ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURA SCIENCES

DEPARTMENT OF STATISTICS

STATISTICAL ANALYSIS OF SPATIAL DISTRIBUTION OF MALARIA IN WEST SHOA ZONE, ETHIOPIA

BY

MITIKU MOSISSA AYANA

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDISA ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

MAY, 2011

ADDISA ABABA, ETHIOPIA

Addis Ababa University

School of Graduate Studies

Department of Statistics

Title of Thesis: Statistical Analysis of Spatial Distribution of Malaria in West Shoa Zone, Ethiopia

By

Mitiku Mosissa Ayana

Approved by the Board of Examiners:

_____	……
Department Head	**Signature**
_____	…….
Examiner	**Signature**
_____	……
Examiner	**Signature**

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF TABLES AND FIGURES

ACRONYMS

CSA	Central Statistical Agency
FHOM	Federal Ministry of Health
ITN	Insecticide treated nets
IRS	Indoor residual insecticide spraying
LLINs	Long-lasting insecticide treated nets
OLS	Ordinary Least Squares
LM	Lagrange Multiplier
RLM	Robust Langrage Multiplier
VIF	Variance inflation factor
WHO	World Health Organization

ACKNOWLEDGMENTS

First, and foremost, I express deepest gratitude to **Dr**. **Butte Gotu,** my advisor for his unlimited constructive advice, suggestions, ideas and comments. My thanks also to Central Statistical Agency and West Shoa Zone Health Bureau for providing me with the data for this thesis.

Grateful acknowledgement is also expressed to the Department of Statistics, Addis Ababa 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 to wards the completion of this study.

ABSTRACT

Malaria is a major cause of illness and death in large parts of the developing world, especially in Africa. Accurate estimates of malaria distribution are required for planning, implementation and evaluation of malaria control programs. The main objective of this study is to examine spatial patterns of malaria distribution in West Shoa zone, Oromia Region, Ethiopia. Malaria incidence data for 2009 from all health centers of the zone, population size and meteorological data were used. The statistical methods used include global and local measures of spatial autocorrelation as well as spatial autoregressive model. The results of this study indicate that malaria incidence varies according to geographical location, with eco-climatic condition and showing significant positive spatial autocorrelation. Significant local clustering of malaria incidence occurs between pairs of neighboring districts (known as Woredas). Malaria incidence was higher in the western part of the zone and lower in the eastern part of the zone. The results of spatial lag model indicate a significant relationship between malaria incidence and meteorological variables (mid-land zone, hot zone, rainfall, minimum temperature and maximum temperature).

Chapter One

Introduction

1.1 Background of the Study

Malaria is the most common tropical disease, remaining 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 report (WHO, 2008), at the end of 2007 there were 109 malaria endemic countries and 3.3 billion people were at the risk of malaria. Malaria causes at least 300 million and possibly as many as 500 million cases of acute illness each year, which result in 1to 3 million deaths each year. Ninety percent of deaths occur in sub-Sahara Africa. The majority of these deaths are among children less than five years of age and pregnant women. These estimates are not reliable because of inadequate malaria case reporting in most endemic countries and lack of national wide malaria distribution pattern. Accurate estimates of malaria distribution are required for planning, implementation and evaluation of malaria control programs. Hence, there is need for precise estimates about the number of people at risk of malaria to optimize the use of limited resources in high-risk areas.

The President's Malaria Initiative (2010) of Ethiopia, national malaria plan indicated that malaria is ranked as the leading communicable disease in Ethiopia, accounting for about 30% of the overall disability adjusted life years lost. Approximately 75% of the country is malarious with about 68% of the total population living in areas at risk of malaria.

Malaria is reported to cause thousands of deaths each year. According to Ethiopia's Federal Ministry of Health (FMOH), in 2009, malaria was the first cause of outpatient visits, health facility admissions and in-patient deaths, accounting for 12% of out-patient visits and 9.9% of admissions. However, as 36% of the population does not have access to health care services, these figures probably under-represent the true burden of malaria in the country. Increasing the understanding of the distribution dynamics of malaria and their relationship could suggest improvements for malaria control efforts.

Control measures are directed at each component involved in the malaria transmission cycle: the human host, the parasite, the mosquito vector and the environmental factor. Complete cure of malaria requires treatment with several drugs over several days and this creates problems of costs and compliance. Prophylaxis drugs have been of great benefit and widely used as a measure of malaria control, but they are no longer effective in many tropical areas because the parasite developed resistance to drugs. Vector control remains, in general, the most effective tool to prevent and control malaria transmission. The principal objective of vector control is to reduce malaria morbidity and mortality by reducing the levels of transmission. Common measures include indoor and outdoor house insecticide spraying, the use of insecticide treated nets (ITN) and environmental measures such as management of water bodies and vegetation clearance. Applications of these techniques, alone or in combination, reduce human-mosquito contact, vector abundance and vector infectivity. ITN's are increasingly being promoted as an efficient method for reducing the incidence of malaria (Lengeler, 2004). These are done via identifying spatial distribution of malaria.

The National malaria control program in Ethiopia assisted by several international programs (such as Global Fund which has been fighting AIDS, Tuberculosis and Malaria (GFATM), World Health Organization (WHO), the President's Malaria Initiative (PMI), United Nations Children's Funs (UNICEF), World Bank, etc.) adopted several key strategies for malaria control including increasing the coverage for anti-malarial treatment, long-lasting insecticide treated nets (LLINs) and indoor residual insecticide spraying (IRS).

Malaria morbidity also depends on occurrences of mosquito/parasite appendance. Plasmodium falciparum malaria is the world's most common parasitic disease and a major cause of morbidity and mortality in Africa. However figures on malaria morbidity and mortality are very uncertain, since reliable maps risk level of the distribution of malaria transmission and the numbers of affected individuals are not available for most of the African countries. Accurate statistics on the geographical distribution of different endemicities of malaria, on the populations at risk, and on the implications of given levels

of endemicity for morbidity and mortality are important for effective malaria control programs. These estimates can be obtained using appropriate statistical models which relate infection, morbidity, and mortality rates to risk factors, measured at individual level, but also to factors that vary gradually over geographical locations (Smith, 2003).

The malaria transmission intensity and temporal variation in Ethiopia is mainly determined by the diverse eco-climatic conditions. Climatic factors such as temperature, rainfall and humidity show high variability mainly as a function of altitude and are the most important variables that influence malaria transmission. Based on this altitudinal variation and the climatic characteristics associated with it, areas of the country are categorized into climatic zones namely: the cold zone, the hot zone, and mid-land zone. The cold zone, which covers areas above 2500m above sea level, has a mean annual temperature of 10-15^{0}C. It is a highland area considered free of local malaria transmission. The mid-land area, ranging in altitude from 1500 – 2500m, with a mean annual temperature ranging between 15-20^{0}C, has a diverse malaria transmission pattern. In the lowland zone, located in areas below 1,500m above sea level, where the mean annual temperature varies from 20-25^{0}C, malaria transmission is endemic and its intensity and duration is mainly dictated by the amount and duration of rainfall. In the mid-land zone, where temperature is a determining factor, malaria transmission often occurs in areas below 2000m, while areas above 2000m are mainly affected during epidemics. Mean annual precipitation, in general, ranges from 800 to 2200-mm in the highlands (>1500 meters) and varies from less than 200 to 800-mm in the lowlands (<1500 meters). Based on these altitudinal ranges and the associated temperature and rainfall patterns, ecological classification has been used to further clarify malaria distributions pattern (FMOH, 2008).

Malaria is the leading cause of morbidity and mortality in Ethiopia, accounting for over nine million cases and thousands of deaths annually. The risks of morbidity and mortality associated with malaria are characterized by spatial variation across the country. Consequently, we recognize the spatial variation of malaria by means of spatial autocorrelation.

1.2 Use of Spatial Autocorrelation

Spatial autocorrelation may be defined as the relationship among values of a single variable that comes from the geographic arrangement of the areas in which these values occur. It measures the similarity of objects within an area; the degree to which a spatial phenomenon is correlated to itself in space (Cliff and Ord, 1981), the level of interdependence between the variables, the nature and strength of the interdependence, i.e. spatial autocorrelation is an assessment of the correlation of a variable in reference to spatial location of the variable.

Spatial autocorrelation is used to test whether the observed value of a variable at one locality is independent of values of the variable at neighboring localities. It may be classified as either positive or negative: Positive if similar values appear together, or negative if dissimilar values appear in close association. When no statistically significant spatial autocorrelation exists, the pattern of spatial distribution is considered random (Chou, 1997).

Spatial autocorrelation measures and analyzes 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, e.g., the distances between neighbors, the lengths of shared border, or whether they fall into a specified directional class such as “west.” Classical spatial autocorrelation compares the spatial weights to the covariance relationship at pairs of locations (Anselin, 1992).

Spatial autocorrelation statistics such as global Moran’s *I* and Geary’s *C* are estimate the overall degree of spatial autocorrelation for a dataset. The possibility of spatial heterogeneity suggests that the estimated degree of autocorrelation may vary significantly across geographic space. Local spatial autocorrelation statistics provide estimates disaggregated to the level of the spatial analysis units, allowing assessment of the dependency relationships across space. *G*etis and Ord statistics compare neighborhoods to a global average and identify local regions of strong autocorrelation.

In these situations, spatial statistics and geographical information systems (GIS) provide methodologies and solutions to analyze the epidemiological and ecological context of malaria and other infectious diseases (Tanser, 2002). Global spatial autocorrelation statistics such as the global Moran's *I* and Geary's *C* describe the overall spatial dependence of malaria over the entire region, local spatial autocorrelation statistics such as the local Moran's *I* (Anselin,1995) and Getis and Ord G_i^* (Getis and Ord ,1992) are useful in identifying local patterns or hot spots.

Resolution enhancement in spatial data offers the opportunity to model disease at finer scales, but it also demands a corresponding enhancement in malaria distributions methodology. Previous approaches to malaria distribution have been largely based on eco-climatic conditions, required model to identify the relation between malaria and covariate variables. The OLS approach assumes that observations are independent of each other, ignoring the possible spatial relationship that may exist between them. This assumption may not be valid with variables representing a geographically varying phenomenon, such as morbidity, because when the spatial scale moves down especially to the district level, the spatial autocorrelation of dependent variable among neighboring locations is likely to become stronger (Brown, 1995). This may lead to the violation of the OLS independence assumption, a problem that needs to be addressed using an improved methodology to ensure a valid model of malaria estimation.

The main goal of this study is to examine spatial patterns or clusters of malaria distribution using district level malaria incidence data. It seeks to identify malaria "hotspot" Woredas by producing map of clustering observation and fit appropriate spatial models for malaria distribution in West Shoa Zone, Oromia Region, Ethiopia. The results of this study may help policy makers and managers at different administrative levels to control and prevent malaria more efficiently.

1.3 Statement of the Problem

Malaria mortality and morbidity are known to vary by geographical location and depend on eco-climatic conditions. Demographic, eco-climatic mortality factors, age and sex vary by geographical location, and many authors recommend that targeting interventions to the high malaria case are omitted due to inconsideration of spatial dependence. According to Smith (2003) regions that are in closer proximity are expected to have similar malaria cases because of similar eco- climatic situation and demographic characteristics. In this study the spatial distributions of malaria is assessed using spatial model along with meteorological and environmental variables of malaria incidence in West Shoa Zone and to identify whether the distribution of malaria is clustered or not.

Spatial models explain malaria morbidity variation by geographical location better than non-spatial models when limited data is available for meteorological variables. Incidences of malaria, which also vary spatially, raise the need for spatial models for covariates. The modeling has implications for malaria net control and risk management. Environmental variation risks can be quantified using spatial models of prevalence and morbidity heterogeneity. Therefore, the research problems include:

- ✓ How to detect and measure spatial dependencies?
- ✓ What is the implication of observing significant spatial dependencies in the data under consideration?
- ✓ How to include spatial dependence in the model?
- ✓ How to evaluate the impact of site-specific factors on malaria distributions?
- ✓ How to interpret fitted spatial lag and spatial error model?
- ✓ How to evaluate the significance of the fitted spatial model?

1.4. Objectives of the Study

The main objective of this study is to assess the spatial distribution of malaria incidence in West Shoa Zone by using spatial statistical methods.

The specific objectives of the study are:

- ✓ To discuss and measure the intensity of spatial autocorrelation in the distributions of malaria incidence
- ✓ To characterize the distribution of malaria incidence
- ✓ To identify factors related with malaria incidence distribution, and model malaria incidence

1.5 Significance of the Study

The results of this study could help classify the woredas of the zone into high and low risk groups so as to give information on how to optimize available resources for malaria control.

1.6 Limitations of the Study

The surveillance data used in this study have obvious limitations related to coverage and thus can underestimate the actual malaria incidence in the population, especially in some remote locations. Nonetheless, this study illustrates the importance of surveillance data in identifying malaria hot spots.

Chapter Two

Review of Literature

From the bulk of literature, only those sources that are more related and relevant to this study and some selected applications of the methods used in this study are presented.

Richard *et al*. (2000) conducted a study on the malaria transmission pattern and reported a marked clustering of persons with malaria parasites and clinical symptoms at particular sites, usually regions. The study shows that in localities of low endemicity the level of malaria risk or case incidence may vary widely between regions because the specific characteristics of climatic conditions and their locations affect contact between humans and vectors. Where endemicity is high, the study indicates, differences in human/ vector contact rates between different households may have less effect on malaria incidences.

Salim (2000) examined the impact of categorical explanatory variables on binary outcomes of owning an ITN, under consideration of spatial correlation in Tanzania. Continuous fixed effect variables are discretisized into several categories to check their linear influence. The results obtained indicate that malaria can best be controlled by combining ITNs and an effective antimalarial therapy.

Jeong and Gluck (2002) employed a triangulation of spatial techniques including griddling the population and spatial aggregation of health services based on the populated areas and spatial statistics to generate information on malaria distribution in Japan.

Guofa *et al.* (2003) conducted a study on spatial distribution of monthly malaria incidence. The objective of the study was to assess the impact of climate in malaria resurgence in the East African highlands, Kenya. The results obtained indicate that monthly rainfall, monthly maximum and minimum temperatures were significantly correlated with monthly malaria incidences.

Smith (2003) argued that in the case of malaria, spatial correlation is present at both, short and large scales, reflecting the transmission of malaria infection by the mosquitoes

over space and the effects of environmental factors that determine mosquito survival over large areas in Africa.

Brooke *et al.* (2004) identified significant spatial clusters of malaria cases in Western Kenya by using spatial autocorrelation statistic.

Chaix *et al.* (2004) analyzed the data from a large follow-up study in an area of high perennial malaria transmission in southern Tanzania in order to describe the spatial effects of bed nets on all-cause child mortality. The fitted spatial models showed significant geographical variations in healthcare utilization. It was also reported that place indicators better explained spatial variations in healthcare utilization when measured across continuous space, rather than administrative areas.

Chaix *et al.* (2005) conducted a study in the distribution of malaria. In the study individual information on health care utilization was plotted across all of mainland in France. The results provide adequate information for modeling variations across continuous space with spatial regression techniques.

Githeko *et al.* (2006) conducted a study on resurgence of malaria in the highlands of western Kenya. The aim of the study was to determine the effects of topography on malaria spatial vector distribution and prevalence by using spatial model. The results show malaria transmission in this area is mainly confined to the valley bottom.

Kazembe *et al.* (2007) examined spatial clustering of malaria risks in northern Malawi. Geodemographic modeling and spatial modeling of demographic data were used in a range of applications. The findings of the study showed a significant association of malaria risk with elevation, annual maximum temperature, rainfall and potential evapotranspiration.

Abellana *et al.* (2008) conducted a study in Mozambique. The main objective of the study was to assess the distribution of malaria incidence and rate of malaria transmission by using spatial statistical methods. The results obtained indicate a clear spatial pattern of malaria incidence, and high transmition rate during the wet season.

Basel (2008) assessed the distribution of malaria incidence in Africa. The focus of the analysis is to identify spatial patterns or trends and to assess association between malaria data and environmental factors that vary gradually over geographical regions. Spatial autocorrelation model was fitted. The results obtained indicate that malaria incidence and environmental conditions are positively associated.

Wimberley *et al.* (2008) examined malaria cases to regional and global assessments of malaria distribution pattern in USA. The study focused on comparing non-spatial and spatial modeling approaches for predicting the geographic distributions of malaria. The results obtained show incorporating either spatial autocorrelation or spatial heterogeneity resulted in substantial difference in estimations of malaria distribution.

Hans (2008) assessed malaria incidence data obtained from different locations in Tanzania. One of the objectives of the study was to estimate the environment-disease relation by using spatial modeling. It was reported that malaria incidence data are spatially correlated.

Gosoniu *et al*. (2008) analyzed malaria risk data obtained from Angola. The main objective of the study was to estimate the degree of spatial correlation and to assess the effect of different covariates in the presence of geographical heterogeneity. Spatial dependence was modeled by assuming non-random effects among the locations. Results show high spatial correlation and strong relationship between malaria incidence and covariates included in the model.

Noor *et al.* (2008) attempted to study malaria distribution in Somalia. The goal of study was to optimize resource allocation for anti-malaria control by using mapping malaria risk level and spatial modeling. The results obtained indicate that precipitation and temperature had significant association with P.falciparum prevalence, but distance to water was found to be insignificant.

Ali *et al.* (2008) modeled the spatial variations in malaria episodes in Iran with the objective of assessing the feasibility of an epidemic early warning system. It was reported that malaria incidence is positively associated with maximum temperature and mean relative humidity.

Zhang *et al.* (2008) conducted a study to detect spatial distribution and clustering of malaria incidence in Anhui province, China. The objective was to identify highly endemic areas for public health to allocate resource by using spatial methods. The study identified environmental factors responsible for the re-emerged malaria risks.

Briet *et al.* (2008) analyzed the intensity of malaria distribution in Sri Lank, the aim was to assess the efficiency of allocation of resources for malaria control by using spatial modeling. The results indicate that both global and local spatial clustering of malaria incidence among the district, used for allocation of resources.

Lili *et al.* (2008) analyzed malaria incidence in western Kenya. The study aimed at providing a better understanding of the distribution and abundance of mosquito vectors. To achieve this objective, spatial and non-spatial methods were employed. Results of the study indicated that distance to high-order streams is an effective predictor for the distribution of adult mosquitoes. It was also reported that spatial methods are more effective in modeling the distribution of adult mosquitoes than the non-spatial methods.

Yeshiwondim *et al.* (2009) examined the global and local patterns of malaria distribution in 543 villages in Ethiopia using individual-level morbidity data collected from six laboratory and treatment centers. It was reported that malaria incidence varies according to gender and age with age less five years and above showing a statistically significant malaria incidence. It was also observed that local clustering of malaria incidence between pairs of villages within distance lags were significant. Furthermore, malaria hot spots were displayed as risk maps that are useful for monitoring and spatial targeting of prevention and control measures.

Chowell *et al.* (2009) conducted a study on the transmission dynamics of falciparum and vivax malaria in Pepu. The main objective of the study was to analyze different spatial scales in conjunction with associated demographic, geographic and climatologically distribution of malaria. It was reported that spatial autocorrelation was observed among regions with slightly higher levels of spatial heterogeneity for P.falciparum and P.vivax.

Ingrid *et al.* (2009) conducted a study in Adama, Ethiopia, by using small-scale spatial dependence. The main goal of the study was to identify foci of malaria transmission in urban communities. The results of the study indicated that proximity to vector breeding site, maximum and minimum temperature, rainfall, were positively associated with malaria incidence.

Tsai *et al.* (2009) employed spatial autocorrelation methodologies, including Global Moran's *I* and Local Getis-Ord statistics. The objective of the study was to describe and map spatial clusters, and areas in which these are situated, for 20 leading causes of death in Taiwan. The results indicate that Cluster mapping helps to illuminate issues such as the spatial aspects of correlations for leading health care events.

Grillet *et al.* (2010) used local spatial statistics and geographically weighted regression (GWR) to determine the spatial pattern of malaria incidence and persistence in northeastern Venezuela. It was reported that the GWR model greatly improved predictions of malaria risk compared with ordinary least squares (OLS) regression models. Results also indicate that disease persistence was associated with greater human population density, lower elevations, and proximity to aquatic habitats.

Riedel *et al.* (2010) conducted a study in Zambia. The main objective of the study was to assess the relationship between malaria incidence and environmental variable. A number of models were fitted to capture the (potential) non-linearity in the malaria-environment relation and to identify the elapsing time between environmental effects and malaria risk. Different model validation methods were used to identify the best fitting model. Model-based risk predictions at unobserved locations were obtained via spatial distributions.

Chapter Three

Data and Methodology

3.1 Source of Data

The data which motivate this study were collected by the national malaria control program under the Oromia Disease Control and Prevention Bureau which was carried out in 2009. The aim of the program was to undertake national malaria control for regional and national family and health planning.

The data were collected from all Woreda health centers and hospitals of West Shoa zone. Here, the results of microscopic examination were recorded including all the malaria cases, malaria cases P. falciparum, malaria case P.vivax, malaria case by age, malaria cases by sex, admitted by malaria case, and clinical cases are treated as malaria. Population size for each Woredas where obtained from the Central Statistical Agency (CSA, 2007 and CSA, 2009).

The Meteorological data were collected by Ethiopian Meteorological Agency. Data were coded and analyzed by using SAS, GeoDa, and ArcGIS software.

This study focuses on meteorological variables that are relevant to malaria incidence distributions. The dependent variable is malaria incidence whereas the independent variables include average annual rainfall, average annual maximum temperature, average annual minimum temperature, percentage of highland areas, percentage of midland areas and percentage of lowland areas of the zone.

3.2 Study Area

West Shoa Zone is one of the 18 Zones in Oromia Region, Ethiopia. It has 18 administrative Woredas and 570 Kebeles with estimated total population of 2.27 million in 2009. The zone covers 14.9 thousand sq. km. The climate conditions can be classified as: highland area (12%), midland area (54%) and lowland area (34%). All 18 woredas were covered in the study. These woredas are: Abuna-Gindeberet, Ada-Berga, Ambo, Bako-Tibe, Cheliya, Dano, Dendi, Ejere, Elfata, Ginde-Beret, Jeldu, Jibat, Midakegn,

Meta-Robi, Nono, Tikur-Enchini, Toke-Kutaye and Walmara. Malaria is the major public health problem and about one million people are at risk of infections of malaria. Plasmodium Falciparum and Plasmodium Vivax are the parasites responsible for more than 95% of malaria infection in the zone. The transmission is greatest during the rainy season and reaches its peak in October.

3.3 Methodology of the Study

3.3.1 The Concept of Spatial Dependence

The essence of spatial analysis is that "space matters", i.e. what happens in one region is related to what happens in neighboring regions. This has been made more precise in what Tobler (1979) refers to as the First Law of Geography: "Everything is related to everything else, but closer things more so". One way to approach this is via the notion of spatial autocorrelation (Anselin and Bera, 1998).

According to Anselin and Bera (1998), spatial autocorrelation can be loosely defined as the coincidence of value similarity with location similarity. In other words, high or low values for a random variable tend to cluster in space (positive spatial autocorrelation) or locations tend to be surrounded by neighbors with very dissimilar values (negative spatial autocorrelation). Of the two types of spatial autocorrelation, researchers usually focus on positive autocorrelation. Negative spatial autocorrelation implies a checkerboard pattern of values and does not always have a meaningful substantive interpretation.

Spatial autocorrelation is something like temporal autocorrelation, but more complicated. The reason is that temporal autocorrelation can only go one way: what happens at one time can be influenced only by what has happened in the past. But spatial autocorrelation can potentially go in any direction: it is like saying that what happens at any one point in time can be influenced by both the past and the future (Anselin, 1992). For this reason, we cannot simply transfer models of temporal autocorrelation to the spatial context.

In this study global and local measures of spatial autocorrelation will be used first to diagnose univariate spatial autocorrelation in the absence of covariates. Next, a standard regression model will be estimated and diagnostics test will be conducted to determine

whether the covariates sufficiently model the spatial dependence in the dependent variable. If they do not, the spatial autoregressive model specification indicated by the diagnostic will be fitted.

3.3.2. Global and Local Measures of Spatial Autocorrelation

Tests for global spatial autocorrelation examine whether the data as a whole exhibit spatial autocorrelation (against H_o: spatial autocorrelation) as well as the strength and direction (positive or negative) of any spatial autocorrelation. Tests for local spatial autocorrelation (again, against H_o: spatial autocorrelation) identify particular observations that are autocorrelated with neighboring observations of the dependent variable of interest and also determine the strength and, depending upon the statistic, also the direction of this spatial autocorrelation (Anselin, 1995). Next, we first present global tests of spatial autocorrelation and subsequently define local tests to decompose the global result.

3.3.2.1. Global Measures of Spatial Autocorrelation

Tests for either global or local spatial autocorrelation in lattice data proceed through the use of a Γ index (Anselin, 1992). In this case spatial data applications, one of the matrices will be a contiguity or distance matrix. Indeed, each element in such a matrix indicates the location similarity between the object corresponding to the row (i location) and the object corresponding to the column (j location). A Γ index consists of the sum of the cross products of the corresponding elements W_{ij}, Y_{ij} of two matrices, **W** and **Y**:

$$\Gamma = \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}Y_{ij} \qquad [1]$$

where

$\mathbf{W_{ij}}$ is the ijth element of spatial weights matrix **W**; $\mathbf{Y_{ij}}$ is the product of the two values $Y_{ij}=Y_iY_j$ or its squared difference $Y_{ij}=(Y_i-Y_j)^2$ of dependent variable at *i* and *j* locations respectively and N is the number of observations.

Measures of spatial autocorrelation are variants of this Γ index, with the $\mathbf{Y_{ij}}$ elements in **Y** reflecting how value (dis)similarity is conceptualized in the particular form of the Γ index.

Methods of Measuring Spatial Autocorrelation

[A]. Defining Spatial Weights Matrix

To asses the nature and degree of spatial autocorrelation, it is necessary to represent the spatial arrangement of observations in order to get a sense of how close or distant they are apart from each other. To express the degree of proximity between observations in space we may attribute a value of one if the observations are nearby (neighbors) and zero otherwise. There are different other options for defining these weights, they may be based on neighborhood which have common boundary and based on distances between centroids (LeSage, 1998). In these cases, pairs of observations might be defined as neighborhoods or entering the actual distances as a measure of the degree of proximity.

Methods of defining spatial weight matrix are:-

- ✓ Distance weight (Threshold Distance and k-nearest Neighbors)
- ✓ Neighborhood /Contiguity weight

Getis and Ord (1992) point out that administrative center of the observation units could adequately represent the location of the observation. Distance has been assumed to be the great circle distance between points (the radius of the circle around point), ignoring barriers and other factors. Distance could further be banded on the basis of the frequencies of inter-point distances, and the furthest nearest neighbor distance for regular lattice and irregular lattice. A typical element of the non-standardized spatial weight matrix *W(d)* for distance *d* is defined as:

$$w_{ij}(d) = \begin{cases} 1 \ \textit{if} \ \textit{hypot}(i,j) \le d, i \ne j \\ 0 \ \ \textit{otherwise} \end{cases}$$

where,

hypot(i,j)= $\sqrt{(y_i - \bar{y})^2 + (y_j - \bar{y})^2}$, y_i and y_j are value of dependent variable at *i* and *j* locations, respectively.

To define the contiguity relation in terms of sets of neighbors of zones or sites having common boundary (Tobler, 1970), these are coded in the form of a spatial weight matrix **W**, with a zero diagonal, and the off- diagonal non-zero elements often scaled to sum to unity in each row (standardized weights matrices), with typical elements:

$$\mathbf{W_{ij}} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}} \qquad \textbf{[2]}$$

where, non-standardized binary spatial weight matrix with typical elements given by:

$$w_{ij} = \begin{cases} 1 & \textit{if } i \textit{ is linked to } j \textit{ and} \\ 0 & \textit{otherwise} \end{cases}$$

Generally, spatial weights matrices are row-standardized so that the sum of the weights for each row equals one. As a result, the spatial influence from neighbors is a weighted average of this influence across the neighbors. Clearly, the definition of neighbors is a critical decision in the modeling of spatial autocorrelation. Closely related is the form and extent of spatial dependence between neighbors. In defining neighbors and the form of spatial dependence between these neighbors, the constraints on potential spatial dependence incorporated in the weights matrix should reflect a priori theoretical expectations. The simplest definition of neighbors is the contiguity case. Here, there are three principal possibilities (Anselin, 1988).

- 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**

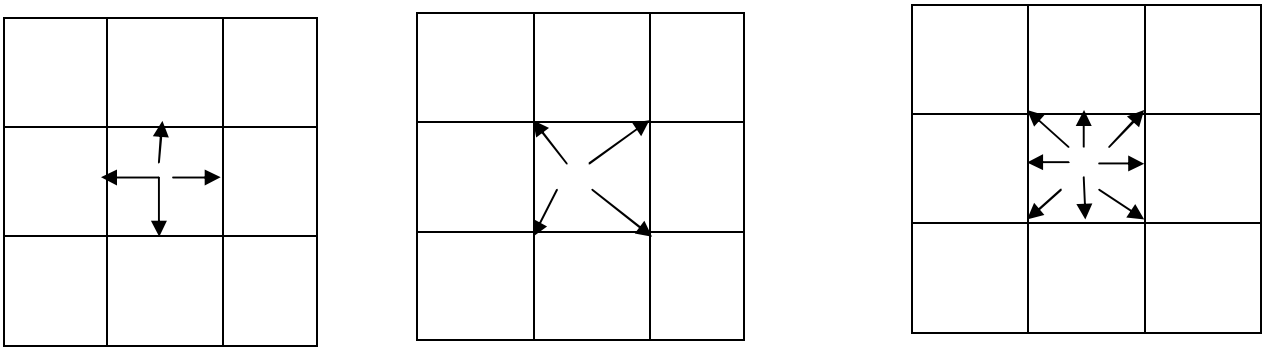


Figure 3.1: Contiguity Case of Representation of Spatial Weight Matrix

[B]. Tests of Spatial Autocorrelation

The two most commonly used measures for spatial autocorrelation are Moran's *I* and Geary's *C* statistics. These tests indicate the degree of spatial association as reflected in the data set as a whole. While Moran's *I* is based on cross products to measure value association, Geary's *C* employs squared differences (Anselin, 1992).

Test for spatial autocorrelation are designed to quantify the extent of clustering and to allow for statistical inference. The null hypothesis (under the normality and independence assumptions) is given by:

H_0 : no spatial autocorrelation (H_0 **:** ρ =0)

Under the alternative hypothesis (H_1**:** $\rho \neq 0$ **)** of spatial autocorrelation (spatial dependence), the interest focuses on instances where large values are systematically surrounded by other large values, or where small values are surrounded by small values.

[i]. Global Moran's *I*

For binary weights Moran (1950) introduced the following coefficient of autocorrelation:

$$\mathrm{I} = \frac{N}{S}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \qquad [3]$$

where N is the number of observations; S is the sum of the elements of spatial weights matrix; y_i and y_j are the values on the dependent variable at locations i and j respectively; $\bar{y}$ is the mean of y.

Under the normal and randomization assumptions, the resulting z-values can be compared to a table of standard normal to assess significance. The null hypothesis (no spatial autocorrelation), will be rejected if the calculated value of $|z| > z_{\alpha/2}$ and the

z-statistic is given by:

$$\mathrm{Z} = \frac{I - E(I)}{\sqrt{Var(I)}},$$

where

$$E(\mathrm{I})_{\mathrm{N}} = \frac{-1}{N-1} = E(\mathrm{I})_{\mathrm{R}}$$

The variance of Moran's I and Geary's C will vary under the assumptions normality and randomization. Under the normality assumption the variance of Moran's I ($\mathrm{Var(I)_N}$) is given as

$$\mathrm{Var(I)_N} = \frac{N^2(N-1)S_1 - N(N-1)S_2 - 2S_0^{\,2}}{(N+1)(N-1)S_o^{\,2}},$$ whereas under randomization ($\mathrm{Var(I)_R}$) is

given by

$$\mathrm{Var(I)_R} = \frac{N\left(S_1\left(N^2 - 3N + 3\right) - NS_2 + 3S_o^{\,2}\right)}{(N-1)(N-2)(N-3)S_o^{\,2}} - \frac{K\left(S_1\left(N^2 - N\right) - 2NS_2 + 6S_o^{\,2}\right)}{(N-1)(N-2)(N-3)S_o^{\,2}} - \left(\frac{1}{N-1}\right)^2$$

$$S_1 = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(W_{ij} + W_{ji}\right)^2, i \neq j,$$

$$K = \frac{N\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^4}{\sum_{i=1}^{N}\left(\left(y_i - \bar{y}\right)^2\right)^2}$$

$$S_2 = \sum_{i=1}^{N}\left(\sum_{j=1}^{N} W_{ij} + W_{ji}\right)^2,$$

$$S_0 = \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}, i \neq j\ .$$

Interpretation: a positive global Moran's *I* that differs significantly from the expected value under the null hypothesis indicates positive spatial autocorrelation and implying the clustering of similar values (i.e, high values are found closer together, and low values are found closer together) on the dependent variable among neighboring observations. A negative global Moran's *I* that differs significantly from the expected value under the null hypothesis indicates negative spatial autocorrelation and implies the clustering of dissimilar values (means high values are found far away from other high values, and low values are found far away from other low values) on the dependent variable among neighboring observations (Anselin, 1992).

[ii]. Global Geary's *C*

The global Moran's *I* defines value (dis)similarity as deviations from the mean, whereas the global Geary's *C* defines value (dis)similarity as the squared difference in values between neighboring observations. For binary weights Geary (1954) introduced the following coefficient of autocorrelation:

$$\mathrm{C} = \frac{N-1}{2S}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}\left(y_i - y_j\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2} \quad , \qquad [4]$$

The z-statistic of Geary's *C* is given by:

$$Z(C) = \frac{C - E(C)}{\sqrt{\text{var}(C)}}$$

where

$$E(C)_N = E(C)_R = 1$$

Variance of Geary's *C* under normality assumption $(Var(C)_N$ is given as

$$Var(C)_N = \frac{\left((2S_1 + S_2)(N-1) - 4S_o^{\,2}\right)}{2(N+1)S_o}$$ and under randomization is given by

$Var(C)_R =$

$$\frac{S_1(N-1)\left(N^2 - 3N + 3 - K(N-1)\right)}{S_o N(N-2)(N-3)} + \frac{\left(N^2 - 3 - k(N-1)^2\right)}{N(N-2)(N-3)} - \frac{(N-1)S_2\left(N^2 + 3N - 6 - K\left(N^2 - N + 2\right)\right)}{4N(N-2)(N-3)S_o^{\,2}}$$

Where all the notations are as in [3].

Therefore, the null hypothesis of no spatial autocorrelation (H_o: $\rho = 0$) will be rejected if the calculated value of $|Z(C)| > z_{\alpha/2}$.

Interpretation: A value of Geary's *C* that is significantly larger than one indicates negative spatial autocorrelation, while a value that is significantly smaller than one indicates positive spatial autocorrelation (Anselin, 1992). Due to the squared term in the numerator in [4], Geary's *C* gives greater weight to extreme values than Moran's *I* . As a consequence, the global Moran's *I* is generally preferred in practice (Cliff and Ord, 1981).

[**C**]. **Moran Scatter Plot**

The Moran Scatter plot enables us to visualize the linear correlation between Y and WY**.** Specifically, WY is plotted against Y and the Moran's *I* coefficient will be the slope of the regression curve (Anselin, 1998).

In additions to this, inspection of global and local spatial instability is carried out by the means of the Moran scatter plot (Anselin, 1996), which plots the spatial lag, WY against

the original values Y. The four different quadrants of the scatter plot correspond to the four types of local spatial association between a region and its neighbors: the first quadrant, (HH) a region with a high value surrounded by regions with high values (top on the right), the second, (LH) a region a with low value surrounded by regions with high values (top on the left),the third (LL) a region with a low value surrounded by regions with low values (bottom on the left) and the last (HL) a region with a high value surrounded by regions with low values (bottom on the right) as shown in the following Figure 3.2 below. The first and the third quadrants refer to positive spatial autocorrelation indicating spatial clustering of similar values whereas the second and the forth quadrants represent negative spatial autocorrelation indicating spatial clustering of dissimilar values. The Moran scatter plot may thus be used to visualize typical localizations, i.e. regions in quadrant two or in the quadrant four.

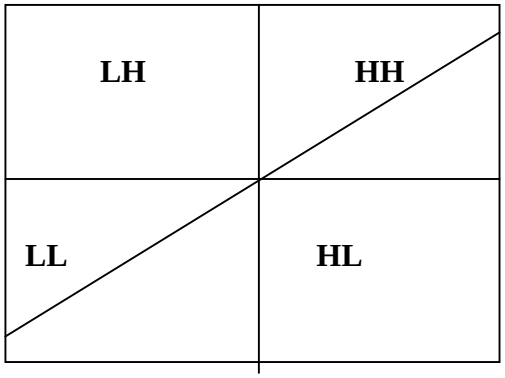


Figure 3.2: Moran's *I* Scatter Plot

3.3.2.2 Local Measures of Spatial Autocorrelation

In order to observe if there is a local spatial cluster of high or low values, and identify the regions that contribute the most to the clustering (spatial autocorrelation), measures of local spatial autocorrelation such as local Getis and Ord statistics and the local indicator of spatial association are used (Anselin, 1995).

Often our interest lies not only in determining whether the data as a whole exhibit spatial autocorrelation, but also, in identifying the specific observations that exhibit spatial

autocorrelation with their neighbors. In this study the following local measures have been used.

- Local Moran's *I*
- Local Ord and Getis' G_i^* statistic

[i]. Local Moran's *I*

According to Anselin (1995), Local Moran's *I* for each observation measures the extent of significant spatial clustering of similar values around that observation.

The null hypothesis which stated no spatial clustering will be rejected if the computed value of $|z_{Ii}| > z_{\alpha/2}$. z_Ii-statistic is given by

$$Z_{Ii} = \frac{I_i - E(I_i)}{\sqrt{Var(I_i)}}$$

$$E(I_i) = \frac{\sum_{j=1}^{N} W_{ij}}{N-1}$$

$$\text{Var}(I_i) = W_i \frac{(N-b_2)}{N-1} + \frac{2W_{i(kh)}(2b_2 - N)}{(N-1)(N-2)} - \frac{W_i^{\,2}}{(N-1)^2} \quad \text{where}$$

$$b_2 = \frac{m_4}{m_2^{\,2}}, \quad m_2 = \sum_{i=1}^{N} \frac{y_i^{\,2}}{N}, m_4 = \sum_{i=1}^{N} \frac{y_i^{\,4}}{N}, \quad W_i = \sum_{j \neq i}^{N} W_{ij}^{\,2}, \quad 2W_{i(kh)} = \sum_{h \neq i}^{N} \sum_{k \neq i}^{N} W_{ik} W_{ih},$$

i, k and *h* are represent i^{th}, k^{th} and h^{th} location respectively.

The local Moran *I* is given by

$$I_i = \frac{\sum_{j=1}^{N} W_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(y_i - \bar{y})^2} \qquad [5]$$

In this formula, the notations are as in (3)

Again, only the *j* neighbors are incorporated in the local Moran's *I* for *i*. The interpretation of values of the local Moran's *I* is analogous to their global counterpart.

In addition to the identification of local spatial clustering, the correspondence between local indicator of spatial association statistics and global spatial autocorrelation measures carries significant additional advantage in decomposing these global measures. Through the estimation of local Moran's *I* scatter plot, we can identify which observations are consistent with the global pattern of positive or negative spatial autocorrelation and which observations run counter to this global pattern (Anselin, 1992).

[ii]. Local Ord and Getis G_i^* Statistic

The Ord and Getis $\boldsymbol{G_i^*}$ test statistic is given by (Ord and Getis, 1995):

$$\mathbf{G^*_i} = \frac{\sum_{j=1}^{N} W_{ij} y_j - \left(\sum_{j=1}^{N} W_{ij} + W_{ii} \right) \bar{y}}{S \left(\left[N S_i^* - \left(\sum_{j=1}^{N} W_{ij} + W_{ii} \right)^2 \right] / (N-1) \right)^{1/2}} \qquad [6]$$

In this formula, $\mathbf{W}_{ij}$ are the elements of the spatial weights matrix **W,**

$\boldsymbol{W_{ii}}$ is a weight in the case in which *i* is in its own neighborhood set,

$\bar{y}$ is the mean of the values on the dependent variable,

$$\mathbf{S^*_i} = \sum_{j=1}^{N} W_{ij}^{\,2} + W_{ii}^{\,2}$$

$\mathbf{S^2}$ is the sample variance

It shows what portion of the total sum of all values is represented by the values at, and near, locations *i* (Getis and Ord, 1992). The G_i^* scores ranges form 0 to 1. The expected values are given by:

$$E(\mathrm{G^*_i}) = \frac{\sum_{j=1}^{N} W_{ij}}{N} \qquad [7]$$

To determine whether or not a particular $\boldsymbol{G_i^*}$ score is significant, we will use cluster map and its significance level (z_i).

The variance of G_i^* is equal to $\frac{W_i^*\left(N - E\left(G_i^*\right)\right)Y_{i2}}{N^2(N-1)Y_{i1}}$, where

$$W_i^* = \sum_{j=1}^{N} W_{ij}$$

$$Y_{i1} = \frac{\sum_{j=1}^{N} y_j}{N},$$

$$Y_{i2} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(y_i y_j\right)^2}{N} - Y_{i1}$$

where

$$z_i = \left(\frac{G_i^* - E\left(G_i^*\right)}{\sqrt{\text{var}\left(G_i^*\right)}}\right) \qquad [8]$$

The null hypothesis of no spatial clustering will be rejected if the computed value of $|z_i| > z_{\alpha/2}$.

Interpretation: Positive values the $\boldsymbol{G_i^*}$ statistic indicate that high values are spatially clustered with other high values of the random variable. Negative values of $\boldsymbol{G_i^*}$ statistic indicate that low values are spatially clustered with other low values of the dependent variable. Note that a consequence of this is that the $\boldsymbol{G_i^*}$ statistic, unlike Moran's *I*, cannot distinguish cases of positive spatial autocorrelation from cases of negative spatial autocorrelation (Getis and Ord, 1992).

The contribution of the $\boldsymbol{G_i^*}$ statistic, as with other measures of local spatial autocorrelation, is that they aid in identification of local pockets of spatial clustering. Such clustering, moreover, may occur even in the absence of global spatial autocorrelation. Local spatial autocorrelation can exist in the absence of global autocorrelation when the clustering at the local level is limited as a proportion of the

overall number of observations or when local patterns are off-setting, producing no global pattern as a consequence (Getis and Ord, 1992).

3.3.3. Diagnostics for Spatial Dependence

Once significant spatial dependence has been identified via global and local tests of spatial autocorrelation, the next step is to model this spatial autocorrelation via covariates. Moran's *I* and Lagrange Multiplier diagnostics are widely applied to determine where the covariates considered in a given study fully model the spatial dependence.

[i]. The Moran's *I* Diagnostic for Spatial Error Dependence

The well-known method of Moran's *I* can be extended to the diagnosis of spatial dependence in the presence of covariates. The Moran's *I* test statistic for spatial error dependence in OLS regression residuals takes the following form (Moran, 1950):

$$I = \frac{N}{S}\frac{e'We}{e'e} \qquad [9]$$

where *N* is the number of observations; *S* is the sum of the weights; **e** vector of the residuals from an OLS regression and **W** is the spatial weights matrix (typically row-standardized). Both numerator and denominator equal the ratio of quadratic forms in OLS residuals and they differ only in the specification of the interconnections between the observation (neighboring locations).

The test statistic *I* (under H_o: $\lambda = 0$) is distributed as *F* (*F*-distribution) with *n* and *m* degree of freedom. Since

$$\frac{e'We/S}{e'e/N} = \frac{\chi^2/n}{\chi^2/m} \sim F \ (\text{the central } F \text{ distribution with } n \text{ and } m \text{ degree of freedom}).$$

Critical value (significance level α): F_α (n, m). Test decision: the null hypothesis (H_o) will be rejected if the computed value of $I > F_\alpha(n, m)$.

[ii]. Lagrange Multiplier Diagnostic for Spatial Lag and Spatial Error Dependence

Anselin and Rey (1991) argue for use of Lagrange Multiplier (LM) diagnostics in OLS specifications. Here, there are two basic LM diagnostics. The first is a Lagrange Multiplier diagnostic for spatial lag dependence in the presence of covariates in an OLS model. Spatial dependence in regression models may not only be inflected in the error. Instead it may be accounted by entering a spatial lag WY in the endogenous variable Y. In this case the regression model reads $Y= X\beta + \mu$ with spatial autocorrelation in the error term as spatially lagged dependent variable $\mu = \rho WY + \varepsilon$,

Under the null hypothesis

$$Ho: \rho = 0$$

the standard regression model $Y=X\beta + \varepsilon$ holds, while under the alternative hypothesis

$$H_1: \rho \neq 0.$$

The LM diagnostic for lag dependence test statistic takes the form:

$$LM_{lag} = \frac{\left(e'WY / S^2\right)^2}{(NJ)} \qquad [10]$$

where $NJ= T(WX\beta)'M(WX\beta)/S^2$

$T=tr[(W+W')W]$

$S^2=\frac{e'e}{N}$

$\mathbf{M} = I - X(X'X)^{-1}X'$

N is the number of observations; ***e*** is the vector of OLS residuals;

tr is the matrix trace operator, and **W** is the spatial weights matrix for the spatial lagged dependent variable

Y is the value of dependent variable.

The test statistic LM_{lag} (under $H_0 : \rho = 0$) is distributed as χ^2**(**chi-square**)** with one degree of freedom. Critical value (significance level α): χ^2**(**1, 1-α). Test decision: the null hypothesis (H_o) will be rejected if the computed value of $LM_{lag} > \chi^2$**(**1,1-α).

The Lagrange Multiplier diagnostic for spatial error dependence in the presence of covariates in an OLS models is based on the estimation of the regression model $Y = X\beta + \lambda W\nu + \varepsilon$ with spatially dependent errors term ν**.**

The null hypothesis that there is no spatial error dependence is

H_o: $\lambda = 0$

This means that OLS estimation of the model $Y = X\beta + \varepsilon$ suffices for conducting the LM error test. The alternative hypothesis claims a spatial autoregressive coefficient

H_1: $\lambda \neq 0$

The test statistic is given as

$$LM_{error} = \frac{\left(e'We / S^2\right)^2}{T} \qquad [11]$$

where **W** is the spatial weights matrix for the spatially lagged error term and the rest notations as in (10).

As in the spatial lag model, the test statistic LM_{error} (Under the null H_o: $\lambda = 0$) is distributed as χ^2**(**chi-square**)** with one degree of freedom. Critical value (significance level α): χ^2**(**1, 1-α). Test decision: the null hypothesis (H_o) will be rejected if the computed value of $LM_{error} > \chi^2$**(**1,1-α).

For each diagnostic, the null hypothesis is the absence of the particular form of spatial dependence. If the null hypothesis cannot be rejected on either diagnostic, the OLS specification is sufficient for modeling the spatial dependence estimated via the global and local measures of spatial autocorrelation. If the null hypothesis will be rejected on the Lagrange Multiplier diagnostic for spatial lag dependence, we should proceed by estimating a mixed regressive, spatial autoregressive (spatial lag) model for the spatially lagged dependent variable. If, alternatively, the null hypothesis will be rejected on the Lagrange Multiplier diagnostic for spatial error dependence, we can either proceed with a more fully specified OLS model or a Maximum likelihood spatial error specification (Gimpel and Cho, 2004). Next, we present Lagrange Multiplier diagnostics that are robust to the alternative form of dependence.

[iii]. Robust Lagrange Multiplier Diagnostics for Spatial Lag and Spatial Error Dependence

The robust Lagrange Multiplier diagnostics for OLS models apply Bera and Yoon's (1996) modified Lagrange Multiplier tests to the diagnosis of spatial lag and spatial error dependence in OLS specifications. The robust Lagrange Multiplier diagnostic for spatial lag dependence in OLS model the Null hypothesis

$$\text{Ho: } \rho = 0$$

The test statistic for this test problem is

$$RLM_{lag} = \frac{\left(\frac{e'WY - e'We}{s^2}\right)^2}{(NJ\rho\beta)^{-1} - T} \qquad [12]$$

where

$(N J\rho\ \beta)^{-1}= [T+ (WX\beta)^{'}M (WX\beta)/S^2]^{-1}$, and the other notations are as in (10)

For robust Lagrange Multiplier diagnostic for spatial error dependence in an OLS model null hypothesis is Ho: $\lambda=0$.

The appropriate test statistic for this test problem is

$$RLM_{Error} = \frac{\left[\left(\frac{e'We}{S^2} - T\right) - (NJ\rho\beta)^{-1}(e'WY)/S^2\right]^2}{\left[T - T^2(NJ\rho\beta)^{-1}\right]} \quad [13]$$

where all the notations are as in [10 and 12].

For both cases, the respective test statistic (under the respective null hypothesis) is distributed as χ^2 **(**chi-square**)** with one degree of freedom. Both null hypotheses will be rejected if $RLM_{error} > \chi^2$ **(**1,1-α) and $RLM_{lag} > \chi^2$ (1, 1-α).

The robust Lagrange Multiplier diagnostic for spatial lag (error) dependence tends to reduce power against spatial lag (error) dependence than the unidirectional Lagrange Multiplier diagnostic for spatial lag (error) dependence in the absence of spatial error (lag) dependence. As a result, if the null hypothesis is rejected for the robust LM diagnostic for spatial lag (error) dependence due to the presence of spatial error (lag) dependence, the null hypothesis will also likely be rejected for the non-robust LM diagnostic for spatial lag (error) dependence (Anselin and Rey, 1991). In general, a likelihood ratio test will be employed after estimation to choose the proper spatial regressions specification.

3.3.4. Modeling Spatial Dependence

Fitting Spatial Autoregressive Models for Malaria Incidence

The methods discussed in Sections 3.3.2 and 3.3.3 are based on measuring and testing for spatial autocorrelation that could reveal spatial pattern in the dependent variable. Modeling is facilitated with spatial autocorrelation specifications. In the case of spatial data, here spatial dependence is detected; it is very unlikely that the standard hypothesis of uncorrelated observation is true. The usual map analysis tools and the Scatter plot can provide the first indications that the observed values are more correlated than would be expected under a condition of independence. In this case, global and local spatial autocorrelation tests on the regression residuals warn of the presence of spatial

autocorrelation. If spatial autocorrelation exists, we must specify a model that takes into account the effect of it.

Spatial autoregressive models are the error generating process and operate with spatial weight matrices that specify the strength of interaction between neighboring sites (Cressie, 1993). We will use a spatial autoregressive model to measure the relationships between malaria incidence rate and meteorological variables obtained at a neighborhood sites. Initially, we will identify spatial dependence by diagnostic checking as in Section (3.3.3) and then to incorporate for spatial autocorrelation effects using the spatial regression methods. There are two ways to incorporate spatial autocorrelation in a spatial autoregressive model (we use the notation presented in Anselin, 1988), depending on where the spatial autoregressive process is believed to occur (Haining, 2003).

The first is the spatial lag model, the value of a dependent variable Y at a location is modeled as a function of the independent variables **X** in that location as well as the values of the dependent variable at the neighboring locations, that is, the spatial lag. A spatial lag is basically the weighted average of the dependent variable values at the neighboring locations (Anselin, 1988), included as an additional explanatory variable in the model as shown in the following Equations (14). On the other hand, spatial lag model assumes that the autoregressive process occurs only in the response variable (‘‘lagged-response model’’), and thus includes a term (ρW) for the spatial autocorrelation in the response variable Y, but also the standard term for the predictors and errors ($X\beta + \varepsilon$) as used in an ordinary least squares (OLS) regression. The spatial lag model takes the form:

$$Y = \rho WY + X\beta + \varepsilon \qquad [14]$$

which is equivalent to $Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$

In this equation, **I** is the identity matrix; ρ is the autoregression parameter; **W** the spatial weights matrix; and $\boldsymbol{\beta}$ a vector representing the slopes associated with the predictors in the original predictor matrix **X**; **ε** *is* the vector of errors

In additions to this, ρWY is a spatial lag term, it is essentially a weighted average of the neighboring values of the dependent variable. If the spatial autoregressive parameter (ρ) is significant, the spatial dependency does exist for the dependent variable. In this case, the spatial lag model can yield a more accurate description of relationship between the dependent variable and independent variables (Anselin, 1998).

The spatial error model addresses the spatial autocorrelation existing in the regression residuals [λ] of the OLS models. The value of the dependent variable Y in a location is redefined as a function of the independent variables ***X*** and the regression residuals of the neighboring location, that is, the spatial error. A spatial error is fundamentally a weighted average of the individual residuals of the neighboring locations (Anselin, 1992), which is added into the model as an additional explanatory variable shown in the following on equation (15).

The spatial error model assumes that the autoregressive process occurs only in the error term and neither in response nor in predictor variables. The model is most similar to the conditional autoregressive model (CAR), with no directionality in the error. In this case, the usual OLS regression model $Y = X\beta + \varepsilon$, is complemented by a term $(\lambda W\nu)$ which represents the spatial structure (λW) in the spatially dependent error term (ν). The spatial error model thus takes the form:

$$\mathrm{Y} = X\beta + \lambda W\nu + \varepsilon \qquad [15]$$

where λ is the spatial autoregression coefficient, and the rest notations as in equation [14].

The matrix $(\mathrm{I} - \rho\,\mathrm{W})^{-1}$ being used in the spatial lag models (Equation (14)) incorporate the influence of higher order neighbors. Unlike the Spatial weight matrix **W** which is a sparse matrix with 0s for all the higher order neighbors, this matrix is no longer sparse as a consequence of the inverse operation. Since most of the elements in this matrix have a non-zero value, the influence of higher order neighbors is implicitly considered in the

spatial autoregressive models (LeSage and Pace, 2009). As mentioned earlier, such a spatial weights matrix requires that testing spatial autocorrelation be based upon the division of the study area into regions so that the spatial contiguity of the observations will not be disrupted by a commonly used spatially random selection scheme. Consequently, spatial autoregressive models were fitted using the other dependent variable and indicator variables.

[i]. Spatial Autoregressive Model Assumptions

- ✓ The error terms across different spatial units are correlated with spatial error in OLS regression (the assumption of uncorrelated error terms is violated).
- ✓ The dependent variable in a specific location is affected by the independent variables with neighboring locations (the assumption of independent observations is violated).
- ✓ All diagonal elements of spatial weight matrix **W** are zero.
- ✓ (I- ρ W) and $(I - \lambda W)$ are *n x n* non-singular matrices.

The Jarque-Bera test for normality of residuals, the independent variables will be tested for heteroscedasticity by using Breusch-Pagan test and Likelihood ratio test for spatial dependence to satisfy the basic assumptions of spatial regression analysis (spatial lag model and spatial error model).

(ii) The Likelihood Ratio (LR) Test

The likelihood ratio test is done to statistically confirm the model diagnostic for spatial lag dependence and spatial error dependence. The LR test statistic for the spatial models is given by:

$$LR = [AIC_{ols}] - [AIC_{spatial}] - 2([k_{ols}] - [k_{spatial}]) \quad [17]$$

with 1 (i.e., $[k_{spatial}] - [k_{ols}]$) degree of freedom, where LR is a likelihood value and k is the number of parameters.

If the computed value of LR> $\chi^2(1,\alpha)$ this indicates a significant spatial lagged dependence and spatial error dependence. This is achieved by having addressed the

spatial autocorrelation in the residuals, which might have been caused by the spatial distributions of dependent variable (Anselin, 1996).

[iii]. Methods of Parameter Estimation

The parameters of the spatial lag and error models could be estimated by means of the maximum likelihood (ML) method (that is, the parameters are estimated by maximizing the probability/likelihood of the sample data).

(a) Maximum Likelihood Spatial Lag Estimation

Ord (1975) gives the maximum likelihood methods for estimating the spatial lag and spatial error models. The logarithm of the determinate of the (*NxN*) asymmetric matrix $(\text{I-}\rho W)$ or $(\text{I-}\lambda W)$ does not tend to zero, it constraints the parameter values to their feasible range between the inverse of the smallest and largest eigenvalues of **W**, since for positive autocorrelation, as $\rho \to 1,\ \ln|\text{I} - \rho W| \to -\infty,$ and analogously for λ. The log likelihood functions for spatial lag models:

$$L(\beta, \rho, \sigma^2) = \frac{-N}{2}\ln(2\pi) - \frac{-N}{2}\ln(\sigma^2) + \ln|I - \rho W| - \frac{1}{2\sigma^2}\left[y'(I - \rho W)'\left(I - X(X'X)^{-1}X'\right)(I - \rho W)y\right] \quad [18]$$

In addition to the above Ord (1975) showed how it can be expressed in function of the eigenvalues ω_i of the spatial weights matrix of $|I - \rho W| is\ equal\ to \prod_{i=1}^{N}(1 - \rho\omega_i)$.

Using this simplification, under the normality assumption, the log-likelihood function for the spatial lag model:

$$\text{L}_{\text{lag}} = \sum_{i=1}^{N}\ln(1 - \rho\omega_i) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{(y - \rho Wy - X\beta)'(y - \rho Wy - X\beta)}{2\sigma^2} \quad [19]$$

where ωi are the eigenvalues of the spatial weights matrix **W.**

The first condition for the ML estimators yield nonlinear (in parameters) equations which are solved by numerical methods. The ML estimate of $\boldsymbol{\rho}$ is obtained from a numerical optimization of the concentrated log-likelihood function (Anselin and Bera, 1998):

$$L^{c}_{Lag} = -\frac{N}{2}\ln\left[\frac{(e_* - \rho e_l)'(e_* - \rho e_l)}{N}\right] + \sum_{i=1}^{N}\ln(1-\rho\omega_i) \qquad [20]$$

where e_* and e_1 are, respectively, the residuals from OLS regressions of Y on **X** and from W*Y* on **X** and the rest notation as in Equation (15). Given the maximum likelihood estimate of $\boldsymbol{\rho}$, the parameters, $\boldsymbol{\beta}$, and the error variance, $\boldsymbol{\sigma}^2$, are then easily computed.

Generally, the estimation of parameters shown as follows. The spatial autoregressive models as defined above given by:

$$Y = \rho WY + X\beta + \varepsilon$$

Here

$$\varepsilon = (I - \rho W)Y - X\beta$$

$$\varepsilon = AY - X\beta \qquad [21]$$

where $A = (I - \rho W)$. The joint likelihood of the ε_i is given by (Mead, 1967):

$$L(\varepsilon) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(\frac{-\varepsilon'\varepsilon}{2\sigma^2}\right) \qquad [22]$$

However, it is the Y_i that are observed and not the ε_i. Thus it is the joint likelihood of the Y_i that needs to be maximized and not the function given in Equation [22]. From [21] and [22] we have as the joint likelihood function for Y=y is given by

$$l(y) = |A|\left(\frac{1}{\sigma^2 2\pi}\right)^{n/2}\exp\left\{-\left(\frac{1}{2\sigma^2}\right)[Ay - X\beta]'[Ay - X\beta]\right\} \qquad [23]$$

where |A| is the Jacobian of the transformation from ε to y**.** The eigenvalues facilitate computation of the Jacobian transformation from an autocorrelated to unautocorrelated mathematical space estimation. Let $\omega = \sigma^2$ then, the log-likelihood function is given by

$$l(y)=\log l(y)=\text{constant}-\left(n/2\right)\ln\omega-\left(1/2\omega\right)\left(y'A'Ay-2\beta'X'Ay+\beta'X'X\beta\right)+\ln|A| \quad [24]$$

Minimizing l(y) gives the following solutions:

$$\hat{\beta}=\left(X'X\right)^{-1}X'Z \quad [25]$$

$$\hat{\omega}=\left(\frac{1}{n}\right)\left(Z'Z-Z'X\left(X'X\right)^{-1}X'Z\right) \quad [26]$$

where $Z=\left(I-\rho W\right)y=Ay.$

Let $M=I-X\left(X'X\right)^{-1}X'$ be the symmetric and idempotent matrix. Now $\hat{\rho}$ maximizes

$$l(y)=l\left(y:\hat{\rho},\hat{\omega},\hat{\beta}\right)=\text{constant}-\left(n/2\right)\ln\hat{\omega}-\ln|A| \quad \textbf{[27]}$$

Using the simplified expression for $\ln|A|$, $\hat{\rho}$ minimizes

$$-\left(2/n\right)\sum_{i=1}^{n}\ln\left(1-\rho\omega_i\right)+\ln\hat{\omega}$$

But $\hat{\omega}=\left(1/n\right)Z'MZ$

$$=\left(1/n\right)y'A'MAy$$

$$=\left(1/n\right)y'\left(I-\rho W\right)'M\left(I-\rho W\right)y$$

$$=\left(1/n\right)\left[y'My-2\rho y'WMy+\rho^2\left(Wy\right)'MWy\right]$$

Thus $\hat{\rho}$ minimizes

$$-\left(\frac{2}{n}\right)\sum_{i=1}^{n}\ln\left(1-\rho\omega_i\right)+\ln\left(\left(\frac{1}{n}\right)\left[y'My-2\rho y'WMy+\rho^2\left(Wy\right)'MWy\right]\right) \quad \textbf{[28]}$$

The value of ρ that minimizes the function can be obtained by a direct search procedure (Keith, 2010). Finally the asymptotic variance-covariance matrix for the estimators of the parameters of the mixed endogenous-exogenous procedures is given by:

$$V(\hat{\omega},\hat{\rho},\hat{\beta}) = \omega^2 \begin{bmatrix} n/2 & \omega tr(B) & 0' \\ \omega tr(B) & \omega^2 tr(B'B) + \omega\beta' X' B' BX\beta - \alpha\omega^2 & \omega X' B' X\beta \\ 0 & \omega X' B' X\beta & \omega X' X \end{bmatrix}^{-1} \quad [29]$$

where $B=A^{-1}W$ and the rest are as defined above.

(b) Maximum Likelihood Spatial Error Estimation

The maximum likelihood estimation for the spatial error model employs the error term into log-likelihood function as follows:

$$L_{Error} = \sum_{i=1}^{N} \ln(1-\lambda\omega_i) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\sigma^2 - \frac{(y-X\beta)'(I-\lambda W)'(y-X\beta)}{2\sigma^2}\right) \quad [30]$$

As in the spatial lag model, the ML estimate can also be solved numerically and the estimates are obtained from the optimization of a concentrated log-likelihood function.
The concentrated log-likelihood in the parameter, λ, is given by:

$$L^{c}_{Error} = -\frac{N}{2}\ln\left[\frac{e'e}{N}\right] + \sum_{i=1}^{N}\ln(1-\lambda\omega_i) \quad [31]$$

In this formula, **e'e** is the residual sum of squares from the regression of the spatially filtered variables $Y - \lambda WY$ and $\mathbf{X} - \lambda\mathbf{WX}$ (Anselin, 1992) and the rest as in (20)**.** The parameters, β, and error variance, σ^2, are then computed, given the maximum likelihood estimate of λ .

Chapter Four

Results and Discussion

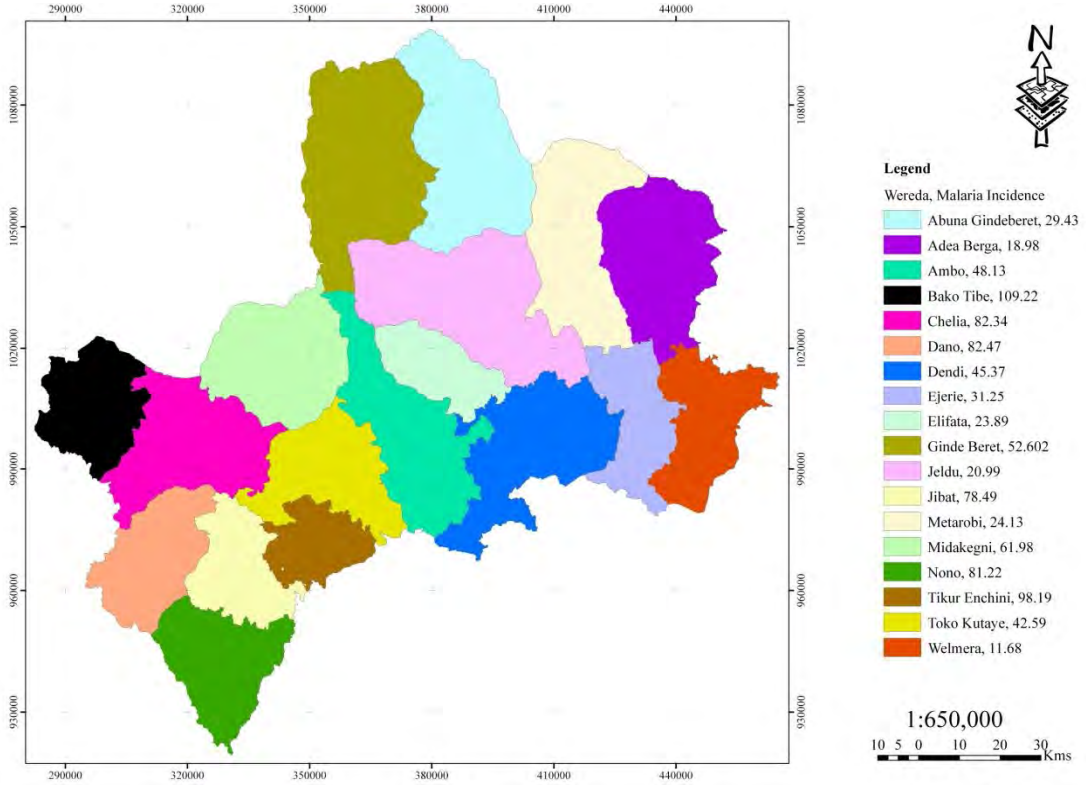
Results for tests of spatial autocorrelation in the malaria incidence rate to determine the distribution pattern of malaria and modeling spatial autoregressive model will be presented in this chapter.

The total number of malaria cases in the study area was 113, 389 with overall malaria incidence rate of 49.97 per 1000 in the year of 2009 (Table 2, Appendix).

4.1 Spatial Distribution of Malaria Incidence by Woreda

Figure 4.1 shows the spatial distribution of the proportion of malaria incidence in the study area. The highest incidence rate was observed in Bako-Tibe (109.22) while the lowest was in Walmara (11.68). In general, a higher proportion of malaria incidence was observed in the West part of the study area while the eastern part has low incidence.

Figure 4.1: Spatial Distribution of the Malaria Incidence rate in West Shoa Zone



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 section (3.3), were used. In this section, our focus is on their application to particular data analysis, the essential task being to seek for spatial pattern. The contiguity spatial weight matrix, in West Shoa Zone is presented in Appendix (Table 3).

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* and local Getis and Ord G_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* and local Getis and Ord G_i^* statistic (Anselin, 1995). First, the

global Moran's *I* and Geary's *C* test statistics were computed to test the null hypothesis ($H_{o:} \rho = 0$) of no significant clustering of malaria incidence in the entire study region ($\alpha = 0.05$). The test was repeated using diagnostic for spatial dependence to validate the consistency of results.

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 incidence, to seek for spatial pattern or to diagnosis for spatial dependence in regression model. The tests are performed under the assumption of normality and the null hypothesis states spatial independence (uncorrelated of error terms) for the data under consideration. 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 incidence (Table 4.1). The test results are also shown in Moran's *I* scatter plot (Figure 4.2). In addition to this, Moran's scatter plot for threshold distance and K-nearest neighbor also indicate significant global spatial autocorrelation (see Appendix Figure B and Figure C).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

Assumption	Coefficient	Observed	Expected	Dev Std	Z	Pr > |Z|
Normality	Moran's *I*	0.6562	-0.0588	0.1004	7.12	< .0001*
Normality	Geary's *C*	0.0596	1.0000	0.1795	-5.24	< .0001*

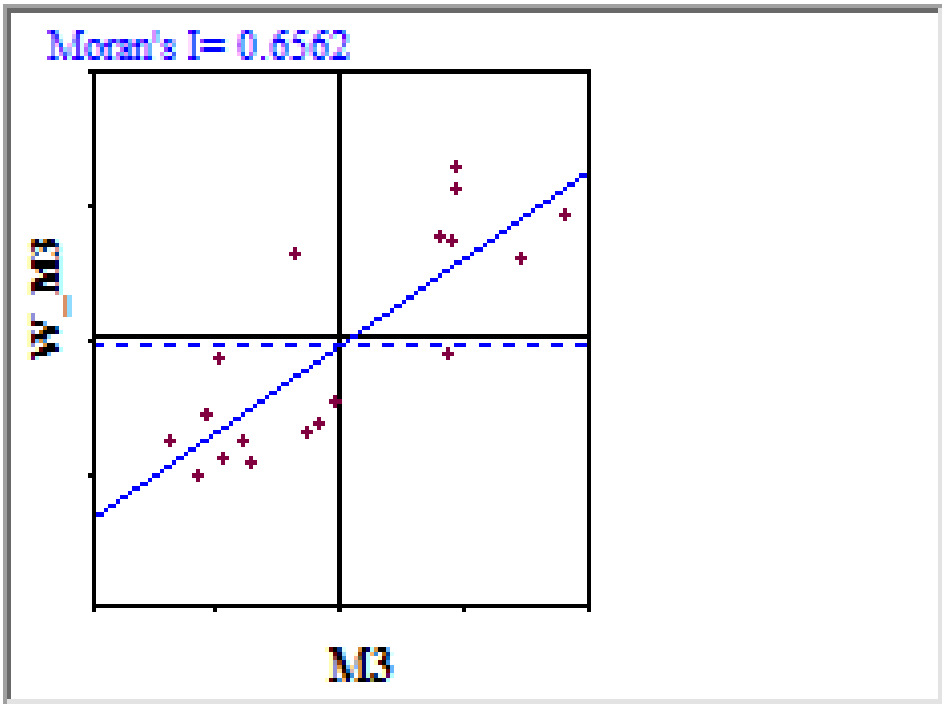
Assumption	Coefficient	Observed	Expected	Dev Std	Z	Pr > |Z|
Randomization	Moran's *I*	0.58036	-0.0588	0.2390	2.680	0.0005*
Randomization	Geary's *c*	0.00318	1.0000	0.346 0	-2.880	0.0003*

***significant at 0.05 level**

Based on the P-values of the reported Moran's *I* and Geary's *C* coefficients, we can reject the null hypothesis of no spatial autocorrelation. Furthermore, the computed Z- statistic for Moran's *I* is positive and for Geary's *C* is negative indicating the existence of significant positive spatial autocorrelation (clustering).

In order to visualize global spatial autocorrelation we use Moran's scatter plot under the assumptions of normality (Figure 4.2). It shows malaria incidence can be assumed to occur with unequal distribution at all locations. This approach is referred to as the non-randomization assumption.

Figure 4.2: Global Moran's *I* Scatter Plot for Malaria Incidence



In Figure 4.2 **M3** represents malaria incidence, while **W-M3** represents spatial lag malaria incidence based on standardized spatial weigh matrix. The figure also shows that malaria incidence is spatially correlated with neighboring values.

4.2.2 Local Moran's *I* Test Statistic for Spatial Autocorrelation

In the global test statistic (as discussed in Section 4.2.1) the result indicates that there is a significant positive spatial autocorrelation (clustering). Local statistics are used to identify where high/low values cluster.

Local Moran's *I* and local Getis and Ord *Gi** statistics are computed to test the null hypothesis of no local spatial clustering among malaria incidence at neighboring Woredas. To find spatial outliers, we use local Moran's *I* (Moran's Scatter plot). It shows us where values cluster spatially and where values are very different from neighbors (outliers).

Table 4.2 shows results of local Moran *I* as a function of neighboring values. The test results indicate that there is statistically significant local clustering of malaria incidence at 5% level of significance (see also Figure 4.3 the first and the third quadrants). Statistically significant local clustering of malaria incidence is detected in all the woredas except in the three Woredas out of a total of 18 woredas (see also Figure 4.3 the second and fourth quadrants). Results indicate that malaria incidence follows spatial pattern (Figure 4.3 and Figure 4.4).

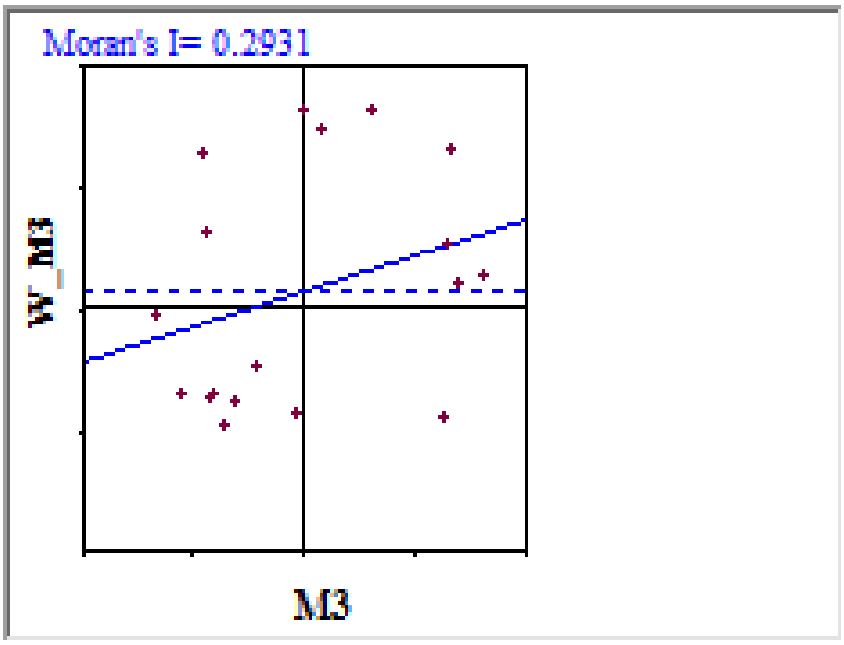
Table 4.2: Results of Local Moran's *I* Test

ID	Woreda	Observed	Expected	Dev Std	Z	P
1	Abuna-Gindeberet	0.5363	-0.17647	0.2546	2.800	0.0051
2	Ada-Berga	0.0097	-0.17647	0.0520	3.580	0.0003
3	Ambo	0.0067	-0.35294	0.1207	2.980	0.0028
4	Bako-Tibe	0.5500	-0.0588	0.0571	10.66	0.0001
5	Cheliya	0.2610	-0.35294	0.0811	7.570	0.0001
6	Dano	0.2700	-0.17647	0.0472	9.460	0.0001
7	Dendii	0.0072	-0.23524	0.0806	3.010	0.0026
8	Ejerie	0.4500	-0.294116	0.1038	7.170	0.0001
9	Elfata	0.2940	-0.17647	0.0485	9.690	0.0001
10	Ginde-Beret	-0.2089	-0.17647	0.6188	-0.0524	0.9580
11	Jeldu	0.6010	-0.41176	0.2944	3.440	0.0006
12	Jibat	0.7310	-0.23530	0.2351	4.110	0.0001
13	Meta-Robi	0.6048	-0.23530	0.2334	3.600	0.0003
14	Midakegn	0.8668	-0.17640	0.2478	4.210	0.0001
15	Tikur-Enchini	0.000119	-0.11765	0.0335	3.510	0.0004
16	Nono	0.5764	-0.11765	0.1385	5.010	0.0001
17	Toke-Kutaye	-0.3063	-0.23529	0.0740	-0.959	0.337
18	Walmara	-0.7014	-0.11765	0.3001	-1.945	0.051

In Toke-Kutaye, Gende-Beret and Walmara Woredas, the spatial correlation is negative (observed < expected). 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 > expected) (Table 4.2).

To better visualize local values, we transform our woreda data into area-based map and draw the scatter plot. Figure 4.4 shows the spatial distribution of standardized local G_i^* for malaria incidence. The map also shows Woredas with significant clustering and Moran's scatter plot values in the first and third quadrants in Raw Umber and Gray colors. In addition to this, from the appendix (Figure A) Moran's scatter plots indicate that malaria incidence is positively spatially correlated with mid-land zone, hot zone, minimum temperature and rainfall, while maximum temperature and cold zone are negatively spatially correlated. These variables may be considered responsible for existence of positive and negative spatial autocorrelation, respectively.

Figure 4.3: Local Moran's *I* Scatter Plot for Malaria Incidence



In Figure 4.3, the four quadrants graph provide a classification of four types of spatial autocorrelation; high-high (upper-right), low-low (lower-left), for positive spatial

autocorrelation; high-low (low-right) and low-high (upper-left), for negative spatial autocorrelation. Those values in the first and third quadrants indicate local clustering of malaria incidence, while the rest did not show local clustering.

4.2.3. Local G_i^* Test for Spatial Autocorrelation

The local Getis and Ord statistic, G_i^*, identified significant local clustering of high (hot spots) or low (cold spots) values of malaria incidence (spatial weighted malaria incidence were standardized by the total malaria incidence at risk in each Woreda) surrounding each Woreda within a nearby location. The spatial weight defined the neighborhood search for each Woreda with nearby locations being expected to have similar values. The observed values were compared with the expected values to indicate if the degree of clustering of malaria cases in the vicinity of a particular Woreda was greater or less than expected by chance. To correct for multiple comparisons when using G^*_i, significance levels were adjusted according to Getis and Ord's criteria. The result of local G_i^* is shown in the Table 4.3 and Figure 4.4. Note that the local cluster map (Figure 4.4) supports the result of Table 4.2.

The significance of the G^*_i statistic is assessed by standardized Z value. A positive and significant Z value for the G_i^* statistic indicates spatial clustering of high values. A negative and significant Z value for the G_i^* statistic indicates spatial clustering of low values (Getis and Ord's, 1992). Accordingly, Bako-Tibe, Nono, Cheliya, Jibat, Midkegy, Tikur-Enchini and Dano show clustering of high values, while the rest except Toke-Kutaye, Walmara and Gende-Berat show clustering of low values (Table 4.3).

Table 4.3: Results of Local G_i^* Test

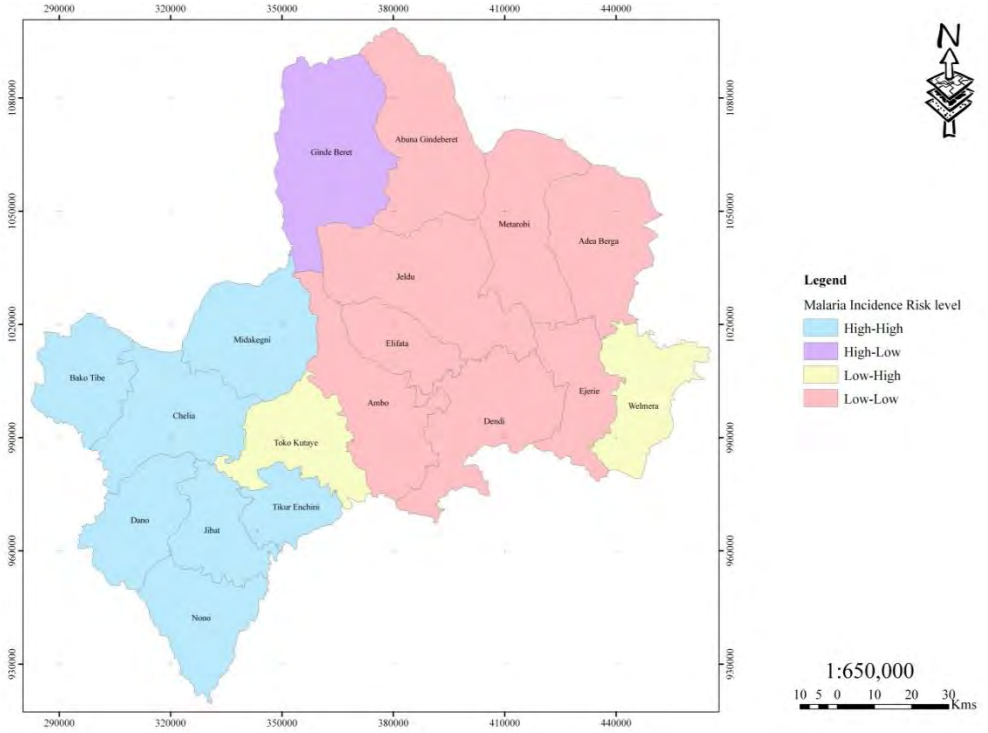
No	Woreda	Observed	Expected	Dev Std	Z	P
1	Abuna-Gindeberet	-0.00108	0.22222	0.08270	-2.7000	0.0070
2	Ada-Berga	-0.8799	0.22222	0.16623	-6.6300	0.0001
3	Ambo	-0.0543	0.38889	0.12043	-3.6800	0.0002
4	Bako-Tibe	0.2663	0.11111	0.05407	2.8700	0.0041
5	Cheliya	0.5280	0.38889	0.01558	8.9300	0.0001
6	Dano	0.2210	0.22222	1.7256	7.0700	0.0001
7	Dendi	-0.7082	0.27778	0.24466	-4.0300	0.0001
8	Ejerie	-0.6991	0.33333	0.3236	-3.1900	0.0014
9	Elfata	-0.4650	0.22222	0.08075	-8.5100	0.0001
10	Ginde-Beret	0.3160	0.22222	0.1044	0.8980	0.3691
11	Jeldu	-0.00318	0.38889	0.13613	-2.8800	0.0039
12	Jibat	0.8534	0.27778	0.05885	9.7800	0.0001
13	Meta-Robi	-0.9330	0.27778	0.12218	-9.9100	0.0001
14	Midakegn	0.5804	0.22222	0.13365	2.6800	0.0075
15	Tikur-Enchini	0.6750	0.16667	0.05699	8.9200	0.0071
16	Nono	0.5270	0.22222	0.0311	9.8000	0.0001
17	Toke-Kutaye	-0.0230	0.27778	0.17129	-1.7560	0.0791
18	Walmara	0.4446	0.16667	0.13172	2.1100	0.0345

Cluster mapping helps in classifying issues such as spatial aspects of both internal and external correlations for leading malaria incidence. This is of great aid in assessing spatial risk factors, which in turn facilitates the planning of the most advantageous pattern of malaria distributions and implantations of effective intervention services (Tsai *et al.*, 2009). This study also has practical utility in making the cluster map that can be used to communicate malaria control easily. The cluster map has been used to define a given Woreda within which interventions are scaled and planned according to malaria distribution intensity. This will involve anti-malarial treatment, long-lasting insecticide

treated nets (LLINs) and indoor residual spraying (IRS) in the high cluster with neighboring area.

The local G_i^* output which indicates the type of statistically significant pattern encountered (HH = cluster of high values; LL = cluster of low values; HL = outlier: a high value surrounded by low values; LH = outlier: a low value surrounded by high values) is presented in the Figure 4.4.

Figure 4.4: Local G_i^* Clustering Map of Malaria Incidence in West Shoa Zone, Ethiopia



From Figure 4.4, it can be observed that Bako-Tibe, Nono, Cheliya, Jibat, Midkegy, Tikur-Enchini and Dano woredas are clustered as high value, while the rest except Toke-kutaye, Walmara and Ginde-Beret are clustered as low value.

4.3 Diagnostic for Spatial Dependence

The global and local test of spatial autocorrelations indicated spatial clustering of malaria incidence. Next, spatial lag model and spatial error model are used to measure the relationships between malaria incidence and meteorological variables obtained at a neighborhood. When the data are spatially structured, OLS scores can be biased and their significance inflated as discussed in the methodology part. A diagnostic statistic indicating problems in OLS regression with spatial data is the degree of spatial non-randomness of residuals; however, a common approach is to filter out or to treat the local spatial information as "noise." Thus, spatial regression has been used under the assumptions of spatial correlation structures that apply equally across the data set and the results of diagnostic are presented in Table 4.4. The table summarizes the results of five test methods that were used to assess the spatial dependence of the model.

First, Moran's *I* score of 0.486243 indicated the existence of a strong spatial autocorrelation of the residuals. Then, the simple LM test for a missing spatially lagged dependent variable (Lagrange Multiplier (lag)) and Robust LM (lag); the simple LM test for spatial error dependence (Lagrange Multiplier (error)) and Robust LM (error) were used.

Table 4.4: Diagnostics for Spatial Dependence

Test	**(Moran's I) /d.f.**	**Value**	**Prob**
Moran's I (error)	0.486243	4.7307713	0.0000022*
Lagrange Multiplier (lag)	1	16.5705634	0.0000469*
Robust LM (lag)	1	10.138098	0.0014524*
Lagrange Multiplier (error)	1	0.3148891	0.5746954
Robust LM (error)	1	2.0790861	0.1493296

***significant at 0.05 level**

The results show that both LM tests of the lag and RLM lag are significant, indicating the presence of spatial lag dependence of malaria incidence, but both LM error and RLM error tests were found to be insignificant. The robust tests help us understand the type of spatial dependence. The robust measure for lag is significant, but the robust error test is insignificant, indicating less evidence for the existence of spatial error model. On the other hand the presence of spatial autocorrelation in the data violates the independence assumption of the OLS regression and suggests explicit treatment with a spatial autoregressive model.

4.4. Fitting Spatial Autoregressive Models

The results of the tests for spatial autocorrelation and diagnostic for spatial dependence indicate the dependence of observations and correlation of error terms at one location with the errors at nearby locations, resulting in the clustering of similar values among nearby locations, or spatial autocorrelation. In such a case, the most commonly used method is the spatial regression model.

This model incorporates spatial autocorrelation effects using the spatial regression methods. There are two ways to incorporate spatial autocorrelation in autoregressive model as briefly discussed in methodology. One is to model spatial autocorrelation in the error term as a spatially lagged dependent variable:

$$\mu = \rho WY + \varepsilon \qquad [32]$$

and the other $\lambda W\nu$ which represents the spatial structure (λW) in the spatially dependent error term ν.

where **W** is the spatial weights matrix characterizing the spatial relationship between every pair of observations; ρ is the spatial autoregressive parameter characterizing spatial autocorrelation; and **ε** is the independent and normally distributed error term with a constant mean of zero and constant variance. The diagnostic for spatial dependence indicates spatial error model is insignificant. The spatial lag model can be fitted for

malaria incidence data considered in this study. The spatial lag model as given in equation [14] can be expressed as:

$$Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon \qquad [33]$$

The OLS method is no longer appropriate for estimating the spatial lag model. Instead, the maximum likelihood estimation method (MLE) should be used. The results obtained by using GeoDa 0.9.5i software are presented in Table 4.5.

As expected, mid-land zone, hot zone, rainfall, maximum and minimum temperature are significant factors and are positively associated with malaria incidence, while the cold zone is insignificant and negatively associated. The spatial patterns of residuals were also analyzed by creating a Moran's *I*. The value of the Moran's *I* test statistic for the OLS residuals is 0.486243 **(**Table 4.4**)** and for the lag residuals is 0.733686 **(**in Appendix Table 4). This implies that the residuals of the spatial lag models are dependent thereby satisfying the fundamental assumption about the correlation of the error terms.

The regression diagnostics reveal considerable normality and heteroscedasticity, as well as high spatial autocorrelation. Following the steps outlined in Table 4.3, we conclude that a spatial lag model is the proper for to indicate relation between malaria incidence and meteorological variables. Both LM lag and RLM lag are significant, while both LM error and RLM error statistic are insignificant. This sets the stage for the estimation of the spatial lag model.

Table 4.5: Results of Spatial Lag Model Estimation

Summary Output: Spatial Lag Model-Maximum Likelihood Estimation

Variable	Coefficient	Std. Error	t-Statistic	Probability
W_Malaria	0.7336859	0.1231336	5.958456	0.000000*
CONSTANT	0.2944724	1.570045	0.1875566	0.851224
Cold	-0.2031315	0.3786051	-0.536526	0.5915950
Mid-land	0.5259689	0.08205635	6.40985	0.000000*
Hot	0.2810949	0.05009784	5.610918	0.000000*
Rainfall	0.3852516	0.1380781	2.790098	0.0005269*
Max.Temperature	0.8358826	0.07850672	10.64728	0.000218*
Min.Temperature	0.5587783	0.0979505	5.704701	0.000000*

***significant at 0.05 level**

From Table 4.5 it can be observed that, mid-land zone, hot zone, rainfall, minimum temperature and maximum temperature may be linked to one of the reasons that caused the similarity in the malaria distributions since they are significant. Accordingly, the fitted spatial lag model equation is given by:-

$$Y = 0.733X_1 + 0.526X_2 + 0.281X_3 + 0.385X_4 + 0.836X_5 + 0.559X_6$$

where Y is the malaria incidence

X_1 is the spatially lagged malaria incidence

X_2 is the mid-land zone

X_3 is the hot zone

X_4 is the rainfall

X_5 is the maximum temperature

X_6 is the minimum temperature

In the model, the distribution of malaria incidence within one province is significant associated with variation of temperature, rainfall, hot zone and mid-land zone of neighboring Woredas as determined by standardized spatial weight matrix. This is

assumed that malaria incidence in a given Woredas is associated with climatic conditions of the neighboring Woredas.

The coefficient for the association between maximum temperature (0.836) and malaria incidence was greater than that for the association between minimum temperature (0.559) and malaria incidence. This indicates that maximum temperature seems to play a more important role in the distribution of the disease than minimum temperature does. A rise of minimum temperature, in some locations, accelerates the distribution dynamics of malaria. Maximum temperature would increase the rate of mosquito emergence from breeding places, and in the presence of rainfall increased humidity results in longer survival of the vector to transmit the parasite (Hay *et al.*, 2000). Spatial shift may occur in both tropical and temperate regions.

Table 4.6: Summary Output of Regression Diagnostics for Spatial Lag Model

Diagnostics for Heteroskedasticity			
Random Coefficients			
Test	**DF**	**Value**	**Prob**
Breusch-Pagan test	**6**	21.01461	0.0018236*
Diagnostics for Spatial Dependence			
SPATIAL LAG DEPENDENCE			
Test	**DF**	**Value**	**Prob**
Likelihood Ratio Test	1	16.5727	0.0000468*

significant at 0.05 level

Note that, besides the information from methodology part of the variables in the study, the spatial lag term of malaria incidence, W-malaria, appeared as additional indicator (its coefficient parameter (ρ) reflects the spatially lagged dependence inherent in malaria incidence, measuring the average influence on observations by their neighboring observations). It has a positive effect and it is highly significant. And in the likelihood ratio test of spatial lag dependence, results indicate that significant spatial lag dependence (Table 4.6).

4.5. Test for Normality of Residuals

To assess departure from normality, the Jarque-Bera test was used (Jarque and Bera, 1980). The null hypothesis of the Jarque-Bera test is that the skewness and kurtosis are jointly zero. In other words the data set tested comes from a normal distribution. This test resulted in a Jarque-Bera statistic of 0.8595699 with a p-value of 0.6506490 indicating that we fail to reject the null hypothesis at the 95 percent confidence level and therefore confirming that the distribution of the OLS residuals is normal and maximum likelihood method can be applied (Figure D in appendix) and also Table 1 in the appendix indicates that there is no series problem with multicollinearity. The primary concern is that as the degree of multicollinearity increases, the regression model estimates of the coefficients become unstable and the standard errors for the coefficients can get wildly inflated. For this reason, we will explore variance inflation factor (VIF) by using SPSS commands that help to detect multicollinearity. As a rule of thumb, a variable whose VIF values are less than 10 may merit further investigation. Tolerance, defined as 1/VIF, is used by many researchers to check on the degree of collinearity.

Table 4.7: Normality Test of Residuals

Regression Diagnostics

Test On Normality of Errors			
Test	**DF**	**Value**	**Prob**
Jarque-Bera	2	0.8595699	0.6506490

Figure A in the Appendix displays the Moran's *I* scatter plot for maximum temperature (-0.4608), hot zone (0.2058), rainfall (0.1094), minimum temperature (0.5179) and mid-land zone (0.4473), calculated using neighborhood spatial weight matrix. The figure also shows that the spatial lagged dependence of the malaria incidence (represent by **W- M3**) is clearly positively spatially correlated with Mid-land zone, hot zone, rainfall and minimum temperature. This implies that mid-land zone and hot zone areas have higher risk levels of malaria incidence. In addition to this, Moran's *I* scatter plot of maximum

temperature (-0.4608) and cold zone (-0.4351) were negatively correlated with spatially lagged malaria incidence (**W-M3**). From these results we suggest that the humidity and higher altitude do have a significant statistical relationship with malaria incidence abundance (since climatic condition varies with elevation). The Woredas with a similar level of humidity and altitude may have conditions that are related to the number of incidence (since rainfall, temperature, hot zone and mid-land zone are significant). As a result, the distribution of malaria in woredas with similar humidity and elevation may be similar.

4. 6. Summary

Using the spatial autocorrelation, our study identified significant spatial clusters where the risk of malaria was higher or lower. Having identified spatial clustering in the distribution of malaria cases, the next step was to investigate the underlying individual, Woreda and meteorological factors that characterize spatial distribution of malaria. Areas characterized by mid-land zone and hot zone were strongly associated with the risk of malaria and risk of spatial clustering. In addition, maximum and minimum temperatures were found to be significant, indicating strong relationship between temperature and malaria incidence.

Furthermore, local spatial statistics were used to test the spatial dependency in the patterns of malaria distribution, detect pockets of disease (Figure 4.3 and Figure 4.4), and identify the relevant spatial scale at which local cluster of malaria occurs. In order to better understand the factors associated with spatial differentials, with malaria incidence distribution were analyzed. The result helps to identify Woredas with malaria burden. The spatial lag model was selected as the appropriate spatial autoregressive model, to account for the spatial autocorrelation (weighted average of the malaria incidence at the neighboring woreda, i.e., spatially lagged malaria incidence). Based on the result it was concluded that, the association between malaria incidence and meteorological variables (minimum temperature, maximum temperature, rainfall, hot zone and mid-land zone) were significant (Table 4.5).

Chapter Five

Conclusions and Recommendations

This study describes the spatial pattern of malaria distribution in West Shoa Zone using routinely collected individual patient morbidity data from health care facilities and meteorological data.

5.1 Conclusions

The results of this study show that the incidence of malaria in West Shoa Zone exhibits a spatial pattern which is dependent on some meteorological variables. The incidence of malaria in the study area is significantly clustered indicating high levels in the western part of the zone and low levels in central and eastern part of the Zone.

The global Moran's *I* and Geary's *C* test statistics show significant clustering (spatial autocorrelation) among neighboring Woredas in the study area. In addition to this, significant local clustering of malaria incidence occurs among Woredas within neighboring woredas. The results of local test statistics suggest that there is significant clustering of malaria incidence. Local risk factors such as temperature, cold-zone, hot-zone, mid-land zone, and rainfall as explained by spatial lag model might all be important in explaining the observed local clustering of malaria incidence.

5.2 Recommendations

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.

References

Abellana**,** R., Carlos, A., John, A., Francisco, S., Delino, N., and Ariel, N. (2008). Spatio-seasonal modeling of the incidence rate of malaria in Mozambique. *Malaria Journal, Vol. 7* ,can be found online at: http://www.malariajournal.com/content/7/1/228

Ali, H., Neal, A. and Jonathan, C. (2008). Modeling of malaria temporal variations in Iran**,** *Tropical Medicine and International Health, vol. 13: 1-4.*

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis, 20: 8-17.*

Anselin, L. (1988). Spatial econometrics: Methods and models. New York, New York.

Anselin, L. (1992). "SpaceStat Tutorial: A Workbook for Using SpaceStat in the analysis of Spatial Data." Typescript. University of Illinois at Urbana-Champaign, pp. 8-67.

Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis, 27: 93-105.*

Anselin, L. (1996). The Moran scatter plot as an exploratory spatial data analysis tool to assess local instability in spatial association. pp. 111-125.

Anselin, L. (1998). ''Interactive Techniques and Exploratory Spatial Data Analysis.'' In Geographical Information Systems: Principles, Techniques, Management and Applications, *pp. 251–64.*

Anselin, L. and Bera, K. (1998). 'Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics', in: A. Ullah and D. Giles (eds), *Handbook of Applied Economic Statistic,s* eds. Aman Ullah and David E.A. Giles. New York: Marcel Dekker.

Anselin, L. and Rey, S. (1991). "Properties of Tests for Spatial Dependence in Linear Regression Models" *Geographical Analysis 23(2): 11-34.*

Basel, P. (2008). Development of Bayesian geostatistical models with applications in malaria epidemiology, *Malaria Journal, vol. 9:4-6.*

Bera., K., Raymond, F., and Yoon, J. (1996). "Simple Diagnostic Tests for Spatial Dependence." *Regional science and Urban Economics, 26(3): 30-76.*

Briet, J., Penelope, V. and Dissanayake, M. (2008). Models for short term malaria prediction in Sri Lanka, *Malaria Journal, 7 (76): 4-7.*

Brooke, K., Sian, C., Joseph, K. and Sarah, P. (2004). Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya: *Tropical Medicine and International Health, Vol. 9: 2-8.*

Brown, D. (1995). Spatial statistics and GIS applied to internal migration in Rwanda, Central Africa. In: S.L. Arlinghaus and D.A. Griffith (eds), Practical handbook of spatial statistics. Boca Raton, Florida: CRC Press.

Chaix, B., Juan, M., John, L. and Pierre, C. (2005). Comparison of a Spatial Perspective with the Multilevel Analytical Approach in Neighborhood Studies: The Case of Mental and Behavioral Disorders due to Psychoactive Substance Use Malmo¨, Sweden, 2001, *American Journal of Epidemiology*, *Vol. 162:2-3.*

Chaix, B., Juan, M. and Pierre, C. (2004). Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilization in France. *Journal of Epidemiology Community Health, Vol. 59:10.*

Chou, Y. (1997). Exploring Spatial Analysis in Geographic Information Systems. Onward Press, Santa Fe.

Chowell, G., Cesar, V., Ananias, A., and Ellis, M. (2009). The spatial and temporal patterns of falciparum and vivax malaria in peru: *Malaria Journal, Vol. 8:2-3.*

Cliff , A. and J.K. Ord (1972). Testing for spatial autocorrelation among regression residuals, *Geographic Analysis, 4:267-284.*

Cliff, A. and J.K.Ord (1981). Spatial processes - models and applications. (London: Pion).

Cressie, N. (1993) . Statistics for spatial data. Revised Edition. John Wiley & Sons, Inc. New York.

Federal Ministry of Health (2008). National malaria indicator survey in Ethiopia

Federal Ministry of Health (2009). National Strategic Plan for Malaria prevention, Control and Elimination in Ethiopia 2010–2015. Addis Ababa: Federal Democratic Republic of Ethiopia Ministry of Health.

Geary, R. (1954). The contiguity ratio and statistical mapping. The Incorporated Statistical 5, 11545.

Getis, A. and Ord, J. (1992).“The Analysis of Spatial Association by Use of Distance Statistics.” *Geographical Analysis 24(3) :189-206.*

Gimpel, F., James, G., and Cho, W. (2004). “The Persistence of White Ethnicity in New England Politics.” *Political Geography 23: 997-1008.*

Githeko, K., John, M., and Peter, K. (2006). Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control. *Malaria Journal, vol. 5:2-3.*

Gosoniu, L., Andre, M., and Penelope, V. (2008). Bayesian Geostistical modeling of malaria Indicator survey data in Angola, *Journal of Epidemology, Vol.9: 8-12.*

Grillet, M., Roberto, B. and Marie, J. (2010): Disentangling the Effect of Local and Global Spatial Variation on a Mosquito-Borne Infection in a Neotropical Heterogeneous Environment, *Am. Journal. Trop. Med. Hyg.,82(2): 1-8.*

Guofa, Z., Noboru, M. and Githeko, K., (2003): Association between climate variability and malaria epidemics in the East African highlands, *Journal of Malaria, vol. 6:1-2.*

Haining, W. (2003). Econometric Analysis, 5th Ed. Upper Saddle River, NJ: Prentice Hall.

Hans, P. (2008). Development of Bayesian geostatistical models with applications in malaria epidemiology in Tanzania. *Pp. 3-5.*

Hay, S. I., Cox, J., Rogers, D. J., Myers, M. F. & Snow, R. W. (*2002*). Nature *415* **:***905, pmid:11859368* CrossRefMedline

Ingrid, P., Luisa, N., Wafaa, E., and Awash, T. **(**2009**).** A Temporal-Spatial Analysis of Malaria Transmission in Adama, Ethiopia, *the American Journal of Tropical Medicine and Hygiene*, *Vol. 86(6): 3-4.*

Jarque, C. and Bera, K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters 6: 255-259.*

Jeong, W., and Gluck, M. (2002). Spatial maps with auditor and haptic display. Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan.

Kazembe, L., Immo K., Timothy, H., and Brian, L. (2006). Spatial analysis and mapping of malaria risk in Malawi using point-Referenced prevalence of infection data, *Internation Journal Health Geogr.,vol.41:3-4.*

Keith, O. (2010). Estimation Methods for models of spatial interaction. *Journal of the Americal Statistical Association*, *Vol. 70:18-20.*

LeSage, J., and Pace, K. (2009). Introduction to spatial econometrics. Boca Raton, Florida: CRC Press.

LeSage, J. (1998). "Bayesian Estimation of Spatial Autoregressive Models." *International Regional Science Review 20: 115-126.*

Lengerler, G. (2004). Operations, costs and cost-effectiveness of insecticide treated net programs and indoor residual spraying programs, *pp. 15-17.*

Li Li, S., Ling, B., and Guiyun, Y. (2008). A study of the distribution and abundance of the adult malaria vector in western Kenya highlands, *Journal of malaria, vol. 6:1-2.*

Mead, R., (1967). "A Mathematical Model for the Estimation of Interplant Competition," *Biometrics, 23: 190-205.*

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika 37:17-23.*

Noor**,** A., Archie, C.**,** Peter, W. and Grainne, M. (2008). Spatial prediction of Plasmodium falciparum prevalence in Somalia, *Malaria Journal, Vol. 7:1-2.*

Ord, J. (1975). Estimation methods for models of spatial interaction, *Journal of the American Statistical Association, 70: 120-126.*

Ord, J. and Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis, 27 (4): 286-306.*

President's Malaria Initiative (2010). Malaria Operational Plan (MOP), Ethiopia.

Richard, C., Kamini, N. and Donald, R. (2000). Spatial targeting of interventions against malaria,*World Health Organization, 78(12): 1*

Riedel**,** N., Penelope, V., John, M., Laura, G., Victor, M., and Rick, W. (2010). Geographical patterns and predictors of malaria risk in Zambia: Bayesian

geostatistical modeling of the 2006 Zambia national malaria indicator survey, *Malaria Journal, Vol.37: 1-2.*

Salim, M. (2000). Malaria control strategies in the kilombero valley, Tanzania, *Journal of malaria, vol. 9: 2-5.*

Smith, T. (2003). Development of spatial statistical methods for modeling point-Referenced spatial data in malaria Epidemiology, *Journal of Malaria, Vol.7:1-25.*

Statistical Abstract (2007). Federal Democratic Republic of Ethiopia Central Statistical Agency

Statistical Abstract (2009). Federal Democratic Republic of Ethiopia Central Statistical Agency

Tanser, L. (2002). The application of geographical information systems to important public health problems in Africa. *Intertional Journal Health Geogr* Volu.4. [PMC freearticle] [PubMed]

Tobler, W. (1979). Cellular geography. In S. Gale and G. Olsson (Eds.), Philosophy in Geography, (Dordrecht, Reidel), *pp. 379.*

Tobler, W. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region" *Economic Geography 46 (2): 234-240.*

Tsai**,** P., Men-Lung, L., Chien-Min, C. and Cheng-Hwang ,P. (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006; *BioMed Central Ltd*

Wimberly**,** M., Adam, D., and Michael, J. (2008). Enhanced spatial Model for predicting the geographic distributions of tick-borne pathogens. *International Journal of Health Geographic's, 7 (15): 4-6.*

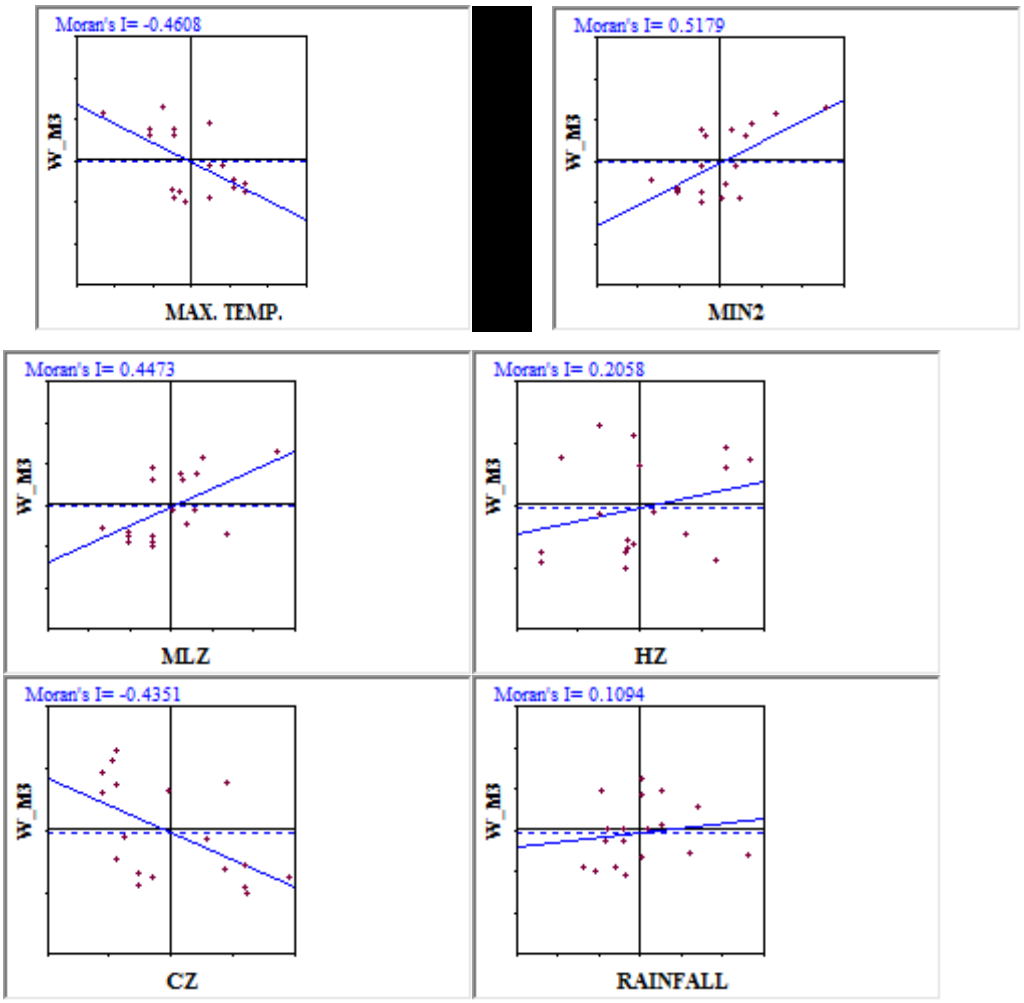
WHO (2008). World Health Organization8. Available from: *http://www.who.int/malaria/wmr2008/malaria2008.pdf*,

Yeshiwondem, A., Sucharita, G.**,** Afework, T. and Hrishikesh, P. (2009). Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia, *International Journal of Health Geographics*, *vol. 8*,can be found online at: http://www.ij- healthgeographics.com/content/8/1/5

Zhang, W., Liping, W., Liqun, F., Jiaqi, M. and Youfu, X. (2008). Spatial analysis of malaria in Anhui province, China, *Malaria Journal, 7:206.*

Appendices

Figure A: Moran's Scatter Plot Based on Neighborhood



Note:- Max.Temp.= Maximum Temperature; MIZ=Mid-land zone
MIN2= Minimum Temperature; HZ= hot zone; CZ= cold-zone;
W-M3= spatially lagged malaria incidence

Figure B: Moran's Scatter Plot Based on Threshold Distance

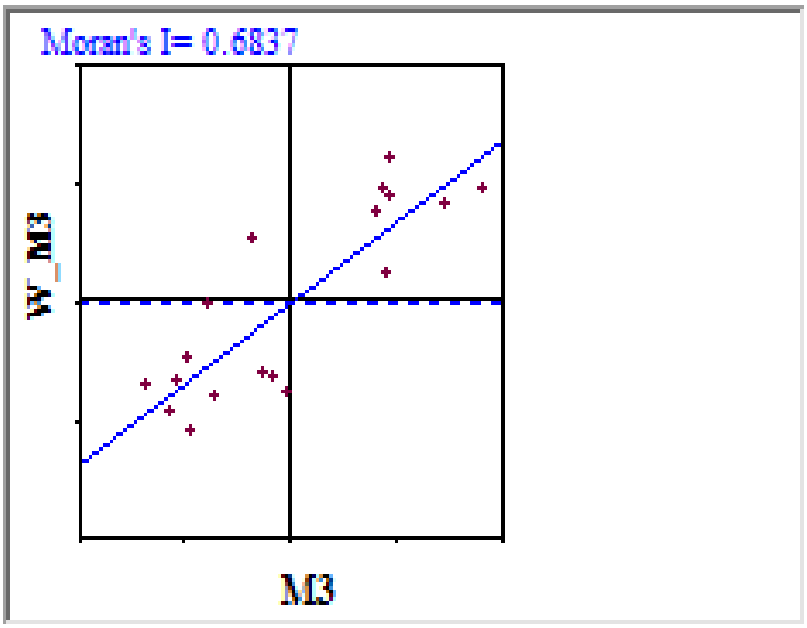
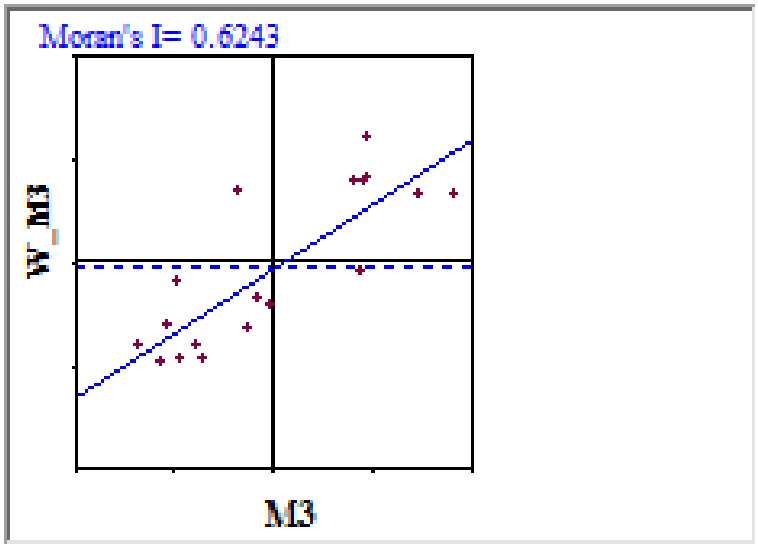


Figure C: Moran's Scatter Plot Based on K-nearest Neighbors



Note:- **M3** represents malaria incidence; **W-M3** spatially lagged malaria incidence

Figure D: Normal Plot

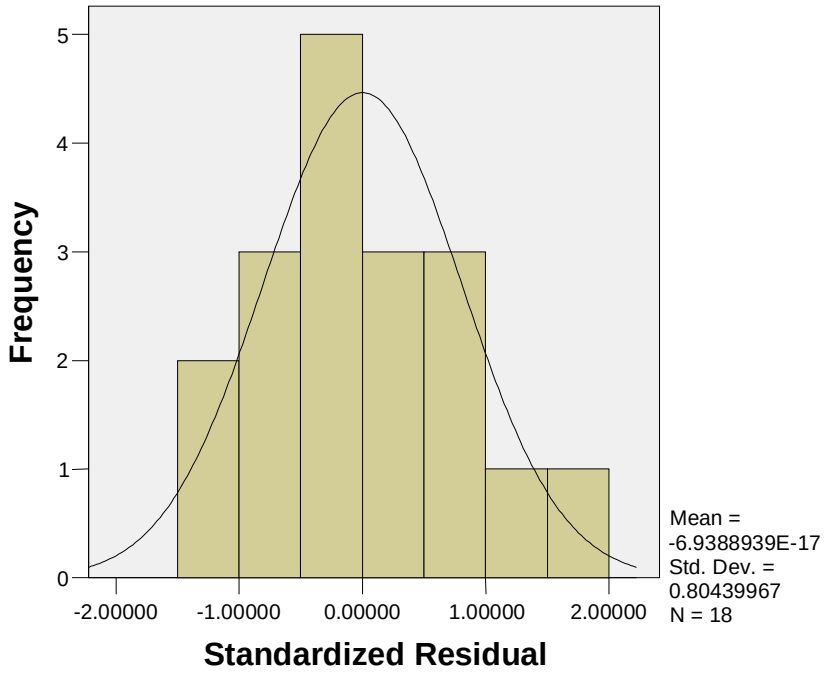


Table 1: Diagnostic for Multicollinearity

Model	Unstandardized coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std.Error	Beta			Tolerance	VIF
(Constant)	-9.990	4.224		-2.365	.037		
Cold	-.033	.072	-.487	-.460	.000	.190	5.268
Mid	.535	.130	.105	4.114	.000	.139	7.193
Hot	.627	.092	.290	6.820	.000	.237	5.297
Rain	.007	.004	.039	3.956	.001	.845	1.184
Max.Temp.	.549	.169	.160	3.249	.008	.580	1.725
Min.Temp.	.273	.055	.188	4.959	.000	.237	4.220

Table 2: Malaria Incidence per 1000 people in each Woreda in West Shoa Zone, Ethiopia

No	Woreda	Population Size (2007) A	Projected population size (2009) B	Malaria case (2009) C	Malaria Incidence rate= $\frac{C}{B}*1000$
1	Abuna-Gindaberet	109,545	119,947	3,530	29.43
2	Ada-Berga	120,177	131,588	2,497	18.98
3	Ambo	161,063	176,357	8,488	48.13
4	Bako-Tibe	123,558	135,291	14,777	109.22
5	Cheliya	156,994	171,901	14,154	82.34
6	Dano	99,361	108,796	8,973	82.47
7	Dendii	170,233	186,398	8,456	45.37
8	Ejerie	89,168	97,635	3,051	31.25
9	Elfata	57,365	62,812	1,501	23.89
10	Ginde-Beret	101,220	110,832	5830	52.60
11	Jeldu	202,655	221,898	4,658	20.99
12	Jibat	71,218	77,981	6,159	78.98
13	Meta-Robi	140,585	153,935	3,714	24.13
14	Midakegn	78,990	86,491	5361	61.98
15	Tikur-Enchini	71,621	78,422	7,701	98.19
16	Nono	84,193	92,188	7,488	81.23
17	Toke-Kutaye	119,932	131,320	5,593	42.59
18	Walmara	114,606	124,806	1,458	11.68
	Total	2,072,485	2,269,282	113,389	49.97

Table 3: Spatial Weighted Matrix W for Neighboring Relation among Woreda in West Shoa Zone, Ethiopia

	Woreda																		
Woreda		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0
	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
	3	0	0	0	0	0	0	1	0	1	1	1	0	0	1	0	0	1	0
	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0
	6	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0
	7	0	0	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
	8	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1
	9	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
	10	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	11	1	0	1	0	0	0	1	1	1	1	1	0	1	0	0	0	0	0
	12	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0
	13	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
	14	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
	15	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
	16	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0
	17	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0
	18	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Where 1,2,. .,18 are the Woredas their name listed as in the Table 2 and location indicated as in the Figure 4.4

Table 4: Results of Maximum Likelihood Estimation (spatial lag model)

Dependent Variable	: malaria incidence	Number of Observations	: 18
Mean dependent var	: 52.3861	Number of Variables	: 8
S.D. dependent var	: 29.0989	Degrees of Freedom	: 10
Lag coeff. (Rho)	: 0.733686		
R-squared	: 0.950716	Log likelihood	: -17.9666
Sq. Correlation	: +	Akaike info criterion	: 59.933
Sigma-square	: 3.34	Schwarz criterion	: 67.056
S.E of regression	: 1.9725		

DECLARATION

The thesis is my original work, has not been presented for a degree in any other university and that all sources of material used for the thesis have been duly acknowledged.

Name: Mitiku Mosissa

Signature: …….

Date: ……...

Place: Faculty of Science, Addis Ababa University

This thesis has been submitted for examination with my approval as a University Advisor.

Name: Dr. Butte Gotu

Signature: ……

Date: ……

Place: Faculty of Science, Addis Ababa University