SciencePG
Science Publishing Group

Spatial Modelling of Malaria Prevalence and Its Risk Factors in Rural SNNPR, Ethiopia: Classical and Bayesian Approaches

Dereje Bekele Dessie

Department of Statistics, Madda Walabu University, Bale Robe, Ethiopia

Email address:

dereje1999@gmail.com

Abstract: The purpose of this study was to assess the spatial distribution of malaria prevalence rates among selected rural part of woredas in SNNPR, Ethiopia. This work is based on data available from the 2011 malaria indicator survey (MIS 2011) of Ethiopian Public Health Institution. ESDA, Spatial regression model and Bayesian Spatial analysis were employed for data analysis. From ESDA, we found positive spatial autocorrelation in malaria prevalence rate. Relying on specification diagnostics and measures of fit; Spatial lag model was found to be the best model for modeling malaria prevalence rate data. The relationship between malaria prevalence and its risk factors was assessed using spatial models. The spatial models also showed an increase of malaria prevalence with a number of factors. From results, increase in the proportion of households sprayed in 12 months and the average altitude in the woreda estimated to decrease the average malaria prevalence. The result also demonstrated that increase in the House hold size of the district, proportion of households having access to piped water, proportion of households having access to radio, proportion of households having access to radio and Main construction material of the room's wall are estimated to raise the average malaria prevalence rate. Finally, the study concluded that malaria is spatially clustered in space and the risk factors exhibit effect on the malaria prevalence in the study area. Based on the results of the study, We recommend for policy makers on the way to reduce malaria prevalence in the rural part of woreda of SNNPR using spatial information.

Keywords: Malaria, Spatial Autocorrelation, Malaria Prevalence, Bayesian Spatial Analysis

1. Introduction

1.1. Background of the Study

Malaria is a mosquito-borne infectious disease of humans and other animals caused by protists of the genus Plasmodium which are introduced into the circulatory system by the bite from an infected female anopheles mosquito. In the human body, the parasites multiply in the liver, and then infect red blood cells. Malaria risk becomes higher in rural areas of developing countries [12]. A large number of malaria causing factors including the proximity to the vector breeding sites, the inadequate use of control measures, low income, illiteracy, land use and the house material play a big role [41]. He mentioned the multiplicity of malaria causing factors in rural areas as the main cause of its persistence as they are difficult to control at the same time [41].

Malaria disease burden and transmission can be assessed using incidence or prevalence in human hosts. Malaria is one of the leading causes of morbidity and mortality in the world, with an estimated 3.3 billion people at risk of malaria [14]. The incidence of malaria worldwide is estimated to be 216 million cases per year, with 81% of these cases occurring in sub-Saharan Africa. Malaria kills approximately 655,000 people per year; 91% of deaths occur in sub-Saharan Africa [14], mostly in children under five years of age.

According to records from the Ethiopian Federal Ministry of Health, 75% of the country is malarious with about 60% of the total population living in areas at risk of malaria. That is, 50.6 million people are at risk from malaria, and four to five million people are affected by malaria annually. The transmission of malaria in Ethiopia depends on altitude

1,000-2,000m above sea level, since malaria is rarely transmitted at higher elevations (unless there are weather abnormalities and widespread epidemics). And rainfall with a lag time varying from a few weeks before the beginning of the rainy season to more than a month after the end of the rainy season (PMI, 2015).

Malaria is the leading public health concern in the Southern Nation Nationalities and People Region (SNNPR) among the regional states of Ethiopia. Epidemiological and ecological data of the region indicate that more than 65% of the area of this region is malarious.

1.2. Statement of the Problem

Malaria cases depend on the Demographic, Environmental, seasonal, climatic and others different socioeconomic factors. Controlling malaria at regional level needs identifications of those factors related to malaria. Malaria is the leading cause of morbidity and mortality in Ethiopia, accounting for over five million cases and thousands of deaths annually. About 75% of the land and 60% of the population is exposed to malaria in Ethiopia. In most part of the country the peak period of the malaria incidence occur from September to December following the main rainy season (June to September)and from March to May during and after the small rainy season (February to March) [5].

In the Southern Nations, Nationalities and People's Region (SNNPR), Ethiopia about 65% of the population is living in malaria endemic areas. In the region, malaria is the primary cause of outpatient and inpatient consultations and hospital deaths (SNNPR, 2013). Hence, considering seasonal and geographical variations malaria transmission has been difficult due to a lack of resource and time as well as usage of inappropriate statistical methods and data. Explaining this distribution is also important, since it provides a rationale for interventions and makes it possible to predict transmission intensity in places in which it has not been measured. 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. In this study, the spatial distribution of malaria prevalence in selected rural part woredas of SNNPR, Ethiopia is explored along with risk factors. Spatial variation can be quantified using spatial models. Therefore, this study is aimed to address the following questions:

a). What are the major significant factors of malaria prevalence rate in rural Southern Nation, Nationalities and People of Ethiopia?
b). How does malaria prevalence rate spatially distributed across woredas in rural Southern Nation, Nationalities and People of Ethiopia?
c). Is there any relation ship between Malaria prevalence rate and its risk factors?

1.3. Objectives of the Study

1.3.1. General Objective

The general objective of this study was to identify and model determinants of Malaria prevalence rate and Its risk factors in rural woredas of Southern Nation, Nationalities and People's region, Ethiopia.

1.3.2. Specific Objectives

a). To identify the significant factors of malaria prevalence rate in rural SNNPR, Ethiopia.
b). To determine the spatial variations of malaria distributions across the districts of rural Southern Nation, Nationalities and People's Region, Ethiopia.
c). To identify and model the relationship between malaria prevalence rate and Its risk factors.
d). To contribute scientific result as information for policy makers and and stake holders.

1.4. Overview of Study

This paper is organized as follows. The first chapter provides a brief Background of the Study, Statement of the problem, Objectives, its Significance and Scope of the study. The third chapter describes the Source of data, Variables of the study and the Methodology used for analysis. The fourth chapter presents Spatial analysis and Bayesian spatial analysis of the data and discussion of each outputs. Finally, Conclusion and Recommendation of the study are dealt within section five of the study.

2. Data and Methodology

2.1. Description of Study Area

The state of Southern Nations, Nationalities and People's region is located in the Southern and South western part of Ethiopia. It is bordered with Kenya in south, Sudan in southwest, Gambella region in northwest and surrounded by Oromia region in northwest, north and east directions. The state of Southern Nations, Nationalities and people of Ethiopia comprises 10% of the total area of the country. It is administratively divided into 13 zones, 133 woredas. The capital city of the region is Hawassa. According to the CSA annual report the state has an estimated area of 110,931.9 sqkm [6]. The total population size of the region is 14,945,992. The rural population of the state accounts for 86.2% of the total population with most rural area region in the country (SNNPR, BoFED, Annual Statistical Abstract 2008 E. C.).

2.2. Data Descriptions

This work is based on data available from the 2011 malaria indicator survey (MIS 2011) of Ethiopian Public Health Institution. The 2011 EMIS was the second malaria indicator survey held in Ethiopia after the 2007 survey. The survey included testing for malaria among all age groups using a finger or heel-prick blood sample. The aim of the program is to undertake national malaria control for regional and national

family and health planning. The main goal of the 2011 EMIS was to measure the progress toward achieving the goals and objectives of the NSP $2007-2015$.

2.3. Study Variables

The dependent variable is Malaria prevalence rate of selected household in the rural parts of woredas of the region where as The independent variables with their corresponding meaning of the symbols which are used throughout the whole description for this study are given in the following table.

***Table 1.** Independent Variables for the Study.*

Mean altitude of a district ASL in meter (Mean ALT)
Proportion of households having access to piped water (DngWtr_100)
Proportion of household having Toilet (Toilet_100)
Proportion of households access to radio (RADIO_100)
Proportion of households Access to phone (PhoneML)
Main construction material of the room's wall (WType)
Main construction material of the room's roof (RType)
Main construction material of the room's Floor (FType)
Use of mosquito nets (Bed nets)
Percentage of households sprayed in the last 12 months (SPd12m)
Average number of household members (AVHHMBR)
Household size (HHsize)

N= number of selected woreda

2.4. Statistical Methods

Spatial data and Spatial dependence

In statistics, spatial data analysis or spatial statistics includes any of the formal techniques which assess entities using their topological, geometric, or geographic properties. Spatial data analysis can be defined as statistical study of phenomenon that manifests them in space. As a result location, area, topology, spatial arrangements, distance and interactions became the focus of attention [2]. Spatial data set consists of a collection of measurements or observations on one or more attributes taken at space [26] or Spatial data contain information about both the attribute of interest as well as its location. The spatial data structures are geographically referenced matrix of uniform size cells from remote sensing (Raster) and Vector (features on the earth's surface that are represented as geographically referenced vectors such as points, line and polygon).

There are three types of spatial data:-

a. Spatial Point Processes:- observes typically complete set of data points in the space that exhibits complete spatial randomness or occurrence of events at locations in space.

b. Geostatistical Data:- observes typically complete set of data points based on data that can be interpolated to unobserved points on continuous surface regularly or irregularly in function of distance.

c. Areal (Lattice) Data:- observes data distributed in to predefined spatial regular or irregular shape regions such as states, countries, census tracts, Zip codes, etc to see discrete variation over space.

According to [3], 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).

Contiguity information is quantified as contiguity (spatial neighbors) matrix which contains elements of 1 and 0; the matrix is denoted by W and defined by:

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \dots & & & \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{pmatrix}$$

Where; n is the number of regions/locations under study, W_{ij} is the element of W representing a weight for $(i,j)^{th}$ location i, $j = 1, 2, ..., n$. As affirmed by [3] and [31] contiguity matrix W is constructed based on the following forms of definitions.

a). A rook contiguity definition considers objects sharing a common edge, defines $W_{ij} = 1$ for regions that shares common edges with the region of interest otherwise $W_{ij} = 0$.

b). A bishop contiguity definition considers objects sharing a common vertex as neighbours, defines $W_{ij} = 1$ for entities that shares common vertex with the region of interest otherwise $W_{ij} = 0$.

c). 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 neighbour, defines $W_{ij} = 1$ for entities that shares common edge or vertex with the region of interest otherwise $W_{ij} = 0$.

In Cartesian space quantification of location contiguity matrix is defined based on distance and inverse distance square between centers of regions or points.

$$W_{ij} = \frac{1}{d_{ij}} \quad \text{or} \quad W_{ij} = \frac{1}{d_{ij}^2} \quad \text{or} \quad W_{ij} = \begin{cases} 1 & \text{if} \quad d_{ij} < r \\ 0 & \text{if} \quad d_{ij} > r \end{cases}$$

where d_{ij} is distance between center of ith and j^{th} location. r is prespecified radius distance. The matrix W is symmetric and by convention the matrix always has zero entries on main diagonal (a lattice/point is not contiguous/distant to itself), but if it is non symmetric has implication for complexity of estimation and testing procedures and it is not sensitive to number of topological transformations. The transformed form of W often used in applied work to convert matrix to have row sum of unit by standardizing each elements of W by row totals to facilitate the interpretation of the model coefficients, however, for distances weights standardization may result in loss of meaningful economic interpretation.

The row standardized contiguity matrix W is called spatial weighted matrix. Spatial weight matrix defines the structure

and intensity of spatial effect. The ultimate objective of the use of spatial weight matrix in spatial data analysis (particularly, spatial econometrics models) is to relate variable at one point in space to the observations for that variable in other spatial units in the system, and it serves as lag operator in analogies to time series context [3] and [31].

Spatial Dependence and Heterogeneity

The spatial heterogeneity is simply structural instability in the form of non constant error variances (heteroskedasticity) and/or spatial varying of model parameters [31]. In a regression context, spatial effects pertain to spatial dependence and spatial heterogeneity. Spatial dependence occurred due to either measurement error or spatial dimension of social process in aspect of modeling. Two broad sources of spatial dependence are generally pointed out. First, It is a result of spatial interaction effects such as spatial aggregation problems, spatial spillovers and factor mobility (spatial externalities). Second, It may be due to the measurement problems resulting from inherit spatial organization and spatial structure of phenomena; the fact that administrative borders may not coincide with the borders of social interaction [3].

2.4.1. Exploratory Spatial Data Analysis (ESDA)

ESDA is a set of techniques aimed at, describing and visualizing spatial distributions, identifying a typical localizations or spatial outliers, detecting patterns of spatial association, clusters or hot spots, and suggesting spatial regimes or other forms of spatial heterogeneity [26] and [3]. These techniques provide measures of global and local spatial autocorrelation. Spatial autocorrelation can be defined as the coincidence of value similarity with location similarity or dissimilarity.

Global and Local Measures of spatial Autocorrelation

Tests for global spatial autocorrelation examine whether the data as a whole exhibit spatial autocorrelation (against Ho: No spatial autocorrelation) as well as the strength and direction (positive or negative) of any spatial autocorrelation. Tests for local spatial autocorrelation (again, against Ho:No 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 [2].

The Moran Scatter plot enables us to visualize the linear correlation between the spatial lag Y and WY. Specifically, WY is plotted against Y and the Moran's I coefficient will be the slope of the regression curve [3]. In additions to this, inspection of global and local spatial instability is carried out by the means of the Moran scatter plot [2], which plots the spatial lag, against the original values.

Measures of Local Autocorrelation

These measures are used when there is no global autocorrelation, and in case where measure of global does not enable us to appreciate the regional structure of spatial autocorrelation. One can wonder which regions contribute more to the global spatial autocorrelation, whether there are

local spatial clusters of high or low values, and finally to what point the global evaluation of spatial autocorrelation masks a typical localizations or pockets of local non stationary. The analysis of local spatial autocorrelation is carried out with two tools. First, the Moran scatter plot which is used to visualize local spatial instability [2], and second local indicators of spatial association which are used to test the hypothesis of random distribution by comparing the values of each specific localization with the values in the neighboring localizations [2].

The four different quadrants of the scatter plot correspond to the four types of local spatial association between a region and its neighbors: (HH) a region with a high value surrounded by regions with high values in Quadrant I, (LH) a region a with low value surrounded by regions with high values in Quadrant II, (LL) a region with a low value surrounded by regions with low values in Quadrant III, (HL) a region with a high value surrounded by with low values in Quadrant IV. Quadrants I and III represents the presence of positive spatial autocorrelation, where as quadrants II and IV represent the presence negative spatial autocorrelation [2]. [3] defines a local indicator of spatial association as any statistics satisfying two criteria. First, the LISA for each observation gives an indication of significant spatial clustering of similar values around that observation, and second, the sum of the LISA for all observations is proportional to a global indicator of spatial association. The local version of Moran's statistics for each region/location is given as:

$$I_i = \frac{(Y_i - \bar{Y})}{m_0} \sum w_{ij} (Y_i - \bar{Y}) \qquad (1)$$

where, $m_0 = \frac{1}{n} \sum ((Y_i - \bar{Y}))^2$ and therefore,

$$I = \frac{1}{n} \sum I_i$$

[2] gives two interpretations for local Moran's statistics (or LMI); first they can be used as indicators of local spatial clusters (or hot spots) which can be identified as locations or sets of neighboring locations for which the LISA are significant and as they can be served diagnostics for local instability. That means for significant outliers with respect to the measure of global spatial autocorrelation, typical localizations or pockets of non stationary. The second interpretation of the LISA statistics is similar to the use of a Moran scatter plot to identify outliers and leverage points for Moran's I.

Diagnostics Tests of Spatial Dependence

Moran's Test for Regression Residuals

A well known test for spatial autocorrelation in the regression residuals is developed by [20] which is similar to Moran's I: The moran's I is given by

$$I = \frac{n}{S_0} \frac{U'WU}{U'U} \qquad (2)$$

where, U is the vector of residuals.

Statistical inference can be based on the assumption of asymptotic normality, but an exact approach depending on the matrix Y is available too, although rather cumbersome to apply. Moran's I for regression residuals is a locally best invariant test and based on moments estimation [20] and [3]. For normally distributed errors the distribution of the standardized Moran's statistic is shown to be asymptotically normal. In order to carry out an operational test, both the expected value and the variance of I are needed.

Lagrange Multiplier (LM) Test

The LM-lag test is introduced by [3] tests the absence of spatial dependence of the endogenous variable. The study demonstrated the similarity of Moran's I for regression residuals to a Lagrange multiplier statistic. This statistic is included within a general framework for testing spatial dependence and spatial heterogeneity (heteroskedasticity) as presented in [3] following the approach of [4]. The LM test statistic for spatial error is defined as:

$$LM_\lambda = \frac{1}{T}(\frac{U'WU}{S^2})^2 \qquad (3)$$

where, S is the maximum likelihood estimate of variance and T is a scalar computed as the trace of a quadratic expression in the weight matrix which is given by $T = tr(W'W + W^2)$. And LM test statistic for spatial lag dependent is expressed as:

$$LM\rho = \frac{1}{J_{\rho\beta}}(\frac{U'WY}{S'})^2 \qquad (4)$$

Where,

$$J_{\rho\beta} = \frac{(Wxb)'M(Wxb) + TS^2}{nS^2}$$

is a part of the estimated information matrix, b the OLS estimated parameter vector, and M the projection matrix expressed as

$$M = IX(X'X)^{-1} X' \qquad (5)$$

Both LM tests for lag and error are asymptotically following a χ^2 distribution with one degree of freedom [3]. LM tests for lag and error are one directional, so that tests do not consider the presence of the alternative form of spatial dependence. As a consequence, they are not robust when the alternative form of dependence is present [3]. In other words tests are labeled robust because the test statistics account for the potential presence of a spatial lag or spatially correlated errors when testing for the presence of spatially correlated errors or a spatial lag, respectively [3].

The test for a spatial error process robust to the local presence of a spatial lag is:

$$LM_\rho^* = \frac{1}{T^2 - T^2(nJ_{\rho\beta}^{-1})}(\frac{U'WU}{S^2} - T(nJ_{\rho\beta}^{-1})\frac{U'WY}{S^2})^2 \qquad (6)$$

This clearly shows the subtraction of a correction factor that accounts for the local misspecification of a spatial lag process [31]. The test for a spatial lag process robust to the local presence of a spatial error is given by:

$$LM_\rho^* = \frac{1}{nJ_{\rho\beta}^{-1} - T}(\frac{U'WY}{S^2} - \frac{U'WU}{S^2})^2 \qquad (7)$$

The robust LM diagnostics have reduced power in comparison to the unidirectional LM diagnostics in the absence of the alternative form of spatial dependence. As a result, they are more prone to Type II errors and are less prone to Type I errors than their non-robust counter parts [3].

The power of the Lagrange Multiplier spatial error (lag) test against a spatial lag (error) model seriously complicates the work of the practitioner of spatial modeling techniques. If both tests are significant, which underlying model is then correct? This issue is addressed in a way one can select model corresponding to test statistic with highest value. Although the Lagrange Multiplier diagnostics assume normally distributed errors, [2] find that the robust LM diagnostics are robust to non normality. However, [3] find that the one directional LM diagnostics are robust to non normality in large samples. In all of these tests discussed above the null hypothesis is non existence of spatial autocorrelation in the OLS residuals and large values of test statistic (χ^2) with degree of freedom one lead to rejection of null hypothesis. If the OLS diagnostics indicate the existence of spatial lag dependence or spatial error dependence, one can estimate autoregressive (spatial lag) or spatial error model via maximum likelihood estimation.

2.4.2. Modeling Spatial Dependence

Spatial Regression Models

In the spatial linear regression model, Spatial dependence can be incorporated in specification in two distinct ways; as an additional regressor in the form of a spatially lagged dependent variable (*WY*) provide spatial lag model, and in the form of spatial lag error structure (W ε) provides spatial error model. In a simultaneous model, the focus is on the explanation of the complete spatial pattern; particularly simultaneous auto regressive models assume that the response at each location is a function not only of the explanatory variable at that location but of the values of the response at neighboring locations as well.

Spatial Lag Model (SLM)

The simultaneous spatial lag regression model for dependent variable of observation i and k independent variable takes the following form:

$$Y_i = \rho \sum w_{ij} Y_j + \sum X_{ir} \beta_r + \varepsilon_i \qquad (8)$$

where, ρ spatial autoregressive coefficient which is scalar, the k explanatory variable and intercept are x_{ir}, $r = 0,1,2,...,k$ with associated coefficient β_r, w_{ij} denote the (i, j) the element of W, and ε_i is the disturbance term.

The matrix notation of the model is $Y = \rho WY + X\beta + \varepsilon$

where, ε is a vector of error terms which is independent and identically multivariate normally distributed with mean vector zero and constant diagonal variance-covariance matrix $\sigma^2 I_n$, Wy is spatial lag operator, Y is vector of dependent variables, X is a designed matrix of explanatory variables including intercept of ones, and β is vector of coefficients of regression model. The scalar coefficient ρ is parameter to be estimated that will determine the strength of the average (overall observations, $i = 1, 2, ..., n$) association between the dependent variable values for regions (observations) and of those values for their neighboring regions [20]. In spatial lag model, the spatial lag term Wy is correlated with the disturbances, even when the errors are identically and independently distributed.

Spatial Error Model (SEM)

With similar set in spatial lag model, the spatial error model for observation i is noted as:

$$Y_i = \sum_{r=0} X_{ir}\beta_r + \lambda \sum_{j=1} W_{ij}\varepsilon_j + \varepsilon_i \qquad (9)$$

where, λ is a spatial autocorrelation coefficient which is scalar, the k explanatory variable and intercept are x_{ir}, $r = 0,1,2,...,k$ with associated coefficient β_r, W_{ij} denote the $(i,j)^{th}$ element of W, and the disturbance term ε_i independently and identically normally distributed with mean zero and constant variance.

The matrix notation of SAR error model is:

$$Y = X\beta + U , \quad U = \lambda WU + \varepsilon \qquad (10)$$

This type of spatial regression is appropriate when the concern is with correcting for the potentially biasing influence of the spatial autocorrelation due to the use of spatial data irrespective of whether the model of interest is spatial or not.

Spatial Model Diagnosis

Breusch-Pagan test of heteroskedasticity for random coefficient is similar to test under OLS regression but here the residuals and estimate of residuals obtained from spatial models (spatial lag or error). The tests asymptotically achieve χ^2 distribution with K degree of freedom. If these statistics are greater than critical value one rejects null hypotheses (Homoskedasticity). If test reject null hypothesis we apply spatial first stage least square estimation or we fit other models like spatial expansion or geographical weighted regression. Likelihood ratio test in spatial model is an alternative the asymptotic significance test on the spatial autoregressive ρ and error coefficient λ. It compares the LR from the OLS model to the LR from the spatial models, and this statistic is asymptotically distributed as χ^2. We reject the null hypothesis of no Spatial autocorrelation if the test statistic exceeds critical value at 5% level of significance. It is similar to traditional specification test (Wald test and LM) that comparing the null model (OLS) with spatial model (either spatial lag or spatial error).

Condition Number (K): It is Eigen system analysis which was used to test near dependence among independents based on Eigen value of the designed matrix (X). K is the ratio of largest Eigen value to the smallest Eigen value, K less than 30 provides a reasonable threshold for no multicollinearity [11]. Nevertheless if sever multicollinearity exist we took measures of omitting some variables, applying principal component analysis, centering or scaling regression data and other measures.

Jarque-Bera Test (JB): In addition to visualize normality of error from histogram and residual plots like normal Q-Q plot, Jarque and Bera develop test that tests the hypothesis of data are come from normal population or error are normally distributed. This type of spatial regression is appropriate when the focus of interest is the assessment of the existence and strength of spatial interaction and suitable to filter out spatial dependence that comes from spatial spillovers. In other words spatially lagged dependent model is appropriate when we believe that the values of dependent in one unit i are directly influenced by the values of dependent variable found in i's neighbors; this influence is above and beyond other covariates specific to i.

Tests of Heteroskedasticity

If the inefficiency of OLS is thought to be serious drawback; so to test heteroskedaticity in residuals often we use residual versus fitted value plot which is descriptive, but Breusch-pagan test, Koenker-Bessett test and White test are quite inference form tests. Both Breusch-pagan and Koenker-Bessett tests are based on random coefficient by assuming specific functional form for heteroskedaticity, and Koenker-Bessett is robust to non normality (i.e though normality is not met we use it than Breusch pagan test). Both tests asymptotically achieve χ^2 distribution with K degree of freedom. If these statistics are greater than critical value one reject null hypotheses (Homoskedasticity).

Measures of Fit for the Spatial Models

Even though some times F test for the significance of overall model shows model is adequate, in addition to adequacy checking tests and analysis for models in the same class but differently specified it is compulsory to utilize other measures of fit to make comparison between or/ and among models. So R^2, Log likelihood, Akaike information criterion, Schwarz criterion and standard error of regression are also used. R^2 refers to coefficient of determination that express proportion of variation explained by the regressors in our model. Higher Log likelihood value signifies good fit. The lower value for AIC and SC show best/better fit of model.

2.4.3. Bayesian Spatial Analysis

The concept underlying Bayesian spatial modeling is Bayes' theorem, a theorem that considers both the distributions of the data and the unknown coefficient estimates [31]. Bayesian spatial modeling embraces most spatial models, such as the spatial lag model, the spatial error model, and geographically weighted regression, as long as the spatial statistical model can be estimated with Bayesian methodology.

The flexibility of the Bayesian spatial Models allows us to

perform an Spatial Regression [15] at the same time as we consider independent variables and spatial random effects. By including covariates in our model we aim to assess and remove the effect of potential confounders or risk factors. The assessment of the importance of a covariate is indicated by the estimated value of its coefficient and its associated probability interval. If, for example, the 95% credible interval does not contain the value 0, we may assume that the coefficient is significant and, if greater than zero, it will indicate a positive relationship between the risk and the variable.

Spatial Correlation

Within spatial applications it is often found that correlation will exist between spatial units. This correlation is geographical and relates to the basic idea that locations close together in space often have similar values of outcome variables while locations far apart are often different. This spatial correlation(or autocorrelation as it's sometimes called) must be allowed for in spatial analyses. This may have an impact on the structure and form of likelihood models that are assumed for spatial data. The assumption made in the construction of conventional likelihoods is that the individual contribution to the likelihood is independent and this independence allows the likelihood to be derived as a product of probabilities. However, if this independence criterion is not met, then a different approach would be required.

Conditional Independence

The idea of inclusion of spatial correlation at a hierarchical level above the likelihood is a fundamental assumption often made in Bayesian spatial modeling. This means that the correlation appears in prior distributions rather than in the likelihood itself. Often parameters are given such priors and it is assumed that conditional independence applies in the likelihood. Note that this approach to correlation does not completely account for spatial effects as there can be residual correlation effects after inclusion of confounders. These effects could be due to unobserved or unknown confounders. Alternatively they could be due to intrinsic correlation in the process. Hence the assumption of conditional independence may only be valid if correlation is accounted for somewhere within the model. A model for spatial dependence in the errors for outcome Y take the form

$$Y = x\beta + e \qquad (11)$$

$$e = \rho We + u$$

where ρ is an unknown correlation parameter, where Y, e and u are vectors of length n, and x is of dimension nxp. Here the u denote spatially unstructured errors, which are frequently taken as homoscedastic $u_i \sim N(0, \sigma^2)$. If the interactions are scaled within rows, then the permissible maximum of ρ is 1 [4] and the permissible minimum is the smallest eigenvalue of W, which is greater than -1 but less than 0. Since spatial correlation is positive in the great majority of econometric or health applications, a prior on ρ constrained to [0, 1] is feasible in many applications.

Conditional priors

All parameters within Bayesian models are stochastic and are assigned appropriate probability distributions. Hence a single parameter value is simply one possible realization of the possible values of the parameter, the probability of which is defined by the prior distribution. The prior distribution is a distribution assigned to the parameter before seeing the data. Note also that one interpretation of prior distributions are that they provide additional 'data' for a problem and so they can be used to improve estimation or identification of parameters. It combines prior knowledge and observation data to obtain posterior distributions for parameters of interest.

Letting Q = Â˘ I- ρ W, the precision matrix E^{-1} of the errors e in Equation (11) may be derived as

$$E^{-1} = \tau Q'Q \qquad (12)$$

where $\tau = \sigma^{-2}$. A full multinormal scheme for the e could be used, with the errors sampled simultaneously fromtheir joint prior

$$e \sim N_n(0, E^{-1}) \qquad (13)$$

However, a conditional scheme is possible, and may be simpler to sample. Modelling of spatially correlated errors may proceed by initially specifying either the joint multivariate distribution of the vector, or the univariate density of each areas error, e_i, conditional on the current estimate of errors in other areas (e_j, j ≠ i). Conditions that ensure the joint density is proper (so that the e_i are identifiable) when the model specification starts with a conditional rather than the joint prior are discussed by [43] and [4]. One possible conditional prior (the conditional autoregressive or CAR prior) expresses e_i in the centred univariate Normal form, $e_i \sim N(M\_i, \sigma^2)$, where mean of each area's error

$$M_i = \rho \sum_j c_{ij}\, e_i \qquad (14)$$

is a weighted average of errors in contiguous areas, and ρ is bounded by the inverses of the minimum and maximum eigenvalues of C. (Note that the interaction matrix for this form of prior needs to be symmetric.)

One may also have a spatial effects or spatial autoregression model, with spatial lags in the outcomes themselves [4], with

$$Y = x\beta + \rho\, WY + u \qquad (15)$$

where u is white noise. Spatial dependence in both Y and e may occur in the same model, for example:

$$Y = x\beta + \rho_1 WY + u; \quad e = \rho_2 We + u$$

where e is multinormal as in Equations above. In some situations, a spatial lag in the outcome might not be substantively sensible. If it were, then a sensitivity analysis might consider both correlation in both spatial effects and

spatial errors, and also consider non-constant variances in the u_i [31]; it might also encompass several forms of spatial interaction matrix, even if these are taken as fixed within the modelling [4]. The means for outcome Y_i are

$$\mu_i = \rho\, Wy + x_{\_i}\,\beta - \rho\, WX\,\beta \qquad (16)$$

Spatial heteroscedasticity may be parameterised in various ways: either one may suppose all areas to have distinct variances, or there may be groups of areas S_r with $\sigma^2{}_i = \sigma^2{}_r$ if i$\in$ S_r.

[31] proposes scale mixtures (with each area having its own variance) to robustify inferences against outlier data points. This applies even after spatial autocorrelation is explicitly modelled, and so would relate to the density assumed for the u_i. Thus, a scale mixture on the σ^2 is equivalent to assuming Student t rather than normal u_i.

Noninformative Priors

Often prior distributions are assumed that do not make strong preferences over values of the variables. These are sometimes known as vague, or reference or flat or noninformative prior distributions. Usually, they have a relatively flat form yielding close-to-uniform preference for different values of the variables. For parameters on an infinite range, such as regression parameters, then a distribution centered on zero with a large variance will usually suffice.

Posterior Distribution

In Bayesian inference, the posterior distribution, links the assumptions made (the prior distribution) with the empirical evidence (the likelihood). The goal is to use the characteristics of this distribution (say the mean or the quantiles) to make inferences about β. Because the parameters are themselves random variables, it is natural to deal with them in a hierarchical way. This means that we are assuming that their distribution may depend on other parameters, called hyperparameters. These hyperparameters are also random variables with their own prior distributions, called hyperpriors.

$$P(\theta / Y) = \frac{L(Y / \theta)g(\theta)}{C} \qquad (17)$$

Where,

$$C = \int L(Y / \theta)g(\theta)d\theta$$

where $g(\theta)$ is the joint distribution of the θ vector. Alternatively this distribution can be specified as a proportionality:

$$p(\theta / Y) \propto L(Y / \theta)g(\theta).$$

Inference for all of the Bayesian Spatial models is based on MCMC simulation, using a combination of Gibbs sampling and Metropolis steps.

Measures of Fit and Diagnostic for Bayesian Spatial Analysis

We can use formal convergence tests for Bayesian Spatial analysis. Briefly, Geweke diagnosis aims to diagnose lack of convergence using a single chain and multiple parallel chains, respectively. These functions also have graphical versions that show how convergence is improved by discarding extra burn-in iterations at the beginning of the series. Geweke can be computed to assess convergence. The DIC criterion is used for comparing the overall fit of multiple models applied to the same data, and lower values indicate a better fitting model [20].

3. Results and Discussion

The objectives of this study was to assess and model the Spatial dependence of malaria prevalence and Its risk factors in the rural part of SNNPR, Ethiopia. The data used for analysis is 2011 EMIS from which the rural area data was filtered and taken. Results for tests of Spatial autocorrelation in the malaria prevalence rate to determine the distribution pattern of malaria and its modeling of spatial Autoregressive model is presented in this chapter. Additionally, Bayesian spatial analysis of the malaria prevalence rate was incorporated in this session to provide inference which included prior information.

3.1. Descriptive Statistics

Results of the descriptive part shows, the lowest malaria prevalence rate of woreda in rural SNNP region was 1.070 per 100 people and the highest was 13.00 per 100 people occurred at Rural area of the region. In the rural woredas of the region on average it was 6.81 Malaria Prevalence per 100 people of a district and standard deviation is 2.513. Malaria prevalence was higher in the region compared to the country level average prevalence 4.5 [16].

A box map in Figure 1 was to identify woredas that had below and above average malaria prevalence rate in Rural SNNP region, It shows the location of every rural area woreda within the overall geographical distribution of malaria prevalence rates. Woreda that had malaria prevalence rates lower than average of the region are indicated by high bright and less bright blue color in the map and they were clustered around the North, Southeast and south west, while the high dark and less dark red color represents woredas that had above average malaria prevalence rate which were concentrated in Western, south and some part eastern rural part of the region.
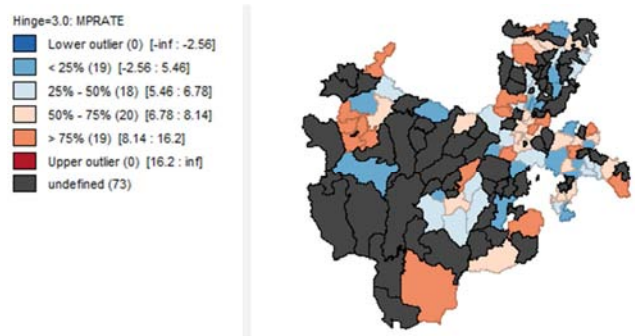

Figure 1. *Box Map for Distribution of Malaria Prevalence Rate in SNNP region.*

This box map provided some indication of spatial clustering in rural area of SNNP region, however the visual inspection of maps has long been recognized by cartographers as unreliable in terms of detecting clusters and patterns in the data as human perception is not sufficiently rigorous to assess 'significant' clusters and indeed tends to be biased towards finding patterns, even in spatially random data [33]. Consequently, we should turn to a consideration of a global statistic for spatial autocorrelation in the next section.

3.2. ESDA of Malaria Prevalence in Rural Part of Woredas of the Region

Moran's I test is perhaps the most common global test, having an argument to use an adjustment for a ranked continuous variable. Therefore, the result for moran's statistic is 0.317 which is significant at 5% significance level, indicating positive spatial autocorrelation across the woredas. The pseudo significance level was obtained from the reference distribution of 999 permutation. From table 3 the Moran's test statistic 0.317 is greater than the theoretical mean of $I(E(I)=-0.0068$). This shows the existence of positive spatial autocorrelation in the malaria prevalence.

Table 2. Results of Moran's I statistics for measuring spatial autocorrelation in Pattern of Malaria prevalence in rural of SNNP region.

Variable Name	Moran's I	Standardized Value	p-value
PVRATE	0.317	-5.6630	0.001

*significant at 5% level

3.2.1. Local Measures of Spatial Association

In the global test statistic, the result indicated a significant positive spatial autocorrelation(clustering). Local statistics are used to identify where high/low values cluster.

As shown in Figure 2 the Moran scatter plot of malaria prevalence rate, which represent a standardized malaria prevalence rate of a woreda in the x-axis versus the weighted average (spatial lag) of standardized malaria prevalence rate of its own woreda in the y-axis, which disaggregate the global spatial autocorrelation into four types of association (HH, HL, LH, LL). Points in quadrant I shows a district with high malaria prevalence rate (i.e., relative to average of the 76 woredas) was surrounded by woredas of high malaria prevalence rate (HH), quadrant II shows districts with low malaria prevalence rate surrounded by woredas with high malaria prevalence rate (LH), quadrant III indicated that districts with low malaria prevalence rate surrounded by woredas with low malaria prevalence rate (LL), and quadrant IV indicated that districts with high malaria prevalence rate were surrounded by districts with low malaria rate (HL). There were more points in quadrant I and III, and this shows in descriptive manner, a positive spatial autocorrelation patterns in the distribution of malaria prevalence rate among woredas of the region. However, this needs a formal statistical test to conclude.
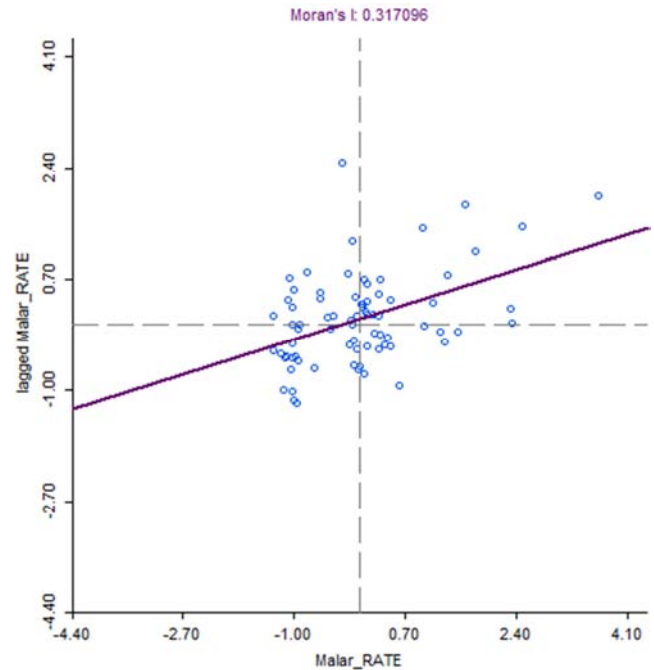

Figure 2. Univariate Moran's Scatter Plot.

The following two figures present univariate LISA significance and univariate LISA cluster maps of malaria prevalence rate in the woredas of rural area SNNP regional state respectively. LISA is a class of statistics that provides woredas specific information and estimates the extent of spatial autocorrelation of the value of malaria prevalence rate in a particular woreda with its neighbors.
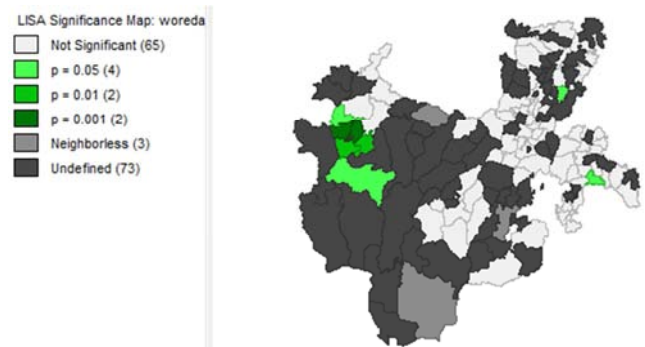

Figure 3. Univariate LISA Significance Map.

Figure 3 shows LISA significance map results of local Moran's I test for local spatial autocorrelation patterns of malaria prevalence rate. In the map, the bright green and green shade corresponds to location of malaria prevalence rate that had significant local spatial autocorrelation at 5% and 1% level of significance respectively. There were about (7) woredas that had significant local spatial autocorrelation patterns of malaria prevalence rate in the regional state at significance level, pointing out that there were presence of a spatial association of malaria prevalence in the rural districts of the region.
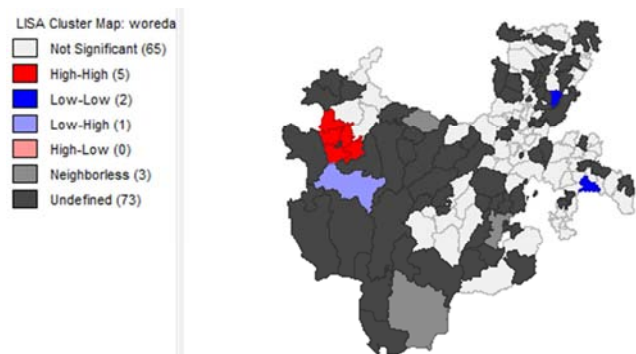
Figure 4. Univariate LISA Cluster Map.

LISA cluster map, useful in identifying the type of spatial association, was presented in Figure (4). The high-high (HH), low-low (LL), low-high (LH) and high-low (HL) association of malaria cases were shown by red, blue, black and grey colors respectively. By saying high or low values, it is to say high or low values relative to neighboring woredas. Positive spatial autocorrelation or clustering of similar values were indicated by high-high or low-low locations where as negative spatial autocorrelation or spatial outlier is indicated by high-low or low-high locations.

Clustering or hot spot locations of high malaria prevalence rate was in the south-west parts of the region in rural woredas of Yeki, sheko, Northern Bench, southern Bench and Minit shasha where as cold spot locations or low clustering of malaria prevalence was around south east parts of the region in rural districts of Dara and Lemmo. As a result, 5 woredas had HH, 2 woredas LL, 2 woredas LH and 1 woredas had HL association of malaria prevalence rate and among rural area of the districts with significant local spatial association. This supports the evidence of positive spatial autocorrelation

pattern in distribution of malaria prevalence rate across the region obtained above in the results of Moran's I statistics for testing of global spatial autocorrelation.

3.2.2. Diagnostic for Spatial Dependence in Residuals of OLS

The Moran's I test is a general test on Spatial correlation without giving precise information on the particular Spatial structure and it provides evidence for a positive and significant spatial dependence in OLS residuals (Moran's I value have p-value of $0.001 < 0.05$). Thus, the null hypothesis of Spatial independence of residuals OLS method was rejected indicating that OLS is inappropriate for the problem under consideration since the assumption of independence is violated. However, It is of great importance to seek for the nature of the spatial autocorrelation (wether the autocorrelation present in the residuals is due to spatial lag or spatial error process) even though it is not provided by Moran's I test. The most commonly used test in identifying the form of spatial dependence in the data is Lagrange multiplier (LM) test. The simple versions of LM test are powerful but not robust in local misspecification of the model, So the LM test for spatial lag dependence can be significant even if the form of the spatial dependence resembles spatial error dependence or vice versa. Thus, it is better to look at their robust part so as to come up with the correct identification of the form of spatial dependence in the data. As suggested by Lagrange multiplier (lag) test which is significant at 5% level of significance, the spatial lag is important to model determinants of malaria prevalence rate. The robust Lagrange multiplier tests, which provides a means of discriminating between the spatial lag or spatial Error model, spatial lag significant at 5% significance level suggest that spatial lag(with larger value= 5.1190) can be used in modeling Malaria Prevalence rate.

Table 3. Diagnostic test of spatial dependence in OLS regression residuals.

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.1977	2.2916	0.0219*
Lagrange Multiplier (lag)	1	9.0526	0.0026*
Robust LM (lag)	1	5.1190	0.0236*
Lagran. Multiplier (error)	1	4.0275	0.0447*
Robust LM (error)	1	0.0940	0.75917

*significant at 5% level

3.3. Fitting Spatial Regression

3.3.1. Spatial Lag Model

Here, we applied a statistical model which incorporates spatial dependence raised from spatial lag of the dependent variable, by adding the spatial lag of malaria prevalence rate on the right hand side of the OLS regression equation to represent the direct influence of the neighboring districts.

Table 4. Maximum Likelihood Estimate for Factors of Malaria Prevalence Rate in Spatial Lag Model.

	Estimate	Std. Error	z value	Pr(> |z|)
CONSTANT	6.0058	1.8556	3.23645	0.0012*
MEANAL	-0.0030234	0.00061444	-4.92058	0.0000*
HHSIZE	5.25473e-00	2.0701e-005	2.53834	0.0111*
AVHHMBR	0.10920	0.14034	0.778143	0.4364
Dng Wtr 100	0.045107	0.0150627	2.99463	0.0027*
Toilet 100	0.0066709	0.009618	0.693586	0.4879

	Estimate	Std. Error	z value	Pr($>$ $|$z$|$)
RADIO 100	0.024650	0.0086856	2.83811	0.0045*
Phone ML	0.012078	0.012499	0.966318	0.3338
Type	0.0001273	0.00737	0.0172639	0.98623
W Type	0.021588	0.008553	2.52406	0.0116*
R Type	0.0043742	0.009282	0.471227	0.6374
SPd12m	-0.024009	0.008945	-2.68401	0.0072*
HB Nets	0.0010373	0.009267	0.111929	0.9108

*significant at 5% level

Adjusted R-squared: 0.494241

ρ : 0.293574, LR test value: 9.6334, p-value: 0.00191

Log likelihood: -152.644 for lag model, AIC: 333.287, Sigma-square =3.15315, S. E of regression: 1.77571

Results from table 4 reveal maximum likelihood estimate for factors of malaria prevalence rate in Spatial Lag Model. Moreover, Estimated coefficient for spatial lag of malaria prevalence rate (ρ) was 0.293574, with p-value (.00191) which was positive and significant at 5% level of significance indicates that malaria prevalence rate in one woreda depends directly on the rate in its neighboring districts. This result supports what we have obtained using Moran's I statistics and cluster map in explanatory data analysis of malaria prevalence rate in the previous section. The importance of including spatial lag effects in our model is supported by the positive and significant value of the the coefficient. Significant coefficients for the variables implies that malaria prevalence rate in a given area depends on the change in explanatory variable in the same area controlling the effect raised due to spatial lag. The variables having negative and statistically significant effect on malaria prevalence are mean altitude and percentage of households sprayed in the last 12 months.

As discussed in the previous section, positive effect means that for a unit change in explanatory variable increase malaria prevalence rate in a given woreda by magnitude of estimate of parameter for that explanatory variable controlling for the effect of neighbor woredas, whereas negative effect is to mean that a unit change in the explanatory variable decreases malaria prevalence rate in a woreda by a magnitude of estimate of parameter for explanatory variable conditioning the neighbor woredas malaria rate and other variables constant.

Proportion of households having access to piped water, Proportion of households having access to radio, House hold size of the woreda and Main construction material of the room's wall have positive and statistically significant effect on malaria prevalence rate of the rural area of the region. The interpretation of the coefficients also, For example 1% increase in Percentage of house-holds having access to radio in certain district increase malaria prevalence rate in that particular district by 0.0246% keeping the effect other variables fixed. The parameter estimate 0.0240 for percentage of households sprayed in the last 12 months indicates that 1% decrease in Percentage household with sprayed insecticide increase the possibility of malaria prevalence infection in that woreda and its neighbors by 0.0240%. Table 4 also revealed measures of fit for models. R-square which equals 0.4942 implies that 49.42% of

variation in malaria prevalence rate was explained due to variation in the explanatory variables in the model and spatial lagged dependent variable. The significance of the over all model is assured by using likelihood ratio test which is analogeous to the F statistic of OLS model. Hence, results from Table 5 indicate that the overall model is significant at 5% level of significance indicating that maximum likelihood estimation was better in the explanation of the geographical variations of malaria prevalence rate than estimation procedure in the presence of spatially lagged dependence.

3.3.2. Spatial Error Model

The spatial error model evaluates the extent to which the clustering of malaria prevalence rate not explained by independent variables can be accounted for with reference to the clustering of the error terms. In this sense, it captures the spatial influence of unmeasured independent variables(i.e variables which are not included in the study but may have significant effect on malaria prevalence rate). Results of maximum likelihood estimates of the spatial error model were displayed in table 6. Spatial Autoregressive term (λ) was 0.378 with p-value 0.01190 and it is significant at 5% level of significance indicating that the spatial autocorrelation present in residuals of OLS was due to a geographic clustering of omitted variables or variables which are not included in the modeling of the factors for malaria prevalence using OLS regression model.

3.3.3. Diagnostic Test Results of Spatial Models

The invertablity or non singularity of the design matrix of explanatory variables, which is one of the basic assumption in multiple linear regression is diagnosed using condition number as a rule of thumb. Values of a condition number larger than 30 is considered to be implication for the existence of multicollinearity. In table 7 the condition number is 29.85 which indicates no existence of multicollinearity, i.e no interdependence among explanatory variables considered in the model. The hypothesis of normality of residuals is not rejected, since the Jarque-Bera test have p-value of 0.77133 which is greater than 0.05. Moreover, the homoskedasticity hypothesis is also not rejected from Koenker-Basset and Breush-Pagan tests, since the p-values for both tests are not significant at the specified level of significance and the tests suggest that the variance of the error is constant.

Table 5. *Diagnostic Tests Results of Spatial Lag model.*

Test & measures of fit	S(Lag)	S(error)
Robust LM (p-value)	0.02366*	0.75917
R-squared	0.494241	0.483217
Log likelihood (LIK)	-152.644	-154.297386
Sigma-Square	1.77571	1.79496
St. Error of Regression	3.15315	3.222
Likelihood Ratio Test	9.6334	6.3260

*significant at 5% level; Conditon number: 29.85

The two spatial regression models discussed so far have nearly similar values for the coefficients of regression and they both have significant LRT. However, the spatial lag rather than the spatial error dependence is supported by the robust measures.

Results in table 7 is regarding measures that are included to maintain comparability with the fit of spatial regression models. They are the log likelihood ratio, the Akaike information criterion and Schwarz criteria. These three measures are based on assumption of multivariate normality and corresponding likelihood for standard regression model. The higher loglikelihood value signifies the better fit. The lower value for Akaike and Scwarz information criteria show best/better fit of model.

The measures of fit imply that Spatial lag model fits good to the data under consideration i.e because LRT equals to 9.6334 and 6.3260 for spatial lag and for spatial error models respectively and the overall models are significant at 5% level of significance, and also when we take log likelihood into consideration, the two models have slightly different estimate.

Table 6. *Tests and Measures of Fit.*

Test	DF	Value	probability
Jarque-Bera	2	0.5193	0.77133
Breusch-Pagan test	12	17.8781	0.11944
Koenker-Bassett test	12	17.3055	0.13846
White test	122	86	0.0421*

Test	DF	Value	probability
LRT	1	9.6334	0.00191*

*significant at 5% level

3.4. Results of Bayesian Spatial Analysis of Malaria Prevalence Rate

The main results interpreted here refer to Bayesian spatial analysis. When performing this task, R software was employed with typically *CARBayes* package which is mainly developed for this purpose.

Prior Specifications

A multivariate Gaussian prior is assumed for β, and the mean μ_β and diagonal variance matrix ε_β are given as default values specified by the software with a constant zero-mean vector and diagonal elements of ε_β equal to 1000. That is β_j, that is, β_j $N(m_j; v_j)$ for $j = 0,...,p$, and the default values specified by the software are ($m_j = 0$; $v_j = 1000$). The scale parameter v^2 for the Gaussian likelihood is assigned a conjugate inverse gamma prior distribution, where the default specification is v^2 : Inverse-Gamma (0.001, 0.001). The variance parameter default prior specification for τ^2 has $a = b = 0.001$ where as σ^2 : Inverse-Gamma (a, b).

The Moran's I statistic equals 0.24425 with a corresponding p-value of much less than 0.05, which suggests that the residuals contain substantial positive spatial autocorrelation. To perform estimations, We generated simulations of the parameters with MCMC using the R software. Inference for this model is based on 24,000 MCMC samples, which were obtained following a burn-in period of 10,000 and thinning the remaining 240,000 samples by 10 to reduce their autocorrelation.

Table 7. *Posterior Quantities and DIC of Model with Covariates.*

	Median	2.5%	97.5%	%accept	Ge.diagnosis
(Intercept)	8.3802	4.4170	12.2172	100.0	-0.5
MEANAL	-0.0035	-0.0049	-0.0021	100.0	0.1
HHSIZE	0.0001	0.0000	0.0001	100.0	-0.5
AVHMBR	0.0749	-0.2509	0.3994	100.0	-0.4
Dng Wtr 100	0.0441	0.0093	0.0790	100.0	-0.3
Toilet 100	0.0123	-0.0097	0.0348	100.0	0.3
RADIO 100	0.0249	0.0045	0.0451	100.0	1.2
Phone ML	0.0075	-0.0216	0.0368	100.0	-0.6
F Type	0.0028	-0.0147	0.0202	100.0	0.3
W Type	0.0237	0.0036	0.0440	100.0	1.4
R Type	0.0081	-0.0135	0.0298	100.0	0.9
SPd12m	-0.0230	-0.0437	-0.0022	100.0	1.1
HB Nets	0.0028	-0.0189	0.0242	100.0	-1.1
nu^2	4.3504	3.1523	6.2530	100.0	-1.3
tau^2	0.0062	0.0018	0.0477	100.0	-1.4
rho	0.4829	0.0593	0.9218	41.8	0.3

DIC = 342.6089;

Table 7 shows the results of the Bayesian spatial estimates for the model parameters and random effects specifications. The residual spatial autocorrelation can be accounted for by adding a set of random effects to the model. Regarding these findings, the posterior median and 95% credible intervals (2.5%, 97.5%) for selected parameters and for a single chain

Geweke's diag criterion were presented.

Regarding these findings, the covariates such as mean woreda altitude, Proportion of households having access to piped water, Proportion of households having access to radio, House hold size of the district and main construction material of the room's wall, proportion of households sprayed in 12 months had statistically significant effect on malaria prevalence, as their 95% credible intervals do not include zero.

We can also observe that Proportion of households having access to piped water, Proportion of households having access to radio, House hold size of the woreda and main construction material of the room's wall had statistically significant and positive relationship with malaria prevalence. This means that increasing in Proportion of households having access to piped water, Proportion of households having access to radio, House hold size of the woreda and main construction material of the room's wall raises malaria prevalence. On the other way round, statistically significant random terms reveal that the covariates explain only some of (don't explain all) the variability in the Malaria prevalence rate of the woredas of rural area of the SNNP region, because none of the credible intervals of the variances include zero. The above output also shows that the random effects have modeled substantial spatial autocorrelation, as the posterior median for the spatial autocorrelation parameter ρ is 0.4829 positive and significant.

Table 8. *Posterior Quantities and DIC of Model without Covariates.*

	Median	2.5%	97.5%	n.sample	%accept	Ge.dig
(Intecpt)	6.8157	6.245	7.378	24000	100.0	-0.6
nu^2	6.2030	4.591	8.6807	24000	100.0	-0.8
tau^2	0.0062	0.002	0.0445	24000	100.0	0.7
rho	0.4884	0.069	0.917	24000	41.8	-0.7

DIC = 357.5705;

Table 8 presented the results of the Bayesian estimates for the model parameters without covariates, incorporating only the random effects. Hence, the random effects exhibit effects on the response Variable, as their 95% credible intervals do not include zero. Both models included an unstructured random effects term and a spatially structured random effects term. The major difference is the first model was modelled with covariates where as the second without covariates. The first model had a DIC of 342.6089; when extended to the second model, the DIC raised to 357.5705, a difference of 15.0384. The higher DIC for the second model indicates a poor fit relative to the fit of the first one. In our study, the presented results demonstrate that the parameters are tested for convergence using Geweke. diag. Hence since the Geweke. diag values for all parameters are in the range ($-1.96, 1.96$), clearly they are converged. Trace plots useful for diagnosing all posterior parameters β's are checked, a phenomenon in which the MCMC sampler covers the support of the posterior distribution which implies the MCMC outputs are a representative sample from the posterior distribution for the parameters. plots are attached on the appendices.
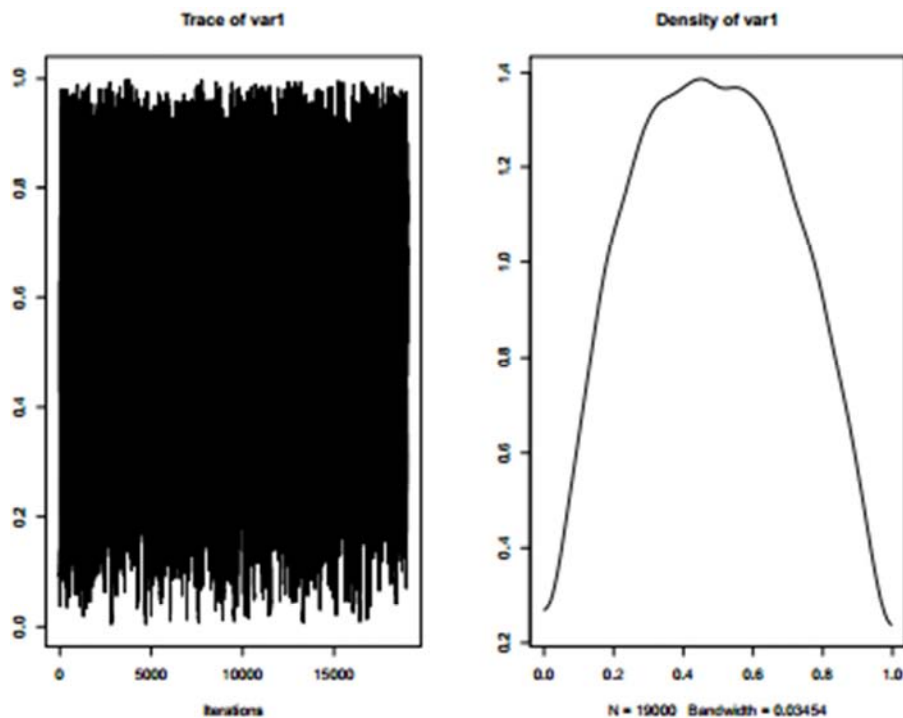

Figure 5. *Posterior Samples and Density Plot for ρ.*

Moreover, the acceptance rate for ρ quantifies the proportion of times the value proposed by the Metropolis updating step was accepted as the new value of the Markov chain. In contrast, due to the conjugacy between the Gaussian likelihood and the prior distributions for (β; ν^2; τ^2), Gibbs sampling is employed for updating these parameters, which is the reason for the 100% acceptance rate. Plotting this object thus yields a trace plot(left panel) and a density estimate (right panel) of the Posterior samples and density plot for ρ.

From figure 5, the first is a trace plot, which shows the evolution of the MCMC output as a time series for ρ, the second is a density plot, which shows a kernel density estimate of the posterior distribution.

When the posterior distribution is very skewed, the posterior median can be a better summary statistic. According to the posterior density of β's shown in Figures, the coefficient of the covariate can be considered as significantly positive given that its posterior mean is greater than 0 and its 95% credible interval is likely not to contain the value 0.

Mapping Posterior prediction for Malaria Prevalence rate

While predicting the malaria prevalence rate in the rural part of woredas of the region using the Bayesian posterior distribution, the Bayesian prediction has smaller standard deviation relative to the observed data ($1.49 < 2.513$) which is in line with study conducted by [37]. The following box map in Figure 6 shows the posterior medians(estimate) of the model fitted for each area, showing the pattern of Malaria prevalence of the rural area of the SNNP region.
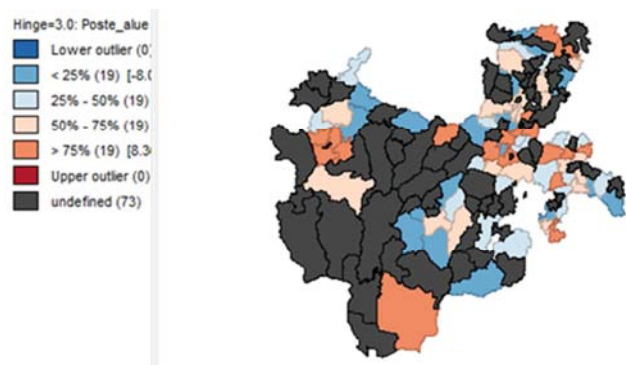

Figure 6. Map of posterior Means of the model fitted (Malaria prevalence).

4. Conclusions and Recommendations

4.1. Conclusion

In this study, ESDA, spatial Regression models and Bayesian Spatial analysis are used to explore the geographical distribution of malaria prevalence rate using EMIS 2011 data for sample data aggregated at Woredas level of data filtered for the rural part of the SNNP regional state, Ethiopia. The spatial analysis results revealed positive spatial autocorrelation pattern of malaria prevalence rates in space. The result of model specification and measures of fits shows that, the spatial lag model was found to be better fit to the data

and explain the geographical variations of malaria prevalence data in the region. The spatial lag model is found to be the correct choice to incorporate this spatial effect and so as to come up with unbiased and consistent parameter estimation.

The Risk Factors Mean altitude of a given woreda, Proportion of households having access to piped water, Proportion of households having access to radio, Proportion of households having access to phone, Main construction material of the room's roof and Proportion of household Sprayed Insecticide in the past 12 months have significant influence in explanation of malaria prevalence rate on the rural area districts of the SNNP regional state. From the Bayesian Spatial result, Significant random terms reveal that the covariates don't explain all the variability in the Malaria prevalence rate of the woredas of rural area of the region, because none of the credible intervals of the variances include zero. Overall, The results obtained from the spatial regression models are consistent with results of Bayesian Spatial analysis.

4.2. Recommendations

The following recommendations are forwarded based on the findings of this study. The study results suggest that there are 'malaria prevalence hot-spots' in the study area. The government and other Health related Non-Governmental Organizations should consider these results when planning malaria control measures. Awareness giving to people regarding the disease and control mechanisms to utilize and use these ways for instance, to apply anti mosquito spraying in 12 months will reduce malaria transmission. The implication of the spatial dependence is that, in cases where the decisions on how to allocate funds for interventions needs to have spatial dimension or woredas with neighborhoods.

We also recommend to study using recent data including all regions in Ethiopia to come up with a broader aspect of malaria prevalence. The local risk factors such as temperature, rainfall and humidity which might be important in explaining local clustering of the disease, are not considered in the analysis.

Acknowledgements

References

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

[2]   Anselin, L. (1995): Local indicators of spatial associations-LISA. Geographical Analysis, 27:93-115.

[3] Anselin, L. (1998): Exploratory Spatial Data Analysis in a Geocomputational Environment, PP. 77-94 in Geocomputation, A Primer, edited by P. A. Longley, S. Brooks, B. Macmillan and R. McDonnell. New York: John Wile.

[4] Anselin, L. (2005): Exploring Spatial Data with GeoDa: A Workbook. Spatial Analysis Laboratory, Department of Geography University of Illinois, Urbana-Champaign, Urban IL 61801.

[5] Asnakew, K., Sucharita, G., Afework, T., Dereje, O., 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, 8:5 doi:10.1186/1476-072X-8-5.

[6] Central Statistical Agency(CSA). 2007: Survey report, Ethiopia.

[7] Charles, D. (2012): Bayesian Hierarchical Approaches to Spatial Analysis of Injury and Disaster Data. Columbia University.

[8] Cressie, N. (1993): Statistics for Spatial Data. In revised ed. Wiley, New York.

[9] Congdon, P. (2003): Applied Bayesian Modelling. Chichester: Wiley.

[10] Chikodzi, D. (2013): Spatial Modelling of Malaria Risk Zones Using Environmental, Anthropogenic Variables and Geogra-Phical Information Systems Techniques. Journal of Geosciences and Geomatics 1, no. 1: 8-14. doi: 10.12691/jgg-1-1-2.

[11] Darper, N. and Smith, H. (1998): Applied regression Analysis, Third edition, John Wiley and Sons, New York.

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

[13] Duncan, L. (2017): An R Package for Spatial Areal Unit Modelling with Conditional Autoregressive Priors. University of Glasgow.

[14] Drissa, C., Stanislas, R., Mark, T., Youssouf, T., Matthew, L., Abdoulaye, K., Karim, T., Ando, G., Issa, D., Amadou, N., Modibo, D., Ahmadou, D., Mody, S., Bourema, K., Nadine, D., Jean, G., Renaud, P., Mahamadou, T., Christopher, P. and Ogobara, D. (2013): Spatio-temporal analysis of malaria within a transmission season in Bandiagara, Mali. Malaria Journal, 12:82.

[15] English, D. (1992): Geographical epidemiology and ecological studies. In Elliott, P., Cuzick, J., English, D., and Stern, R., editors, Geographical and Environmental Epidemiology. Methods for Small-Area Studies. Oxford University Press, Oxford, pp 3-13.

[16] Federal Ministry of Health (FMOH). (2014): National Strategic Plan for Malaria Prevention, Control and Elimination in Ethiopia:2014-2020. FMOH;

[17] Faraway, J. J. (2004): Linear Models with R. Chapman Hall, Boca Raton.

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

[19] Gemperli, A., Vounatsou, P., Sogoba, N., Smith, T. (2006): Malaria mapping using transmission models: Applications to survey data from Mali. Am J Epidemiology, 163:289-297.

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

[21] Ghebreyesus, T., Haile, M., Witten, H., Getachew, A., Yohannes, M., Yohannes, M., Teklehaimanot, D., Lindsay, W., Byass, P. (1999): Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey. BMJ, 319:663-666.

[22] Ghebreyesus, T., Haile, M., Witten, H., Getachew, A., Yohannes, M., Lindsay, W., Byass, P. (2000): Household risk factors for malaria among children in the Ethiopian highlands. Trans R Soc Trop Med Hyg, 94(1):17-21.

[23] 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.

[24] Githeko, K., John, A., Peter, O., Francis, A., Bryson, N., John, G. and Guiyun, Y., 2006: Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control Malaria Journal, 5:107

[25] 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.

[26] Haining, R. (2010): Spatial Data Analysis in the Social and Environmental Sciences, Cambridge University Press: Cambridge.

[27] Ingrid Peterson, Luisa N. Borrell, Wafaa El-Sadr and Awash Teklehaimanot. (2009): Individual and Household Level Factors Associated with Malaria Incidence in a Highland Region of Ethiopia: The American Society of Tropical Medicine and Hygiene: pp. 103-111.

[28] Tuyishimire, J. (2013): Spatial Modelling of Malaria Risk Factors in Ruhuha sector, Rwanda.

[29] Kiszewski, E., Tekelehaimanot, A. (2002): A review of the clinical and epidemiologic burdens of epidemic malaria.

[30] Kleinschmidt I, Bagayoko M, Clarke GPY, Craig M, Le Sueur D. (2000): A spatial statistical approach to malaria mapping. Int J Epidemiol, 29:355-61.

[31] LeSage, J., and Pace, K. (2009): Introduction to Spatial Econometrics, Boca Raton, Florida: CRC Press.

[32] Lubetzky-Vilnai, A., Ciol, M., & McCoy, S. W. (2013): Statistical Analysis of Clinical Prediction Rules for Rehabilitation Interventions: Current State of the Literature. Archives of Physical Medicine and Rehabilitation.

[33] Matheron, G. (1971): The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau.

[34] Merkle, E., Shev, C. and Trisha, G. (2005): Simulation Based Bayesian Inference Using Winbugs. Winbugs Tutorial Outline.

[35] Mitiku M. and Bute G. (2011): Statistical Analysis of Spatial Distribution of Malaria in West Shoa Zone, Ethiopia. Journal of Ethiopian Statistical Association, AA.

[36] Molla, E., and Ayele, B. (2015): Prevalence of Malaria and Associated Factors in Dilla Town and the Surrounding Rural Areas, Gedeo Zone, Southern Ethiopia. J Bacteriol Parasitol, 6:242.

[37] Omumbo, J., Hay, I., Snow, W., Tatem, J., Rogers, J. (2005): Modeling malaria risk in East Africa at high-spatial resolution. Trop Med Int Health, 10:557-566.

[38] Raso, G., Schur, N., Utzinger, J., Koudou, B., Tchicaya, E., Rohner, F., Nâ€™ Goran, E., Siluã©, D., Matthys, B., Assi, S., Tanner, M. (2012): Mapping malaria risk among children in Cã´te d'Ivoire using Bayesian geo-statistical models. Malaria Journal, 11:160.

[39] Rosas, A, Oscar, J., Gabriel, C., Niko, S., Juan, C., Dionicia, G., Edwar, P., Socrates, H. and Alejandro, L. (2015): Spatial clustering and risk factors in a low endemicity urban area of the northwestern Peruvian coast, *M*alaria Journal, 14:176.

[40] Robert, W., Marsh, K. (2002): The consequences of reducing transmission of Plasmodium falciparum in Africa.

[41] Stratton, L., O'Neill, M. S., Kruk, M. E., & Bell, M. L. (2008): The persistent problem of malaria: Addressing the fundamental causes of a global killer. Social Science & Medicine, 67(5), 854-862.

[42] Thomas, J., Lindsay, W. (2000): Local-scale variation in malaria infection amongst rural Gambian children estimated by satellite remote sensing. Trans R Soc Trop Med Hyg, 94(2):159-63.

[43] Wakefield, J., N. Best, and L. Waller. (2000): Bayesian Approaches to Disease Mapping. In Spatial Epidemiology Methods and Applications, edited by P. Elliott, J. Wakefield, N. Best, and D. J. Briggs. Oxford: Oxford University Press.

[44] Yamamoto, S., Louis, R, A., Sauerborn, R. (2010): Household risk factors for clinical malaria in a semi-urban area of Burkina Faso: a caseâ€"control study. Transactions of the Royal Society of Tropical Medicine and Hygiene, 104(1), 61-65.