,** 54.

HULME, M., DOHERTY, R., NGARA, T., NEW, M. & LISTER, D. 2001. African climate change: 1900-2100. *Climate research,* 17**,** 145-168.

IKEMOTO, T. 2008. Tropical malaria does not mean hot environments. *Journal of medical entomology,* 45**,** 963-969.

IMBAHALE, S., PAAIJMANS, K., MUKABANA, W., VAN LAMMEREN, R., GITHEKO, A. & TAKKEN, W. 2011. A longitudinal study on Anopheles mosquito larval abundance in distinct geographical and environmental settings in western Kenya. *Malaria Journal,* 10**,** 81.

ISHWARAN, H. & RAO, J. S. 2005. Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics***,** 730-773.

JAWARA, M., PINDER, M., DRAKELEY, C. J., NWAKANMA, D. C., JALLOW, E., BOGH, C., LINDSAY, S. W. & CONWAY, D. J. 2008. Dry season ecology of Anopheles gambiae complex mosquitoes in The Gambia. *Malar J,* 7.

KARTHE, D. 2010. Geographic Determinants of Malaria Transmission.

KAZEMBE, L., KLEINSCHMIDT, I., HOLTZ, T. & SHARP, B. 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. *International Journal of Health Geographics,* 5**,** 41.

KAZEMBE, L. N. 2007. Spatial modelling and risk factors of malaria incidence in northern Malawi. *Acta Tropica,* 102**,** 126-137.

KELLY-HOPE, L. A. & MCKENZIE, F. E. 2009. The multiplicity of malaria transmission: a review of entomological inoculation rate measurements and methods across sub-Saharan Africa. *Malaria Journal,* 8**,** 19.

KELLY, G. C., HALE, E., DONALD, W., BATARII, W., BUGORO, H., NAUSIEN, J., SMALE, J., PALMER, K., BOBOGARE, A. & TALEO, G. 2013. A high-resolution geospatial surveillance-response system for malaria elimination in Solomon Islands and Vanuatu. *Malar J,* 12**,** 108.

KIANG, R., ADIMI, F., SOIKA, V., NIGRO, J., SINGHASIVANON, P., SIRICHAISINTHOP, J., LEEMINGSAWAT, S., APIWATHNASORN, C. & LOOAREESUWAN, S. 2006. Meteorological, environmental remote sensing and neural network analysis of the epidemiology of malaria transmission in Thailand. *Geospat Health,* 1**,** 71-84.

KLEINSCHMIDT, I., BAGAYOKO, M., CLARKE, G., CRAIG, M. & LE SUEUR, D. 2000. A spatial statistical approach to malaria mapping. *International Journal of Epidemiology,* 29**,** 355-361.

KLEINSCHMIDT, I., SCHWABE, C., SHIVA, M., SEGURA, J. L., SIMA, V., MABUNDA, S. J. A. & COLEMAN, M. 2009. Combining indoor residual spraying and insecticide-treated net interventions. *The American journal of tropical medicine and hygiene,* 81**,** 519-524.

KLEINSCHMIDT, I., SHARP, B. L., CLARKE, G. P. Y., CURTIS, B. & FRASER, C. 2001. Use of Generalized Linear Mixed Models in the Spatial Analysis of Small-Area Malaria Incidence Rates in KwaZulu Natal, South Africa. *American Journal of Epidemiology,* 153**,** 1213-1221.

KRUGER, A. & SHONGWE, S. 2004. Temperature trends in South Africa: 1960–2003. *International Journal of Climatology,* 24**,** 1929-1945.

KUMAR, V., MANGAL, A., PANESAR, S., YADAV, G., TALWAR, R., RAUT, D. & SINGH, S. 2014. Forecasting Malaria Cases Using Climatic Factors in Delhi, India: A Time Series Analysis. *Malaria research and treatment,* 2014.

KUO, L. & MALLICK, B. 1998. Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B***,** 65-81.

LIAW, A. & WIENER, M. 2002. Classification and Regression by randomForest. *R news,* 2**,** 18-22.

LINDBLADE, K. A., WALKER, E. D., ONAPA, A. W., KATUNGU, J. & WILSON, M. L. 2000. Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda. *Tropical Medicine & International Health,* 5**,** 263-274.

LINDSAY, S. & MARTENS, W. 1998. Malaria in the African highlands: past, present and future. *Bulletin of the World Health Organization,* 76**,** 33.

LOEVINSOHN, M. E. 1994. Climatic warming and increased malaria incidence in Rwanda. *The Lancet,* 343**,** 714-718.

LUTJEHARMS, J., MONTEIRO, P., TYSON, P. & OBURA, D. 2001. The oceans around southern Africa and regional effects of global change: START Regional Syntheses. *South African Journal of Science,* 97**,** p. 119-130.

LYIMO, I. N. & FERGUSON, H. M. 2009. Ecological and evolutionary determinants of host species choice in mosquito vectors. *Trends in Parasitology,* 25**,** 189-196.

MABASO, M., CRAIG, M., VOUNATSOU, P. & SMITH, T. 2005. Towards empirical description of malaria seasonality in southern Africa: the example of Zimbabwe. *Tropical Medicine & International Health,* 10**,** 909-918.

MABASO, M., VOUNATSOU, P., MIDZI, S., DA SILVA, J. & SMITH, T. 2006. Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe. *International Journal of Health Geographics,* 5**,** 20.

MABASO, M. L., CRAIG, M., ROSS, A. & SMITH, T. 2007. Environmental predictors of the seasonality of malaria transmission in Africa: the challenge. *The American journal of tropical medicine and hygiene,* 76**,** 33-38.

MABASO, M. L. H., SHARP, B. & LENGELER, C. 2004. Historical review of malarial control in southern African with emphasis on the use of indoor residual house-spraying. *Tropical Medicine & International Health,* 9**,** 846-856.

MACHAULT, V., VIGNOLLES, C., BORCHI, F., VOUNATSOU, P., BRIOLANT, S., LACAUX, J.-P. & ROGIER, C. 2011. The use of remotely sensed environmental data in the study of malaria. *Geospatial Health,* 5**,** 151-168.

MAHARAJ, R., MORRIS, N., SEOCHARAN, I., KRUGER, P., MOONASAR, D., MABUZA, A., RASWISWI, E. & RAMAN, J. 2012. The feasibility of malaria elimination in South Africa. *Malaria Journal,* 11**,** 423.

MAHARAJ, R., RAMAN, J., MORRIS, N., MOONASAR, D., DURRHEIM, D., SEOCHARAN, I., KRUGER, P., SHANDUKANI, B. & KLEINSCHMIDT, I. 2013. Epidemiology of malaria in South Africa: From control to elimination. *SAMJ: South African Medical Journal,* 103**,** 779-783.

MALERA 2011. A research agenda for malaria eradication: vector control. *PLoS medicine,* 8**,** e1000401.

MARSH, K., FORSTER, D., WARUIRU, C., MWANGI, I., WINSTANLEY, M., MARSH, V., NEWTON, C., WINSTANLEY, P., WARN, P., PESHU, N., PASVOL, G. & SNOW, R. 1995. Indicators of Life-Threatening Malaria in African Children. *New England Journal of Medicine,* 332**,** 1399-1404.

MARTENS, P. & HALL, L. 2000. Malaria on the move: human population movement and malaria transmission. *Emerging infectious diseases,* 6**,** 103.

MARTENS, W., NIESSEN, L. W., ROTMANS, J., JETTEN, T. H. & MCMICHAEL, A. J. 1995. Potential impact of global climate change on malaria risk. *Environmental health perspectives,* 103**,** 458.

MATTINGLY, P. 1969. The Biology of Mosquito-Borne. *The science of biology series No1*.

MIDEKISA, A., SENAY, G., HENEBRY, G. M., SEMUNIGUSE, P. & WIMBERLY, M. C. 2012. Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. *Malar J,* 11**,** 165.

MILLER, L. H., BARUCH, D. I., MARSH, K. & DOUMBO, O. K. 2002. The pathogenic basis of malaria. *Nature,* 415**,** 673-679.

MILLER, L. H., MASON, S. J., CLYDE, D. F. & MCGINNISS, M. H. 1976. The resistance factor to Plasmodium vivax in blacks: the Duffy-blood-group genotype, FyFy. *New England Journal of Medicine,* 295**,** 302-304.

MINAKAWA, N., MUNGA, S., ATIELI, F., MUSHINZIMANA, E., ZHOU, G., GITHEKO, A. K. & YAN, G. 2005. Spatial distribution of anopheline larval habitats in Western Kenyan highlands: effects of land cover types and topography. *The American journal of tropical medicine and hygiene,* 73**,** 157-165.

MMBANDO, B. P., KAMUGISHA, M. L., LUSINGU, J. P., FRANCIS, F., ISHENGOMA, D. S., THEANDER, T. G., LEMNGE, M. M. & SCHEIKE, T. H. 2011. Spatial variation and socio-economic determinants of Plasmodium falciparum infection in northeastern Tanzania. *Malar J,* 10**,** 10.1186.

MOONASAR, D., NUTHULAGANTI, T., KRUGER, P., MABUZA, A., RASISWI, E., BENSON, F. & MAHARAJ, R. 2012. Malaria control in South Africa 2000-2010: beyond MDG6. *Malaria Journal,* 11**,** 294.

MOUCHET, J., MANGUIN, S., SIRCOULON, J., LAVENTURE, S., FAYE, O., ONAPA, A. W., CARNEVALE, P., JULVEZ, J. & FONTENILLE, D. 1998. Evolution of malaria in Africa for the past 40 years: impact of climatic and human factors. *Journal of the American Mosquito Control Association,* 14**,** 121-130.

MPUMALANGA PROVICIAL GOVERNMENT 2009. Mpumalanga Economic Profile

MPUMALANGA PROVICIAL GOVERNMENT 2013. About Mpumalanga Province.

MPUMALANGA PROVINCE DEPARTMENT OF FINANCE 2013. Socio-economic review and outlook of Mpumalanga.

MULAMBA, C., IRVING, H., RIVERON, J. M., MUKWAYA, L. G., BIRUNGI, J. & WONDJI, C. S. 2014. Contrasting Plasmodium infection rates and insecticide susceptibility profiles between the sympatric sibling species Anopheles parensis and Anopheles funestus ss: a potential challenge for malaria vector control in Uganda. *Parasit Vectors,* 7**,** 71.

MUNGA, S., MINAKAWA, N., ZHOU, G., MUSHINZIMANA, E., BARRACK, O.-O. J., GITHEKO, A. K. & YAN, G. 2006. Association between land cover and habitat productivity of malaria vectors in western Kenyan highlands. *The American journal of tropical medicine and hygiene,* 74**,** 69-75.

MUNGA, S., YAKOB, L., MUSHINZIMANA, E., ZHOU, G., OUNA, T., MINAKAWA, N., GITHEKO, A. & YAN, G. 2009. Land use and land cover changes and spatiotemporal dynamics of anopheline larval habitats during a four-year period in a highland community of Africa. *The American journal of tropical medicine and hygiene,* 81**,** 1079-1084.

MUSA, M. I., SHOHAIMI, S., HASHIM, N. R. & KRISHNARAJAH, I. 2012. A climate distribution model of malaria transmission in Sudan. *Geospatial Health,* 7**,** 27-36.

MUSHINZIMANA, E., MUNGA, S., MINAKAWA, N., LI, L., FENG, C.-C., BIAN, L., KITRON, U., SCHMIDT, C., BECK, L. & ZHOU, G. 2006. Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. *Malar J,* 5**,** 13.

NAFCOC. 2014. *Provinces:Mpumalaga* [Online]. Available: http://www.nafcoc.org.za/en/mpumalanga.html.

NÁJERA, J. A., KOUZNETSOV, R. & DELACOLLETTE, C. 1998. *Malaria epidemics: detection and control, forecasting and prevention*, World Health Organization Division of Control of Tropical Diseases.

NDOEN, E., WILD, C., DALE, P., SIPE, N. & DALE, M. 2010. Relationships between anopheline mosquitoes and topography in West Timor and Java, Indonesia. *Malaria Journal,* 9**,** 242.

NDOH 2009. National Malaria Programme Review 2009. *South African National Department of Health*.

NDOH 2010. Guidelines for the Treatment of Malaria in South Africa. *South African National Department of Health*.

NGOMANE, L. & DE JAGER, C. 2012. Changes in malaria morbidity and mortality in Mpumalanga Province, South Africa (2001–2009): a retrospective study. *Malar J,* 11**,** 19.

NICODEMUS, K. K., MALLEY, J. D., STROBL, C. & ZIEGLER, A. 2010. The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC bioinformatics,* 11**,** 110.

NOOR, A., CLEMENTS, A., GETHING, P., MOLONEY, G., BORLE, M., SHEWCHUK, T., HAY, S. & SNOW, R. 2008. Spatial prediction of Plasmodium falciparum prevalence in Somalia. *Malaria Journal,* 7**,** 159.

NUR, E. T. M. 1993. The impact of malaria on labour use and efficiency in the Sudan. *Social Science & Medicine,* 37**,** 1115-1119.

O'MEARA, W. P., BEJON, P., MWANGI, T. W., OKIRO, E. A., PESHU, N., SNOW, R. W., NEWTON, C. R. J. C. & MARSH, K. 2008. Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *The Lancet,* 372**,** 1555-1562.

O'MEARA, W. P., MANGENI, J. N., STEKETEE, R. & GREENWOOD, B. 2010. Changes in the burden of malaria in sub-Saharan Africa. *The Lancet Infectious Diseases,* 10**,** 545-555.

OKOGUN, G. R., NWOKE, B. E., OKERE, A. N., ANOSIKE, J. C. & ESEKHEGBE, A. C. 2003. Epidemiological implications of preferences of breeding sites of mosquito species in Midwestern Nigeria. *Annals of Agricultural and Environmental Medicine,* 10**,** 217-222.

OLLIARO, P. 2005. Drug resistance hampers our capacity to roll back malaria. *Clinical infectious diseases,* 41**,** S247-S257.

OSTFELD, R. S., GLASS, G. E. & KEESING, F. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in ecology & evolution,* 20**,** 328-336.

OYEWOLE, I. & AWOLOLA, T. 2006. Impact of urbanisation on bionomics and distribution of malaria vectors in Lagos, southwestern Nigeria. *Journal of vector borne diseases,* 43**,** 173.

PAAIJMANS, K. P., BLANFORD, S., BELL, A. S., BLANFORD, J. I., READ, A. F. & THOMAS, M. B. 2010. Influence of climate on malaria transmission depends on daily temperature variation. *Proceedings of the National Academy of Sciences,* 107**,** 15135-15139.

PARHAM, P. E. & MICHAEL, E. 2010. Modeling the effects of weather and climate change on malaria transmission. *Environmental health perspectives,* 118**,** 620.

PATZ, J. A., OLSON, S. H., UEJIO, C. K. & GIBBS, H. K. 2008. Disease emergence from global climate and land use change. *Medical Clinics of North America,* 92**,** 1473-1491.

PLUESS, B., TANSER, F. C., LENGELER, C. & SHARP, B. L. 2010. Indoor residual spraying for preventing malaria. *Cochrane Database Syst Rev,* 4.

RAGHAVENDRA, K., BARIK, T. K., REDDY, B. N., SHARMA, P. & DASH, A. P. 2011. Malaria vector control: from past to future. *Parasitology research,* 108**,** 757-779.

RBM 2004. Malaria in Africa. *visto en Internet en www. rbm. who. int el,* 24.

RBM 2005. World malaria report 2005. *World Health Organization and UNICEF*.

RBM 2013. Progress and Impact Series - Focus on South Afria.

REID, H., HAQUE, U., CLEMENTS, A. C., TATEM, A. J., VALLELY, A., AHMED, S. M., ISLAM, A. & HAQUE, R. 2010. Mapping malaria risk in Bangladesh using Bayesian geostatistical models. *Am J Trop Med Hyg,* 83**,** 861-7.

RENAUD, O. & VICTORIA-FESER, M.-P. 2010. A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference,* 140**,** 1852-1862.

RIEDEL, N., VOUNATSOU, P., MILLER, J. M., GOSONIU, L., CHIZEMA-KAWESHA, E., MUKONKA, V. & STEKETEE, R. W. 2010. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). *Malar J,* 9**,** 37.

ROBERT, V., LE GOFF, G., ANDRIANAIVOLAMBO, L., RANDIMBY, F. M., DOMARLE, O., RANDRIANARIVELOJOSIA, M., RAHARIMANGA, V., RAVELOSON, A., RAVAONJANAHARY, C. & ARIEY,

F. 2006. Moderate transmission but high prevalence of malaria in Madagascar. *International journal for parasitology,* 36**,** 1273-1281.

ROBERT, V., MACINTYRE, K., KEATING, J., TRAPE, J.-F., DUCHEMIN, J.-B., WARREN, M. & BEIER, J. C. 2003. MALARIA TRANSMISSION IN URBAN SUB-SAHARAN AFRICA. *The American Journal of Tropical Medicine and Hygiene,* 68**,** 169-176.

ROGERS, D. J., RANDOLPH, S. E., SNOW, R. W. & HAY, S. I. 2002. Satellite imagery in the study and forecast of malaria. *Nature,* 415**,** 710-715.

ROLLAND, E., CHECCHI, F., PINOGES, L., BALKAN, S., GUTHMANN, J.-P. & GUERIN, P. J. 2006. Operational response to malaria epidemics: are rapid diagnostic tests cost-effective?

Réponse opérationnelle aux épidémies de malaria: les tests de diagnostic rapides ont-ils un bon rapport coûts-efficacité?

Respuesta operativa a epidemias de malaria: ¿'son costo-efectivos los test de diagnóstico rápido? *Tropical Medicine & International Health,* 11**,** 398-408.

ROSSI, A., AMADDEO, F., SANDRI, M. & TANSELLA, M. 2005. Determinants of once-only contact in a community-based psychiatric service. *Social psychiatry and psychiatric epidemiology,* 40**,** 50-56.

RUSSELL, P. F. 1946. Lessons in Malariology from World War II The Charles Franklin Craig Lecture, 1945. *The American journal of tropical medicine and hygiene,* 1**,** 5-13.

RUSSELL, S. 2004. The economic burden of illness for households in developing countries: a review of studies focusing on malaria, tuberculosis, and human immunodeficiency virus/acquired immunodeficiency syndrome. *The American journal of tropical medicine and hygiene,* 71**,** 147-155.

SACHS, J. & MALANEY, P. 2002. The economic and social burden of malaria. *Nature,* 415**,** 680-685.

SADASIVAIAH, S., TOZAN, Y. I. & BREMAN, J. G. 2007. Dichlorodiphenyltrichloroethane (DDT) for indoor residual spraying in Africa: how can it be used for malaria control? *American Journal of Tropical Medicine and Hygiene,* 77**,** 249-263.

SANBI. 2007. *MBCP: Vegetation* [Online]. Available: http://bgis.sanbi.org/MBCP/vegetation.asp [Accessed 12 March 2014 2014].

SCHAPIRE, R. E., FREUND, Y., BARTLETT, P. & LEE, W. S. 1998. Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of statistics***,** 1651-1686.

SCHMIDT, E., LOTTER, M. & MCCLELAND, W. 2002. *Trees and shrubs of Mpumalanga and Kruger National Park*, Jacana Media.

SCHULZE, R. & MCGEE, O. 1978. Climatic indices and classifications in relation to the biogeography of southern Africa. *Biogeography and ecology of southern Africa.* Springer.

SCHULZE, R. E., KIKER, G. A. & KUNZ, R. P. 1993. Global climate change and agricultural productivity in southern Africa. *Global Environmental Change,* 3**,** 330-349.

SEXTON, A. 2011. Best practices for an insecticide-treated bed net distribution programme in sub-Saharan eastern Africa. *Malaria Journal,* 10**,** 157.

SHARP, B. L., KLEINSCHMIDT, I., STREAT, E., MAHARAJ, R., BARNES, K. I., DURRHEIM, D. N., RIDL, F. C., MORRIS, N., SEOCHARAN, I. & KUNENE, S. 2007. Seven years of regional malaria control collaboration—Mozambique, South Africa, and Swaziland. *The American journal of tropical medicine and hygiene,* 76**,** 42-47.

SHAUKAT, A. M., BREMAN, J. G. & MCKENZIE, F. E. 2010. Research Using the entomological inoculation rate to assess the impact of vector control on malaria parasite transmission and elimination.

SHIH, S. 2011. Random forests for classification trees and categorical dependent variables: an informal quick start R guide. *Random forests for categorical dependent variables: an informal quick start R guide.[Online] Available from http://www. stanford. edu/~ stephsus/R-randomforest-guide. pdf [Accessed 25th July 2012]*.

SHIRAYAMA, Y., PHOMPIDA, S. & SHIBUYA, K. 2009. Geographic information system (GIS) maps and malaria control monitoring: intervention coverage and health outcome in distal villages of Khammouane province, Laos. *Malaria journal,* 8**,** 217.

SILAL, S. P., BARNES, K. I., KOK, G., MABUZA, A. & LITTLE, F. 2013. Exploring the Seasonality of Reported Treated Malaria Cases in Mpumalanga, South Africa. *PLoS ONE,* 8**,** e76640.

SLUTSKER, L. & MARSTON, B. J. 2007. HIV and malaria: interactions and implications. *Current Opinion in Infectious Diseases,* 20**,** 3-10 10.1097/QCO.0b013e328012c5cd.

SMALL, J., GOETZ, S. J. & HAY, S. I. 2003. Climatic suitability for malaria transmission in Africa, 1911–1995. *Proceedings of the National Academy of Sciences,* 100**,** 15341-15345.

SNOW, R. W. & GILLES, H. M. 2002. The epidemiology of malaria. *Essential malariology,* 4**,** 85-106.

SNOW, R. W. & OMUMBO, J. A. 2006. Disease and Mortality in Sub-Saharan Africa.

SPIELMAN, A. 2003. The behavioural and social aspects of malaria and its control.

SPILLINGS, B. L., BROOKE, B. D., KOEKEMOER, L. L., CHIPHWANYA, J., COETZEE, M. & HUNT, R. H. 2009. A new species concealed by Anopheles funestus Giles, a major malaria vector in Africa. *The American journal of tropical medicine and hygiene,* 81**,** 510-515.

STRESMAN, G. H. 2010. Beyond temperature and precipitation: Ecological risk factors that modify malaria transmission. *Acta Tropica,* 116**,** 167-172.

STROBL, C., BOULESTEIX, A.-L., KNEIB, T., AUGUSTIN, T. & ZEILEIS, A. 2008. Conditional variable importance for random forests. *BMC bioinformatics,* 9**,** 307.

STROBL, C., MALLEY, J. & TUTZ, G. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods,* 14**,** 323.

TANSER, F. C., SHARP, B. & LE SUEUR, D. 2003. Potential effect of climate change on malaria transmission in Africa. *Lancet,* 362**,** 1792-8.

TAYLOR, T., BORGSTEIN, A. & MOLYNEUX, M. 1993. Acid-base status in paediatric Plasmodium falciparum malaria. *Qjm,* 86**,** 99-109.

TEKLEHAIMANOT, A. 2005. *Coming to grips with malaria in the new millennium*, Earthscan.

THOMSON, M. C., CONNOR, S. J., D'ALESSANDRO, U., ROWLINGSON, B., DIGGLE, P., CRESSWELL, M. & GREENWOOD, B. 1999. Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. *The American journal of tropical medicine and hygiene,* 61**,** 2-8.

TUNO, N., OKEKA, W., MINAKAWA, N., TAKAGI, M. & YAN, G. 2005. Survivorship of Anopheles gambiae sensu stricto (Diptera: Culicidae) larvae in western Kenya highland forest. *Journal of medical entomology,* 42**,** 270-277.

TYSON, P. D. 1986. *Climatic change and variability in Southern Africa*, Oxford University Press.

VER HOEF, J. M., CRESSIE, N., FISHER, R. N. & CASE, T. J. 2001. Uncertainty and spatial linear models for ecological data. *Spatial Uncertainty in Ecology.* Springer.

VITTOR, A. Y., PAN, W., GILMAN, R. H., TIELSCH, J., GLASS, G., SHIELDS, T., SÁNCHEZ-LOZANO, W., PINEDO, V. V., SALAS-COBOS, E. & FLORES, S. 2009. Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, Anopheles darlingi. *The American journal of tropical medicine and hygiene,* 81**,** 5-12.

WARRELL, D. A. & GILLES, H. M. 2002. *Essential malariology*, Arnold.

WAYANT, N. M., MALDONADO, D., ROJAS DE ARIAS, A., COUSIÑO, B. & GOODIN, D. G. 2010. Correlation between normalized difference vegetation index and malaria in a subtropical rain forest undergoing rapid anthropogenic alteration. *Geospatial Health,* 4**,** 179-190.

WHITE, G. 1974. Anopheles gambiae complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 68**,** 278-298.

WHITE, N. 2012. Review of statistical methods for disease mapping.

WHO 1986. Ottawa charter for health promotion.

WHO 2006a. *Guidelines for the treatment of malaria*, World Health Organization.

WHO 2006b. Pesticides and their application: for the control of vectors and pests of public health importance.

WHO 2014. World Malaria Report 2014.

WILLIAMS, H. A. & BLOLAND, P. B. 2002. *Malaria control during mass population movements and natural disasters*, National Academies Press.

WOYESSA, A., DERESSA, W., ALI, A. & LINDTJØRN, B. 2013. Malaria risk factors in Butajira area, south-central Ethiopia: a multilevel analysis. *Malar J,* 12**,** 273.

YAMANA, T. K. & ELTAHIR, E. A. 2013. Incorporating the effects of humidity in a mechanistic model of Anopheles gambiae mosquito population dynamics in the Sahel region of Africa. *Parasites & vectors,* 6**,** 235.

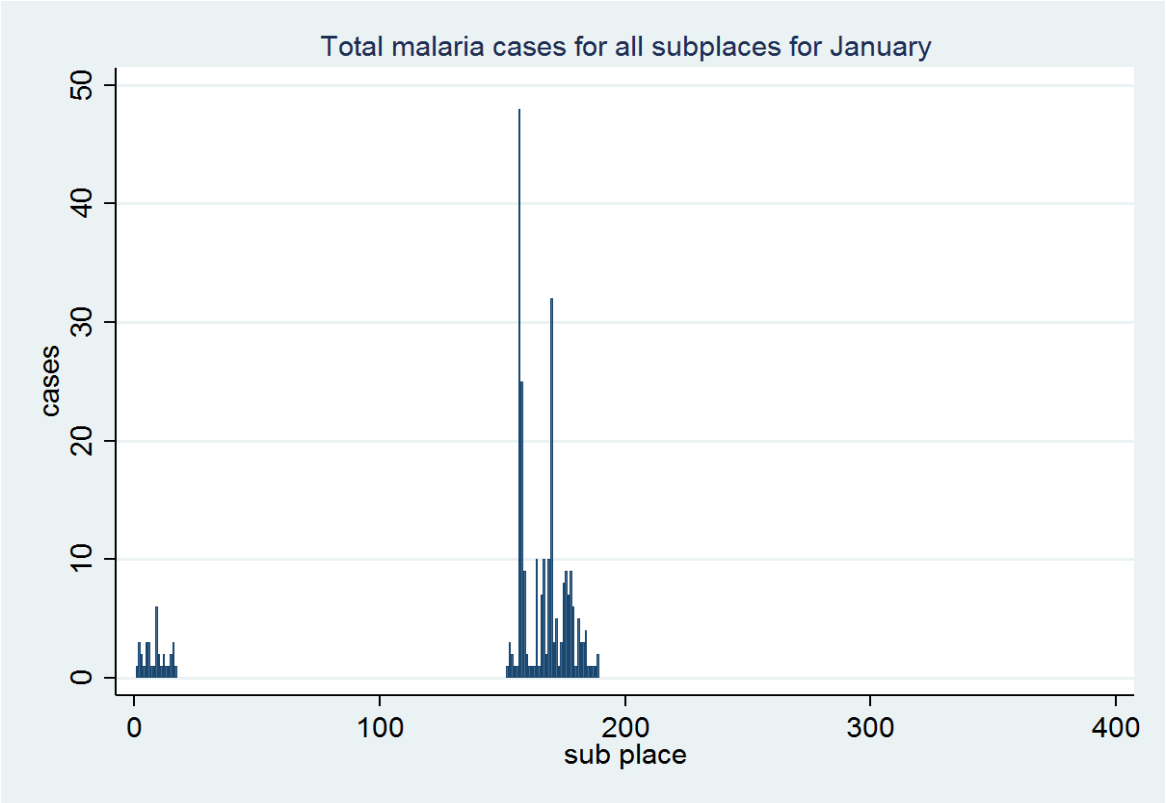
YANG, H., MACORIS, M., GALVANI, K., ANDRIGHETTI, M. & WANDERLEY, D. 2009. Assessing the effects of temperature on dengue transmission. *Epidemiology and infection,* 137**,** 1179-1187.

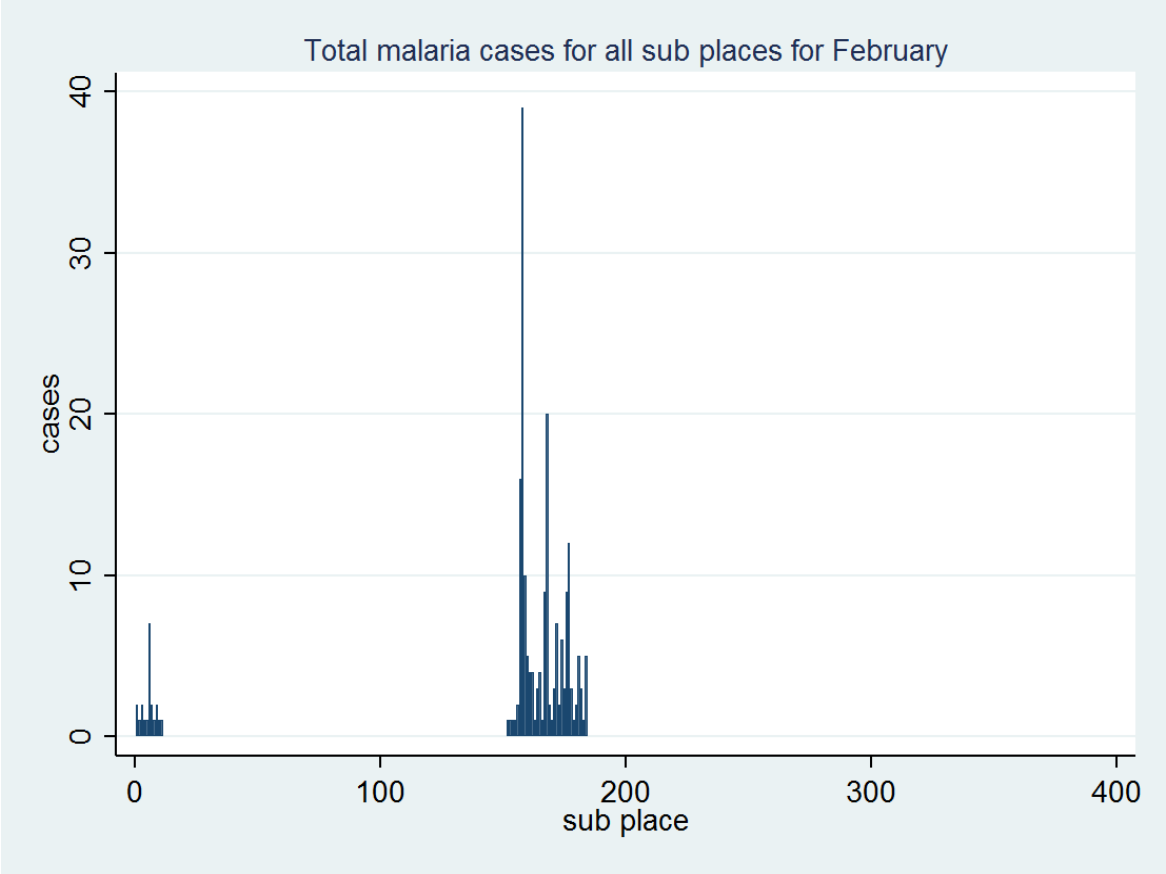
YÉ, Y., LOUIS, V. R., SIMBORO, S. & SAUERBORN, R. 2007. Effect of meteorological factors on clinical malaria risk among children: an assessment using village-based meteorological stations and community-based parasitological survey. *BMC Public Health,* 7**,** 101.

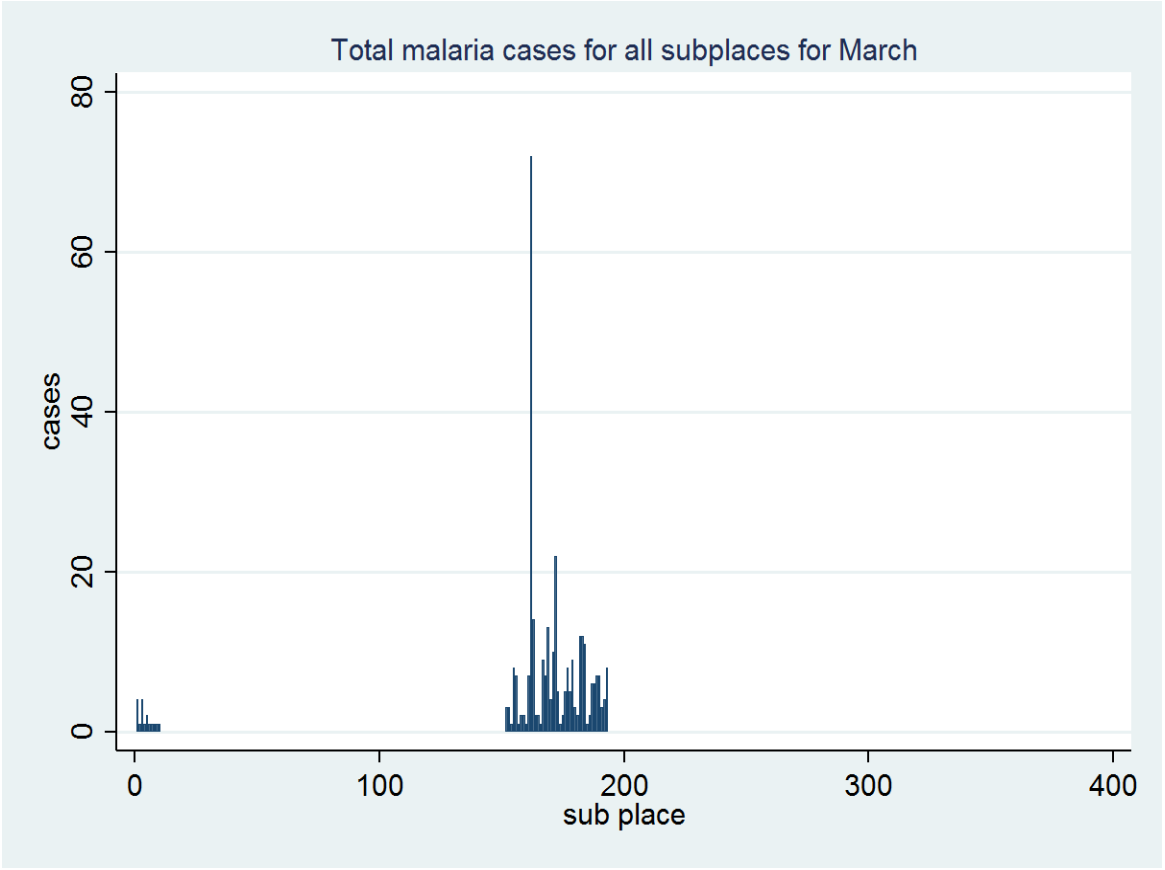
ZHOU, G., MINAKAWA, N., GITHEKO, A. K. & YAN, G. 2004. Association between climate variability and malaria epidemics in the East African highlands. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 2375-2380.

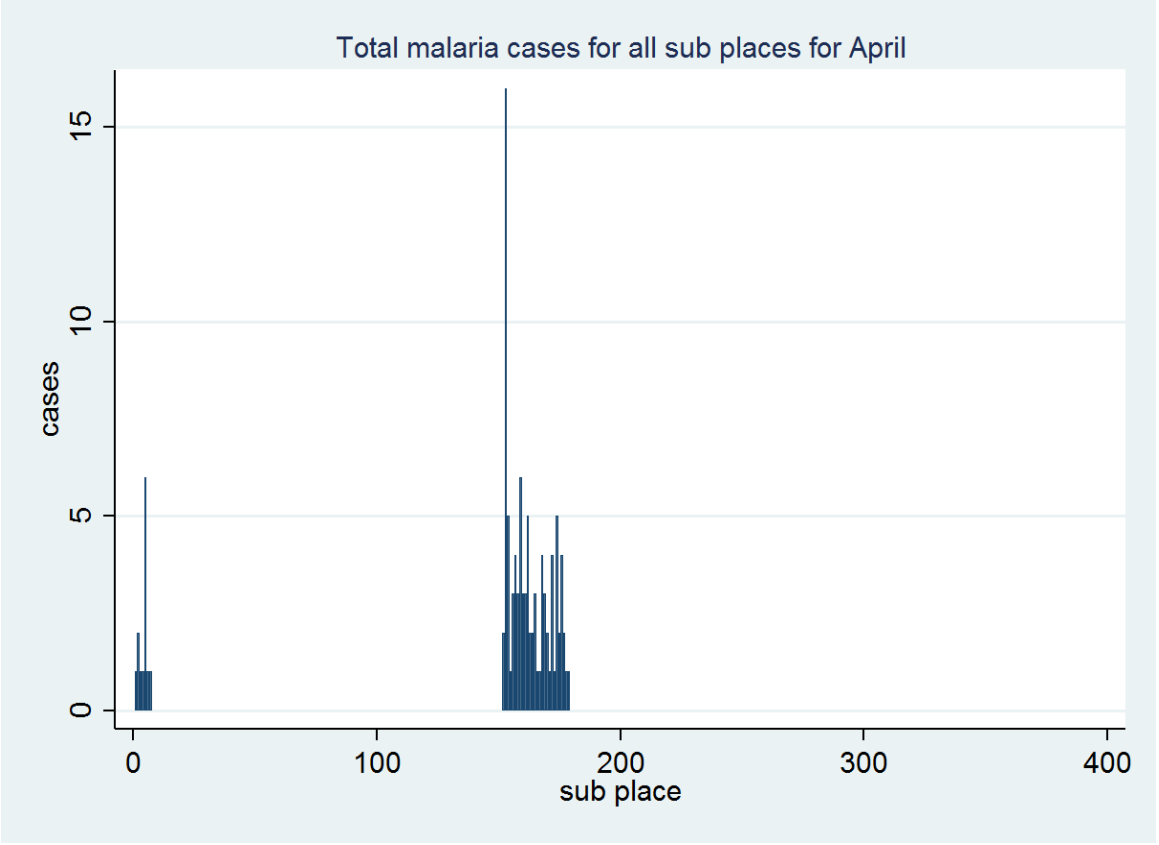
CHAPTER 8. APPENDICES

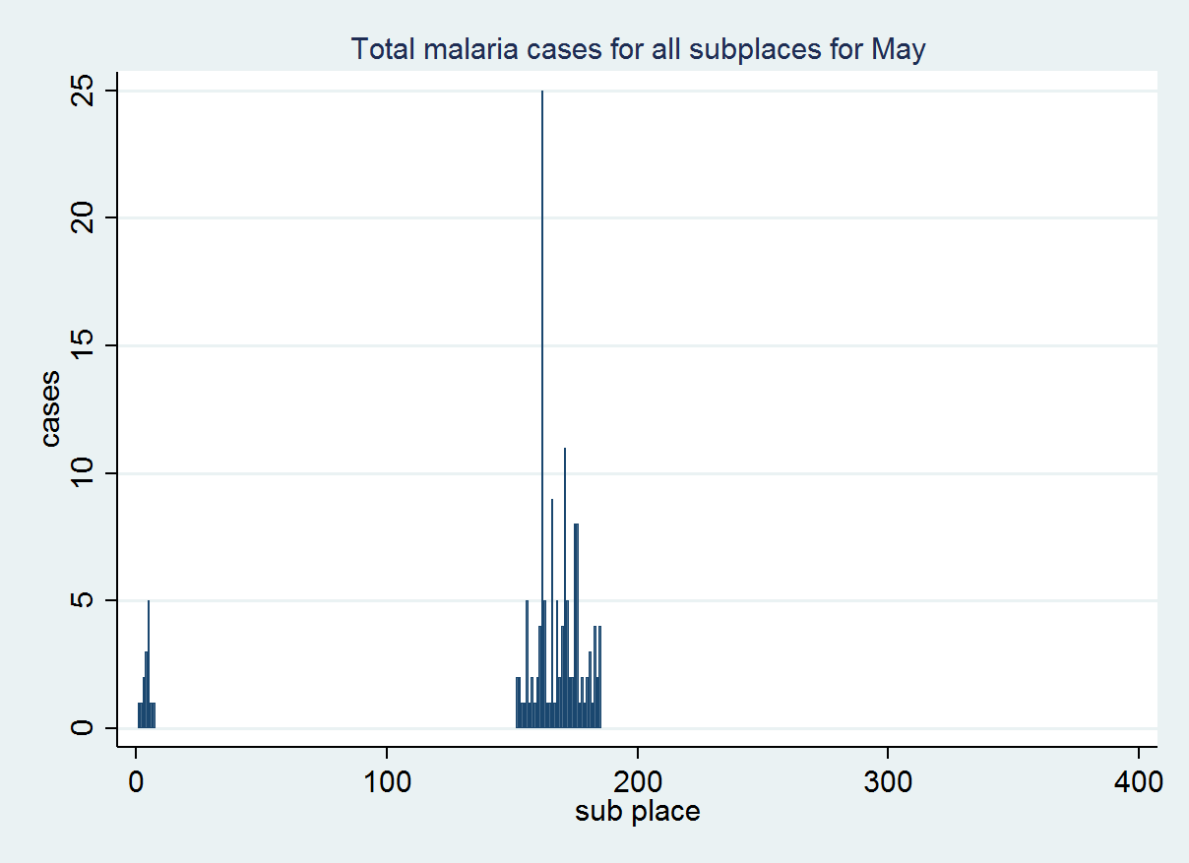
Appendix 1: Graphs of total malaria cases for all sub places by month

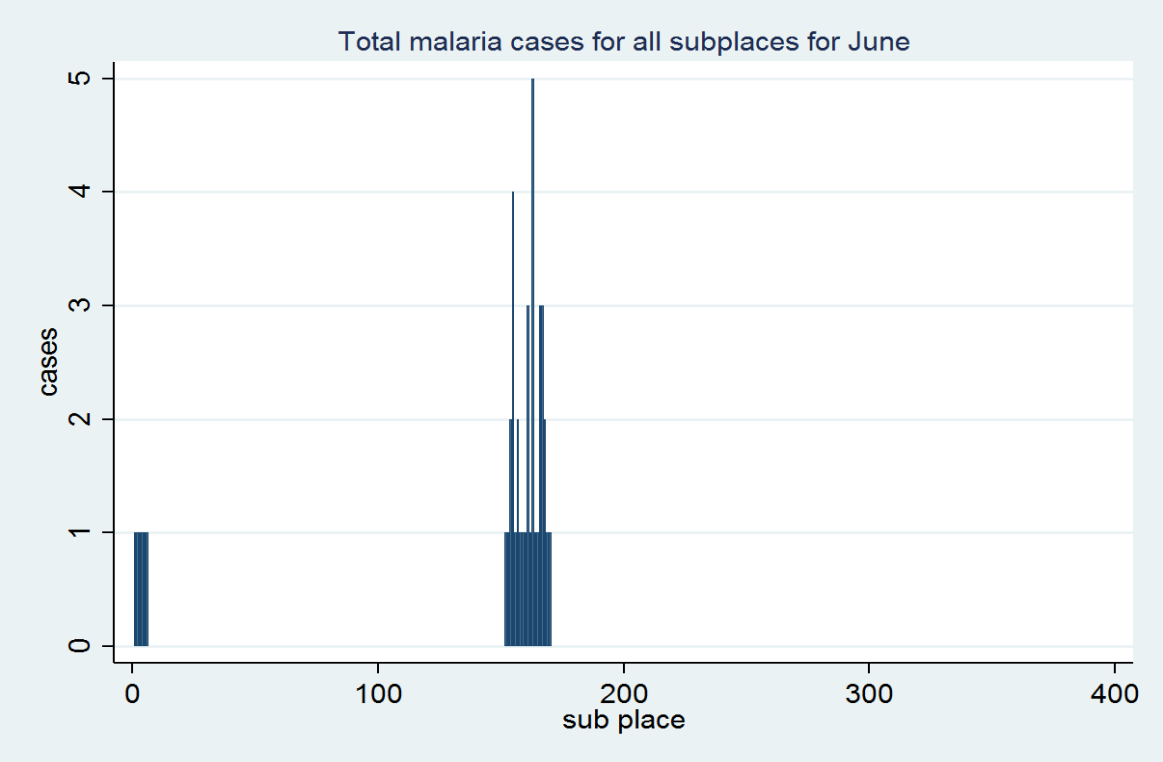


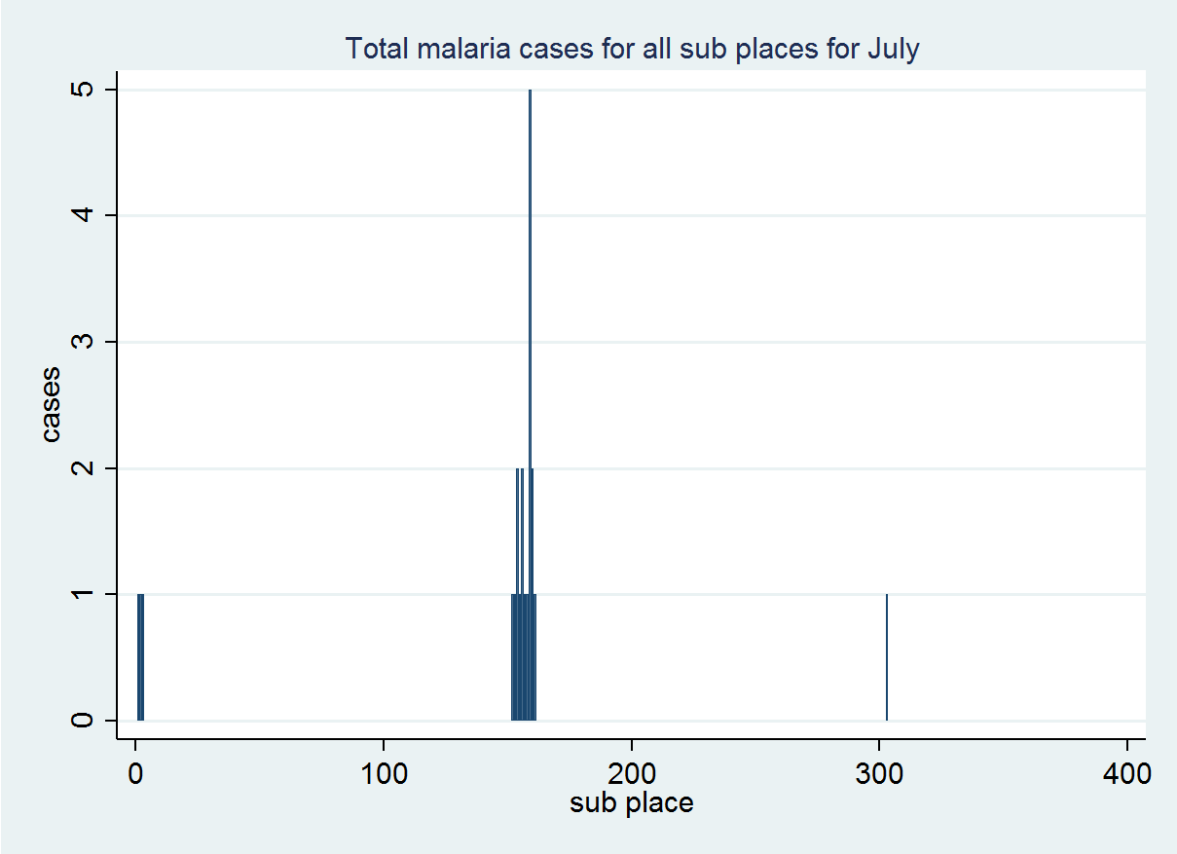


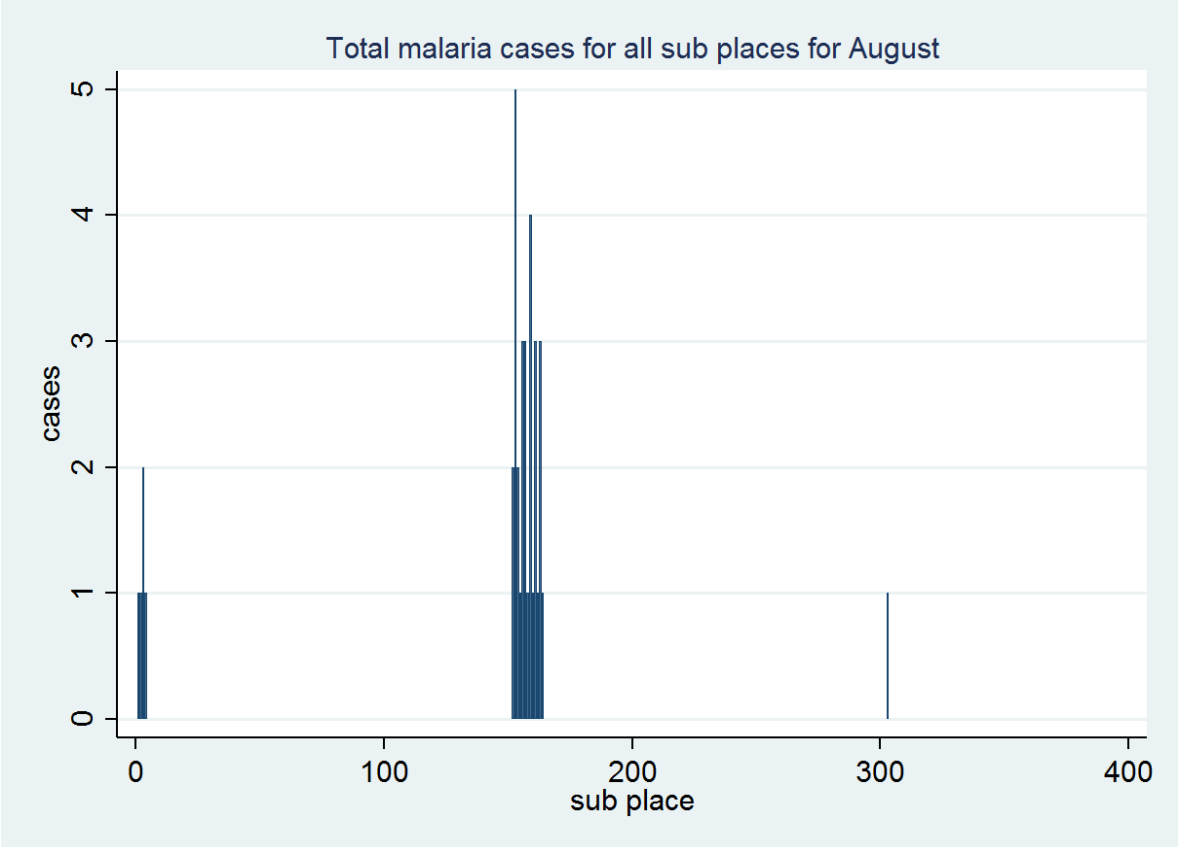

Total malaria cases for all subplaces for March
cases
0
20
40
60
80
0
100
200
300
400
sub place

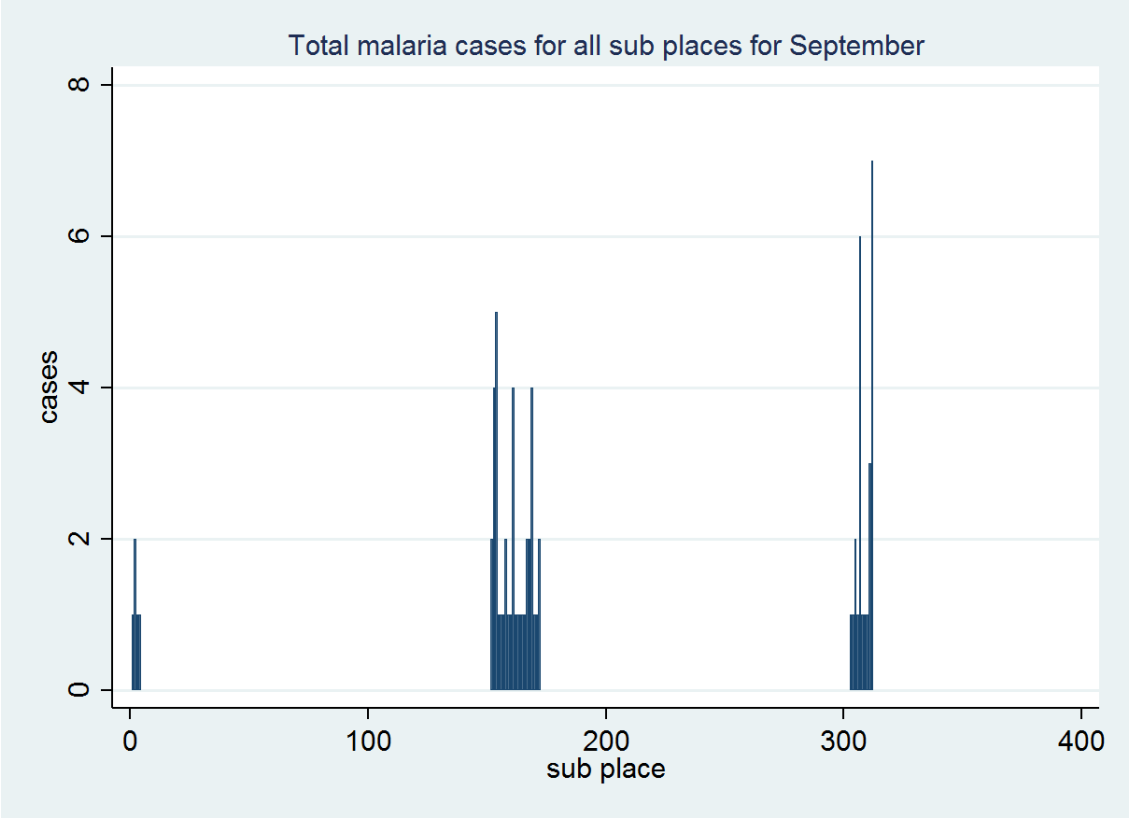

Total malaria cases for all sub places for April
cases
0
5
10
15
0
100
200
300
400
sub place

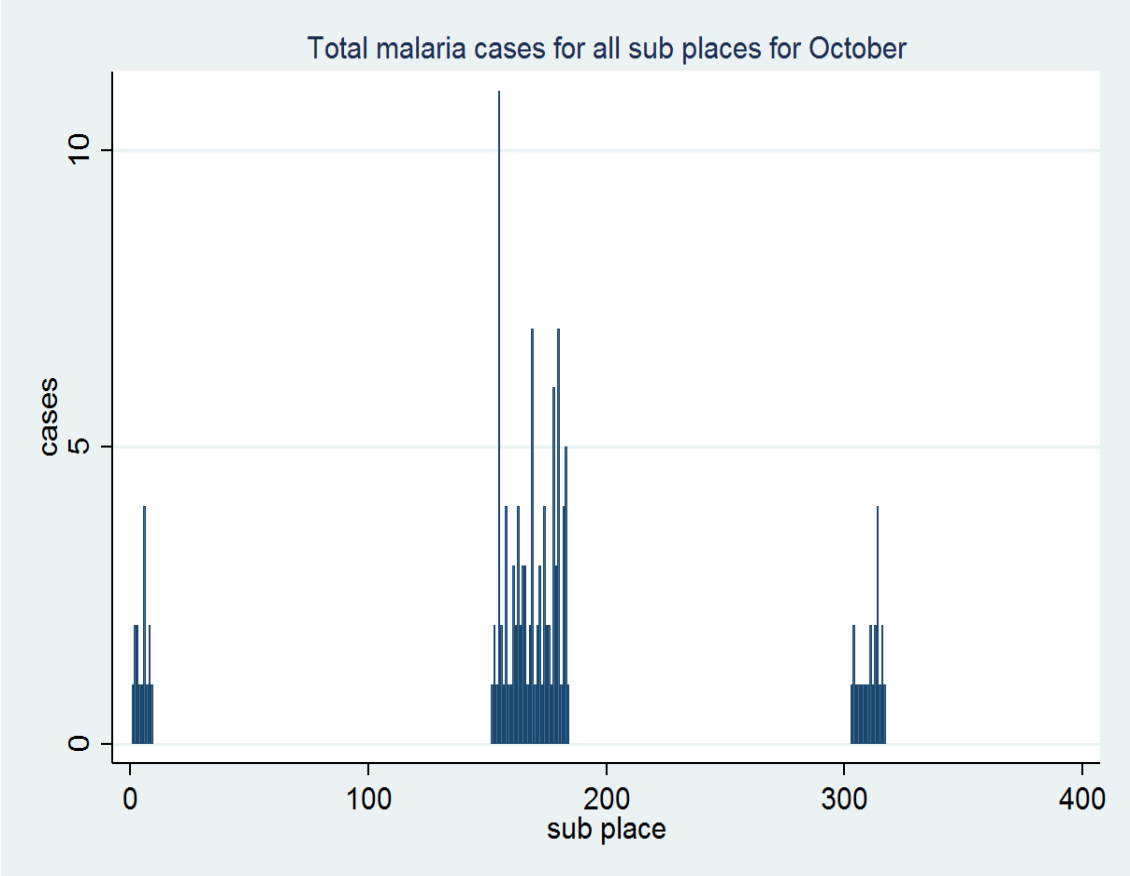

Total malaria cases for all subplaces for May
cases
0
5
10
15
20
25
0
100
200
300
400
sub place

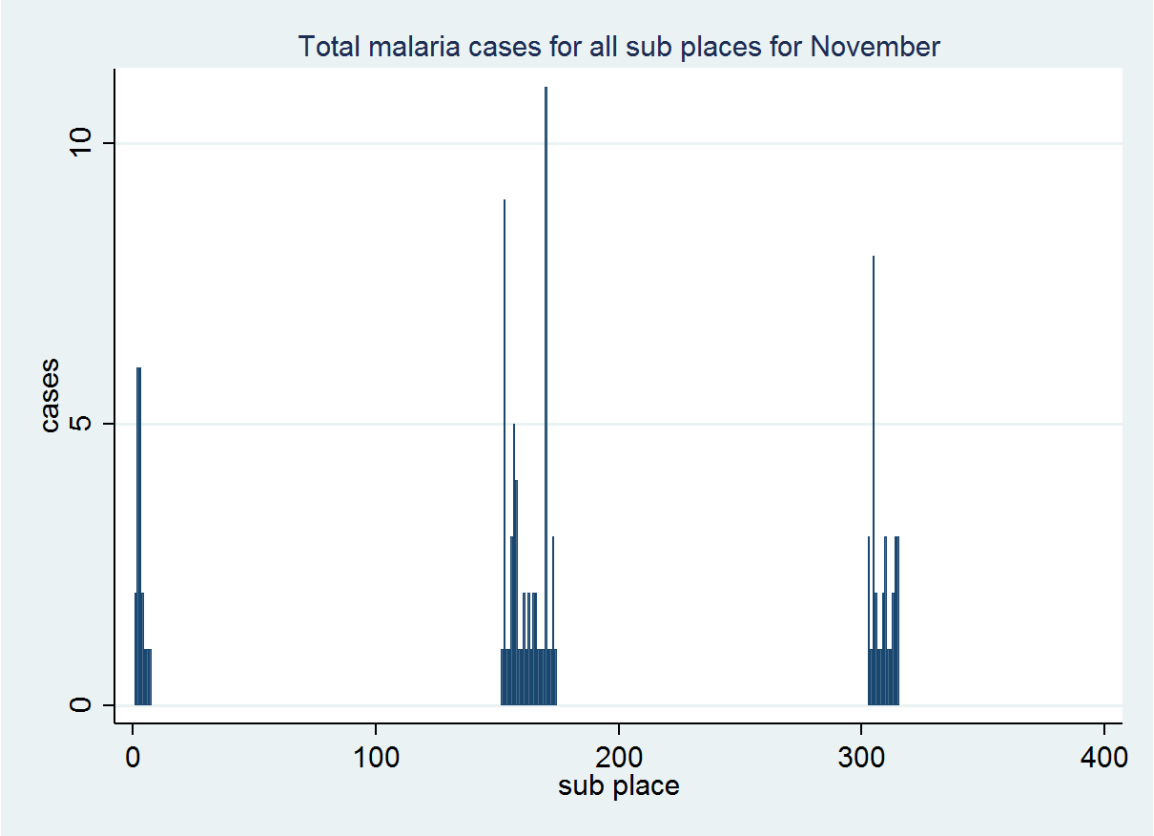

Total malaria cases for all subplaces for June
cases
0
1
2
3
4
5
0
100
200
300
400
sub place

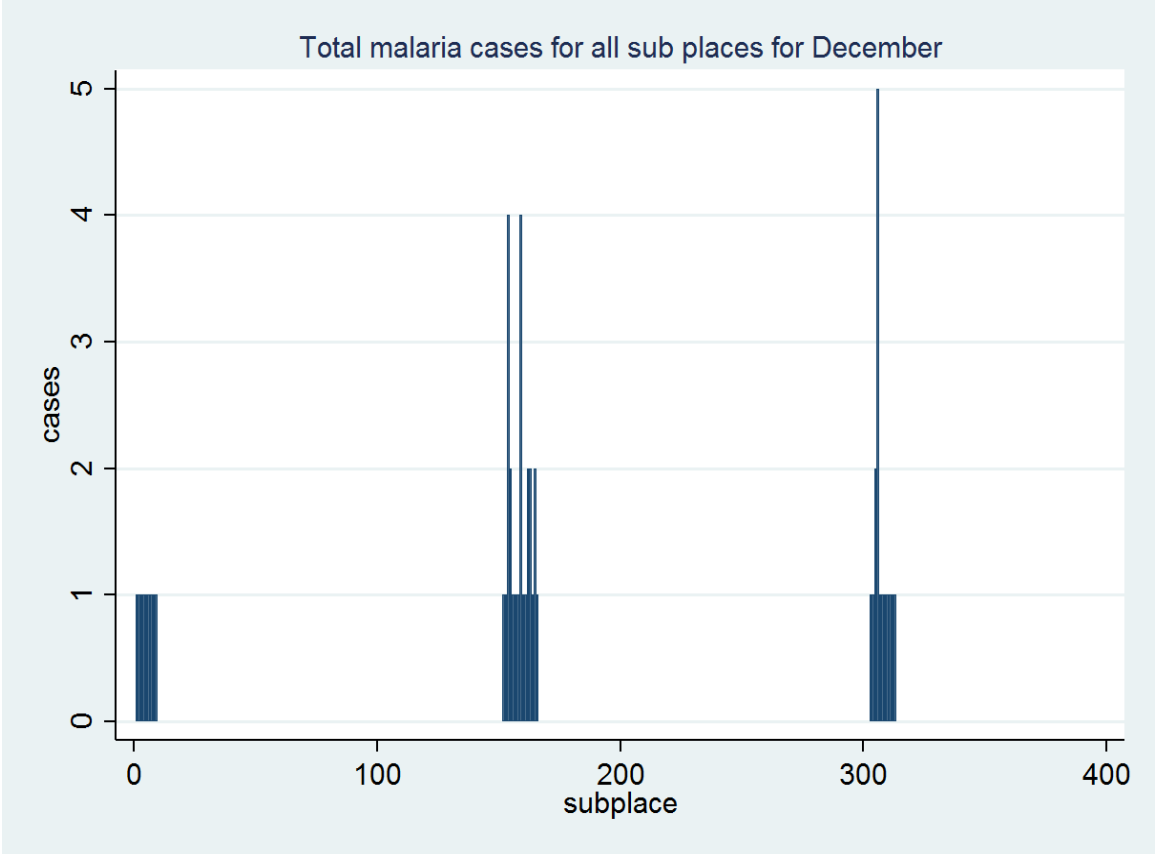

Total malaria cases for all sub places for July
cases
0
1
2
3
4
5
0
100
200
300
400
sub place


Total malaria cases for all sub places for August
cases
0
1
2
3
4
5
0
100
200
300
400
sub place


Total malaria cases for all sub places for September
cases
0
2
4
6
8
0
100
200
300
400
sub place


Total malaria cases for all sub places for October
cases
0
5
10
0
100
200
300
400
sub place


Total malaria cases for all sub places for November
cases
0
5
10
0
100
200
300
400
sub place


Total malaria cases for all sub places for December
cases
0
1
2
3
4
5
0
100
200
300
400
subplace