

Spatial Modelling of Malaria Risk in Bayesian Setting: A Case Study of Wolaita and Dawuro Zones in SNNPR, Ethiopia

Desalegn Dargaso Dana (MSc)* Bereket Tessema Zewude (MSc.)

Department of Statistics, College of Natural and Computational Sciences, Wolaita Sodo University, Po.box 138, Wolaita Sodo, Ethiopia

Abstract

Background: Malaria is a mosquito-borne infectious disease affecting humans and other animals caused by parasitic protozoans. The main objective of the study was to develop Bayesian spatial model for malaria risk in Wolaita and Dawuro zone of Southren Regional State, Ethiopia.

Methods:

In this study, malaria data obtained from seven woreda Health Centers of Wolaita and Dawuro zones at 345 spatial locations were used. At these locations, about three hundred twelve respondents had malaria in their blood samples out of 5, 062 respondents were tested for malaria infection. In the analysis, Generalized Linear Mixed Model was fitted to estimate Generalized Linear Mixed Model parameters to identify significant explanatory variables analyzed by using statistical softwares (STATAversion 12 and SPSS version 16).

Result: The overall malaria prevalence in the study area was about 19.87%. Results indicate that malaria incidence follows spatial pattern because the test result indicates that there is statistically significant local clustering of malaria incidence at 5% level of significance. Statistically significant local clustering of malaria incidence is detected in all the woredas except in the two woredas (Kido Didaye and Loma woredas). In Kido Didaye and Loma woredas, the spatial correlation is negative that means the observed is less than that of expected value. The rest of the woredas exhibit positive spatial correlation since the observed value is greater than expected value. For Bayesian spatial models; the environmental factor elevation was negatively associated with malaria risk. This is to mean that as elevation above sea level of the study area increases, the chance of being a candidate of malaria decreases. A negative relation of maximum temperature with malaria risk reveals that the lower the maximum temperature the higher malaria risk.

Conclusions: Spatial modelling of malaria risk was the basis for differentiation of predicted malaria prevalence from high level to low on a map. The differentiation may allow effective use of limited financial and human resources. It also helps to identify priority areas to control malaria in case of change of climatic variables.

Keywords: Spatial autocorrelation, GLMM, Local risk factor, Global Local risk factor

1. Introduction

Malaria is a mosquito-borne infectious disease affecting humans and other animals caused by parasitic protozoans or a group of single-celled microorganisms belonging to the *Plasmodium* type. Malaria cannot be prevented by vaccination but it can be prevented by medications, mosquito elimination, and prevention of bites. Prevention of malaria may be more cost effective than treatment of the disease in the long run. Use of mosquito nets assist mosquitoes reduce infection rates and transmission of malaria. Malaria is one of infectious disease affecting humans and other animals and the disease regularly transmitted by a contaminated female anopheles mosquito. A large number of children in Africa were passed away due to malaria. Out of two hundred seven million malaria cases, estimated amount of four hundred seventy three thousand to seven hundred eighty nine thousand were executed due to malaria disease [1]. Ethiopia is one of the Sub-Saharan African countries in which malaria leftovers is a leading communicable disease [2].

Malaria is relatively meso-endemic, a severing infectious disease caused by the protozoan parasite called Plasmodium and which shows seasonal fluctuations due to seasonal variations in climatic conditions. Climates are highly related with the risk of malaria distribution and the transmission prototype is usually unstable and seasonal distinguished by cyclic prevalent epidemics. There are four main types of malaria communicate a disease to humans: Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae and Plasmodium ovale. In Ethiopia, Plasmodium falciparum and Plasmodium vivax are the majority leading parasites [3]. When we compare malaria death in Africa with America, death due to malaria in Africa is very massive than malaria death in America. The number of death due to malaria in all age group was internationally declined from 2000 to 2015. However, a huge number of children killed by malaria in sub-Saharan Africa. In this report the large number of death were recorded in Africa (about 90%) and remaining 10% encompassed from other country [4].

In study conducted in Butajira, Ethiopia, a total of 19,207 individual samples in six survey on Prevalence of Plasmodium infection in about $0.93\% \left(\frac{178}{19,207} \right) * 100\%$ were malaria positive. In the study, individuals with lowest wealth status affected higher compared with middle and higher wealth index (1.74%, 0.62% and 0.42%) respectively. At 5% level of significance, age category, gender, household, wealth index status, house status,

altitudinal strata and seasons all have significant influence on malaria infection. Due to seasonal variation, October to November 2009 category has highest influence on malaria prevalence and January to February has lowest influence. This is due to Malaria shows a strong seasonal pattern. The researcher had used multilevel mixed-effects logistic regression analysis and found that increased malaria prevalence in individuals categorized under five age group and when age group increases the risk of malaria decreases (1.77%, 1.68%, 0.74% and 0.45% for age category under 5 age category, 5-9, 10-14 and more than or equals to 15 respectively) [5].

Prevalence and risk factors of malaria study conducted in Ethiopia, the researcher examined age, sex, family size, region, toilet facility, main source of drinking water, availability of television, number of rooms (persons), main material of room's wall, main material of room's roof, main material of room's floor, anti-malarial spraying, number of nets (persons) and use of mosquito nets were statistically significant at 5% level of significance. Base on the result investigated via fitting a generalized linear model by using statistical software (SAS version 9.2) the increment of the age (OR = 0.970, CI = 0.319 – 2.505) of respondent inversely related with the odds of malaria prevalence [6].

Today, Ethiopia has made progress in reducing the number of malaria nationally, but the observed changes are not sufficient enough compared to the desired goals of the response against the epidemic. Given the size of population and the magnitude of the damage inflicted, it will take us a number of years to see significant declines in malaria prevalence and incidence. Malaria death and birth are known to vary geographical location and depend on eco-climatic conditions. Demographic, eco-climatic death factors, age, 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. In this study the spatial distributions of malaria is assessed using spatial model along with meteorological and environmental variables of malaria incidence in Wolaita and Dawuro zones 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. Environmental variation risks can be quantified using spatial models of prevalence and morbidity heterogeneity. Furthermore, modeling the probability of contracting malaria is helpful to identify a group with higher chance of getting the disease and take evidence based on prevention measure.

2. Methodology

2.1 Source of Data

In this study, malaria data obtained from seven woreda Health Centers of Wolaita and Dawuro zones at 345 spatial locations were used. At these locations, about three hundred twelve respondents had malaria in their blood samples out of 5, 062 respondents were tested for malaria infection.

2.1.1 Measurement

The dependent variable in this study was malaria positive counts of individuals at household locations during the malaria season at the study area. Predicting whether an event will or will not occur and identifying the variables in making the prediction is an important step in carrying out the study. The independent variables: Minimum Temperature (MIT), Maximum Temperature (MAT), Rainfall (RF), Distance to nearest Water Body (DWB), Economic Status (Poor or Medium Status) Elevation (ELN), Normalized Difference Vegetation Index (NDVI) and Relative humidity (RH).

2.2 Statistical Methods

Statistical models are useful for quantifying the relation between malaria risk and environmental factors and upon this relation predicting malaria risk at locations without observed malaria data.

2.2.1 Generalized Linear Mixed Model (GLMM)

The generalized linear mixed model is an additional room of the generalized linear model, complicated by random effects. It has gained significant popularity in recent years for modeling binary/count, clustered and longitudinal data. The origin of the likelihood function is also straightforward for generalized linear mixed models. However, numerical methods are needed in most cases to obtain the Maximum Likelihood Estimators. The two scholars give a brief review of some numerical techniques, such as a restricted pseudo-likelihood, the Gauss-Hermite quadrature, and Bayesian methods [7].

The malaria prevalence data were treated as binomial data and modeled by means of the logistic regression. Malaria positive counts of individuals at closer geographical locations are under similar climatic and environmental conditions. Analysis of spatially correlated data under the assumption of independence leads to overestimation of the statistical significance of the covariates. Based on the existing geographical information, spatial model incorporate spatial correlation. Spatial correlation is commonly taken as a function of the distance between locations for geo-statistical data [8]. At each spatial household location i ; $i = 1, 2, 3, \dots, n$ malaria test of

individuals may result either malaria positive or negative individuals in their blood samples and $X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{bmatrix}$ be

the vector of p associated environmental predictors observed at location i . Suppose Y_i is binomially distributed random variable for all i and it is assumed to be the number of observed malaria cases between N_i number of individuals in a household located at i . Thus, $Y_i \sim \text{Bin}(N_i, P_i)$, with parameter P_i measuring malaria risk at location i . $E(Y_i) = N_i P_i$, $g(P_i) = X_i \beta$, Where $g(\cdot)$ represents a link function which is given to be a logit function in malaria risk study, $X_i = (1, X_{i1}, X_{i2}, \dots, X_{iq})'$ is a vector of q associated environmental predictors

observed at location i and $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix}$ is a vector of regression coefficients.

3. Results

3.1.1 Descriptive Statistics for the Variables used in the Study

In this study, malaria data obtained from seven woreda Health Centers of Wolaita and Dawuro zones at 345 spatial locations were used. At these locations, about three hundred twelve respondents had malaria in their blood samples out of 5, 062 respondents were tested for malaria infection. The overall malaria prevalence in the study area was about 19.87%. The spatial data used for the spatial stationary model and validation are displayed in the given figure in appendix (figure 1).

3.1.2 The Moran's I and Geary's C coefficient

Moran's I is similar but not equivalent to a correlation coefficient. It varies from -1 to $+1$. In the absence of autocorrelation and regardless of the specified weight matrix, the expectation of Moran's I statistic is $\frac{-1}{n-1}$, which tends to zero as the sample size increases. For a row-standardized spatial weight matrix, the normalizing factor S_0 equals n and the statistic simplifies to a ratio of a spatial cross product to a variance. A Moran's I coefficient larger than $\frac{-1}{n-1}$ indicates positive spatial autocorrelation, and a Moran's I less than $\frac{-1}{n-1}$ indicates negative spatial autocorrelation. Geary's C ranges from 0 (maximal positive autocorrelation) to a positive value for high negative autocorrelation. Its expectation is 1 in the absence of autocorrelation and regardless of the specified weight matrix [9]. If the value of Geary's C is less than 1, it indicates positive 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. In this section, our focus is on their application to particular data analysis, the essential task being to seek for spatial pattern. First, the global Moran's I and Geary's C test statistics were computed to test the null hypothesis.

Null hypothesis

- There is no significant clustering of malaria incidence in the entire study region.
 $(H_0: \rho = 0)$

Alternative hypothesis

- There is significant clustering of malaria incidence in the entire study region at 5% level of significance
 $(H_1: \rho \neq 0)$. The test was repeated using diagnostic for spatial dependence to validate the consistency of results.

3.1.3 Moran's I and Geary's C Test Statistics for Global Spatial Autocorrelation

The aim of estimating Moran's I and Geary's C test statistics for global spatial autocorrelations to measure the strength of spatial autocorrelation surrounded by neighboring woreda of malaria incidence, to seek for spatial pattern or to diagnosis for spatial dependence in regression model.

In the above table (table 2), the test results indicate the presence of significant global spatial autocorrelation of malaria incidence. These global results in the distribution of malaria need to be further explored using local spatial statistics. Based on the P-values of the Moran's I and Geary's C coefficients output, we can reject the null hypothesis of there is no spatial autocorrelation or there is no significant clustering of malaria incidence in the entire study region. Additionally, the calculated Z - statistic for Moran's I is positive and for Geary's C is negative showing the existence of significant positive spatial autocorrelation (clustering).

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

Local statistics are used to identify where high or low values cluster. Local Moran's *I* and local Getis statistics are computed to test the null hypothesis of no local spatial clustering among malaria incidence at neighboring woredas [10]. As can be seen from the results of local Moran's *I* test (table 3), it shows results of local Moran *I* as a function of neighboring values. Results indicate that malaria incidence follows spatial pattern because the test result indicates that there is statistically significant local clustering of malaria incidence at 5% level of significance. Kindo Koysha, Damot Woyide, Duguna Fango, Tercha and Marka woredas are statistically significant local clustering of malaria incidence where as Kido Didaye and Loma woredas are statistically insignificant at 5% level of significance. In these woredas which are categorized under insignificant at a given level that implies high value is surrounded by low values and vice-versa is true for neighboring woredas.

3.1.5 Inference for Spatial Non-stationary Model

In appendix (table 1), Coefficients whose the 95% dependable interval does not include zero with in interval are considered to be significant and Coefficients whose the 95% dependable interval does not include zero with in interval are considered to be significant. Explanatory variables ELN, MAT, MIT, NDVI and RF are statistically significant at 5% level of significance.

The 95% credible intervals for the parameters in spatial non-stationary modelling of malaria risk in Bayesian setting. Moreover, the median of the spatial correlation parameter in sub-region one and two were 0.21089 with the 95% dependable interval (0.03254, 0.26249) and 0.00720 with the 95% dependable intervals (0.00985, 0.00986) respectively. For Bayesian spatial models; the environmental factor elevation was negatively associated with malaria risk. This is to mean that as elevation above sea level of the study area increases, the chance of being a candidate of malaria decreases. A negative relation of maximum temperature with malaria risk reveals that the lower the maximum temperature the higher malaria risk.

3.1.6 Fitting Spatial Autoregressive Models

Spatial autoregressive models are the error generating process and operate with spatial weight matrices that specify the strength of interaction between neighboring sites [11]. We will use a spatial autoregressive model to measure the relationships between malaria incidence rate and meteorological variables obtained at a neighborhood sites. In the given table below (table 4), results of Spatial Lag Model-Maximum Likelihood Estimation output analysis of malaria risk, all the variables except cold zone were found to have significant effect on malaria distributions at 5% level of significance.

Consequently, the fitted spatial lag model equation is given by: -

$$\hat{Y}_i = \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5 + \hat{\beta}_6 X_6$$

Estimated Malaria risk = 1.260 (spatially lagged malaria incidence) + 1.261 (mid-land zone) + 2.688 (hot zone) + 0.697 (rainfall) + 1.459 (maximum temperature) + 1.084 (minimum temperature).

As the values of coefficients of variables (table 4) implies the statistical association between malaria incidence and minimum temperature (1.084) was less compared with that of association between malaria incidence and maximum temperature (1.459). When the amount of temperature increases it would increase the rate of mosquito emergence from breeding places. The increased temperature allows the development of parasites to occur in the mosquitoes, and the mosquito population also increases as the temperature rises.

4. Discussion

This study is undertaken to develop and apply Bayesian spatial model for risk in Wolaita and Dawuro Zone of SNNP Regional State, Ethiopia. Secondary data used for analysis was taken from selected health centers of two woredas mentioned in above statement. Malaria disease is predisposed by both environmental and climatic factors. The relation of malaria prevalence and explanatory variables can be established by statistical models.

According to this research finding minimum temperature, maximum temperature, rainfall, hot zone and mid-land zone are statistically significant at 5% level. Other similar study conducted in Switzerland shows that NDVI, maximum temperature, minimum temperature and rainfall were significantly associated with malaria prevalence. As statistical output presented in table 4, areas characterized by minimum temperature, maximum temperature, rainfall, hot zone and mid-land zone were strongly associated with the risk of malaria and risk of spatial clustering. In adding up, maximum and minimum temperatures were found to be significant, indicating strong relationship between temperature and malaria risk [8]. The main concern of spatial study of malaria risk was to demonstrate spatial patterns and levels of malaria prevalence. The findings of the study showed that malaria prevalence exhibits spatial patterns with varying levels of malaria prevalence and a significant association with climatic and environmental variables. Global Moran's *I* computed over the study area suggested the presence of spatial cluster of malaria positive individuals. The presence of clustering was in agreement with the studies carried out by [13]. In this study, rainfall was associated negatively with malaria prevalence. The outcome was not in agreement with no association of rainfall with malaria prevalence outcome obtained [13]. Again, it was not in agreement with the study carried out in Switzerland in SNSM [8]. The possible explanation for the outcome of the study could be the more likely get ride off larva of infectious mosquitoes by high and

relatively consistent levels of rainfall across the study area over the malaria season. On the contrary to the positive association of distance to nearest water body to malaria prevalence done in Switzerland [8], it was not significantly associated with malaria prevalence in this study. This may be due to seasonal or episodic nature of the water bodies or people awareness in keeping the water bodies clean so as to avoid the possible occurrence of malaria. High rainfall may be a manifestation of high relative humidity. According to this study, the higher the rainfall, there could be less likely occurrence of malaria. Similar relation may hold for relative humidity. However, it was not related with malaria prevalence in this study. This is possibly the insignificant variation of relative humidity over the study area. Moreover, local spatial statistics were used to test the spatial dependency in the patterns of malaria distribution, detect pockets of disease 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 risk distribution were analyzed. The result helps to identify woredas with malaria burden.

5. Conclusions

This study was performed for Spatial Modelling of Malaria Risk in Bayesian Setting in a Case Study of Wolaita and Dawuro Zones in SNNPR, Ethiopia. The results of the analysis show that the risk of malaria in the study area exhibits a spatial pattern which is dependent on some meteorological variables. Non-spatial analysis was performed by fitting Generalized Linear Mixed Model using the explanatory variables in cooperation. As a result, minimum temperature and NDVI were positively and significantly related with malaria prevalence. The results of the analysis also obscured negative and significant effects from elevation above sea level, maximum temperature and rainfall on malaria risk. Distances to nearest water body and relative humidity were not significantly related with malaria prevalence. In addition to this, significant local clustering of malaria risk occurs among woredas within neighboring woredas. Spatial modelling of malaria risk was the basis for differentiation of predicted malaria prevalence from high level to low on a map. The differentiation may allow effective use of limited financial and human resources. It also helps to identify priority areas to control malaria in case of change of climatic variables.

List of abbreviations

DWB	Distance to nearest Water Body
ELN	Elevation
FMoH	Federal Ministry of Health
GLMM	Generalized Linear Mixed Model
MAT	Maximum Temperature
MIT	Minimum Temperature
NDVI	Normalized Difference Vegetation Index
RH	Relative humidity
SNNPRS	Southern Nation Nationalities and People Regional State
SPSS	Statistical Package for Social Sciences
WHO	World Health Organization

Acknowledgements

We would like to thank Wolaita Sodo University and selected health centers of both Wolaita and Dawuro zones for providing relevant data and, Health center managers for their support during data collection. We thank Paulos Lukas (Lecturer at department of Geography in College of Social Science and Humanity) for his professional support of the paper.

References

- [1] World Health Organization, World Malaria Report 2014.
- [2] Federal Democratic Republic of Ethiopia Ministry of Health, *National Strategic Plan for Malaria Prevention, Control and Elimination in Ethiopia 2011–2015*, Addis Ababa, Ethiopia, 2010
- [3] FMoH, National Five Year Strategic Plan for Malaria Prevention and Control in Ethiopia, 2006–2010, Ministry of Health, Addis Ababa, Ethiopia, 2006.
- [4] World Health Organization, World Malaria Report 2015.
- [5] Adugna Woyessa, Wakgari Deressa, Ahmed Ali and Bernt Lindtjorn, 2013. Malaria risk factors in Butajira area, south-central Ethiopia: a multilevel analysis. *Malaria Journal* 2013, 12:273
- [6] Dawit G Ayele, Temesgen T Zewotir and Henry G Mwambi, 2013. Prevalence and risk factors of malaria in Ethiopia. *Malaria Journal* 2012, 11:195
- [7] Antonio, K. and J. Beirlant, 2006. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, in press.
- [8] Gosoniu L., Vounatsou P., Sogoba N., Smith.T, 2006, Bayesian Modelling of Geostatistical Malaria Risk Data: Swiss Tropical Institute, Basel, Switzerland, Malaria Research and Training Center, University of

- Mali, Bamako, Mali Geospatial Health, vol. 1, pp.127-130.
- [9] Sokal, R.R. & Oden, N.L. (1978) Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society*, **10**: 199-228.
- [10] Getis, A. and Ord, J. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* **24(3)**: 189-206.
- [11] Cressie, N. A. 1993. Statistics for spatial data. Revised edition. John Wiley and Sons, New York, New York, USA.
- [12] Yeshe wondim A., Gopal S., Hailemariam A., Dengela D. and Patel H., 2009, Spatial Analysis of Malaria Incidence at the Village Level in Areas with Unstable Transmission in Ethiopia, *International Journal of Health Geographics*, 8:5
- [13] Reid H., Haque U., Clements A., Tatem A., Vallely A., Ahmed S., Islam A., Haque R., 2010, Mapping Malaria Risk in Bangladesh Using Bayesian Geostatistical Models, *American Journal of Tropical Medicine and Hygiene*, vol.83(4), pp. 861-867, Bangladesh

Appendix

Table 1: Posterior Estimates in Spatial Non Stationary Model.

Variables	Median	95% Dependable interval
INT	54398.1876	(-8589.02894, 87691.5219)
DWB	0.53241	(-0.39871, 0.26548)
ELN	-0.69850	(-0.89241, -0.41259)*
MAT	-69.26927	(-96.35281, -13.86101)*
MIT	91.12528	(16.25907, 102.59109)*
NDVI	0.85691	(0.00896, 0.99895)*
RF	-95.36879	(-159.02951, -26.00818)*
RE	-0.52044	(-0.79337, 0.22011)
RH	-85211.52648	(-109.23658, 12529.02009)
λ_1	0.21089	(0.03254, 0.26249)
λ_2	0.00720	(0.00985, 0.00986)
δ_1^2	0.60123	(0.65009, 0.98502)
δ_2^2	0.00347	(0.00013, 0.03095)

Table 2: Results of Global Moran's *I* and Geary's *C* Statistics

Assumption	Coefficient	Observed	Expected	Dev.Std	Z	Prob > Z
Normality	Moran's I	0.7341	-0.0215	0.2001	3.78	<.0003*
Normality	Geary's C	0.0413	0.9872	0.3511	-2.69	<.0020*

Asterisked *p*-values indicate the significance of the covariates at 5% level.

Table 3: Results of Local Moran's *I* Test

S. No.	Woreda	Observed	Expected	Std Dev.	Z	P-value
1	Kindo Koyisha	0.8412	-0.3553	0.5322	2.25	0.0005
2	Kindo Didaye	-0.3291	-0.1063	0.0950	-2.35	0.1009
3	Damot Woyide	0.7121	-0.3391	0.3587	2.93	0.0005*
4	Duguna Fango	0.2160	-0.2579	0.1072	4.42	0.0004*
5	Tercha	0.4572	-0.2930	0.0936	8.01	0.0001*
6	Loma	-0.5057	-0.1173	0.1405	-2.76	0.0607
7	Marka	0.5481	-0.6305	0.4069	2.90	0.0038*

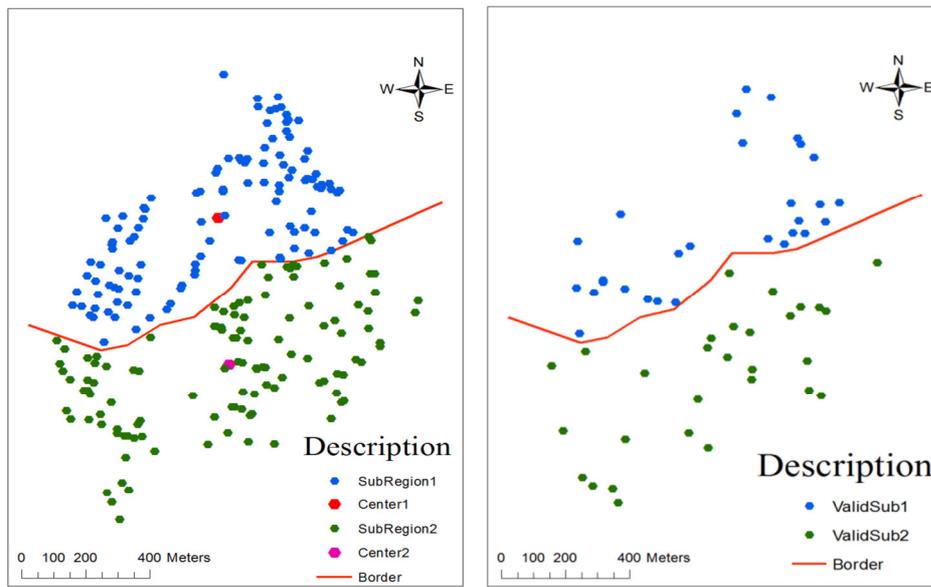
Asterisked *p*-values indicate the significance of the covariates at 5% level.

Table 4: Results of Spatial Lag Model Estimation

Variable	Coefficient	St.d error	t-statistic	P-value
Constant	0.517	0.315	1.64127	0.101
W-malaria	1.260	.349	3.61032	.0000*
Cold	-0.263	0.230	-1.1435	0.253
Mid-land	1.261	.443	2.84650	.0041*
Hot	2.688	.445	6.04045	.0000*
Rainfll	.697	.336	2.07440	.0381*
Max temp.	1.459	.513	2.84405	.0040*
Min.temp.	1.084	.367	2.95368	.0030*

Asterisked p-values indicate the significance of the covariates at 5% level

Figure 4.1: Spatial Data for Spatial Stationary Model



(a) Spatial Data for Estimation

(b) Spatial Data for Model Validation