Assessment of Environmental Determinants of Acute Gastro Enteritis using Geographically Weighted Regression Analysis

Vidyalakshmi R[1], Pawlin Vasanthy Joseph[2] , Balasubramanian S [3] and M Prashanthi Devi [1]
[1]Department of Environmental Management, Bharathidasan University, Tiruchirappalli-24
[2]Department of Zoology, Nirmala College, Coimbatore
[3]JSS University, Mysore.

About the Author:

R.Vidyalakshmi
UGC Project Fellow
Department of Environmental Management
Bharathidasan University
Tiruchirappalli-620 024
Mobile: 8220360953
Mail: vidyalakshmi07@gmail.com

Academic Profile

- Pursuing Doctoral Degree under the Guidance of Dr. M.Prashanthi Devi, Assistant Professor, Department of Environmental Management, Bharathidasan University, Tiruchirappalli-620 024
- Working as a Junior Project Fellow in UGC MRP project "Assessing drinking water quality in schools of Tiruchirappalli region and mapping common flu and diarrhea among school children"
- MBA in Environmental Management, Bharathidasan University.

Research Interest: Medical Geography, Spatial Statistics

1.0 Introduction

Disease mapping helps to investigate the geographical distribution of a diseases burden on a certain population. Area specific estimates of risk may be informative to health managers by estimating the disease burden in a specific area and the informal comparison of risk maps with exposure may provide clues to etiology or generate hypothesis. When addressing cases of variability in residual spatial risk and its variability, hotspot analysis in a similar vein, regression will be aided with prior information on magnitude and forms of non spatial and spatial variability.

Geographically Weighted Regression differs from disease mapping in which it estimates the association between risk and co-variates, rather than to provide area specific relative risk estimates. This distinction has important implications for modeling both the mean function and the residual variability. The variable of interest is typically available at only a limited number of spatial locations, whereas the independent variables are mapped across the entire landscape. Statistical methods such as logistic regression are used to develop predictive equations and these equations are then applied to the unsampled areas to generate a predictive map (Yabsley et al., 2005).

Geographically Weighted Regression (GWR) is a local version of spatial regression that generates parameters disaggregated by the spatial units of analysis. This allows assessment of the spatial heterogeneity in the estimated relationships between the independent and dependent variables. The use of Markov Chain Monte Carlo (MCMC) methods can allow the estimation of complex functions, such as Poisson-Gamma-CAR, Poisson-lognormal-SAR, or Over dispersed logit models. Geographically-Weighted Regression models were constructed to identify spatial variation of relationships across rural Bangladesh to predict diarrheal disease risk with tube well density. Zero inflated negative binomial regression models were run to simultaneously measure the likelihood of disease events (Carrel et.al, 2011). A step-wise regression technique was used to fit the statistical model for climatic factors affecting Dengue Hemorrhagic fever (Promprou et.al, 2005)

Wakefield (2006) provides critical reviews of methods suggested for the analysis of aggregate count data in the context of disease mapping and spatial regression for male lip cancer. In disease mapping studies, hierarchical models can provide robust estimation of area level risk parameters, though care is required in the choice of covariate model and it is important to assess the sensitivity of estimates to the spatial model chosen.

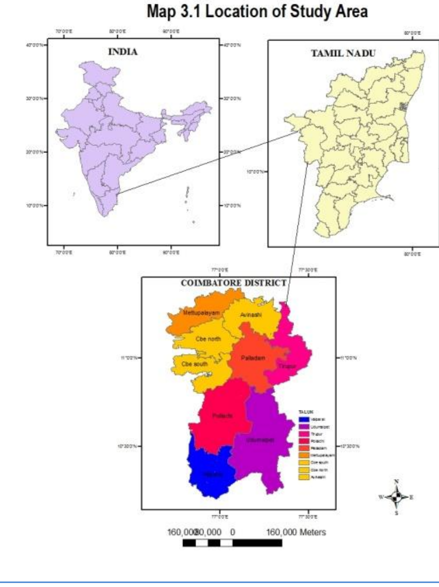
2.0 Objectives

Earlier spatial statistics was approached through external statistical software and the results were mapped in ArcGIS. The present ArcGIS10.0 facilitates spatial statistics whereby GWR and OLS regression analysis can be performed in the software through the input of accurately attributed maps. With the help of such facility, for this present study, we have analysed the influence of environmental parameters on the exposure risk of AGE in the Coimbatore region of Tamil Nadu using GWR 3.0 model in ArcGIS10.0.

3.0 Study area description

The Coimbatore district covers an area of 7805 sq.km falls between 10°13'4" North to 11° 24'5" North latitude and 76° 39' 25" East to 77° 18' 26" East longitude. Coimbatore is situated in the extreme west of Tamilnadu, near the state of Kerala. The district has 9 taluks and 481 revenue villages.

Map 3.1 Location of Study Area



4.0 Data and Methods

For the present study, the epidemiology data, climatic data, water quality data and socio demographic data were obtained for the period 2000-2009 for each revenue village.

4.1 Acute gastroenteritis incidence data

The Acute gastroenteritis data used in this study was collected from the Directorate of Health, Coimbatore under the Government of the Tamilnadu Health Ministry and from five major Private Hospitals located in the urban centres of Coimbatore North and South taluks of Coimbatore district. The recordings of incidences were collected from January 2000 to December 2009 (both months inclusive). Each registered case; outpatient and inpatient who required treatment were treated as incidence figures. The incidence data was collected from nine Government hospitals from the Directorate of Health, Coimbatore and from five major private hospitals of Coimbatore district.

4.2 Environmental data

The climatological data namely rainfall (in mm), Temperature (in ºC), socio demographic variables such as drinking water facilities and sanitation and the water quality for a period of ten years (2000 to 2009) for each village from the respective governmental departments**.**

4.2.1 Dependent Variable

The disease infection data was prepared according to monthly seasons as dry summer, (March, April, May), Wet Summer (June, July, August) winter (September October, November) and Dry Winter (December January February). The population was added as an offset variable in the analysis. Therefore, the incidence rate per area is considered as the dependent variable.

4.2.2 Co-variates

The independent variables included several factors that are related to the spatial variation of diseases risk and agents that are pre-seemed to promote disease prevalence. For this study we have used Rainfall and Temperature and local sanitation as the independent variables. The observed Mean Maximum Temperature and Mean Maximum Rainfall data for the study period was subjected to season wise analysis. Therefore the data were prepared as maximum temperature for dry winter, dry summer, wet winter and wet summer. Similar preparation was done for the maximum rainfall data also.

5.0 Results

In this study, the AGE annual cumulative incidence rates (IR), given as cases per existing populations, was used as the measure of disease severity, and as the dependent variable; independent variables were Wet Winter Rainfall (RWW), Wet Summer Rainfall (RWS), Pit latrines and Open drainages . A summary of the variables in both Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models are shown in Table 5.1. OLS regression was first applied, in an attempt to explain the global relations between dependent and independent variables. The model was set as: $IR = \beta0 + \beta1_{RWW} + \beta2_{RWS} + \beta3_{\text{Pit latrines}} + \beta4_{\text{Open Drainages}} + \varepsilon$. β0 and β1..n were the regression coefficients whereas ε was the model random error.

Table 5.1 Summary of dependent and independent variables used in OLS and GWR

Variable		Numerator	Denominator
Dependent	IR	Incidences	Population Density
Independent	RWW	Total Rainfall in (Sept, Oct, Nov)	Number of months
	RWS	Total Rainfall in (June, July Aug)	Number of months
	Pit Latrine	Number of Pit latrines	House holds
	Open drainage	Number of Open drains	Total number of drains

The diagnoses of an OLS model were approached by assessing multicollinearity and the residuals. The multicollinearity was assessed through variance inflation factor (VIF) values, and if VIFs were greater than 10, this indicated multicollinearity existed (Menard 2002). The spatial independency of residuals was evaluated by the spatial autocorrelation coefficient, Moran's I, which was expressed as:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \overline{y})(y_j - \overline{y})}{\left(\sum_{i \neq j}\sum w_{ij}\right)\left(\sum_{i=1}^{n}(y_i - \overline{y})^2\right)} \qquad \text{………….5.1}$$

Where, n was the total number of cases in the study. i and j represented different villages. yi was the residual of i, and y was the mean of residuals. wij was a measure of spatial proximity pairs of i and j (Wong and Lee, 2005). We used the inverse of the distance between i and j for specifying the relationship between them.

A GWR local model was applied to analyze how the IR-RWW, RWS, Pit latrines and open drainage relationships changed from one village extent to another. It was a localized multivariate regression that allowed the parameters of a regression estimation to change locally. Unlike conventional regression, which produced a single regression equation to summarize global relationships among the independent and dependent variables, GWR detected spatial variation of relationships in a model and produced maps for exploring and interpreting spatial non-stationarity (Fotheringham et al, 2002). GWR was calibrated by multiplying the geographically weighted matrix $w_{(g)}$ consisting of geo-referenced data.

The spatial variability of an estimated local regression coefficient was examined to determine whether the underlying process exhibited spatial heterogeneity. The regression model can be rewritten as

$$IRi(g) = \beta_0 i(g) + \beta_{1\ RWW}(g) + \beta_{2\ RWS}(g) + \beta_{3\ \text{Pit latrines}}(g) + \beta_{4\ \text{Open Drainages}}(g) + \varepsilon i, \text{ ……5.2}$$

Where (g) indicated the parameters that were estimated at each village in which the coordinates were given by vector g; i represented each village.

By applying GWR modeling, the spatial influences among neighborhoods could be assessed, which was not able to be achieved through traditional OLS methods. The local collinearity as well as the independency and normality of residuals of GWR model to evaluate the fit of the model was also estimated. The local collinearity was assessed by scatter plots of the local coefficient estimates for RWW. RWS, Pit Latrine and Open drainage and Condition number. The condition number is the square root of the largest eigenvalue divided by the smallest eigenvalue. If the condition numbers are greater than 30, multicollinearity would be a very serious concern. The adjusted coefficient of determination (Adjusted R2) and ANOVA were used for comparing OLS and GWR models. Akaike Information Criterion (AIC) generated for OLS and corrected Akaike Information Criterion (AICc) calculated for GWR were also used for model comparison (Fotheringham et al., 2002.)

All analyses were implemented using ESRI Arc GIS 10.0 and GWR 3.0 with 0.05 significance level. In the GWR model, the adaptive kernel with AICc estimated bandwidth was chosen. The adaptive kernel was chosen because the distribution of AGE was inhomogeneous in the study area.

5.1. OLS Regression

The spatial distributions of the Incidence rates (IR), RWW, RWS, Pit latrines and Open drainages were mapped in Map 5.1. The map of cumulative IR showed high values in Ikkaraiboluvampatti and moderates rates scattered in some northern regions. The north eastern areas generally had lower IR values than middle and southern areas in the Coimbatore. The results of applying OLS regression showed that holding the variable of Incidences and Rainfall at wet and dry seasons vary at low levels. However, the pit latrine and open drainages contribute lesser to the diseases (Table 5.2). The VIF values indicated OLS estimations were not biased from multicollinearity. However, this global regression model explained only 18 percent of the total variance of IR with the AIC -665.36. Since the existence of dependent residuals violates the model assumptions, GWR model was employed to fit the data. GWR was used to present the spatial diversities of the IR-rainfall and pit latrines and open drainages relationships.

Table 5.2 Ordinary Least Squares (OLS) Estimates

Parameter	Estimated Value	Standard Error	P-value
Intercept	-5.686	4.23	0.182
RWW	0.00063	0.013	0.5011
RWS	-0.0172	0.000935	0.1956
Adjusted R^2	0.108		
AIC	-665.36		

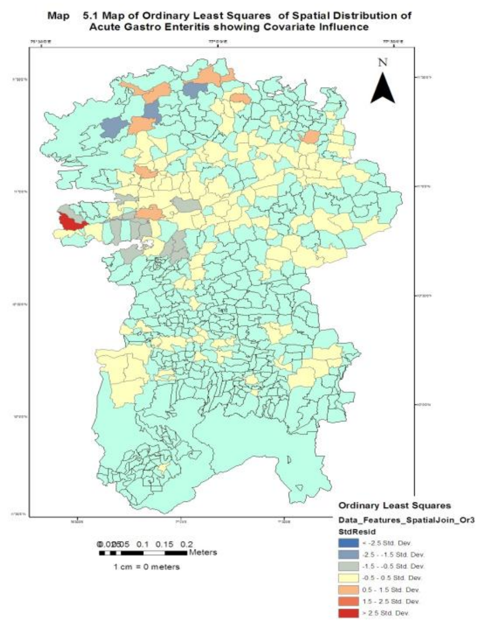
5.2 GWR Model and Spatial Variations

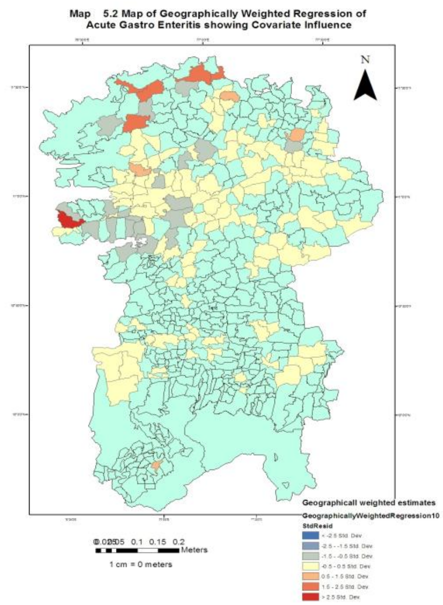
The summary results of GWR are listed in Table 5.3 and showed the GWR was more similar to the OLS model and GWR could explain 14 percent of the total model variation with the decreased AICc -668.64. However, the ANOVA comparison results also showed the GWR local model was significantly more appropriate than the OLS global model ($F = 2.89$, $p < 0.05$).

Table 5.3 Geographically Weighted Regression (GWR) Estimates

Parameter	Minimum	Maximum	Standard error
Intercept	-0.04083	-0.04085	0.012
RWW	0.00007	0.0008	0.00003
RWS	0 .0003	0.00038	0.00109
Condition Number	17.98		
Adjusted R^2	0.149		
AIC	-668.647		

Map 5.2 shows more regions of high incidences than that of map 5.1 indicated. This shows how well the GWR model replicated the incidences rates with rainfall. It was obvious that the value was not homogeneously distributed in all villages, and the overall GWR regression fitted best in districts Ikkaraiboluvampatti, Marudur, Chikkasampalayam, Odanthurai, Irrumburai, and Muduthurai. This model did not fit well in other regions, and this could imply additional covariates were needed to explain the IR in district 12. Map 5.2 helped us realize whether additional explanatory factors were required and where could those factors be applied.





The GWR models have high explanatory power with the parameters being very significant and a residual deviance value close to the number of degrees of freedom (d_f). The diagnosis of the parameters shows that significant independent variables and dependent variables exhibit high spatial variability and more geographical heterogeneity. The overall map of GWR and AGE shows that rainfall, influence the risk of AGE in these regions and the influence of pit latrines and open drainages have to be further probed.

6.0 Conclusion

This study provides further indications that the relationships of Incidence rates and rainfall were spatially non-stationary Coimbatore region. In regression maps, it is clear that the intensity and directions of the influence of rainfall during summer and winter on AGE incidence were different in the study area. This result gives the policy makers more ideas how to better adopt specific control and prevention strategies to specific areas. The use of the available models in Arc GIS makes it more easier for decision making and related health interventions

7.0 References

Carrel M., Escamilla V, Messina J, Glebultowicz S, Winslon J, Yernus M, Streatfield K, Emch M(2011), Diarrhoeal disease risk in rural Bangladesh decreases as tubewell density increases: a zero inflated and Geographically Weighted Analysis, Int.J.Health Geogr. 10:41

Fotheringham A.S., Brindson C., Charlton M., Geographically Weighted Regression: The Analysis of spatial varying relationships: Wiley, New York, NY, USA, 2002

Menard S. Applied Logistic Regression Analysis, 2nd ed.; Sage: Newbury Park, CA, USA, 2002.

Pamprou S, Jaroensustasinee M, Jaroensustasinee K (2005), Climatic factors affecting Dengue haemorrhagic fever incidence in Southern Thailand. Dengue Bulletin 29: 41-48

Wong, D.W.S.; Lee, J. Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS; John Wiley and Sons: Hoboken, NJ, USA, 2005.

Yabsley M.J, Wimberly M.C, Stallknecht D.E, Little S.E, Davision W.R (2005) Spatial analysis in the distribution of *Ehrilichia Chaffeensis*, causative agent of human monocytotrophic ehrilichiosis across a multi state region. Am J Tro Med Hyg 72:840-850.