International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Estimation of Environmental and Geographical Determinants of Acute Gastro Enteritis Using Geographically Weighted Regression Analysis

Pawlin Vasanthi Joseph[1], Prashanthi Devi[2], Balasubramanian S[3]

[1]Department of Zoology, Nirmala College for women (Autonomous), Redfields, Coimbatore – 641018, Tamilnadu, India

[2]Department of Environmental Management, Bharathidasan University, Tiruchirapalli – 620024, Tamilnadu, India

[3]JSS Medical University, Srivratheeswararnagar, Mysore -570015, Karnataka, India

Abstract: ***Disease mapping helps to investigate the geographical distribution of a disease burden on a certain population. Spatial regression differs from disease mapping in that the aim is to estimate the association between risk and co-variates, rather than to provide area specific relative risk estimates. Geographically Weighted Regression (GWR) is a statistical technique that allows variations in relationships between predictors and outcome variable over space to be measured within a single modeling framework. In the present paper we have combined rainfall and sanitation data to explain the exposure risk of Acute Gastroenteritis in Coimbatore district of Tamilnadu, India, using Geographically weighted regression model. 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. Ordinary Least Square regression (OLS) showed high incidence rates in Ikkaraiboluvampatti and moderates rates scattered in northern regions.GWR regression fitted best in villages Ikkaraiboluvampatti, Marudur, Chikkasampalayam, Odanthurai, Irrumbarai, and Muduthurai. This study provides further indications that the relationships of Incidence rates and rainfall were spatially non-stationary in Coimbatore region.***

Keywords: Spatial regression, co-variates, Geographically Weighted Regression, Ordinary Least Square regression, spatially non-stationary

1. Introduction

Disease mapping helps to investigate the geographical distribution of a disease 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

Spatial regression differs from disease mapping in that the aim is to estimate the association between risk and co-variates, rather than to provide area specific relative risk estimates. 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[1].

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. Geographically Weighted Regression (GWR) is a statistical technique that allows variations in relationships between predictors and outcome variable over space to be measured within a single modeling framework[2, 3]. Applications of GWR include studies in a wide variety of demographic fields including the analysis of health and disease[4,5,6,7], health care delivery[8], environmental equity[9], population density and housing[10].

Geographically-weighted regression models were constructed to identify spatial variation of relationships across rural Bangladesh to predict diarrheal disease risk with tube well density[11].The relationship of the number of weekly non-cholera cases were examined with rainfall and temperature using generalized linear regression models[12]. In the present paper we have combined rainfall and sanitation data to explain the exposure risk of Acute Gastroenteritis in Coimbatore district of Tamilnadu, India, using Geographically weighted regression model.

2. Data and Methods

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.

a) 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 sanitation information as the independent variables.

b) Climate

The Climatic variables rainfall was recorded as Mean Maximum Rainfall and Mean Minimum Rainfall. The observed Mean Maximum Rainfall data for the study period was subjected to season wise analysis. Therefore the data were prepared as maximum rainfall for dry winter, dry summer, wet winter and wet summer.

c) Sanitation

The parameters that define good sanitation of a residential area such as number of households (bathroom facilities, pit latrines, Water closets, number of latrines, closed/open drainages, and number of drainages) and the bathroom, toilet and drainage facilities for each village were taken as co-variates and subjected to analysis.

3. Geographically Weighted Regression (GWR)

GWR is the term introduced by Fotheringham, Charlton and Brunsdon [13] to describe a family of regression models in which the coefficients, β, are allowed to vary spatially. GWR uses the coordinates of each sample point or zone centroid, t_i, as a target point for a form of spatially weighted least squares regression. The GWR model can be expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{j} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i \{1\}$$

where $y_i$is the value of the outcome variable at the coordinate location i where (u_i,v_i)denotes the coordinates of i, β_0 and β_i represents the local estimated intercept and effect of variable j for location i, respectively. To calibrate this formula, a bi-square weighting kernel function is frequently used[14] to account for spatial structure. The locations near to i have a stronger influence in the estimation of $\beta_i(u_i, v_i)$ than locations farther from i. In the GWR model localized parameter estimates can be obtained for any location i which in turn allows for the creation of a map showing a continuous surface of parameter values and an examination of the spatial variability (non-stationary) of these parameters. The result is a model of the form:

$$y = X\beta(t) + E \{2\}$$

The coefficients β (t) are determined by examining the set of points within a well-defined neighborhood of each of the sample points. The functions utilized in the GWR software package are of the form:

$$f(d) = e^{-d^2/2h^2}(or)\ f(d) = e^{-d/h}(or) f(d) = \left[1 - \frac{d^2}{h^2}\right]^2, d < r, f(d) = 0\ otherwise \{3\}$$

Using a selected kernel function and bandwidth, *h*, a diagonal weighting matrix, W(*t*), may be defined for every sample point, *t*, with off-diagonal elements being 0. The parameters β(*t*) for this point can then be determined using the standard solution for weighted least squares regression:

$$\hat{\beta}(t) = (X^T W(t) X^{-1} X^T W(t) y)$$

or, letting $D = (X^T W(t) X)^{-1} X^T W(t)$

$$\hat{\beta}(t) = D_y \text{ and } var\left(\hat{\beta}(t)\right) = DD^T\sigma^2 \{4\}$$

The standard errors of the parameter estimates can be computed as the square root of these variances and used in *t*-tests to obtain estimates of the significance of the individual components. In this model the variance component, σ^2, is defined by the normalized Residual Sum of Squares (RSS) divided by the degrees of freedom. The latter are defined by the number of parameters, *p*, in a global model, or the *effective number of parameters* in the GWR model. This value is approximated by the authors as the trace of a matrix S, tr(S) (the sum of the diagonal elements of S) defined by the relation: y= S_y

4. Results

In this study, the Acute Gastroenteritis 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 1, OLS regression was first applied, in an attempt to explain the global relations between dependent and independent variables. The model was set as:

$$IR = \beta0 + \beta1_{RWW} + \beta2_{RWS} + \beta3_{Pit\ latrines} + \beta4_{Open\ Drainages} + \varepsilon.$$

β0 and β1..n were the regression coefficients whereas ε was the model random error.

Table 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 No. of Pit latrines	House holds
	Open drainage No. of Open drains	Total No. 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[15]. The spatial independency of residuals was evaluated by the spatial autocorrelation coefficient; Moran's I, which was expressed as:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i\neq j}\sum w_{ij}\right)\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)} \quad \{5\}$$

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[16]. 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 regression model can be rewritten as
$IRi(g) = \beta_0 i(g) + \beta_{1RWW}(g) + \beta_{2RWS}(g) + \beta_{3Pit\ latrines}(g) + \beta_{4\ Open\ Drainages}(g) + \varepsilon i,$
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[2].

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 Acute Gastroenteritis was inhomogeneous in the study area.

5. OLS Regression

The spatial distributions of the Incidence rates (IR), RWW, RWS, Pit latrines and Open drainages were mapped in Map 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 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 (Table2). 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.

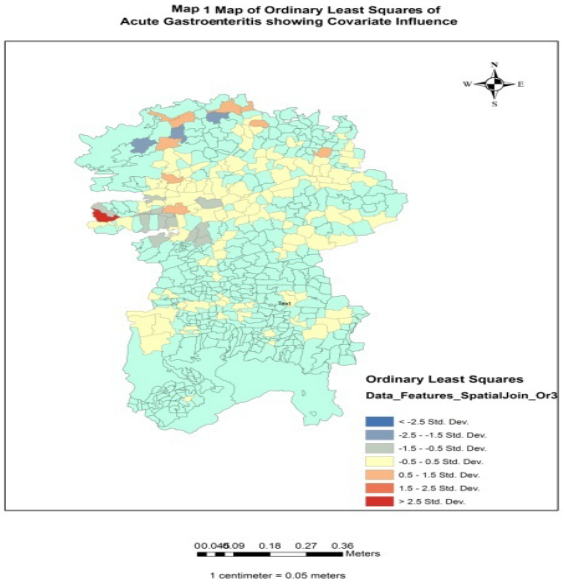


Table 2: Ordinary Least Squares (OLS) Estimates

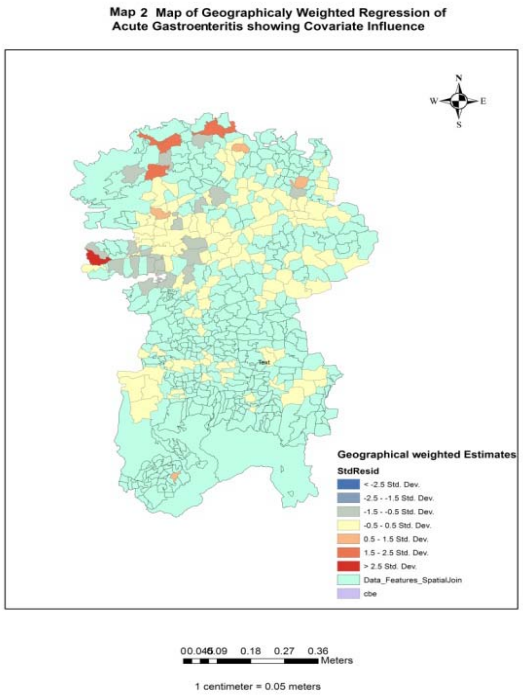
Parameter	*Estimated value*	*Standard error*	*P-value*
Intercept	-5.686	4.23	0.182
RWW	0.00063	0.013	0.5011
RWS	-0.0172	0.000935	0.1956
Adjusted R^2	0.108		
AIC	-665.36		

GWR Model and Spatial Variations

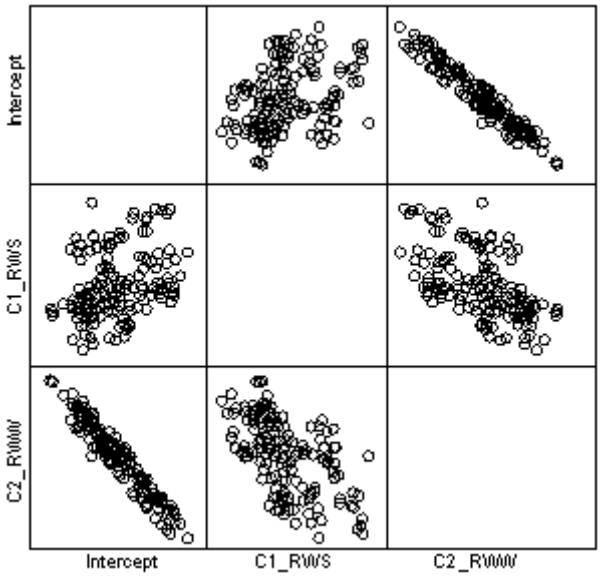
The summary results of GWR are listed in Table3 and showed the GWR was more similar tothe OLS model and GWR could explain 14 percent of the total model variation with the decreased AICc -668.64. However, the ANOVA comparison results also showed the GWR local model was significantly more appropriate than the OLS global model ($F = 2.89$, $p < 0.05$).

Table 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		



Map2 shows more regions of high incidences than that of Map 1. 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 villages Ikkaraiboluvampatti, Marudur, Chikkasampalayam, Odanthurai, Irrumbarai, and Muduthurai. This model did not fit well in other regions, and this could imply additional covariates were needed to explain the IR in Coimbatore district. Map.2 helped us realize whether additional explanatory factors were required and where could those factors be applied. The condition number shown in Table3 and the matrix scatter plot of the GWR coefficients suggested multicollinearity was not serious (Figure 1).



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 Acute Gastroenteritis shows that rainfall, influence the risk of Acute Gastroenteritis in these regions and the influence of pit latrines and open drainages have to be further probed.

6. Conclusion

This study provides further indications that the relationships of Incidence rates and rainfall were spatially non-stationary in Coimbatore region. In regression maps, it is clear that the intensity and directions of the influence of rainfall during summer and winter on Acute Gastroenteritis 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.

7. Acknowledgement

The authors would like to thank the District Directorate of Health and Preventive Medicine, Coimbatore for providing the disease data and the District Collectorate of Coimbatore for providing the rainfall and sanitation data.

References

[1] M.J. Yabsley, M. C. Wimberly, D. E. Stallknecht, S. Little, W. R. Davision, “Spatial analysis in the distribution of Ehrilichia Chaffeensis, causative agent of human monocytotrophicehrilichiosis across a multi state region,” Am J Trop Med Hyg, 72: pp. 840-850, 2005
[2] A. S. Fotheringham, C. Brunsdon, M. E. Charlton, “Geographically weighted regression: The analysis of spatially varying relationship,” New York, NY: Wiley, 2002
[3] National Center for Geocomputation, Maynooth, Ireland: National University of

Ireland.http://ncg.nuim.ie/ncg/GWR/software.htm (Febuary 1, 2012), 2009.
[4] P. Gooaverts,"Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson krigging,"Int J Health Geog, 4: 31, 2005
[5] T. Nakaya, A.S. Fotheringham,C.Brunsdon,M. Charlton,"Geographically weighted Poisson regression for disease association mapping,"Statistics in Medicine,24(17): 2695-2717. doi:10.1002/sim.2129, 2005
[6] T.C. Yang,H.W. Teng,M. Haran,"The impacts of social capital on infant mortality in the U.S.: A spatial investigation,"Applied Spatial Analysis and Policy,2(3): 211–227. doi:10.1007/s12061-009-9025-9, 2009
[7] Y.Y.J. Chen,P.C.Wu,T.C. Yang,H.J. Su,"Examining non-stationary effects of social determinants on cardiovascular mortality after cold surges in Taiwan,"Science of the Total Environment, 408(9): 2042–2049. doi:10.1016/j.scitotenv.2009.11.044, 2010
[8] C. Shoff,T.C. Yang,S.A. Matthews,"What has geography got to do with it? Using GWR to explore place-specific associations with prenatal care utilization,"GeoJournal,doi:10.1007/s10708-010-9405-3, 2012
[9] J.L. Mennis,L.M. Jordan, "The distribution of environmental equity: exploring spatial nonstationarity in multivariate models of air toxic releases,"Annals of the Association of American Geographers,95(2):249-268. doi:10.1111/j.1467-8306.2005.00459.x, 2005
[10] J.L. Mennis,"Mapping the results of geographically weighted regression,"The Cartographic Journal,43(2): 171-179. doi:10.1179/000870406X114658, 2006
[11] M. Carrel,V. Escamilla,J. Messina,S. Glebultowicz, J. Winslon,M. Yernus,K. Streatfield,M. Emch,"Diarrhoeal diseaserisk in rural Bangladesh decreases as tubewell density increases: a zero inflated and Geographically Weighted Analysis,"Int.J.HealthGeogr, 10:41, 2011
[12] M. Hashizume,B. Armstrong,S. Hajat,Y. Wagatsuma,S.G. Faruque,T. Hayashi,D.A. Sack,"Association between climate variability and hospital visits for noncholeradiarrhoea in Bangladesh: effects and vulnerable groups,"Int J Epidem, 36:1030-1037, 2007
[13] A.S. Fotheringham,C. Brunsdon,M.E. Charlton,"Two techniques for exploring non-stationarity in geographical data,"Geographical Systems,4: 59-82, 1997
[14] C. Brundson,A.S. Fotheringam,M. Charlton,"Geographically Weighted Regression Modelling spatial Non-stationarity,"The statistician, 47(3): 431 – 443, 1998
[15] S. Menard, " Applied Logistic Regression Analysis," 2nd ed.; Sage: Newbury Park, CA, USA, 2002
[16] D.W.S. Wong, J. Lee,"Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS, John Wileyand Sons: Hoboken, NJ, USA, 2005

Author Profile

Dr Pawlin Vasanthi Joseph is an Associate Professor in the Department of Zoology, Nirmala College for women, Coimbatore. Her Ph.D. work was on Epidemiological and spatial statistical approaches in the study of Acute Gastroenteritis in Coimbatore district. She has published five papers in National and International journals.

Dr Prashanthi Devi M is an Assistant Professor in the Department of Environmental Management, Bharathidasan University, Tiruchirapalli. Her specialization is in Hyperspectral analysis of vegetation indices, Image processing analysis and GIS analysis. She has published more than 25 papers.

Dr Balasubramanian S is the Director of Research in JSS Medical University, Mysore, Karnataka. He has guided 26 Research Scholars and published more than 70 research papers. He has organized a number of training programmes in GIS and Remote sensing and is a member of many Professional bodies.