Statistical Analysis of Spatial Distribution of Malaria in Ilu Aba Bor Zone, southwest Ethiopia

By: Ebsa Gelan

A Thesis Submitted to Department of Statistics, College of Natural Sciences, Jimma University as a Partial Fulfillment for the Requirements of the Degree of Master of Science (MSc) in Biostatistics

February 2020

Jimma,Ethiopia

Statistical Analysis of Spatial Distribution of Malaria in Ilu Aba Bor Zone, southwest Ethiopia

By: Ebsa Gelan

Advisor: Akalu Banbeta (PhD Scholar)

Co-advisor: Reta Habtamu (MSc)

February 2020

Jimma, Ethiopia

STATEMENT OF THE AUTHOR

As author of this research study, I declare that the thesis is a result of my work, support of my supervisors and help hands of other individuals. Thus, all those had who participated in the study and sources of materials used for writing this thesis have been duly acknowledged. I have submitted this thesis to Jimma University as a partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The library directorate of Jimma University can deposit the copy of the thesis in the university library so that students and researchers can refer it. Moreover, I declare that I have not so far submitted this thesis to any other institution anywhere for the award of any academic degree, diploma or certificate and/or to get prove of society's problems. Any brief quotations from this thesis are allowed without requiring special permission if an accurate acknowledgement and citation (after publication) of the source is made. In all other instances, however, permission must be obtained from the author.

Ebsa Gelan

Date: _____

Signature: _____

February 2020

Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis research advisors, we here by certify that we have read the thesis prepared by Ebsa Gelan under our guidance, which is entitled "**Statistical Analysis of Spatial Distribution of Malaria in Ilu Aba Bor Zone, southwest Ethiopia**", in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready.

Akalu Banbeta (PhD Scholar)	_____	_____
Advisor	Signature	Date
Reta Habtamu (MSc)	_____	_____
Co-Advisor	Signature	Date

As the members of the board of examiners of M.Sc. thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____	_____	_____
Name of Chairman	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of Co-Advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

ACKNOWLEDGEMENTS

Above all, I would like to thank the almighty God for the gift of health, wisdom and strength throughout any steps for the achievement of my thesis and more of my entire life.

My deep gratitude goes to my advisor Mr. Akalu Banbata (PhD scholar at Hasselt University) for his professional and deep comments, advice and guidance at any stage of this thesis. I appreciated his friendly contact and his punctuality for commented this paper. Simply, without his frequent supervision, this thesis may not reach at the stage it is now.

I sincerely appreciate my co-advisor Mr. Reta Habtamu (MS.c) for his valuable suggestions and comments for the successful realization of this thesis. It is a great pleasure to express his commitment to share his experience, knowledge and his smooth contacts is also appreciable.

My thanks also to Ethiopian metrological Agency and Ilu Aba Bor Zone Health Bureau for providing me with malaria case counts data and metrological data for this thesis; Grateful acknowledgement is also expressed to the Department of Statistics, Jimma University for kind assistance in many ways.

Furthermore, I want to thank all my friends, my family members and others who have provided me a great help and moral support directly and indirectly contributed towards the completion of this study.

Table of Contents

LIST OF TABLES

LIST OF GRAPHS

ACRONYMS

AIC……Akaike Information Criterion

ARDL……Autoregressive Distributed Lag model

BIC……Bayesian Information Criterion

CI……...Confidence Interval

DLNM……...Distributed Lag Non-Linear Models

ENMIS…….....Ethiopia National malaria indicator survey

FMOH……....Federal Ministry of Health

FY…….....Fiscal year

GCM…….....Global Climate models

GEE ……..Generalized Estimating Equation

GLM…….....Generalized Linear Model

GLMM…….....Generalized Linear Mixed Model

GWR ……Geographically Weighted Regression

MCMC……..Markov Chain Monte Carlo

MLE……..Maximum Likelihood Estimation

MOP……..Malaria Operational Plan

PQL……Penalized Quasi-Likelihood

RDT……...Rapid Diagnosis Test

SNNPR…….Southern Nation, Nationalities and People's region.

WHO……....World Health Organization

Abstract

Introduction*: Malaria is the major public health problem, widespread throughout the tropical and subtropical regions, including parts of Africa, Asia and America. According to the World Health Organization report an estimated 219 million cases of malaria occurred worldwide in 2017.The WHO African Region takes largest burden of malaria morbidity, with 200 million cases (92%) in 2017. In Ethiopia 68 percent of population lives in malarious areas, and 75 percent of country's landmass is favorable for malaria transmission.*

Objectives: *The main objective of this study was to find spatial distribution of malaria and to identify variables that are associated with malaria distribution in Ilu Aba Bor Zone, southwest Ethiopia.*

Methods: *The study has been conducted in Ilu Aba Bor zone of entire districts and the data is basically secondary which is obtained from Ilu Aba Bor zone health office and Ethiopian metrological agency. Spatial distribution of malaria was identified using global and local measures of spatial autocorrelation. After spatial pattern is identified, the counts of malaria case have been analyzed with covariates like average annual maximum temperature, average annual minimum temperature, average annual rainfall, percentage of highland area, percentage of midland area and percentage of lowland area. The author of this study extended the application performed using generalized linear model to generalized linear mixed model by adding random effect and correlation structure to account spatial dependence in the model.*

Results: *The value of global and local measures of spatial autocorrelation shows that malaria varies according to geographical location and shows significant positive spatial autocorrelation. The results of negative binomial regression model with spatial dependence shows that statistically significant relationship between malaria case counts and independent variables (rainfall, maximum temperature, midland area and lowland area).*

Conclusions: *There is evidence of significant malaria clustering in Ilu Aba Bor zone, southwest Ethiopia. Significant hot spots clusters were identified in three woredas and cold spots of malaria clusters were identified in eight woredas. Clustering of dissimilar value identified in three woreda. There is significant relationship between malaria and covariates (rainfall, maximum temperature, midland area and lowland area).*

Key words: Malaria, Spatial Autocorrelation, Poisson Regression Model, Negative Binomial Regression Model, Generalized Linear Model and Generalized Linear Mixed Model.

Chapter One

1. Introduction

1.1 .Background of the Study

Malaria is a mosquito borne infectious disease of other animals and human caused by protists of the genus plasmodium which are introduced into the circulatory system by the bite from an infected female anopheles mosquito. It is the most common widespread throughout the tropical and subtropical regions, including parts of Africa, Asia and America. It is a major cause of illness and death in large area of the developing world, especially Africa. According to the World Health Organization world malaria report (WHO, 2018) an estimated 219 million cases of malaria occurred worldwide in 2017. The WHO African Region takes 200 million cases (92%) of largest malaria morbidity in 2017, followed by the WHO South-East Asia Region (5%) and the WHO Eastern Mediterranean Region (2%). The cases were increased by 3 million as compared to 2016 report.

In 2017 there were an estimated 435 000 deaths from malaria globally, from this 266 000(61%) malaria deaths were children aged under 5 years. According to world health organization world malaria report, in every 2 minutes, a child under five dies by malaria disease (WHO, 2018). Most of these deaths occurred in Sub-Saharan Africa. Malaria is severe socioeconomic impact on populations (WHO, 2017). It causes household poverty because it results in absenteeism from the daily activities of productive living and income generation. It prevents many school children from attending school due to illness, diminishing their capacity to realize their full potential.

Malaria is the most severe public health problem affecting many people each year and it is a leading cause of death in Africa, where young children and pregnant women are the groups that most affected by this disease. Malaria in Africa shows variation from year to year and from country to country. A large number of environmental factors affect the distribution, seasonality and transmission intensity of malaria in Africa. Climatic conditions in Africa affect the number and survival of mosquitoes. Those climatic factors that affect distribution and transmission of malaria are rainfall patterns, temperature and humidity (WHO, 2017). In many places in Africa,

malaria transmission is seasonal, with the peak during and just after the rainy season. Malaria epidemics in Africa can occur when climate and other conditions suddenly favour transmission in areas where people have little or no immunity to malaria. They can also occur when people with low immunity move into areas with intense malaria transmission, for instance to find work, or as refugees. Human immunity is another important factor, especially among adults in areas of moderate or intense transmission conditions. In 2017 African region was home to 403 000(93%) deaths from malaria (WHO, 2018).

In Ethiopia about 75 percent of country landmass is malarious area and 68 percent of the total population lives in areas at risk of malaria (Eniyew,S, 2018). Malaria morbidity and mortality have been significantly decreased in Ethiopia and worldwide in the past decade (WHO, 2017). However, malaria still remains a major public health problem in Ethiopia (WHO, 2017).Malaria transmission in Ethiopia occurs mainly at altitudes less than 2000 m, although endemic regions greater than 2000 m (Yalew *et al.*, 2017).

The levels of malaria risk and transmission intensity shows spatial variability, inter-annual and seasonal because of altitude and climatic factors, with the exception of the southwestern international border lowland area where transmission is year-around (Zhou *et al.*, 2016). The combination of climatic factors such as temperature, rainfall and humidity shows high variability as a function of altitude are the most important variables that influence malaria transmission. Based on altitudinal variation and eco-climatic condition associated with it, area of the country are categorized into three climatic zones those are: highland zone, lowland zone and mid-land zone. In most regions of the country, the major transmission season is from september to december, following the main rainy season from june to september (Taffese *et al.*, 2018*).*

In 2016 there were an estimated 2,927,266(95% CI, 525,000-6,983,000) new malaria cases in Ethiopia and it caused an estimated 4,782(95% CI 122.5–12,750) deaths with a crude death rate of 4.7/100,000 and Age-standardized death rate (ASDR) of 4.9/100,000 Population (Girum *et al.*, 2019).Generally, the diverse ecology of the country supports a wide range of transmission intensities ranging from low-seasonal to high-perennial transmission.

The Ethiopian President's Initiative has a regional focus with priority to the most populous and malaria-prone regions including oromia .In oromia three-quarters of the administrative Weredas

(242 out of 261) and 3932 Kebebles out of 6107 are considered as malarious. 17 million people are at risk in Oromia with annual clinical cases numbering between 1.5 and 2 million, this accounts for 20 – 35% of outpatient visits, and 16% of hospital admissions in the region, where 18-30% of annual deaths are caused by malaria (MOP FY, 2008).

In Ilu Aba bor zone malaria is one of the leading causes of public health problems affecting population of the zone each year. Malaria case in this zone varies from district (woreda) to district (woreda) and from year to year (Ilu Aba Bor Zone Health Bureau, unpublished source).

1.2. Statement of the Problem

Malaria is one of the leading causes of mortality and morbidity worldwide especially in sub Saharan African countries including Ethiopia, around 3.3 billion people are at risk of malaria worldwide (Coulibaly *et al.,* 2013). The incidence of malaria worldwide is estimated to be 216 million cases per year and 655,000 people died by malaria disease per year, from this 81% of malaria cases and 91% of deaths occur in sub-Saharan Africa countries (Coulibaly *et al*., 2013). In Ethiopia malaria is one of the common diseases that cause illness and death. About 75% of the country landmass is malarious area and more than 54 million peoples are vulnerable (Alelign, A., & Dejene, T, 2016).Malaria in Ethiopia varies from location to location and varies from season to season (Graves *et al.*, 2009).

Malaria mortality and morbidity are known to vary by geographical location and depend on eco-climatic conditions .Targeting interventions to high malaria case are omitted due to inconsideration of spatial dependence. Including spatial dependence is essential to understand and decrease malaria (Ayele, D. G ,*et al*, 2016). Hence malaria mortality and morbidity vary according to geographical location and depend on eco-climatic condition including spatial dependence is very important. In spatial dependence the regions that are in closer proximity are expected to have similar malaria cases because of similar eco-climatic situation and demographic characteristics. Measures of Spatial autocorrelation analyze the degree of dependency among observations in a geographic space. It requires measuring a spatial weights matrix that reflects the intensity of the geographic relationship between observations in a neighborhood. Controlling malaria at woreda(districts) level needs identifications of spatial pattern of malaria and identification of environmental factors that related with malaria distribution among the districts of Ilu Aba Bor zone .The researcher measure spatial dependence using global and local measures

of spatial autocorrelation to find spatial pattern of malaria in districts of Ilu Aba Bor zone, southwest Ethiopia.

Different studies conducted on malaria finds only spatial pattern without finding covariates associated with malaria distribution(Berga, B) and those studies find covariates associated with malaria distribution does not find spatial pattern of malaria (Boateng *et al*.,2015; Mohammadkhani *et al*., 2019; Midekisa *et al*., 2015).Also some study find spatial distribution using explanatory spatial data analysis and they used spatial lag and spatial error model to find variables associated with malaria (Dessie, D. B. 2017, Mosissa, M., & Gotu, B, 2011),but in this study the author identify spatial pattern of malaria in districts of Ilu Aba Bor zone, southwest Ethiopia and model malaria case counts by extending application of generalized linear model to generalized linear mixed model by adding random effects and correlation structure. Offset variable is also considered to adjust the number of events and population size. So far there is no published research or study has been conducted on spatial pattern indicator of malaria and its environmental factor indicators in Ilu Aba Bor Zone, southwest Ethiopia.

Therefore, this study is aimed to address the following questions:

- ✓ Does the spatial distribution of malaria spatially random or clustered in districts of Ilu Aba Bor zone?
- ✓ How to include spatial dependence in the model?
- ✓ Which predictor variables significantly affect the distribution of malaria in this study area?

1.3. Objectives of the Study

1.3.1. General Objective

The main objective of this study is to find spatial distribution of malaria and to identify variables that are associated with malaria distribution in Ilu Aba Bor Zone, southwest Ethiopia.

1.3.2. Specific Objective

The specific objectives of this study are:

- ✓ To identify spatial distribution of malaria in districts of Ilu Aba Bor zone.
- ✓ To include spatial dependence in the model.
- ✓ To identify covariates that significantly affects the distribution of malaria.

1.4. Significance of the Study

The organizations as well as individuals who work in this area get a clue on the spatial distribution of malaria in the districts of Ilu Aba Bor zone. The other basic significance of this study was to further assist other researchers interested in explanatory spatial data analysis and they may use it as a benchmark for their future works in identifying spatial distribution of disease, therefore, researchers would benefit by getting familiar with this method and may further help in advertising explanatory spatial data analysis, Therefore, the result of this study has the following importance:

- Provide information to government and other concerned organizations on spatial pattern of malaria in this study area.
- Provide information to researchers for further studies on explanatory spatial data analysis.
- Help to identify factors that are related to malaria distribution in this study area.
- Help to identify woredas under hot spot and cold spot.

1.5 Limitation of the Study

In this study the author used only some climatic factors and altitude that determine spatial pattern of malaria in this study area. As limitation relative humidity that determines spatial pattern of malaria is not included in this study.

1.6 Organization of the Study

This paper is organized as follows. The first chapter of this study describes brief Introduction about the Study, Statement of the problem, Objectives of the study, Significance of the study, limitation of the study. The second chapter provides overview of malaria case, determinants of malaria distribution, model based and explanatory spatial data analysis literatures .The third chapter describes the study area, Source of data, Variables in the study and Methodology that used for data analysis. The fourth chapter provides outputs of explanatory spatial data analysis, outputs of model based data analysis and discussion of each output. Chapter five provides Conclusion and Recommendation of the study depending on the results.

Chapter Two

2. Review of Literature

In this chapter only those sources that are more related and relevant to this study are presented.

2.1. Overview of Malaria Cases

Malaria is a life-threatening disease caused by parasites and is transmitted to people through the bites of infected female Anopheles mosquitoes. It is preventable and curable. Infected mosquitoes carry the plasmodium parasite. When this mosquito bites humans, the parasite is released into bloodstream. Once the parasites are inside human body, they travel to the liver, where they mature. Within 48 to 72 hours, the parasites inside the red blood cells and continue to infect red blood cells. There are four kinds of malaria parasites that can infect humans those are: Plasmodium vivax, P. ovale, P. malariae, and P. falciparum.Malaria found mainly in the tropical and subtropical climates countries around the world where the parasites can live. It is found in large parts of sub-Saharan Africa and Asia, Central and South America, Haiti and the Dominican Republic, some Pacific islands, such as Papua New Guinea and some parts of Middle East.

2.2. Determinants of Malaria Disease.

Transmission of malaria is very complicated. It can be determined by climatic or non-climatic factors. Climatic factors greatly influence the pattern and level of malaria transmission in the world, in Africa and in Ethiopia. Different study said transmission of malaria is based on climatic variability (M'Bra *et al*., 2018; Zayeri *et al*., 2011; Hussein,H.H.,2019; Nkurunziza *et al.*, 2010).Malaria transmission patterns are spatially and temporally heterogeneous due to the strong association with environmental conditions. The relationship between malaria transmission intensity and climate and ecological parameters has been validated by both entomological models and patterns of human case incidence (Mabaso *et al.*, 2007).The most important climatic factors that affect malaria transmissions are minimum temperature, maximum temperature and rainfall(Adeola *et al*., 2017; Srimath-Tirumula-Peddinti *et al*.,2015; Alemu *et al*., 2011).

The relationship between malaria and environmental factors is strongly varied spatially in different regions (Hasyim *et al.*, 2018) the aim of their study was to investigate the spatial

association between malaria occurrence and environmental risk factors. The researchers used GWR (geographically weighted regression) model and find that altitude and rainfall were significantly associated with malaria cases.

Arab *et al*. (2014) conducted research on modeling the effects of weather and climate on malaria distributions in West Africa. The objective of the study was to present a hierarchical Bayesian statistical modeling framework that can be used to analyze the effect of multiple climate factors on the distribution of malaria by taking into account spatiotemporal dependencies. In their study annual malaria and climate data from ten West African countries (Benin, Burkina Faso, Côte d'Ivoire, Gambia, Ghana, Liberia, Mali, Senegal, Sierra Leone, and Togo) during the period 1996-2006 is used. The result of the researchers showed that there is a statistically significant correspondence between malaria rates and the climate variables.

Nkurunziza *et al*. (2011) conducted research on Geo-additive modeling of malaria in Burundi. The goal of this study was to understand the dependence of malaria cases on climatic variables and spatial (correlated and uncorrelated) effects in Burundi. The semi parametric models were used to model the effects of both climatic covariates and spatial effects on malaria distribution in Burundi. The spatial analysis was based on a geo-additive model in which the province is the geographic unit of analysis. The spatial effect was split into smooth structured and unstructured (random) components. Inference was fully Bayesian and was based on (MCMC) Markov chain Monte Carlo techniques. The effects of climatic covariates and the effects of other spatial determinants were estimated simultaneously, in a unified regression framework. The obtained results suggest that malaria incidence in a given month is positively associated with the minimum temperature of the same and previous month. In contrast, malaria incidence is negatively associated with rainfall and maximum temperature of the same month.

Temperature

Malaria is more sensitive to minimum temperature in cool climates and maximum temperature in warm climates (Xiang *et al.,* 2018).The ranges of minimum and maximum temperature greatly affect the development of the malaria parasite and its mosquito vector. There is significant association between malaria and maximum temperature, minimum temperature (Chuang *et al*.,

2017). They used seasonal autoregressive integrated moving average models and distributed lag non-linear models (DLNM) the result of their study showed that maximum temperature and minimum temperature associated with malaria distribution.

Altitude

Altitude or elevation above sea level is one of the most important factors that determine the pattern of malaria transmission in Ethiopia. There is significant relationship between malaria and altitude (Drakeley *et al.*, 2004).Altitude in Ethiopia varies from 100 metres below sea level to more than 4,000 metres above sea level. Altitude influences the distribution and transmission of malaria indirectly, through its effect on temperature. As altitude increases, temperature decreases, so highlands are colder and lowlands are warmer. According to Kipruto *et al*, 2017 Malaria cases increase significantly in the highland and midland zones they used negative binomial regression model with lagged climate variables to model long-term monthly malaria cases.

Rainfall

Rainfall was an important meteorological factor for prediction of malaria incidence. There is significant relationship between rainfall and malaria (Kifle *et al.*, 2019, Lingala, M.A, 2017, Kumar Ra, P., Nathawat, M. S., & Onagh, M. 2014). Anopheline mosquitoes breed in water. So the right amount of rainfall is often important for them to breed. Different anopheline mosquitoes prefer different types of water bodies in which to breed. In Ethiopia, water collections that support vector breeding appear mainly after the rains, therefore malaria transmission is highest following the rainy season.

Samadoulougou *et al*. (2014) conducted research on multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. The goal of this study was to examine individual, household, community characteristics and climatic/environmental factors associated with malaria infection. The researchers used Multilevel and geo-statistical models to explore determinants of malaria using nationally representative database the result of their study showed that malaria prevalence was associated with monthly rainfall.

Model based and explanatory spatial data analysis literatures

In this section different literature that shows spatial pattern of malaria and different statistical models that identify determinants of malaria was presented.

Laguna *et al*. (2017) conducted research on modeling malaria incidence by an autoregressive distributed lag model with spatial component .The main goal of their study was to fit a spatio-temporal model predicting malaria incidence taking into account a heterogeneous geography and the non-availability of spatially varying climatic variables across the whole region of an endemic-epidemic-prone region in Northern South America. They develop a two step methodology to model the spatio-temporal dependence of malaria incidence on local rainfall, temperature, and humidity as well as the regional sea surface temperatures (SST) in the northern coast of Venezuela. First they fit an autoregressive distributed lag model (ARDL) to the weekly data, and then, they adjust a linear separable spatial vectorial autoregressive model (VAR) to the residuals of the ARDL. The result of the researchers showed that the best model to account for the variations of malaria incidence from 2001 to 2008 in 10 endemic Municipalities in North-Eastern Venezuela is a logit model that included the accumulated local precipitation in combination with the local maximum temperature of the preceding month as positive regressors. Additionally they showed that malaria dynamics is highly heterogeneous in space.

Hurtado *et al*. (2018) conducted research on Analysis of spatial distribution of malaria in the department of Chocó for the year 2016. The main objective of this study was to implement spatial statistical model that best explain the dependence between observations of malaria in 2016 and to identify explanatory variables that have an impact on the occurrence of the malaria, as well as the generation of maps that allow to stratify the zones according to the number of occurrence and the risk that the spreading of the infection in the department of Chocó. The study result showed that phenomenon transmitted by malaria vectors has a tendency of self-spatial correlation, where high-value groupings occur and another one at the opposite extreme with low values.

Umer *et al*. (2018) conducted research on Spatiotemporal Clustering Analysis of Malaria Infection in Pakistan the study result showed that malaria incidence varies by species among the districts and over the years.

Mohammadkhani *et al*. (2019) conducted research on the Relation Between Climatic factors and malaria incidence in Sistan and Baluchestan, Iran .Changes in incidence of malaria with climatic factors were analyzed by negative binomial regression model .The incidence of malaria had a significant positive correlation with the average minimum, and maximum monthly temperatures and a negative correlation with rainfall.

Kleinschmidt *et al*. (2001) conducted research on Use of Generalized Linear Mixed Models in the Spatial Analysis of Small Area Malaria Incidence Rates in KwaZulu Natal, South Africa the researchers describe a method of adjusting the regression analysis results for strong spatial correlation in the rates by using generalized linear mixed models and variograms. The results of the spatially adjusted, multiple regression analysis showed that malaria incidence was significantly positively associated with higher winter rainfall and a higher average maximum temperature and was significantly negatively associated with increasing distance from water bodies.

Boateng *et al*. (2015) conducted research on Analysis of Malaria Incidence using Quasi-Poisson Regression Model: Evidence from Obuasi Municipality, Ghana they use quasi-Poisson regression model to determine the incidence of malaria in Obuasi Municipality, Ghana the result of this study showed that temperature and water for breeding habitats are important primary ecological factors that impact the distribution of malaria vectors and the rate at which mosquito and parasite develop.

Akinbobola A, *et al.* (2018) conducted research on Determining Malaria Hotspot Using Climatic Variables and Geospatial Technique in Central Urban Area of Ibadan, southwest, Nigeria. The objective of this study was to evaluate the environmental risk factors affecting malaria prevalence and the spatial-temporal distribution of malaria incidence. The results of this study showed that there is a strong relationship between climatic factors and malaria incidence especially maximum temperature. It was also observed that the intensity of clustering of high

values (hot spot) and low value (cold spot) throughout the study period. The statistically significant hotspots of malaria were consistently detected in southern and eastern part of Ibadan central urban area (Ibadan South-West, Ibadan South-East and Ibadan North-East LGAs) and Ibadan North and Ibadan North-West LGAs (local government areas) remained malaria cold spot throughout the study period. They researchers also argued that the results of the study help government agencies, health practitioners and policy makers to plan in the prevention and control of malaria prevalence in the area.

Eunice. (2018) conducted research on statistical Modeling of Malaria Incidences in Apac District, Uganda. The objective of this study was to model factors associated with malaria incidences in Apac Districts. The researcher used Poisson and negative binomial regression models to analyze the data. Negative binomial model provided a better fit as compared to the Poisson regression model. The result of this study showed that there is significant relationship between monthly malaria incidence and climate variables that is rainfall and temperature.

Ihantamalala *et al*. (2018) conducted research on Spatial and temporal dynamics of malaria in Madagascar, the study result showed that the crude incidence of malaria in Madagascar increased between 2010 and 2016 from 14 per 1000 to 20 per 1000. The highest incidence was in 2015 with a value of 32 per 1000 and the lowest was 12 per 1000 in 2011. The incidence is also spatially heterogeneous with different seasonal patterns across the country.

Ouédraogo *et al*. (2018) conducted research on spatial distribution and determinants of asymptomatic malaria risk among children under 5 years in 24 districts in Burkina Faso. The objective of this study was to produce maps at the 24 health districts that guide malaria control programme in Burkina Faso and to examine the determinants of malaria asymptomatic infection in children under five. The result of this study showed that the prevalence of asymptomatic malaria is influenced by the geography, climatic and socio demographic/economic characteristics. Its Eradication will require not only biological interventions but also social interventions (distal). This disease mapping technique could be used systematically in the planning and evaluation of malaria elimination efforts.

Tewara *et al*. (2018) conducted research on Small-area spatial statistical analysis of malaria clusters and hotspots in Cameroon; 2000–2015.The objectives of this study are; to use the spatial autocorrelation technique to analyze malaria spatial pattern and to identify environmental factors associated with the distribution of malaria cases. The study result showed that the spatial distribution of malaria showed statistically significant clustered pattern for the year 2000 and 2015 .There exist varying degrees of malaria clusters and statistically significant hotspots in the urban-rural areas of the 12 administrative regions. They used Pearson's Correlation analysis to identify associative environmental factors and find that malaria cases were associated with rainfall.

DePina *et al*. (2019) conducted research on Spatio temporal characterization and risk factor analysis of malaria outbreak in Cabo Verde in 2017. The aim of this study was to examine the spatial and temporal epidemiological profile of malaria across the country during the 2017 outbreak and to analyze the risk factors which influence malaria. The study result showed that there is strong positive correlation between malaria and temperature.

Rouamba *et al*. (2019) Conducted research on Socioeconomic and environmental factors associated with malaria hotspots in the Nanoro demographic surveillance area, Burkina Faso. The aim of this study was to define accurately the different transmission (or incidence) periods of malaria at a fine scale rural area and to estimate the lag times between meteorological variables and high malaria incidence period. The researchers used statistical cross correlation to quantify the temporal association between weekly malaria incidence and meteorological factors. Local spatial autocorrelation analysis was performed and restricted to each transmission period using Kulldorff's elliptic spatial scan statistic. Univariate and multivariable analysis were used to assess the principal socioeconomic and meteorological determinants of malaria hotspots using a Generalized Estimating Equation (GEE) approach. The result of the study showed that Rainfall and temperature were positively and significantly associated with malaria incidence, with a lag time of 9 and 14 weeks, respectively. Result of spatial analysis showed that there is significant malaria hotspot throughout the study period. Furthermore, low socioeconomic status of households was strongly associated with malaria hotspots (aOR = 1.21, 95% confidence interval: 1.03–1.40).

Yankson *et al*. (2019) conducted research on Geo statistical analysis and mapping of malaria risk in children under 5 using point referenced prevalence data in Ghana. The aim of this study was to analyze and map malaria risk in children under 5 years old, with the ultimate goal of identifying areas where control efforts can be targeted. The risk map indicates spatial heterogeneity of malaria prevalence.

Ayele *et al*. (2013) conducted research on spatial distribution of malaria problem in three regions of Ethiopia. The objective of this study was to undertake statistical analysis of malaria incidence with important socio-economic, demographic and geographic variables associated with malaria and to draw map that used to show variation in malaria risk. They used generalized linear mixed model with spatial covariance structure to analyze the data where the response variable was the presence or absence of malaria using the RDT (rapid diagnosis test). The results of the researchers showed that households in the SNNP region were found to be at more risk than Amhara and Oromia regions. Moreover, households which have toilet facilities ,clean drinking water, and a greater number of rooms and mosquito nets in the rooms, have less chance of having household members testing positive for RDT and the researchers also suggested that incorporating spatial variability is necessary for understanding and reducing the risk of malaria.

Mosissa, M., & Gotu, B,(2011) Conducted research on Statistical Analysis of Spatial Distribution of Malaria in West Shoa Zone, Ethiopia .The main objective of the researcher was to identify spatial patterns of malaria in West Shoa zone, Oromia region, Ethiopia. The statistical methods the researcher used were global and local measures of spatial autocorrelation as well as spatial autoregressive model. The result of the researcher showed that there are significant positive global measures of spatial autocorrelation and significant local clustering of malaria incidence occurred between pairs of neighboring districts. Measures of spatial autocorrelation also showed that Malaria incidence was higher in the western part of the zone and lower in the eastern part of the zone. Spatial lag model is best fit model and the results of spatial lag model indicate that there is significant relationship between malaria incidence and independent variables (midland zone, hot zone, rainfall, minimum temperature and maximum temperature).

Midekisa *et al*. (2015) conducted research on seasonal associations of climatic drivers and malaria in the highlands of Ethiopia. They assessed the interannual variability of malaria occurrence from 2001 to 2009 in the Amhara region of Ethiopia. They tested for associations of climate variables summarized during the dry (January–April), early transition (May–June), and wet (July–September) seasons with malaria incidence in the early peak (May–July) and late peak (September–December) epidemic seasons using generalized linear models. They find that both early and late peak malaria incidence had the strongest associations with meteorological conditions in the preceding dry and early transition seasons. Temperature had the strongest influence in the wetter western districts, whereas moisture variables had the strongest influence in the drier eastern districts. There is significant correlation between malaria incidence in the early and the subsequent late peak malaria seasons, and the addition of early peak malaria incidence as a predictor substantially improved models of late peak season malaria in both of the study sub-regions.

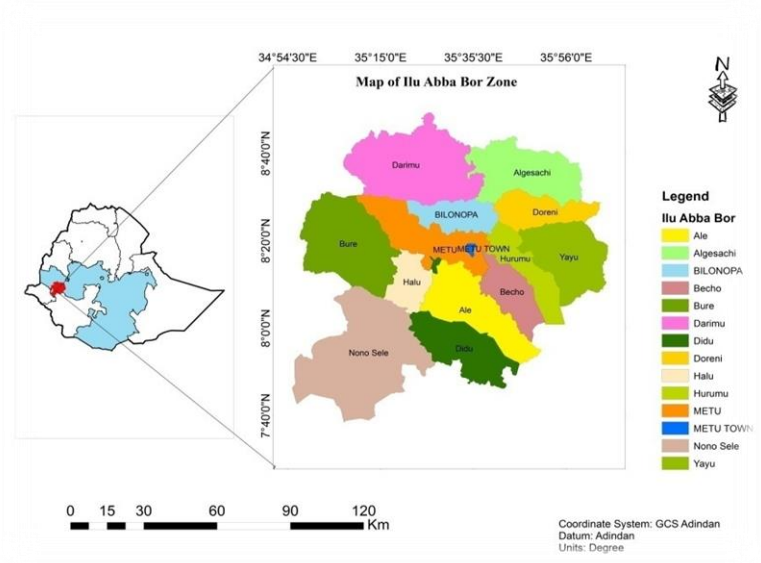
Dessie, D. B. (2017). Conducted research on Spatial Modeling of Malaria Prevalence and Its Risk Factors in Rural SNNPR, Ethiopia: Classical and Bayesian Approaches . The objective of this study was to identify and model determinants of Malaria prevalence rate and its risk factors in rural woredas of Southern Nation, Nationalities and People's region, Ethiopia. The result of spatial analysis revealed positive spatial autocorrelation pattern of malaria prevalence rates in space. The researcher used spatial lag and spatial error model. The result of model specification and measures of fits shows that, the spatial lag model was found to be better fit to the data and explain the geographical variations of malaria prevalence data in the region.

Chapter Three

3. Data and Methodology

3.1. Study Area

The study has been conducted in Ilu Aba bor zone, Oromia regional state southwest Ethiopia. There are 14 woredas in this Zone. The capital city of the zone, Mettu town is located in south west of Ethiopia at 600 Km far from Addis Ababa. This zone is bordered on the south by the Southern Nations,Nationalities and Peoples Region,on the southwest by the Gambela Region, on the west by Kelem Welega Zone, on the north by west Welega Zone ,on the east by bunno bedele Zone .All 14 woredas has been covered in this study.Those woredas are:Algesachi, Darimu, BiloNopha,Doreni ,Hurumu,Yayu,Ale,Bure, Nono sale ,Becho,Mettu town , Mettu rural ,Halu , Didu.the location of Ilu Aba Bor zone and each woreda of this zone is shown in the following map.



3.2 .Source of Data

The data for this study is discrete secondary data or count data obtained from Ilu Aba bor zone health Bureau which was carried out in 2010 Ethiopian calendar; the data were collected from all Woreda health centers, hospitals of Ilu Aba Bor zone and the Meteorological data has been obtained from Ethiopian Meteorological Agency. This study was focused on meteorological variables and altitude, because meteorological variables and altitude are the most important factors that are relevant to malaria Distributions.

3.3. Variable of study

The variables included in this study are listed as follows

Dependent variable: Malaria case counts.

Independent Variables: Average annual rainfall, Average annual maximum temperature, Average annual minimum temperature, Percentage of highland areas, Percentage of midland areas and Percentage of lowland area.

3.4. Statistical analysis

In this section, exploratory spatial data analysis (Moran's *I*, Geary's *C* and Local Indicators of spatial autocorrelation mainly Moran scatter plot) was used to know the distribution of events. To identify covariates related to malaria distribution, Poisson regression model and negative binomial regression model was used. Analysis is implemented using R, Geoda and ArcGIS software.

3.4.1. Spatial statistical analysis

The development of many of statistical methods and models has been linked to the study of specific applications within various scientific research fields. The analysis of spatial and spatio-temporal data is currently of great interest to statistical modeling. Problems related to meteorology, environmental pollution, ecology, epidemiology or economics, demand the use of statistical models for spatial and spatio-temporal data.

Spatial statistical analysis provides useful insights about the causes and patterns of malaria. There are different methods available to display disease distributions and analyze spatial patterns. By considering a variety of linkages or looking at the patterns of clustering of a malaria distribution, it is possible to investigate the factors at large or small scale.

Spatial statistics can be divided into three methods. These are: point pattern analysis, methods for lattice data and geo statistics (Schabenberger and Gotway, 2005, Cressie, 1993). Point referenced data: - is often called geo coded or geo statistical data. Areal data: - is often called lattice data. Some spatial data sets feature both point and areal-level data. Point pattern data: - The response occurrence of the event is often fixed and only the locations where it occurs are thought of at random. Of these, the geo statistical approach is most relevant to epidemiological analysis which is conducted at the landscape scale and based on remote sensing (Goovaerts, 1997).

3.4.1.1. Spatial Dependence

A basic property of spatially located data in a set of values (xi) is likely to be related over space. Many authors in various disciplines discuss presence of dependence among observations related on diseases in space. Some as cited in Cliff and Ord (1981) are summarized as follow.

Tobler (1970) has referred to "The first law of geography**:** everything is related to everything else but near things are more related than distant things". Stephan (1934) gives the following remark: "data of geographic units are tied together like bunches of grapes, not separated like balls in an urn." These ideas explain why spatial dependence has to be an issue in determining the distribution of Malaria disease.

Spatial dependence is a key concept in understanding and analyzing a spatial phenomenon. Pattern is that characteristic of the spatial arrangement of objects given by their spacing in relation to each other. Patterns might consist of clusters of points, a more regular than random arrangement, and trends across real and statistical surfaces and so on. Spatial dependence indicates that near places are more likely to be related than distant ones and usually most geographical patterns of interest involve groupings of similar values in clusters.

Spatial autocorrelation analysis is a technique used to detect disease patterns and measures the extent to which the occurrence of an event in a real unit contains or makes more probable to the occurrence of an event in neighboring areal unit. It is defined as the relation among values of a single variable that is attributable to the geographic arrangement of areal units on a map. Spatial

Autocorrelation is a measure of interdependence between values of a variable at different geographic locations and can be used to identify the degree of clustering (Goodchild, 1987).

Spatial autocorrelation uses a measure known as spatial autocorrelation coefficient to measure and test how clustered or dispersed points are in space with respect to their attribute values. In addition to this, spatial autocorrelation (as a concept that is applied to detect the patterns of points) of a set of points is concerned with degree to which points or things happening at these points are similar to other points or phenomena happening there. If significantly positive spatial autocorrelation exists in a point distribution, points with similar characteristics tend to be near each other. If spatial autocorrelation is weak or non-existent, adjacent points in a distribution tend to have different characteristics. The most widely used measures for the proximity of locations and the similarity of the characteristics of these locations are Moran's I and Geary's C statistics. These statistics mainly measure the strength of spatial autocorrelation among neighboring areal units and are used for testing the assumption of spatial independence or randomness.

3.4. 1.2. Method of testing spatial randomness

Testing for complete spatial randomness is the first step in the analysis of spatial point pattern data. Basically, the main question here is: are locations randomly distributed through the study Area or do the locations indicate some structure? There are several methods and algorithms that is used to answer the scientific question of spatial randomness or clustering of cases. However Quadrant count and nearest neighbor methods are commonly used to test the spatial randomness or clustering of events.

3.4.1.3. Quadrant Count Method

The basic idea of quadrant method is to divide the region D into subsets often rectangular shape (but other shapes are used as well), and then counting the number of events in each of the subsets. The use of quadrant counts can be used to access whether there is any spatial pattern in the data. If clustering is present in the data, then one would expect quadrants with higher counts to be located near each other. On the other hand, if the quadrant counts are spread out over the region, then there is evidence of uniformity. The quadrant count method can be described simply as partitioning the data set into n equal sized sub regions; these sub regions are called quadrants.

In each quadrant the number of events that occur will be counted and it is the distribution of quadrant counts that will serve as an indicator of pattern. The choice of the quadrant size can greatly affect the analysis, where large quadrants produce a coarse description of the pattern. If the quadrant size is too small then many quadrants may contain only one event or they might not contain any events at all.

3.4.1.4. Nearest neighbor method

Identification of polygons (woredas in this study) which are nearest to each other is a primary concern in exploratory spatial data analysis. In this approach, the nearest neighbor distance method is used to define spatial weight matrix that helps to develop the statistical method used in testing randomness.

Nearest neighbor method gives a flexible weight matrix that represents spatial dependence based on a decay relationship and the number of neighbors. The measurements of areal units that are nearer to each other tend to be similar. When the measurements are independent, then no spatial pattern is expected. The first stage to implement a spatial pattern analysis is the construction and estimation of the weight matrix, given the spatial arrangement of the observations (Anselin and Hudak, 1998).

Weight Matrix

A general spatial weight matrix can be defined as a symmetric binary contiguity matrix, which can be generated from topological information based on either adjacency or distance criteria. A fundamental characteristic distinguishing spatial data from time series data is the spatial arrangement of the observations. The spatial linkages or proximity of the observations are measured by defining a spatial weight matrix, denoted by W_{nxn}. The spatial weight matrix represents the strength of the potential interaction between locations. However, it has to be noted that the determination of the proper specification for the elements of a spatial weight matrix is one of the difficult and controversial methodological issues in spatial data analysis (Odland,1987).

There are two methods that are used in computing spatial weight matrix, namely the Euclidean distance method and the proximity method. The most common method is to consider two or more regions as neighbors if they share a common border or vertex. The proximity or adjacent

matrix W **,** also referred to as the weighting matrix, provides the mechanism for introducing spatial structure in spatial data.

According to the adjacency criteria, the element of spatial weight matrix is 1 if location i is adjacent to location j and 0, otherwise. These will result in nXn matrix with zeros and ones if the study region has n sites to be investigated. Spatial contiguity for polygons is the property of sharing a common boundary or vertex. Contiguity analysis is an important method for assessing unusual features in the connectivity distribution. The following examples show how weight matrix is constructed. Neighborhood relations are defined as either Rooks case, Bishop's case, or Queen's (King's) case.

- Rook's case: A rook contiguity definition considers objects sharing a common edge as neighbors (as shown in the Figure 1 below)
- A bishop contiguity definition considers objects sharing a common vertex as neighbors (as shown in the Figure 2).
- A queen contiguity definition incorporates both the rook and bishop definitions as any object sharing either a common edge or vertex to be considered as a neighbor (as shown in the Figure 3).

.

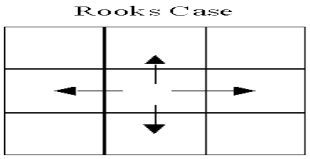


Figure 1

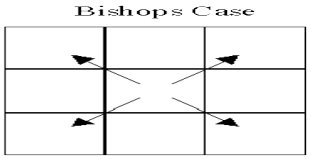


Figure 2

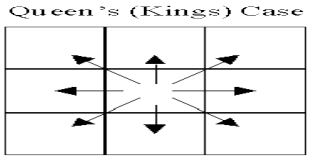


Figure 3

There is no universal type of constructing a spatial weight matrix that can be used in spatial data analysis. Constructing spatial weight matrix is obligatory to account for spatial dependence between different polygons and to know the distribution of events using exploratory spatial data analysis. In this study, Global Moran's *I*, Geary's *C* and local measures of spatial autocorrelation such as moran scatter plot and local Moran I are used to test for significance of spatial clustering.

3.4.1.5. Global Measures of Spatial Clustering

Moran's *I* global measures summarize spatial association with respect to the whole region. One of the Global measures of spatial autocorrelation is Moran's *I* given by:

$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} wij} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{wij(yi-\bar{y})(yj-\bar{y})}{\sum_{i=1}^{n}(yi-\bar{y})^2}$ ..... [1]

where n is the number of polygons (woredas in this study), wij is the element in the spatial weight matrix corresponding to the observation pair i, j or wij is 1 if the geographic areas associated to yi and yj are neighbours and 0 otherwise, and yi and yj are observations for areas i and j, with mean $\bar{y}$. The index *I* take on large values if there is a high correlation between neighboring values of the spatial variable i.e. if either large values or small ones (or both) are spatially clustered. The above index has a positive value in case of positive spatial autocorrelation, i.e. when the pairs of deviations from the mean for contiguous locations having the same sign are prevalent. In contrast, when the pairs of deviations from the mean have prevalently opposite sign the index has a negative value, therefore showing negative spatial autocorrelation. However, Moran's *I* indicates a departure from independent observations but doesn't tell where this departure occurs or even whether large or small values or both are affected since it is applied globally. The observed value of *I* can be compared to its distribution, Under the null hypothesis of no spatial autocorrelation or no clustering i.e. when the values of yi are independent of the values $yj(i \neq j)$ at neighboring locations .This is equivalent to say that under the reference null distribution, data are randomly distributed over locations. Therefore, Inference can be based on the standardized version of *I*.

The variance of Moran's *I* and Geary's *C* varies under normality and randomization assumption.

$Z(I) = \frac{I-E(I)}{\sqrt{Var(I)}}$ ..... [2]

With, $E(I)_N = E(I)_R = \frac{-1}{n-1}$

$Var(I)_N = \frac{n^2(n-1)s_1 - n(n-1)s_2 - 2{s_0}^2}{(n+1)(n-1){s_0}^2}$ .....Under normality assumption

$Var(I)_R = \frac{n(s_1(n^2-3n+3) - ns_2 + 3{s_o}^2}{(n-1)(n-2)(n-3){s_o}^2} - \frac{k(s_1(n^2-n) - 2ns_2 + 6{s_o}^2}{(n-1)(n-2)(n-3){s_o}^2} - \left(\frac{1}{n-1}\right)^2$ …Under randomization assumption.

$s_0 = \sum_{i\neq j}^{n} wij \, , s_1 = \sum_{i\neq j}^{n}(wij - wji)^2 \, , s_2 = \sum_{k=1}^{n}\left(\sum_{j=1}^{n} wkj + \sum_{i=1}^{n} wi\,k\right)^2$ where k represents districts.

Interpretation: We can use Moran index for identification of spatial distribution as dispersion, random or cluster patterns. Indices close to zero indicate the presence of random pattern. Indices Close to +1 indicate a tendency toward clustering. Besides the fact that Moran's *I* takes the usual form of autocorrelation, its distribution is well studied so that it can be used for testing the significance of spatial autocorrelation in neighboring plots or counties in a study area (Prince, 2010).

Geary's *C*

Geary's *C* interactions are not the cross product of the deviations from the mean, but the deviations in intensities of each observation location with one another**.** Geary's *C* is given by:-

$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} wij(yi-yj)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} wij\{(yi-\bar{y})^2\}} \text{……………………………………………………………………………. [3]}$$

where the notation is the same as in Equation 1. Usually the values of *C* range between 0 and 2. Values of *C* between 1 and 2 indicate presence of negative spatial autocorrelation while values between 0 and 1 indicate presence of positive spatial autocorrelation. Moran's *I* gives a more global indicator, whereas the Geary's coefficient is more sensitive to differences in small neighborhoods**.**

Testing the significance is done by using the standardized version of *C*, namely

$$Z(C) = \frac{C-E(C)}{\sqrt{Var(C)}} \text{……………………………………………………………………………. [4]}$$

$With\ E(C)_N = E(C)_R = 1\ \ , Var(C)_N = \frac{((2s_1+s_2)(n-1)-4{s_0}^2}{2(n+1)s_0}$

$Var(C)_R = \frac{s_1(n-1)\left(n^2-3n+3-k(n-1)\right)}{s_o n(n-2)(n-3)} - \frac{\left(n^2-3-k(n-1)^2\right)}{n(n-2)(n-3)} - \frac{(n-1)s_2\left(n^2+3n-6-k\left(n^2-n+2\right)\right)}{4n(n-2)(n-3){s_o}^2}$ ….. Under randomization assumption, where the notation are the same as [2] and E(C), Var(C) are the expectation and variance of C coefficients respectively.

Interpretation: The interpretation of Geary's C is analogous to that of Moran'sI. The only difference is that when C lies in (1, 2) indicates the presence of negative spatial autocorrelation (Malaria clustering of dissimilar values), whereas (0, 1) indicates positive spatial autocorrelation representing the presence of malaria clustering of similar values. Smaller p-values correspond to

stronger autocorrelation for both I and C statistics. Based on the preceding remarks, we have positive spatial autocorrelation when $ZI > 0$ or $ZC < 0$ and we have negative autocorrelation when $ZI < 0$ or $ZC > 0$

3.4.1.6. Local Indicators of Spatial Autocorrelation

While the strength of Moran's I lies in its simplicity, its major limitations is that it tends to average local variations in the strength of spatial autocorrelation. This has encouraged researchers to develop local indices of spatial association. This category of tools examines the local level of spatial autocorrelation in order to identify areas where values of the variable are both extreme and geographically homogeneous. This approach is useful when, in addition to global trends in the entire sample of observations, there exist also pockets of localities exhibiting homogeneous values that do not follow global trend. This leads to identification of hot spots, cold spots and clustering of dissimilar values. Anselin (1995) defines a LISA (local indicators of Spatial autocorrelation) statistic satisfying the following two conditions:

- The LISA for each observation measures the extent of sign, when positive spatial clustering of similar values around the observation and when negative spatial clustering of dissimilar values around observations.
- The sum of LISAs for all observation is proportional to a corresponding global indicator of spatial autocorrelation.

The local value of a LISA is computed as:

$$Ii = \frac{\sum_{j=1}^{n} wij(zi-\bar{z})(zj-\bar{z})}{(zi-\bar{z})^2} \quad \text{[5]}$$

From the proportionality condition we get**:**

$$\sum_{i=1}^{n} Ii = \gamma I \quad \text{[6]}$$

Where Ii is the LISA statistic for each observation, γ is a scale factor and *I*, is a corresponding global spatial autocorrelation measure. The sum of the LISAs is proportional to a global analog up to a scaling factor. These specific configurations can be first identified from a Moran scatter plot showing observed values against the averaged value of their neighbors. Once significance level is set, values can also be plotted on a map to display the specific locations of hot spots and potential outliers.

The Moran scatter plot is a useful visual tool for exploratory spatial analysis because it enables us to assess how similar an observed value is to its neighboring observations. Its horizontal axis is based on the values of the observations and is also known as the response axis. The vertical Y axis is based on the weighted average of the corresponding observation on the horizontal X axis**.** The Moran scatter plot provides a visual representation of spatial association (dependence) in the neighborhood around each observation. Depending on their position in the plot, the Moran scatter plot data points express the level of spatial association of each observation with its neighboring ones.

The Moran scatter plot can be divided into four quadrants. The top right and the bottom left quadrants contain observations showing positive spatial autocorrelation respectively with high-high and low-low data values indicating presence of clusters. The top left quadrant contains low values in a neighborhood of high values (low high), while the bottom right quadrant contains high values in a neighborhood of low values (high low). In both cases they are showing values of dissimilar clustering (Anselin, 1996).

4.4.2 Generalized Linear Model

Linear models are not the best approximation for modeling events related to the counting of cases of a certain disease, a Poisson distribution was assumed, which is part of the generalized linear models. Generalized Linear Model (GLM) is an extension of the linear modeling process that allows models to be fitted to data that follow probability distributions other than the Normal distribution. Generalized linear model helps to include response variables that follow any probability distribution in the exponential family of distributions. The exponential family of distributions includes Normal distribution, Binomial distribution, Poisson distribution, Multinomial distribution, Gamma distribution, Negative Binomial distribution and others.

Hypothesis tests applied to the Generalized Linear Model do not require normality of the response variable and do not require homogeneity of variances.

3.4.3. Poisson Regression Model

Hence our dependent variable in this study is count (malaria cases counts) Poisson regression model has been used. If the assumption of standard Poisson regression model that assumes the variances of malaria case counts equal with the mean of malaria case counts (randomness assumption) does not hold true. In such cases over dispersed Poisson regression and negative binomial regression models are more relevant**.** It is known that Poisson regression model is appropriate and is useful when the outcome is a count type with large counts of rare events (Mc Cullagh and Nelder, 1989).

In a Poisson regression model, observed counts y_i are assumed to have a Poisson distribution, with expected values depending on k predictor variables $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots\ldots\ldots \boldsymbol{x_k})^{\boldsymbol{T}}$ .The Poisson regression model aims at modeling a case count dependent variable y, which follows a Poisson distribution with a parameter μ_i. The probability that the number of cases takes the value on the entity is y_i, and is given by:-

$$f(Y = yi) = \frac{\mathrm{e}^{-\mu i}\mu i^{yi}}{yi!} \quad yi = 0,1,2,\ldots.. \qquad \mu i > 0 \qquad i = 1,2,\ldots..n$$

The systematic portion of the model involves the explanatory Variables $x_1, x_2, \ldots x_k$. Suppose that we want to let the mean μi and the variance depends on a vector of explanatory variables xi. Then, we can obtain a simple linear model of the form:

$$\mu i = xi'\beta$$

The above model has a disadvantage that the linear predictor on the right hand side can assume any real value; whereas the Poisson mean μi which represents an expected count has to be non negative straight forward solution to this problem is to model the logarithm of the mean using a linear model (Czado, 2008).Thus, the most common formulation of this model is the log linear model which can be written as:

$\log \mu i = xi'\beta$ …….. [a]

Which implies that μi is the exponential function of independent variables and can be expressed as: $\mu i = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_k x_k}$

The Poisson distribution depends on a single parameter μi. Although there is no theoretical upper bound for Poisson distribution, in practice these probabilities could be considered as negligible

when yi is very large. Thus, one can determine the magnitude of yi based on the values of μi (Bauer et al, 2007).

The major assumption of the Poisson regression model is

$$E(yi/xi) = \mu i = e^{xi'\beta} = \text{var (yi/xi)}$$

This assumption is based on the premise that successive events occur independently and at the same rate, which is implausible in many applications especially when events occur in clusters. In this model, the regression coefficient βi represents the expected change in the log of the mean per unit change in the predictor variable xi. In other words, increasing xi by one unit is associated with an increase of βi in the log of the mean. The exponentiated regression coefficient $e^{\beta i}$ represents multiplicative effect of the i^{th} predictor on the mean. Increasing xi by one unit multiplies the mean by a factor $e^{\beta i}$ **(**Bryan and Manfred, 2008**)**.

3.4.3.2. An over dispersed Poisson regression model

In the standard Poisson regression model the mean and variance of yi are assumed to be equal. If this assumption is violated. The standard Poisson regression would have been misspecified under the mean variance equality assumption (Sturman, 1999, Cameron and Trivedi, 1986). It is possible to account for over dispersion with respect to the Poisson model by introducing a dispersion parameter $\emptyset$ into the relationship between the variance and the mean and use alternative models such as over dispersed Poisson and Negative Binomial Regression models (Winkelmann, 2003).In an over dispersed model we assume that the variance is a linear function of the mean (McCullagh and Nelder, 1989):

$$\text{Var (yi)} = \emptyset(\mu i) \quad \text{[b]}$$

Where $\emptyset$ denotes the dispersion parameter. Equation [b] is known as the linear variance function since in it the variance of yi increases as a linear function of μi. McCullagh and Nelder (1989) suggested estimating the dispersion parameter $\emptyset$ as the ratio of Pearson's chi square to its associated degree of freedom. In other words, the over dispersion parameter $\emptyset$ of Equation[b] can be estimated by: $\widehat{\emptyset} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(yi-\hat{\mu}_i)^2}{\hat{\mu}_i}$

Where $\hat{\mu}_i = e^{xi'\beta}$ and $\hat{\beta}$ is the maximum likelihood estimator of β under the null hypothesis of over dispersed Poisson model. We call this Poisson regression model with correction to the estimated variance of the coefficients as over dispersed Poisson regression model.

3.4.3.3. Maximum Likelihood Estimation

Let yi be an observed count and assumed to be drawn from the Poisson distribution with a conditional mean μi on a given explanatory variable xi , for area i. Then, the density function of yi can be expressed as (McCullagh and Nelder, 1989)**:**

$$\mathrm{f(yi/xi)} = \frac{\mathrm{e}^{-\mu i}\mu i^{yi}}{yi!}$$

The model parameters $\mu i = e^{xi'\beta}, i = 1,2, \ldots . k$ (k explanatory variables) are estimated by the maximum likelihood method. The log likelihood is given by the equation:

$$\ln L = \sum_{i=1}^{n} [-\mu i + yi\, xi'\beta - \ln yi!]$$

$$\hat{\beta} = \sum_{i=1}^{n} (-\mu i + yi)xi$$

Once we obtain the parameter estimates, i.e. estimates of β, we can calculate the conditional mean $\hat{\mu} i = e^{xi'\hat{\beta}}$ which gives us the expected relative risk per areal unit. The 95% confidence interval of β is given by :($\widehat{\beta} \pm (z_{\alpha/2}) s.e(\hat{\beta})$))

Where is $z_{\alpha/2}$ a critical value on the standard normal distribution for a given predictor variable With a level of 95% confidence, we would say that we are 95% confident that upon repeated trials, the confidence interval would include the true population Poisson regression coefficient

3.4.3.4. Method of goodness of fit for over dispersed Poisson regression model

Having fitted a statistical model to the data, diagnostic tests are needed to assess the fit of the Model. Goodness of fit tests use the properties of a hypothesized distribution to assess whether or not observed data are generated from a given distribution (Read and Cressie, 1988). As we had mentioned, a major assumption in an over dispersed Poisson regression model is the variance of the count data are larger than the mean. Here the diagnostic tests are concerned with checking for this assumption. The most well known goodness of fit test of statistics used is the Deviance function and Pearson's chi square.

Pearson's Chi square Statistic.

One assumption of an over dispersed Poisson regression model is the variances are larger than the mean. In practical terms we assume proportionality, i.e. var (yi) $= \emptyset\mu i$ the statistical Hypothesis associated to this is:

H_0:$\emptyset = 1$ and H_1: $\emptyset > 1$

The appropriate test of statistic for this hypothesis is Pearson's chi square

$$\chi^2 = \sum_{i=1}^{n} \frac{(yi - \hat{\mu}_i)^2}{\widehat{\sigma\, i}}$$

Where yi the observed data, μi is the true mean from the model, and δi is the error and is usually represented by the standard deviation of yi. In large samples the distribution of Pearson's statistic is approximately chi squared with $n - p$ degrees of freedom where n the number of observations and p is the number of parameters. Several authors have proposed estimating $\emptyset$ using Pearson's chi squared statistic divided by its degree of freedom.

$$\hat{\phi} = \frac{\chi^2}{n - p}$$

Nelder (1989) interpreted the dispersion parameter in the analysis of parameter estimates as, when $\hat{\phi} = 1$, we have a standard Poisson regression model and when $\hat{\phi} > 1$, we have an over dispersed Poisson regression model. If there is no over dispersion that represents standard Poisson regression model, the Pearson's chi square will roughly equal with difference between the numbers of observations in the data set minus the number of parameters in the model.

Deviance function for testing over dispersed Poisson regression model

The goodness of fit of Poisson regression model can also be evaluated using the deviance function (McCullagh and Nelder, 1989). A measure of discrepancies between observed and fitted Values are the deviance. The deviance, denoted by G^2, is calculated as twice the difference between the log likelihood under the maximum model and the log likelihood under the reduced (Or unsaturated) model and is given by (Wood, 2002):

$$G^2 = 2\sum_{i=1}^{n}[yi \ \ln\left(\frac{yi}{\mu i}\right) - (\text{yi} - \ \mu i)]$$

The hypothesis is:

H_0:$\emptyset = 1$ (The fitted model is a standard Poisson regression model)

H_1: $\emptyset > 1$ (The fitted model is an over dispersed Poisson regression model)

For large samples the distribution of the deviance is approximately a chi squared with $n - p$ degrees of freedom. The goodness of fit tells us that if the ratio of the deviance to its degree of freedom is larger than 1 an over dispersed Poisson regression model is adequate or the data fits the model well.

3.4.4. Negative Binomial Regression Model

Negative binomial regression model, is another alternative method used to deal with over dispersion in count data (Bauer et al, 2007). The condition that the mean and variance must be equal is nullified in the presence of over dispersion in count data. Count data often exhibit over dispersion with variance larger than the mean. Another count model which allows for over dispersion is the negative binomial model. The negative binomial distribution can be derived from the Poisson when the mean parameter is not identical for all members of the population, but itself is gamma distributed. Unlike the Poisson distribution, the negative binomial adds a quadratic term to the variance in representing an over dispersion (Prince, 2010).

The probability density function of negative binomial random variable is given by**:**

$$\text{f (yi}; \mu i; \emptyset) = \frac{\Gamma(\phi + yi)}{\Gamma(\phi)\Gamma(yi + 1)} \left(\frac{\emptyset}{\mu i + \emptyset}\right)^{\emptyset} \left(\frac{\mu i}{\mu i + \emptyset}\right)^{yi}$$

The mean and variance of negative binomial regression model are given by

$$\text{E(yi}; \mu i; \emptyset) = \ \mu i \ \text{ And } \ \text{var (yi}; \mu i; \emptyset) = \ \mu i + \frac{\mu i^2}{\emptyset}$$

Where, $\mu i = e^{xi'\beta}$, $\Gamma(.)$ is a gamma function and $\emptyset$ is the dispersion parameter that must be estimated. Fitting negative binomial regression model is very similar with over dispersed Poisson regression model. That is, the log of the mean μi is a linear function of explanatory variables.

$$\log \ \mu i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon i$$

This implies that μi is an exponential function of independent variables and the denotation of the

Coefficients, explanatory variables and assumptions are the same as in over dispersed Poisson regression model. The relationship between the variance and the mean of negative binomial distribution is presented by

$$\text{Var (yi)} = \mu i + \phi^{-1}\mu i^2$$

The dispersion parameter is usually assumed to be fixed and can be estimated from observed data using the method of moments by using the maximum likelihood. For an over dispersed count variable, negative binomial regression model is better than Poisson Regression. If $\phi^{-1} \rightarrow 0$ in value (when ∅ gets large) we obtain the fitted model is a standard Poisson regression model. if $\phi^{-1} > 0$, then the variance is larger than the mean. Thus, the negative binomial distribution is over dispersed relative to the Poisson. One important characteristic of the negative binomial distribution is that it naturally accounts for over dispersion due to its variance always being greater **(** $\phi^{-1} > 0$) than the variance of a Poisson distribution with the same mean μi (Zhang et al, 2006).

3.4.4.1. Maximum Likelihood Estimation Method

As in Poisson regression model, the negative binomial regression coefficients $\beta_0, \beta_1, \beta_2, \ldots$.are estimated by maximum likelihood method. The estimation of the model parameters can be done by minimizing the log likelihood function. For the negative binomial distribution, the log likelihood is given by the equation (Bauer et al 2007).

$\ln L(\phi;\beta) = \sum_{i=1}^{n}\{\sum_{j=1}^{y-1} \ln\ (j+\phi) - \ln yi\,! - (yi+\phi)\ln\ (1+\phi^{-1}\mu i) + yi \ln\ \phi^{-1} +$

$yi\ \ln\ yi\}$..... [c], where $\mu i = e^{xi'\beta}$ from this log likelihood function it is possible to obtain the parameter estimates.

3.4.4.2. Test for Goodness of Fit

In the presence of over dispersion an alternative approach to model is negative binomial regression model. For negative binomial distribution the major assumption considered is

$\text{Var (yi)} = \mu i + \phi^{-1}\mu i^2$, $\quad\quad$ $\text{E (yi)} = \mu i$

Where $\phi^{-1} > 0$ indicates dispersion parameter. When $\phi^{-1} \rightarrow 0$, the negative binomial distribution reduces to Poisson. Thus, in testing an over dispersion the null hypothesis is defined as Ho: $\phi^{-1} \rightarrow 0$ and the alternative hypothesis is H_1: $\phi^{-1} > 0$

The null hypothesis defined indicates that the fitted model is a standard Poisson regression model. The appropriate test statistics are the deviance function and the Pearson's chi square as in an over dispersed Poisson regression model. A decision about whether the Poisson form is appropriate can be based on one of several statistics. Thus, Pearson's χ^2 statistic is given as

$$\chi^2(\mu i: n) = \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{[yi - \mu i]^2}{\mu i + \frac{\mu i^2}{\phi}}$$

If the value of the Pearson's chi square greatly in excess of $\chi^2(n-p)$ then the model is over dispersed due to non Poisson form. Thus, when Pearson's chi squared statistic is divided by Degree of freedom

$$\frac{\chi^2(\mu; n)}{n-p}$$

If it is larger than 1, this indicates over dispersion. Likewise, the deviance model based on the definition of the scaled deviance (Wood, 2002), the G^2statistic for a negative binomial model is given by

$$G^2(\mu i, \phi, n) = 2\sum_{i=1}^{n}[\,\phi \log\left(\frac{\mu i + \phi}{yi + \phi}\right) + yi \log\left(\frac{yi(\mu i + \emptyset)}{\mu i(yi + \emptyset)}\right)]$$

A value of the deviance greatly in excess of $\chi^2(n-p)$ suggests that the model is an over dispersed due to non Poisson form. Thus, when deviance is divided by degrees of freedom

$$\frac{G^2(\mu i, \phi, n)}{n-p}$$

is larger than one, this is an indication of over dispersion. Pearson's chi- square and deviance Statistic should be approximately chi-square distributed with $n-p$ degree of freedom.

Test of over dispersion

Deviance and Pearson's Chi-square divided by degree of freedom are used to detect over dispersion in Poisson regression. Values greater than 1 indicate over dispersion that is the true variance bigger than the mean. We can test the significance of over dispersion with a Likelihood Ratio Test which follows chi square ($\chi^2(1-2\alpha, 1d.f)$) distribution with 1 degree of freedom based on Poisson and Negative Binomial distributions. This test tests equality of the mean and

the variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean. For the negative binomial distribution, Var (yi) $= \mu i + \phi^{-1}\mu i^2$ the negative binomial distribution reduces to the Poisson when $\phi^{-1} \rightarrow 0$.

Therefore, in testing over dispersion the hypothesis is given by: Ho: $\phi^{-1} \rightarrow 0$ and H1: $\phi^{-1} > 0$

The Likelihood Ratio Test statistic for this hypothesis is given as:

$LR = -2[LL\ (Poisson) - LL\ (negative\ binomial)]$

Reject: H_0: if $LR > \chi^2(1 - 2\alpha, 1d.f))$

In all the above tests whenever over dispersion exists in the Poisson regression model (i.e.: H_0 *is* rejected), it is recommended to use negative binomial regression model.

3.4. 5 Methods of model selections

Statistical comparisons between Poisson and negative binomial regression models in most cases confirm that negative binomial regression model better represents observed counts that are over dispersed than Poisson regression model (Hausman et al. (1984)).Negative binomial regression is the extension of Poisson with modification in variance assumption and will be equal to Poisson regression when the dispersion parameter gets large or ϕ^{-1} is equal to zero. This important fact provides a possibility to make comparison between Poisson and negative binomial regression models. First of all, we can look at the value of chi square statistic of dispersion parameter to assess the significance of over dispersion. Then, a Likelihood ratio (LR) test, which follows chi square distribution with 1 degree of freedom, between two regressions can be used to determine the preferred model for the data. The likelihood ratio test is calculated as minus twice the difference between the log likelihood under the maximum model (Poisson) and the log likelihood under the reduced (or unsaturated) model (negative binomial):-

$$LR = -2(log\ Poisson - log\ Negative\ bin)$$

A large value of log likelihood indicates that the model is a preferable model. Generally, in Choosing between two models the information will be based on mainly AIC and BIC.

3.4.5.1. Akaike Information Criterion (AIC)

The AIC is another measure of fit that can be used to assess models. This measure also uses the Log likelihood, but adds a penalizing term associated with a number of variables. It is well known that by adding variables, one can improve the fit of models. Thus, the AIC tries to balance the Goodness of fit versus the inclusion of variables in the model. The AIC is given by $AIC = -2\ln L + 2\,p$, where p is the number of unknown parameters included in the model and $ln\ L$ is the log Likelihood described in Equation [d]. Smaller values of AIC indicate the best models to be selected.

3.4.5.2. Bayesian Information Criterion (BIC)

Similar to AIC, BIC also employs a penalty term associated with the number of parameters p and the sample size *n.* This measure is also known as the Schwarz Information Criterion and is computed as: $BIC = -2\ln L + p\ln n$, again, model which has smaller value of BIC is best fit model.

3.4.6 Generalized Linear Mixed Models

In situations where responses are both non-normal and correlated, we can use generalized linear mixed models for inference. As linear mixed models provide an extension for the linear model Generalized linear mixed models (GLMMs) provide an extension for the GLM (generalized linear model) by the addition of random factors. This extension is however not as seamless as the extension from the LM to the LMM (linear mixed model) by mismatch of distributions. As the name implies, GLMMs combine elements from both linear mixed models and generalized linear models. These models add a random effects component to the GLM, yielding the following model: $g(\mu) = X\beta + Zb$

Where μ is still the mean response through the link function, meaning E(y/b)= $g(\mu) = X\beta + Zb$ (Breslow and Clayton 1993). The random effects, b, are still assumed to be unobserved random variables that follow some distribution (typically a normal distribution) with unknown variance components.

The Generalized Linear Mixed Model (GLMM) is a type of model that is general enough to be used for modeling data from discrete as well as continuous distributions and to allow for different sources of variability in the mean response. The first feature is achieved by specifying

the mean of the response as a function (usually nonlinear) of some explanatory variables. The second feature is achieved by modeling the mean of the response as a function of random variables called the random effects.

When the response was normally distributed, there was no issue with estimation because we could use maximum likelihood to obtain estimates of the fixed effects and use the restricted likelihood for the variance components. Similar to GLMs, we cannot derive a closed form solutions for any estimates in the GLMM as the likelihood function still is nonlinear in the parameters. The addition of the random effects also complicates matters because with each random effect b, the dimensionality of the integral also increases, adding to the intractability of the likelihood function:

$$L(\beta, D) = \prod_i \int f(y_{ij}|b_i) f(b_i) db_i$$

Where $f(y_{ij}|b_i)$ is the probability distribution of the j^{th} response in the i^{th} random effect and $f(b_i)$ is the distribution of the i^{th} random effect. No closed form solution means that other methods of obtaining the likelihood function must be used, and this is where most of the study in GLMMs has been.

Broadly, inference for GLMMs can be conducted either through approximations of the likelihood or with Bayesian methods. Approximations of the likelihood either focus on approximating the integrand (the function being integrated) like penalized quasi-likelihood or to approximate the integral itself as in Laplace approximation or Gaussian-Hermite quadrature. Bayesian methods rely on approximating the posterior distribution of the parameters in the GLMM, which is usually accomplished with Markov chain Monte Carlo (MCMC) methods.

3.4.6.1 Methods of Estimating GLMMs

The theory and implementation of generalized linear mixed models are a fairly new compared to others like ANOVA or general linear regression, with the most development in computational methods for GLMMs occurring in the past twenty-five years. The first wide spread introduction to these models came from Mccullagh and Nelder (1989). Since then, a range of methods from both frequentist and Bayesian perspectives have been developed for inference on GLMMs.

Likelihood-based Approaches

Likelihood based (used interchangeably with frequentist) approaches to GLMMs rely on the likelihood; usually the log-likelihood for simplicity for use in estimation. The use of the full likelihood function with GLMMs is not possible as there are no closed form solutions. To side step this issue, approximations of the likelihood function need to be used. In practice, there are three approximation methods used: Laplace approximation, adaptive Gaussian Hermite Quadrature, and Penalized Quasi-Likelihood. These three likelihood-based methods are based on of Laplace's method for approximation.

Penalized Quasi-Likelihood

Penalized quasi-likelihood (PQL) was the first widely used method of estimating parameters for a GLMM, which works through approximating the likelihood integrand. Penalized quasi-likelihood was introduced by Breslow and Clayton (1993) as a modified Laplace approximation of the marginal quasi-likelihood, yielding a normally distributed approximate likelihood. The "penalized" in PQL refers to the random effects being biased towards zero in the approximation. To fit the model from the approximated likelihood, linear mixed models using REML are iteratively fit to the approximate likelihood until convergence is met.

Compared to the previous methods, PQL is the fastest and most flexible approach for estimating GLMMs. The number of random effects and their structure are not as restricted as in AGHQ and should have asymptotic properties as the sample size increase (Breslow and Clayton 1993). Since penalized quasi-likelihood is based on the quasi-likelihood, standard likelihood-based tests such as likelihood ratio tests cannot be performed (Tuerlinckx *et al*.,2006, Zhang *et al*.,2011).

Chapter Four

4. Results and Discussions

In this study, both generalized linear model and generalized linear mixed model were used to see the significance and type of relationship that exists between the dependent and independent variables after distribution pattern of malaria cases counts are identified using Global Moran's I and Geary's C statistics under both randomization and normality assumption. In addition to that, to visualize global spatial autocorrelation the researcher used Moran's scatter plot under the assumption of normality. Local Moran's I were used to identify local spatial clustering mainly to identify clustering of high values and low values. The explanatory variables included in this study are expected to have significant effect on the dependent variable (malaria case counts).

In figure 4.1.spatial distribution of malaria by percents in districts of Ilu Aba Bor zone was presented. In table 4.1 test of spatial autocorrelation (results of global moran's *I* and geary's *C* statistics under randomization and normality assumption**)** was presented. In figure 4.2 moran scatter plot for malaria in districts of Ilu Aba Bor Zone was presented. In figure 4.3 woreda under hot spot, woreda under cold spot and clustering of dissimilar value was presented.

In table 4.3 results of parameter estimates for Poisson regression model by MLE method of estimation was presented , In table 4.4 results of over-dispersion test after fitting a Poisson regression model was presented , In table 4.5 results of parameter estimates for negative binomial regression model by MLE method of estimation was presented, In table 4.6 results of goodness of fit for the negative binomial regression model was presented ,In table 4.7 results of test of over dispersion in Poisson model and negative binomial model was presented, In table 4.8 results of model comparison between Poisson and negative binomial regression model was presented, In table 4.9 parameter estimates for negative binomial regression model with spatial dependence by PQL method of estimation was presented.

4.1. Spatial Distribution of Malaria by Woreda

The total number of malaria cases in Ilu Aba Bor zone was 3005 in the year 2010 Ethiopian calendar. This malaria case count was collected from 14 woredas of this zone. Figure 4.1 presents the percentages of malaria case counts in each districts of Ilu Aba Bor zone, From these

map the light blue color reveals woredas accounted 1.9 to 4.3 percents of malaria case, where as the dark magenta color shows woredas accounted 6.7 to 10.9 percents of malaria case and the red color reveals woredas accounted >=12.7 percents of malaria. From total malaria case of 3005 sale Nono and Didu woreda accounted 1.9 percent and 2.09 percent of malaria case respectively .whereas Yayu, Doreni, Becho, Hurumu, Halu and Ale woredas accounted 4.03, 3.56, 4.3, 3.63, 3.76 and 4.02 percents of malaria case respectively. On the other hand Mettu town, Bilonopha ,Alge sachi ,Bure, Darimu woreda and Mettu woredas accounted 10.9,6.7,13.8,10.71,17.9 and 12.7 percents of malaria case respectively.

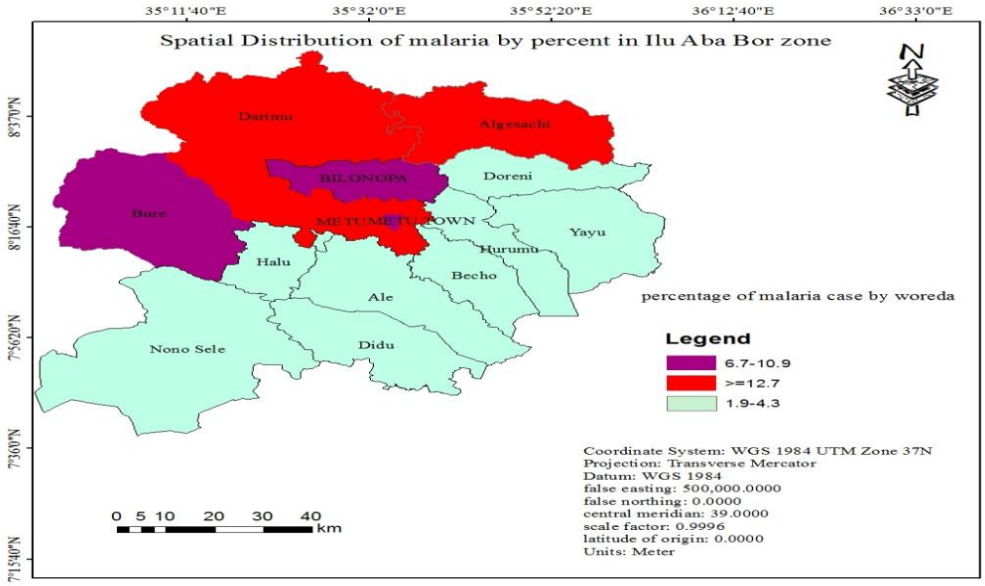


Figure 4.1: Spatial Distribution of Malaria by Woreda

4.2. Testing for Spatial Autocorrelation.

The Moran's *I* and Geary's *C* coefficient, both being among the most widely implemented measures of spatial autocorrelation between neighboring districts as briefly discussed in the methodology part in chapter 3, were used. In this section, our focus is on their application to particular data analysis. The spatial weight matrix, in Ilu Aba Bor Zone is presented in Appendix (Table 5.1). The element of the Contiguity weight matrix was calculated by using Queen's method because Queen's method considers common edge and/ or vertex in defining the spatial dependence.

Spatial autocorrelation analysis includes tests and visualization of both global test (moran's *I* and geary's *C*) and local test for clustering (local Moran's *I)* statistics. The global test is visualized by means of Moran scatter plot**,** in which the slope of the regression line corresponds to moran's *I.* Local analysis is based on the local Moran's *I* .First, the global moran's *I* and Geary's *C* test statistics were computed to test the null hypothesis (Ho:$\rho = 0$) of no significant clustering of malaria case counts in the entire study region ($\alpha = 0.05$).

4.2.1. Moran's *I* and Geary's *C* Test Statistics for Global Spatial Autocorrelation

The main objective of estimating spatial autocorrelation coefficient (global and local) is to measure the strength of spatial autocorrelation amongst neighboring Woreda of malaria case counts to seek for spatial pattern. Hence variance of global moran's I and geary's C vary under both normality and randomization assumption the tests are performed under both normality and randomization assumption .The null hypothesis states spatial independence (uncorrelated of error terms) for the data under consideration and the alternative hypothesis states spatial dependence (correlated of error terms). The estimated results of moran *I* and geary's *C* are also used for model specification.

The test results indicate the presence of significant global spatial autocorrelation of malaria case counts in Ilu Aba Bor zone (Table 4.1). The test results are also shown in moran's *I* scatter plot (Figure 4.2).These global results in the distribution of malaria need to be further explored using local spatial statistics.

Table 4.1 : Results of Global Moran's *I* and Geary's *C* Statistics under randomization and Normality assumption

Assumption	**Coefficient**	**Observed**	**Expected**	**Std Dev**	**Z**	**P**
Normality	Moran's *I*	0.33367024	-0.07692308	0.16597	2.4739	0.006683
Normality	Geary's *C*	0.60338506	1.00000000	0.19983	-1.9848	0.02358

Assumption	**Coefficient**	**Observed**	**Expected**	**Std Dev**	**Z**	**P**
Randomization	Moran's *I*	0.33367024	-0.07692308	0.16797	2.4444	0.007254
Randomization	Geary's *C*	0.60338506	1.00000000	0.1947	-2.0365	0.02085

Significant at 0.05 levels

Hypothesis test under the global measure of spatial autocorrelation is H_0 : no spatial autocorrelation ($H_0 : \rho = 0$) versus under the alternative hypothesisH_1:$\rho \neq 0$ (There is spatial autocorrelation (spatial dependence). The variance of moran's *I* and geary's *C* varies under the assumptions of normality and randomization.

Interpretation: Based on the P-values of Moran's *I* and Geary's *C* coefficients, we reject the null hypothesis of no spatial autocorrelation. Furthermore, the computed Z- statistic for moran's *I* is positive under both normality and randomization assumption and Geary's *C* is negative under both normality and randomization assumption indicating the existence of significant positive spatial autocorrelation (malaria case counts clustering of similar value).

In order to visualize global spatial autocorrelation moran's scatter plot under the assumptions of normality was used (Figure 4.2). It shows malaria case counts can be assumed to occur with unequal distribution at all cluster (woreda).

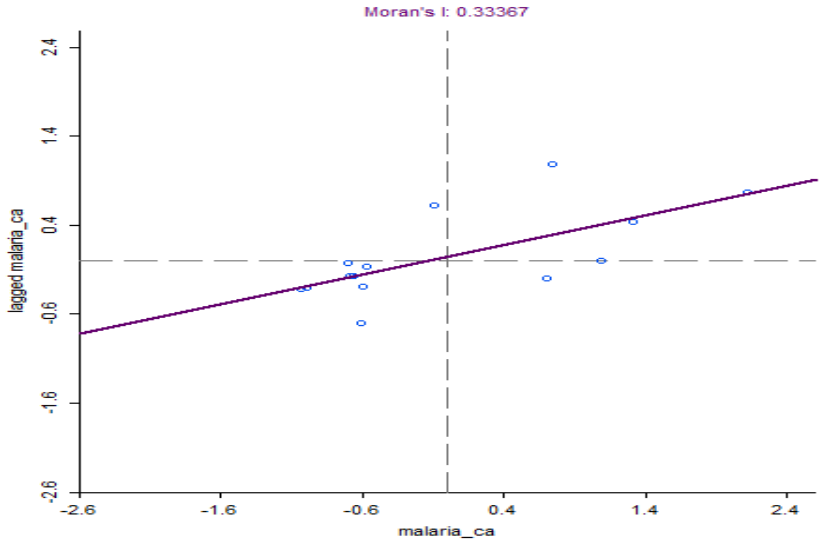


Figure 4.2: Moran Scatter Plot for Malaria case counts in Ilu Aba Bor Zone

From above figure 4.2 we conclude that malaria case counts in Ilu Aba bor zone of each woreda is spatially correlated with neighboring values. From the first quadrant (upper right) of Moran scatter plot we understand that in three woredas the distribution of malaria case counts are highly clustered. This result indicates that in these three woredas, there are high malaria case counts

clustering of similar values (hot spots). From the 3rd quadrant (lower left) we see that the distribution of malaria case counts in eight woredas is less clustered. This indicates that in these eight woredas the distribution of malaria case counts are cold spots (low low). On the other hand, as it is seen from the second and fourth quadrant (lower right and upper left) of the Moran scatter plot there is malaria case counts clustering of dissimilar values in three woreda(either high low or low high value).

However, identification of woredas for the presence of significant malaria case counts clustering is done based on local measures of spatial autocorrelation and depicted in Figure 3 as High-High, High-Low, Low-High and Low –low.

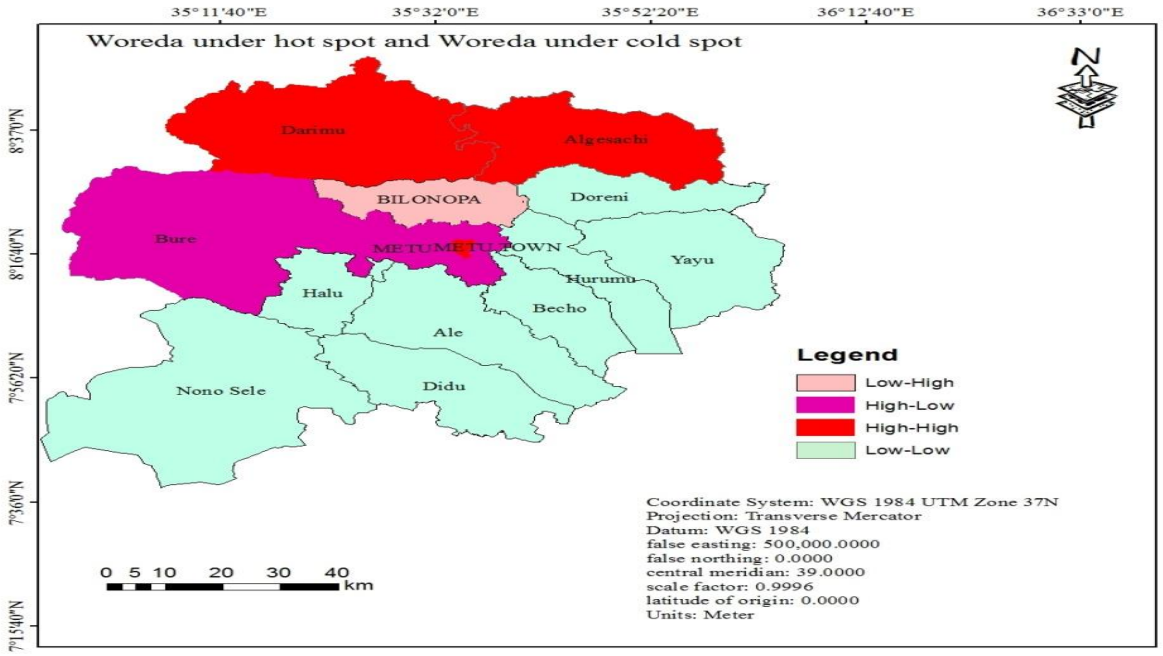


Figure 4.3: Significant malaria case clustering

In Figure 4.3, the red color indicates the presence of hot spots (high-high malaria case counts clustering) in three woredas as indicated in Moran scatter plot ,while the light blue color indicates the presence of cold spots (low-low malaria case counts clustering) in the eight woredas .The pink and dark magenta color indicates clustering of dissimilar value . The map reveals that in Darimu, Metu town and Alge sachi woredas there is malaria case counts

clustering of high high values (hotspots). In Nono sale,Didu,Ale,Halu,Becho,Yayu,Doreni and Hurumu there is malaria case counts clustering of low-low values (cold spots) is observed. On the other hand in Bilonopha there is malaria case counts clustering dissimilar values (low-high values) and in Bure and Mettu woreda there is malaria case counts clustering of high –low values. In general, a higher malaria case counts was observed in the northen and western part of the study area while the southern and eastern part has low malaria case counts.

Table 4.2: Results of Local Moran's *I* Test

ID	Woreda	Ii	E.Ii	Var.Ii	Z.Ii	Pr(z > 0)
1	Darimu	5.27514232	-0.23076923	2.278258	3.64777318	0.0001322615
2	Algesachi	1.86662747	-0.23076923	2.278258	1.38956599	0.0823303551
3	Yayu	0.92545938	-0.15384615	1.650027	0.84023129	0.2003893581
4	Metu Woreda	-0.05230678	-0.69230769	3.292673	0.35270068	0.3621564238
5	Ale	0.73705144	-0.30769231	2.775299	0.62712613	0.2652882802
6	Bure	-0.46227579	-0.23076923	2.278258	-0.15337758	0.5609497388
7	Nono Sale	1.06594689	-0.23076923	2.278258	0.85909958	0.1951427899
8	Becho	0.12416911	-0.23076923	2.278258	0.23515353	0.4070447849
9	Bilonopha	-0.30077218	-0.38461538	3.141151	0.04730678	0.4811343614
10	Hurumu	0.66188670	-0.38461538	3.141151	0.59046704	0.2774387888
11	Didu	1.29690263	-0.30769231	2.775299	0.96318683	0.1677268872
12	Halu	0.59914741	-0.38461538	3.141151	0.55506770	0.2894241778
13	Metu town	0.87379848	-0.07692308	0.890608	1.00741876	0.1568667902
14	Doreni	0.06769112	-0.30769231	2.775299	0.22533062	0.4108610396

In Mettu Woreda, Bure woreda and Bilonopha woreda the spatial correlation is negative because observed value is less than zero($Ii < 0$). This indicates that in these Woredas high value is surrounded by low values or low value is surrounded by high values of neighboring Woredas. The rest of the Woredas exhibit positive spatial correlation since observed value is greater than zero ($Ii > 0$) indicates a grouping of similar values (higher or lower value)

4.3. Model Based Data Analysis

Generalized linear model

When data is counts of events (or items) then a discrete distribution is more appropriate. In generalized linear model the coefficients are calculated using methods such as Maximum Likelihood Estimation (MLE) or Maximum quasi-likelihood but in this study the researcher used maximum likelihood. Generalized linear models (GLM) are used to fit fixed effect models to certain types of data that are not normally distributed and most commonly used to model binary or count data, so in this section the researcher focused on models for these types of data.

4.3.1. Poisson Regression Analysis

In order to fit Poisson regression model the researcher considers the offset variable to adjust the number of events and variation in population size. The offset is the special type of variable that was widely applicable when the observation was assumed to have Poisson distribution with the known slope of 1 that helps to adjust the problem due to variation in population size from one district to the other. The Poisson regression model says mean and variance of the response variable is equal if this assumption is violated there is over dispersion or under dispersion. The expected malaria case counts were modeled using Poisson regression and the results are presented in the following table.

Table 4.3: Parameter Estimates of Poisson Regression by MLE

Coefficients:	Estimate	Std. Error	Pr (>|z|)
Intercept	-51.902232	3.259061	< 2e-16 ***
Max.temp	0.537289	0.044709	< 2e-16 ***
Min.temp	0.229752	0.019972	< 2e-16 ***
Rainfall	0.035231	0.004464	2.95e-15 ***
Highland	0.125173	0.020446	9.23e-10 ***
Midland	0.234226	0.021565	< 2e-16 ***
Lowland	0.256246	0.021775	< 2e-16 ***

The residual deviance for the fitted Poisson regression was given as 168.64 on 7 degrees of freedom. The fitted Poisson model is given as

$$log\mu = -51.9 + 0.537x_1 + 0.23x_2 + 0.035x_3 + 0.1251x_4 + 0.2342x_5 + 0.256x_6$$

Where μ is the expected malaria case counts, x_1 stands for average annual maximum temperature, x_2 stands for average annual minimum temperature, x_3 stands for average annual

rainfall,x_4, stands for percentage of highland area,x_5,stands for percentage of midland area ,x_6,stands for percentage of lowland area .where average annual maximum temperature, average annual minimum temperature ,average annual rainfall, percentage of highland area, percentage of midland area and percentage of lowland area are statistically significantly affect malaria.

To check the fit of the fitted Poisson regression model, the value of the residual deviance 168.64 on 7 degrees of freedom, it was observed to be far greater than the number of degrees of freedom. This implies that the ratio$\frac{168.64}{7} = 24.09$ which is the dispersion parameter. The value 24.09 is far greater than one. Therefore it can be concluded that the model has lack of fit. If the mean and variance were equal, the residual deviance should be approximately equal to the degree of freedom for error. The assumption of mean equal to variance of the Poisson random variable hence was violated since the dispersion parameter was not approximately equal to 1, an indication of over dispersion in the data. This means that the parameters of the model had been over estimated and the standard errors had been under estimated which did not give a true reflection of the model that could provide appropriate expected malaria case counts. The fitted Poisson model had an AIC value of 280.25 and a null deviance of 602.75 on 13 degrees of freedom. The variance is much greater than the mean, which suggests that over-dispersion in the model.

Table 4.4: The results of over-dispersion test after fitting a Poisson regression model

Statistics	Value	Degree of freedom	$\frac{\text{Value}}{degree\ of\ freedom}$	P-value
Deviance	**168.64**	**7**	**24.01**	**0**
Pearson chi-square	**189.1285**	**7**	**27.01**	**0**

The test of over-dispersion, Deviance statistics and Pearson Chi-square Statistic divided by their corresponding degrees of freedom, are greater than one indicating over-dispersion. As mentioned in chapter three, both Poisson and negative binomial regression models were used for analyzing malaria case counts. Even though both models are recommended when cases are over dispersed, they are different in the assumption about mean variance equality.

4.3.2. Negative Binomial Regression Estimator

Table 4.5: Parameter Estimates for Negative Binomial Regression Model by MLE.

Coefficients	Estimate	Std. Error	Pr(>|z|)
Intercept	-56.33647	8.60905	5.99e-11 ***
Max.temp	0.68621	0.14026	9.96e-07 ***
Min.temp	0.17839	0.06861	0.009324 **
Rainfall	0.05717	0.01555	0.000235 ***
Highland	0.09307	0.05389	0.084179
Midland	0.19444	0.05405	0.000321 ***
Lowland	0.24605	0.05498	7.62e-06 ***

An alternative method for analyzing the relationship between predictor variables and response variable is the negative binomial regression model. The result obtained from this model indicates that average annual maximum temperature, average annual minimum temperature, average annual rainfall, percentage of midland area and percentage of lowland area are significantly affect malaria at $\alpha = 0.05$. But percentage of highland area is insignificant at alpha=0.05

Accordingly, the fitted model is given by:

$$log\mu = -56.34 + 0.68x_1 + 0.178x_2 + 0.057x_3 + 0.194x_4 + 0.246x_5$$

Where μ indicates malaria case counts per woreda. Where x_1 stands for average annual maximum temperature, x_2 stands for average annual minimum temperature**,** x_3 stands for average annual rainfall andx_4,x_5 represent percentage of midland area, percentage of lowland area respectively.

4.3.2.1. Results of goodness of fit for the negative binomial regression model

We can test the adequacy of this model and the significance of the predictor variables using deviances and Pearson's chi square test statistics. The results of the tests are presented in the following table and these indicate that the fitted model is good.

Table 4.6: Results of goodness of fit for the negative binomial regression model

Statistics	Value	Degree of freedom	$\frac{\text{Value}}{degree\ of\ freedom}$
Deviance	13.546	7	1.94
Pearson chi-square	13.39447	7	1.913

Specifically, the values of the deviance and Pearson's chi square are nearly the same (13.54 and 13.39 respectively). Furthermore, the ratio of values of deviance's and Pearson's chi square to their corresponding degree of freedom are larger than one.

4.3.2.2. Results of test of over dispersion in Poisson model and negative binomial model

For the Poisson model, the Pearson Chi-square and deviance values divided by the degrees of freedom are sufficiently larger than 1. This is a possible indication that the fit is over dispersed However, this is also justified by applying a formal statistical test of dispersion.

Table 4.7: Results of test of over dispersion in Poisson model and negative binomial model

Criteria	Estimate	Poisson model	Negative binomial model
Deviance	$Value/\ d.f$	24.01	1.94
Pearson's Chi-square	$Value/d.f$	27.01	1.913
Log likelihood	$Value$	-133.13	-71.18

To carry out the LR test for significance of over dispersion, that is to test the hypothesis:

$$H_0: \emptyset = 1 \quad \text{And the } H_1: \emptyset > 1$$

The result obtained from -2(LL (Poisson) – LL (negative binomial) is 123.9, which corresponds to P-value = 0. Hence, we reject H_0 and conclude that the mean and variance are not equal as a result the assumption of standard Poisson regression model is violated. This clearly, shows that presence of significant over dispersion. It was observed that the Negative Binomial regression model was the best model which fit the expected malaria case because the ratio of deviance to degree of freedom and Pearson chi-square to degree of freedom given by Poisson Regression Model had been reduced from 24.01 to 1.94 and 27.01 to 1.913 respectively.

4.3.2.3. Comparisons of Poisson and negative binomial regression models

Three different methods are used to compare Poisson and negative binomial regression models those methods are log likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC). The results obtained indicate there is observed difference in values between the two models. Since negative binomial regression model has smaller value in AIC and BIC. Consequently, we conclude that in this study negative binomial regression model is better than Poisson for modeling an over dispersed data.

Table 4.8: Results of comparison of Poisson and negative binomial regression models

Methods of comparisons	Poisson regression model	Negative binomial regression Model
Log Likelihood Ratio test	-133.13	-71.18
AIC	280.2516	158.37
BIC	284.7250	163.4831

In the selection and testing of goodness of fit of the estimated parameters, the Log likelihood statistics, the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) are used to choose the best model out of the candidate models. The high values of log-likelihood statistics and minimum AIC and BIC values chooses the Negative Binomial distribution. This shows that the unknown true model fits the Negative Binomial Distribution for all malaria case counts data. We can, therefore, argue that the malaria case count data for all woreda(cluster) in Ilu Aba Bor zone which follow a Negative Binomial Distribution. This means that malaria case counts data on the various woreda are better modeled by negative binomial.

4.3.2.4. Accounting spatial dependence to the model

The above fitted model is by ignoring spatial dependence between observations and used to fit fixed effect only. Hence from explanatory spatial data analyses there is positive spatial autocorrelation. Fitting Poisson and negative binomial regression model by ignoring spatial dependence under estimate the standard error due to this reason to account spatial dependence the researcher extend application of generalized linear model to generalized linear mixed model by adding random effect and correlation structure. In generalized linear model the researcher used maximum likelihood method of estimation but in generalized linear mixed model the researcher used penalized quasi likelihood method of estimation.

It is important to remember that directly comparison of the parameters estimated by the generalized linear model and generalized linear mixed model could lead to erroneous conclusions, as the estimation methods are fundamentally different. Generalized linear mixed model are accessible, flexible methods that enable to account for residual spatial correlation.

Generalized linear models focus on analyzing data under the independence assumption. Often, data from diverse areas such as climatology, ecology, environmental monitoring, and health science are spatially correlated. This implies that the dependence structure underlying the data is some function of location information. Typically, observations located closer together will be more similar than those farther apart and thus the data exhibit spatial variability. It is well established that ignoring spatial dependence in the data when working with regression models will result in biased estimates of variation and inefficient statistical inference (Cressie, 1993). Therefore, in order to accurately assess the association between response and covariates, it is important to allow for spatial dependence when developing regression models for data that is spatially correlated.

Spatial count data are more challenging for statistical modeling. This type of data commonly presents the problems of over dispersion and spatial effects (spatial autocorrelation and heterogeneity) (Haining et al., 2009).

If the over dispersion in the count data is ignored, it results in underestimating the standard errors of regression coefficients and leading to biased hypothesis testing (Myers, 2002). In the above out put the best fit model is negative binomial but there is over dispersion in the data because residual deviance divided by degree of freedom is greater than one. Spatial autocorrelation is concerned with the dependence of a response variable among adjacent neighboring units in a geographical area. Thus, one of the model assumptions, i.e., spatial dependence is violated, which has a strong impact on the standard error estimation of regression coefficients. The impact is complicated depending on the sign and magnitude of spatial autocorrelation. Positive spatial autocorrelation causes the underestimation of the standard errors, while negative spatial autocorrelation leads to their overestimation (Anselin, 1996).

4.3.2.4.1. Model fitted by accounting spatial dependence.

The researcher account spatial dependence to negative binomial regression model because when we used penalized quasi likelihood method of estimation it is difficult for model comparison hence it doesn’t give AIC, BIC and loglik

Table 4.9: Parameter Estimates for Negative Binomial Regression Model with spatial dependence by PQL method of estimation.

Coefficients	Estimate	Std.Err	95% C.I
Intercept	-56.3364	11.908737	(-84.5, -28.18)
Max.temp	0.68621	0.194021	(0.227, 1.145)
Min.tem	0.17839	0.094914	(-0.046, 0.403)
Rainfall	0.05717	0.021505	(0.006, 0.108)
Highland	0.09307	0.074552	(-0.08, 0.269)
Midland	0.19444	0.074762	(0.017, 0.371)
Lowland	0.24605	0.076048	(0.066, 0.426)

An alternative method for analyzing the relationship between predictor variables and response variable is by accounting spatial dependence to negative binomial regression model. The result shown in above table indicates the presence of positive relationship between dependent and explanatory variables. The values of the coefficients suggest positive relationship between average annual maximum temperature, average annual rainfall, midland area, lowland area and malaria case counts. The result obtained from this model also indicates that average annual maximum temperature, average annual rainfall, midland area and lowland area are significantly affect malaria distribution because their confidence interval does not include zero, but average annual minimum temperature and percentage of highland area are insignificant because their confidence interval include zero.

Accordingly, the fitted model is given by:

$$log\mu = -56.33 + 0.686x_1 + 0.057x_2 + 0.194x_3 + 0.2461x_4$$

Where x_1 Stands for average annual maximum temperature, x_2 Stands for average annual rainfall, x_3 stands for midland area, x_4 stands for lowland area. In the above model, the distribution of malaria case counts within a woreda(cluster) is statistically significantly associated with maximum temperature, rainfall, lowland area and mid-land area of neighboring woredas.

4.4. Discussions

From explanatory spatial data analysis both global and local measure of spatial autocorrelation is used and spatial weight matrix is constructed depending on queen methods. The main goal of estimating spatial autocorrelation coefficient (global and local) is to measure the strength of spatial autocorrelation amongst neighboring cluster of malaria case counts and to find for spatial pattern. The global tests are performed under the assumption of normality and randomization, the null hypothesis tested states spatial independence (uncorrelated of error terms) and alternative hypothesis tested states spatial dependence (correlated of error terms) for the data under consideration. Based on the P-values of Moran's *I* and Geary's *C* coefficients, we reject the null hypothesis of no spatial autocorrelation and accept alternative hypothesis of there is spatial autocorrelation. Furthermore, the computed Z- statistic for Moran's *I* is positive under both Normality and Randomization assumption and Geary's *C* is negative under both Normality and Randomization assumption indicating the existence of significant positive spatial autocorrelation (malaria case counts clustering of similar values).

The result of local measures of spatial autocorrelation showed that malaria case counts in Ilu Aba bor zone of each woreda is spatially correlated with neighboring woreda(cluster). The result of local moran'I shows positive spatial correlation in 11 woreda (observed is greater than zero) and negative spatial autocorrelation in three woreda(observed is less tha zero). The first quadrant (upper right) of Moran scatter plot showed in three woredas(cluster) the distribution of malaria case counts are highly clustered those woreda are Darimu, Metu town and Algesachi . This result indicates that in these three woredas, there are high malaria case counts clustering of similar values (hot spots). The 3rd quadrant (lower left) of Moran scatter plot showed that the distribution of malaria case counts in eight woreda is less clustered those woreda are Nono sale,Didu, Ale,Halu,Becho,Doreni,Yayu and Hurumu there is malaria case counts clustering of low-low values (cold spots) is observed. On the other hand the upper left and lower right of moran scatter plot showed that the distribution of malaria case counts in three woreda of Ilu Aba bor zone is clustering of dissimilar value those woreda are Metu woreda,Bilonopha woreda and Bure. In Bilo nopha woreda there is malaria case counts clustering (low-high values) and in Bure woreda and Metu Woreda there is malaria case counts clustering of high –low values.

Generally measures of spatial autocorrelations shows malaria case counts in Ilu Aba bor zone of each woreda(cluster) is spatially correlated with neighboring woreda(cluster),Where high-high (hot spot) value grouping occurred and another one at the opposite extreme with low-low (cold spot) values occurred. Previous studies are also consistent with this result that malaria is spatially correlated with neighboring woreda(cluster),Where high (hot spot) value grouping occurred and another one at the opposite extreme with low (cold spot) values grouping occurred(Hurtado *et al*., 2018; Akinbobola *et al***.,** 2018; Tewara *et al*., 2018, Mosissa, M., & Gotu, B, 2011).

For model based data analysis the Generalized Linear Model (GLM) is used because generalized linear model allows us to model responses with other distributions rather than the Normal distribution. Hence our data is counts of events (or items) then a discrete distribution is more appropriate by considering offset variable to adjust event and population size. From exponential family of distribution the researcher used Poisson and negative binomial distribution because Poisson and negative binomial distribution is used to model such types of data (malaria case counts data).

The result of fitted Poisson model showed that residual deviance is 168.64 with 7 degrees of freedom; it is far greater than the number of degrees of freedom. This implies that the ratio$\frac{168.64}{7} = 24.09$ which is far greater than one. Therefore it can be concluded that the model has lack of fit. If the mean and variance were equal, the residual deviance should be approximately equal to the degree of freedom for error. The assumption of mean equal to variance of the Poisson random variable was violated since the dispersion parameter was not approximately equal to one, an indication of over dispersion in the model. If there is over dispersion in the model coefficients of the model is over estimated and standard error of the model is under estimated. The test of over-dispersion, Deviance statistics and Pearson Chi-square Statistic divided by their corresponding degrees of freedom, are greater than one indicating over-dispersion. Hence there is over dispersion in the model negative binomial regression models were used for analyzing malaria case counts with covariate. Count data often exhibit over dispersion with variance larger than the mean. Another count model which allows for over dispersion is the negative binomial model. The negative binomial distribution can be derived from the Poisson when the mean parameter is not identical for all members of the population, but

itself is gamma distributed. Unlike the Poisson distribution, the negative binomial adds a quadratic term to the variance in representing an over dispersion.

Results of model comparison between Poisson and negative binomial regression model choose negative binomial because the high values of log-likelihood statistics and minimum AIC and BIC values chooses the negative binomial regression model. This shows that the unknown true model fits the negative binomial distribution for all malaria case counts data. We can, therefore, argue that the malaria case count data for all woreda in Ilu Aba Bor zone which follow a Negative Binomial Distribution, the closest to the true model. This means that malaria case count data on the various woreda are better modeled by negative binomial and covariate considered in this model have positive impact on malaria case counts in Ilu Aba Bor zone.

Ignoring spatial dependence between observations is it own impact on our results. Hence from explanatory spatial data analyses there is positive spatial autocorrelation. Fitting Poisson and negative binomial regression model by ignoring spatial dependence under estimate the standard error due to this reason to account spatial dependence the researcher extend application from generalized linear model to generalized linear mixed model by adding random effect and correlation structure. In generalized linear model the researcher used maximum likelihood method of estimation but in generalized linear mixed model the researcher used penalized quasi likelihood method of estimation.

It is important to remember that directly comparison of the parameters estimated by the generalized linear model and generalized linear mixed model could lead to erroneous conclusions, as the estimation methods are fundamentally different. Generalized linear mixed model are accessible, flexible methods that enable to account for residual spatial correlation. Generalized linear models focus on analyzing data under the independence assumption. Often, data from diverse areas such as climatology, ecology, environmental monitoring, and health science are spatially correlated. This implies that the dependence structure underlying the data is some function of location information. Typically, observations located closer together will be more similar than those farther apart and thus the data exhibit spatial variability. It is well established that ignoring spatial dependence in the data when working with regression models

will result in biased estimates of variation and inefficient statistical inference (Cressie, 1993). Therefore, in order to accurately assess the association between response and covariates, it is important to allow for spatial dependence when developing regression models for data that is spatially correlated.

In generalized linear model the best fit model is negative binomial we account spatial dependence to negative binomial and the result of this model showed that there is positive relationship between average annual maximum temperature, average annual rainfall, midland area and lowland area i.e average annual maximum temperature, average annual rainfall, midland area and lowland area are effect to increase malaria, but average annual minimum temperature and percentage of highland area are insignificant. Previous studies are also consistent with this result that malaria is statistically significant associated with average annual maximum temperature and rainfall (Aulia *et al.*, 2018; Kumar *et al*., 2014; Gemechu *et al*., 2015; Hannah. 2016; Athuman, A., 2013; Simple *et al*., 2018).On the other hand malaria transmission is significantly affected by lowland area and midland area this finding is consistent with Mosissa, M., & Gotu, B, 2011.

Chapter Five

5. Conclusions and Recommendations

This study describes the spatial pattern of malaria in Ilu Aba Bor Zone, southwest Ethiopia using routinely collected individual patient morbidity data from health care facilities and model malaria case counts with associated covariates.

5.1. Conclusions

The results of this study showed that malaria case counts in Ilu Aba Bor Zone exhibits a spatial pattern which is dependent on some meteorological variables and altitude (i.e. midland and lowland area). Malaria case counts in the study area is significantly clustered indicating high levels in the Northern part of the zone and low levels in southern and eastern part of the zone, clustering of dissimilar value in western part of this zone.

Geographical clusters of malaria case counts were identified through exploratory spatial data analysis, using Global Moran's *I*, Geary's *C* and also local indicators. The results obtained reveal that the distribution of malaria case counts in Ilu Aba Bor Zone is clustered. Furthermore, the Moran scatter plot also depicts the presence of spatial clustering of high values in three woredas (hot spots) of which in Darimu, Mettu town and Alge Sachi woredas are malaria case counts clustering of high values (hot spots) was observed. Malaria case counts clustering of low values were observed in eight woredas those are Nono sale, Didu,Ale,Halu,Becho,Yayu,Doreni and Hurumu, woredas from among the woredas shown in Moran scatter plot. On the other hand, in Bure Woreda ,Mettu Woreda and Bilonopha Woreda there was malaria case counts clustering of dissimilar values. For count data with the evidence of over dispersion, negative binomial regression model is preferred rather than Poisson regression model. To account spatial dependence generalized linear mixed model is better than generalized linear model.

Based on negative binomial regression model with spatial dependence average annual maximum temperature, average annual rainfall, midland and lowland area were significantly associated with malaria distribution .i.e. maximum temperature, rainfall; midland area and lowland area are effect to increase malaria distribution, but minimum temperature and highland area are insignificant.

5.2. Recommendations

This paper has endeavored to analyze the spatial distribution of malaria in the Ilu Aba Bor Zone .The results of these study shows that in the study area malaria case counts pattern varies from woreda to woreda and is clustered. The presence of spatial dependence between woredas was also established. Based on the results obtained, the study recommends that interventions should be facilitated in highly clustered malaria distribution areas by giving special attention in targeting intervention and health services to the highly risk exposed woredas and neighboring Woredas.

Based on the results obtained the following recommendations can also be made.

- Interested researchers are recommended to include clinical and lab diagnostic related variables, and extend this work in order to identify covariates associated with malaria distribution in Ilu Aba Bor Zone.
- Interested researchers are recommended to extend this work to all regions of Ethiopia to find spatial distribution of malaria Ethiopia.
- Expand and strengthen medical facilities especially in woredas identified as hot spots.
- Interested researchers are recommended to extend this work from spatial distribution of malaria to spatiotemporal distribution of malaria in Ilu Aba Bor Zone.

Reference

Adeola, A., Botai, J., Rautenbach, H., Adisa, O., Ncongwane, K., Botai, C., & Adebayo-Ojo, T. (2017). Climatic variables and malaria morbidity in Mutale Local Municipality, South Africa: A 19-year data analysis. *International journal of environmental research and public health*, *14*(11), 1360.

Akinbobola, A., & Ikiroma, I. A. (2018). Determining Malaria Hotspot Using Climatic Variables and Geospatial Technique in Central Urban Area of Ibadan, Southwest, Nigeria. *J Climatol Weather Forecasting*, *6*(225), 2.

Alelign, A., & Dejene, T. (2016). Current status of malaria in Ethiopia: evaluation of the burden, factors for transmission and prevention methods. *Acta Parasitologica Globalis*, *7*(1), 01-6.

Alemu, A., Abebe, G., Tsegaye, W., & Golassa, L. (2011). Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia. *Parasites & vectors*, *4*(1), 30.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis*, *27*(2), 93-115.

Anselin, L. (1996). The Moran Scatter plot as an ESDA (explanatory spatial data analysis) Tool to assess local instability in spatial association. *Spatial Analytical Perspectives on GIS 28,111-125.*

Anselin, L. and Hudak, J. (1998). Spatial econometrics in practice and reviews of soft ware options. *Regional Science and Urban Economics 22, 509-536.*

Arab, A., Jackson, M. C., & Kongoli, C. (2014). Modelling the effects of weather and climate on malaria distributions in West Africa. *Malaria journal*, *13*(1), 126.

Athuman, A. (2013). *Relationship Between Malaria Cases AND Weather Parameters Over Morogoro Municipality* (doctoral dissertation, university of nairobi).

Aulia, D., Ayu, S. F., & Matondang, A. (2018, January). Analysis of forecasting malaria case with climatic factors as predictor in Mandailing Natal Regency: a time series study. In *IOP Conference Series: Materials Science and Engineering* (Vol. 300, No. 1, p. 012035). IOP Publishing.

Ayele, D. G., Zewotir, T. T., & Mwambi, H. G. (2013). Spatial distribution of malaria problem in three regions of Ethiopia. *Malaria journal*, *12*(1), 207.

Ayele, D. G., Zewotir, T., & Mwambi, H. (2016). Spatial analysis of malaria on The Geo-Additive Bayesian Model. *Cancer Prog Diagn*, *1*(106), 27-32.

Bauer .L., Greibe, P., Hua, L. and Liang, L. (2007). Statistical Models of Accidents on interchange ramps and speed change lines, FHWA- RD – 97-106, U.S. Department of Transportation.

Berga, B. Spatial distribution of malaria indicators in Ethiopia in 2009.

Boateng, A., Lesaoana, M., Darikwa, T., Belete, A., & Siweya, H. Analysis of Malaria Incidence using Quasi-Poisson Regression Model: Evidence from Obuasi Municipality, Ghana.

Bryan, T. and Manfred, M. (2008).Regression Models for Count Data. *Journal of Pediatric Psychology 33(10), 1076-1034*

Cameron, A. and Trivedi, K. (1986)**.**Econometric models based on the count data, comparisons and applications of some estimators and tests. *Journal of Applied Econometrics 1, 29-53***.**

Chuang, T. W., Soble, A., Ntshalintshali, N., Mkhonta, N., Seyama, E., Mthethwa, S., ... & Kunene, S. (2017). Assessment of climate-driven variations in malaria incidence in Swaziland: toward malaria elimination. *Malaria journal*, *16*(1), 232.

Cliff, A. and Ord, J. (1981).Spatial processes, Modeling and application. Pion, London.

Coulibaly, D., Rebaudet, S., Travassos, M., Tolo, Y., Laurens, M., Kone, A. K., ... & Daou, M. (2013). Spatio-temporal analysis of malaria within a transmission season in Bandiagara, Mali. *Malaria journal*, *12*(1), 82.

Cressie, N. (1993). Statistics for spatial data, Wiley, New York.

Czado, C. (2008).Modeling Count Data by spatial random effects. *Statistical papers49(3), 531-552.*

DePina, A. J., Andrade, A. J. B., Dia, A. K., Moreira, A. L., Furtado, U. D., Baptista, H., ... & Seck, I. (2019). Spatiotemporal characterisation and risk factor analysis of malaria outbreak in Cabo Verde in 2017. *Tropical medicine and health*, *47*(1), 3.

Deressa, W., Chibsa, S., & Olana, D. (2004). The distribution and magnitude of malarian in Oromia, Ethiopia. *Ethiopian Journal of Health Development*, *18*(3), 163-170.

Dessie, D. B. (2017). Spatial modelling of malaria prevalence and its risk factors in rural SNNPR, Ethiopia: classical and bayesian approaches. *American Journal of Theoretical and Applied Statistics*, *6*(6), 254.

Drakeley, C. J., Carneiro, I., Reyburn, H., Malima, R., Lusingu, J. P., Cox, J., ... & Riley, E. M. (2005). Altitude-dependent and-independent variations in Plasmodium falciparum prevalence in northeastern Tanzania. *Journal of Infectious Diseases*, *191*(10), 1589-1598.

Eniyew, S. (2018). Modelling of malaria hotspot sites using geospatial technology in the north-western highlands of Ethiopia.

Eunice, A. (2018). *Statistical Modeling of Malaria Incidences in Apac District, Uganda* (Doctoral dissertation, JKUAT).

Gemechu, T., Samuel, A., & Yewhalaw, D. (2015). Ten years trend analysis of malaria prevalence and its correlation with climatic variables in Sibu Sire District, East Wollega Zone, Oromia Regional state, Western Ethiopia: A retrospective study. *Science, Technology and Arts Research Journal*, *4*(4), 99-105.

Girum, T., Shumbej, T., & Shewangizaw, M. (2019). Burden of malaria in Ethiopia, 2000-2016: findings from the Global Health Estimates 2016. *Tropical Diseases, Travel Medicine and Vaccines*, *5*(1), 11.

Goodchild, M. F. (1987). A spatial analytical perspective on geographical information systems. *International journal of geographical information system*, *1*(4), 327-334.

Goovaerts ,P. (1997). Geostatistics for Natural Resources Evaluation Oxford University Press, New York.

Graves, P. M., Richards, F. O., Ngondi, J., Emerson, P. M., Shargie, E. B., Endeshaw, T., ... & Zerihun, M. (2009). Individual, household and environmental risk factors for malaria infection in Amhara, Oromia and SNNP regions of Ethiopia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *103*(12), 1211-1220.

Haining, R., Law, J., & Griffith, D. (2009). Modelling small area counts in the presence of over dispersion and spatial autocorrelation. *Comput. Stat. Data. Anal 53, 2923–2937.*

Hasyim, H., Nursafingi, A., Haque, U., Montag, D., Groneberg, D. A., Dhimal, M., ... & Müller, R. (2018). Spatial modelling of malaria cases associated with environmental factors in South Sumatra, Indonesia. *Malaria journal*, *17*(1), 87.

Hausman, J., Hall, B. and Griliches, Z. (1984)**.** Econometric models for count data with an application to the patents-R and D relationship. *Econometrica 52, 909-938*.

Hiscox, A., Otieno, B., Kibet, A., Mweresa, C. K., Omusula, P., Geier, M., ... & Takken, W. (2014). Development and optimization of the Suna trap as a tool for mosquito monitoring and control. *Malaria journal*, *13*(1), 257.

Hurtado, L., Aguilar, M., & Ávila, M. (2018). Analysis of spatial distribution of malaria in the department of Chocó for the year 2016. *Ingeniería y competitividad*, *20*(2), 57-68.

Hussien, H. H. (2019). Malaria's association with climatic variables and an epidemic early warning system using historical data from Gezira State, Sudan. *Heliyon*, *5*(3), e01375.

Ihantamalala, F. A., Rakotoarimanana, F. M., Ramiadantsoa, T., Rakotondramanga, J. M., Pennober, G., Rakotomanana, F., ... & Wesolowski, A. (2018). Spatial and temporal dynamics of malaria in Madagascar. *Malaria journal*, *17*(1), 58.

Kazembe, L. N. (2007). Spatial modelling and risk factors of malaria incidence in northern Malawi. *Acta Tropica*, *102*(2), 126-137.

Kifle, M. M., Teklemariam, T. T., Teweldeberhan, A. M., Tesfamariam, E. H., Andegiorgish, A. K., & Azaria Kidane, E. (2019). Malaria Risk Stratification and Modeling the Effect of Rainfall on Malaria Incidence in Eritrea. *Journal of environmental and public health*, *2019*.

Kipruto, E. K., Ochieng, A. O., Anyona, D. N., Mbalanya, M., Mutua, E. N., Onguru, D., ... & Estambale, B. B. (2017). Effect of climatic variability on malaria trends in Baringo County, Kenya. *Malaria journal*, *16*(1), 220.

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*(12), 1213-1221.

Kumar Ra, P., Nathawat, M. S., & Onagh, M. (2014). Application of multiple linear regression model through GIS and remote sensing for malaria mapping in Varanasi District, INDIA.

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*.

Laguna, F., Grillet, M. E., León, J. R., & Ludeña, C. (2017). Modelling malaria incidence by an autoregressive distributed lag model with spatial component. *Spatial and spatio-temporal epidemiology*, *22*, 27-37.

Legesse, Y., Tegegn, A., Belachew, T., & Tushune, K. (2007). Knowledge, attitude and practice about malaria transmission and its preventive measures among households in urban areas of Assosa Zone, Western Ethiopia. *Ethiopian Journal of Health Development*, *21*(2), 157-165.

Lingala, M. A. (2017). Effect of meteorological variables on Plasmodium vivax and Plasmodium falciparum malaria in outbreak prone districts of Rajasthan, India. *Journal of infection and public health*, *10*(6), 875-880.

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*(1), 33-38.

M'Bra, R. K., Kone, B., Soro, D. P., N'krumah, R. T., Soro, N., Ndione, J. A., ... & Schindler, C. (2018). Impact of climate variability on the transmission risk of malaria in northern Côte d'Ivoire. *PloS one*, *13*(6), e0182304.

McCullagh, P. and Nelder, J. (1989)**:** Generalized Linear Models (2^{nd} edition), London: Chapman and Hall.

Midekisa, A., Beyene, B., Mihretie, A., Bayabil, E., & Wimberly, M. C. (2015). Seasonal associations of climatic drivers and malaria in the highlands of Ethiopia. *Parasites & vectors*, *8*(1), 339.

Mohammadkhani, M., Khanjani, N., Bakhtiari, B., Tabatabai, S. M., & Sheikhzadeh, K. (2019). The Relation Between Climatic Factors and Malaria Incidence in Sistan and Baluchestan, Iran. *Sage Open*, *9*(3), 2158244019864205.

Mosissa, M., & Gotu, B.(2011). Statistical Analysis of Spatial Distribution of Malaria in West Shoa Zone, Ethiopia. *Journal of the Ethiopian Statistical Association*, 1.

Nkurunziza, H., Gebhardt, A., & Pilz, J. (2010). Bayesian modelling of the effect of climate on malaria in Burundi. *Malaria Journal*, *9*(1), 114.

Nkurunziza, H., Gebhardt, A., & Pilz, J. (2011). Geo-additive modelling of malaria in Burundi. *Malaria Journal*, *10*(1), 234.

Odland, M. (1987): Spatial Autocorrelation. Sage: Beverly, CA.

Ouédraogo, M., Samadoulougou, S., Rouamba, T., Hien, H., Sawadogo, J. E., Tinto, H., ... & Kirakoya-Samadoulougou, F. (2018). Spatial distribution and determinants of asymptomatic malaria risk among children under 5 years in 24 districts in Burkina Faso. *Malaria journal*, *17*(1), 460.

President's Malaria Initiative. (2008). Malaria Operational Plan (MOP) Ethiopia.

Prince, O. (2010). Statistical methods of disease mapping. Journal of Royal Statistical Society 154(3), 421-441.

Read, M. and Cressie N. (1988). Goodness of fit Statistics for discrete multivariate data. Springer Verleg, New York

Rouamba, T., Nakanabo-Diallo, S., Derra, K., Rouamba, E., Kazienga, A., Inoue, Y., ... & Ouédraogo, B. (2019). Socioeconomic and environmental factors associated with malaria hotspots in the Nanoro demographic surveillance area, Burkina Faso. *BMC public health*, *19*(1), 249.

Samadoulougou, S., Maheu-Giroux, M., Kirakoya-Samadoulougou, F., De Keukeleire, M., Castro, M. C., & Robert, A. (2014). Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. *Parasites & vectors*, *7*(1), 350.

Schabenberger O, Gotway CA (2005). Statistical Metods for Spatial Data Analysis, Chapman and Hall/CRC, New York.

Shimaponda-Mataa, N. M., Tembo-Mwase, E., Gebreslasie, M., Achia, T. N., & Mukaratirwa, S. (2017). Modelling the influence of temperature and rainfall on malaria incidence in four endemic provinces of Zambia using semiparametric Poisson regression. *Acta tropica*, *166*, 81-91.

Simple, O., Mindra, A., Obai, G., Ovuga, E., & Odongo-Aginya, E. I. (2018). Influence of Climatic Factors on Malaria Epidemic in Gulu District, Northern Uganda: A 10-Year Retrospective Study. *Malaria research and treatment*, *2018*.

Srimath-Tirumula, R. C. P. K., Neelapu, N. R. R., & Sidagam, N. (2015). Association of climatic variability, vector population and malarial disease in district of Visakhapatnam, India: a modeling and prediction analysis. *PLoS One*, *10*(6), e0128377.

Stephan, F. F. (1934). Sampling errors and interpretations of social data ordered in time and space. *Journal of the American Statistical Association*, *29*(185A), 165-166.

Sturman, M. (1999).Multiple approaches to analyzing count data in studies of individual differences. The propensity for the type one errors, illustrated with the case of absenteeism prediction. *Educational and Psychological Measurement 59, 414-430.*

Taffese, H. S., Hemming-Schroeder, E., Koepfli, C., Tesfaye, G., Lee, M. C., Kazura, J., ... & Zhou, G. F. (2018). Malaria epidemiology and interventions in Ethiopia from 2001 to 2016. *Infectious diseases of poverty*, *7*(1), 1-9.

Tewara, M. A., Mbah-Fongkimeh, P. N., Dayimu, A., Kang, F., & Xue, F. (2018). Small-area spatial statistical analysis of malaria clusters and hotspots in Cameroon; 2000–2015. *BMC infectious diseases*, *18*(1), 636.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*(sup1), 234-240.

Tuyishimire, J., Kateera, F., Mugisha, J., Amer, S., & Mens, P. (2016). Spatial modelling of malaria risk factors in Ruhuha sector in the east of Rwanda. *Rwanda Journal*, *1*(1S).

Umer, M., Zofeen, S., Majeed, A., Hu, W., Qi, X., & Zhuang, G. (2018). Spatiotemporal Clustering Analysis of Malaria Infection in Pakistan. *International journal of environmental research and public health*, *15*(6), 1202.

Winkelmann, N. (2003)**.**Recent developments in count data Modeling: *Theory and application, Journal of Economic Surveys 9(1), 1-24.*

Wood, G. (2002)**.**Generalized Linear Accidents Models and goodness of fit testing. *Accident Analysis and Prevention 34(4), 417-427.*

World Health Organization. (2017). World Malaria Report 2017.

World Health Organization. (2018). World malaria report 2018. Geneva: World Health Organization; 2018. *Licence: CC BY-NC-SA*, *3*.

Xiang, J., Hansen, A., Liu, Q., Tong, M. X., Liu, X., Sun, Y., ... & Weinstein, P. (2018). Association between malaria incidence and meteorological factors: a multi-location study in China, 2005–2012. *Epidemiology & Infection*, *146*(1), 89-99.

Yalew, W.G., Pal, S., Bansil, P., Dabbs, R.A., Tetteh, K.K., Guinovart, C., Kalnoky, M., Serda, B., Tesfay, B.H., Beyene, B.B., Seneviratne, C., Littrell, M., Yokobe, L., Noland, G.S., Domingo, G.J., Getachew, A., Drakeley, C.J., & Steketee, R.W. (2017). Current and cumulative malaria infections in a setting embarking on elimination: Amhara, Ethiopia. *Malaria Journal*.

Yankson, R., Anto, E. A., & Chipeta, M. G. (2019). Geostatistical analysis and mapping of malaria risk in children under 5 using point-referenced prevalence data in Ghana. *Malaria journal*, *18*(1), 67.

Zayeri, F., Salehi, M., & Pirhosseini, H. (2011). Geographical mapping and Bayesian spatial modeling of malaria incidence in Sistan and Baluchistan province, Iran. *Asian Pacific journal of tropical medicine*, *4*(12), 985-992.

*Zhang, V., Zhirui, N and Lord, L. (2006)***.** *Investigating goodness of fit test statistics for generalized linear crash models with low means.*

Zhang, H., N. Lu, C. Feng, S. W. Thurston, Y. Xia, L. Zhu, and X. M. Tu (2011). On fitting generalized linear mixed effects models for binary responses using different statistical packages. Statistics in Medicine 30 (20), 2562-2572.

Zhou, G., Yewhalaw, D., Lo, E., Zhong, D., Wang, X., Degefa, T., ... & Yan, G. (2016). Analysis of asymptomatic and clinical malaria in urban and suburban settings of southwestern Ethiopia in the context of sustaining malaria control and approaching elimination. *Malaria journal*, *15*(1), 250.

Appendices

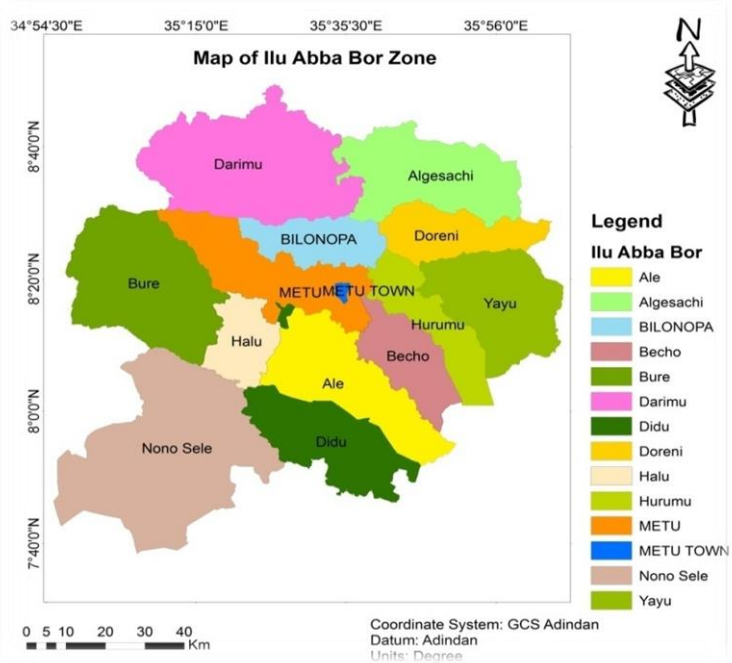


Figure 5.1: Map of Ilu Abba Bor Zone

Table A: Spatial Weight Matrix for Neighboring Relation among Woreda in Ilu Aba Bor Zone, southwest Ethiopia. Construction of spatial weight matrix is based on Queen's method because Queen method share the edge and vertex as from below map and their spatial weight matrix is given below this map that show neighborhoods

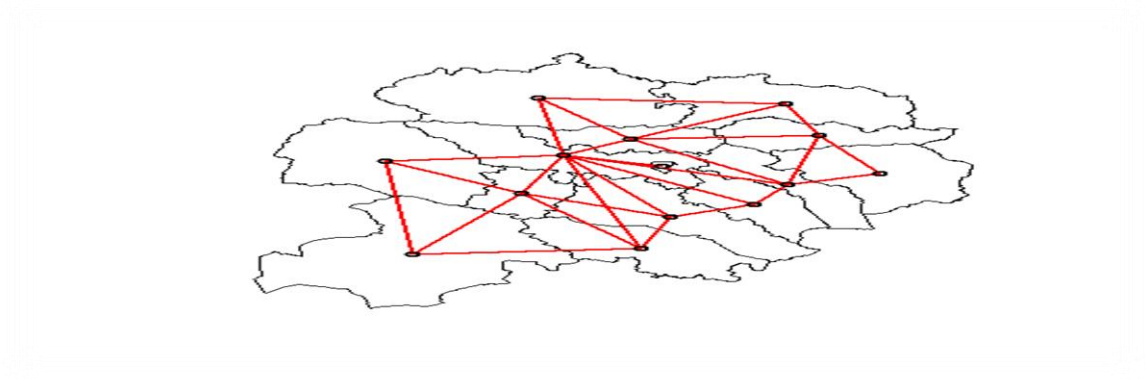


Table 5.1.Spatial weight matrix

	Woreda														
	ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	1	0	0	0	0	1
	3	0	0	0	0	0	0	0	0	0	1	0	0	0	1
	4	1	0	0	0	1	1	0	1	1	1	1	1	1	0
Woreda	5	0	0	0	1	0	0	0	1	0	0	1	1	0	0
	6	0	0	0	1	0	0	1	0	0	0	0	1	0	0
	7	0	0	0	0	0	1	0	0	0	0	1	1	0	0
	8	0	0	0	1	1	0	0	0	0	1	0	0	0	0
	9	1	1	0	1	0	0	0	0	0	1	0	0	0	1
	10	0	0	1	1	0	0	0	1	1	0	0	0	0	1
	11	0	0	0	1	1	0	1	0	0	0	0	1	0	0
	12	0	0	0	1	1	1	1	0	0	0	1	0	0	0
	13	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	14	0	1	1	0	0	0	0	0	1	1	0	0	0	0

Remark: Identification numbers for woredas are given below:

1=Darimu,2=Algesachi,3=Yayu,4=Metuworeda,5=Ale,6=Bure,7=NonoSale,8=Becho,9=Bilonopha,10=Hurumu,11=Didu,12=Halu,13=Metu town and 14=Doreni.

Table 5.2: Malaria case counts by percents from each woreda of Ilu Aba Bor Zone in 2010 Ethiopian calander

Woreda	Malaria case counts in 2010	By percent	Projected population size(2010)
Darimu	538	17.9	194581
Sale Nonoo	57	1.9	30978
Yayu	121	4.03	72340
Mettu Woreda	381	12.7	82661
Ale	123	4.02	71794
Bure	322	10.71	69156
Alge sachi	414	13.8	103992
Becho	128	4.3	49797
Bilonopha	201	6.7	38925
Hurumu	109	3.63	58019
Didu	63	2.09	43829
Halu	113	3.76	22930
Mettu Town	328	10.9	45352
Doreni	107	3.56	48973
Total	3005		