RESEARCH ARTICLE

Geo-statistical Dengue Risk Model using GIS techniques to identify the risk prone areas by linking rainfall and population density factors in Sri Lanka

Nirosha Sumanasinghe[1], Armin Mikler[2], Chetan Tiwari[3] and Jayantha Muthukudage[4,*]

[1,2]*Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203, USA*
[3]*Deparment of Geography and the Environment, University of North Texas, Denton, TX 76203, USA*
[4]*Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka*

Abstract: Frequent dengue outbreaks is one of the main health related problems in Sri Lanka. The biggest outbreak occurred in 2014 with 47,246 dengue cases identified. An effective analysis of the epidemic is a vital part in controlling the outbreak. There is an uncertainty in identification of the relationship of dengue outbreak and influencing factors such as rainfall and population density. Hence, a careful study of these factors is needed. Ordinary Least Square (OLS) regression was first applied to find its suitability in identification of the linear relationship. OLS analysis conducted under this study revealed OLS is not a good method to model the relationship between dengue incidence and influencing factors. Then Geographically Weighted Regression (GWR) analysis was conducted and it outran OLS in modeling the relationship. For explanatory variables rainfall and population density, OLS can only explain 33.2% of the variance of dengue incidence while GWR can explain 56.3% of the same. GWR can identify the spatially non-stationary behavior of influencing factors on dengue incidence. These analyses revealed the influence of rainfall and population density is location dependent and hence need local analysis over conventional global analysis. All 25 districts in Sri Lanka were selected as the study area of this study. Rainfall and temperature data were prepared by applying pre-processing on data obtained from GSMaP remote sensing data archive. Dengue incidence data was obtained from Epidemiology Unit of Sri Lanka. The geo statistical risk model generated can be used to identify high risk areas in Sri Lanka. The high risk area map can be used to cater dengue control programs to effectively address the dengue epidemic.

Keywords: Dengue Risk Model, Dengue epidemic analysis, Geographically Weighted Regression.

INTRODUCTION

Dengue is a major public health burden to most of the countries in the tropical and subtropical regions of Asia, Australia, and Africa (WHO/TDR 2009). In 2012, dengue ranked as the most important mosquito borne viral disease in the world. Dengue Fever (DF) outbreaks were recorded as early as in late 1940s in South East Asia. During 1950-1970s, dengue was eliminated from most of the countries in the American region (Gubler 1998). Nevertheless, dengue has re-emerged globally with intensified epidemic and geographical expansion since 1980s, and has rapidly become a major epidemiological threat in Asia Pacific and South America. According to the World Health Organization (WHO 2012), DF is currently endemic to more than 100 countries with an estimated 2.5 billion population at risk. The annual number of infections is estimated at around 50 - 100 million globally, with 500,000 severe cases accounting for the majority of the approximately 12,500 deaths. Asia Pacific shoulders about two thirds of the global burden of dengue and is home to about 70% of the 2.5 billion world populations at risk (WHO/TDR 2009, WHO/WPRO 2013). In past decades, nearly all the countries in Southeast Asia reported dengue cases annually.

Since 1962, Sri Lanka experienced outbreaks of dengue. Dengue became endemic in 1989. Since then there has been large number of cases reported every year. By the end of May 2016, the total number of cases reported was 14,204. The mortality rate of dengue is 3% of the number of cases reported. Dengue experts and medical professionals agree that there is an urgent need for a comprehensive management plan to curtail the impact of the disease. Currently, a Presidential Task Force has been setup to effectively address the need of dengue mitigation needs. Government agencies and other national and international organizations are

cooperatively working to keep the mortalities at a minimum.

Since no successful treatment or vaccinations available, the methods of control dengue are controlling the vector population and avoid human- vector contact. Education and awareness programs can keep human mosquito contact at a minimum. Mosquito population can be controlled by destroying breeding sites. Identification of high risk areas is also a key success factor in controlling vector population. Analysis of spatial pattern of the vector population distribution plays a major role in concentrating in the most needed area.

The timing of the epidemics in these dengue endemic countries tends to vary with seasonal cycles. Dengue is mainly an urban disease; however, studies have suggested dengue has expanded its territory from urban to rural (Muhammad *et al.*, 2011). Dengue mosquitoes are container breeders. There is an increase in mosquito population soon after heavy rains. After rain, more breading sites are created. But, heavy rainfall may lower dengue cases as it reduces the survival rate of Aides (Yang *et al.*, 2009, Fouque *et al.*, 2006, Tran *et al.*, 2004). Heavy rainfall creates abundant outdoor breeding sources for Aedes. The study conducted by authors (Pathirana, 2009) presented that there is a strong relationship between rainfall and dengue incidence. In Sri Lanka, most of dengue incidence is reported in western province. Around 60% of cases are reported in Western Province. Twenty five percent of Sri Lankan population lives in Western Province which is made up of three districts, namely Colombo, Gampaha and Kaluthara. Population density and dengue incidence have a positive relationship as presented n WHO (1997) which revealed that areas with a high population density report higher cases as it causes many people to be exposed to the outbreak, even if the mosquito house index shows a low-density value. Majority of the researches done in the past focused on global factors when modeling the relationship between population density, rainfall and dengue incidence. The relationship of population density and rainfall on dengue incidence are not well understood. Studies based on spatial variation also present confusing results about dengue transmission. A study conducted in South Thailand (Thammapalo *et al.*, 2005) found that there was a negative correlation between rainfall

and Dengue Hemorrhagic Fever (DHF) incidence; whereas Central Thailand showed positive correlation between the same factors (Wiwanitkit, 2005). Authors of (Picardal and Elnar, 2012) performed a study in Central Visayas (Philippines) and that revealed there is no correlation between the two.

Geographically Weighted Regression (GWR) model is the most suitable model than global regression model where spatial heterogeneity of data is present (Brunsdon *et al.*, 1998)**.** The main goal of this study was set to analyze dengue incidence distribution pattern by developing a geostatistical dengue risk model that can identify risk areas. Rainfall data, population density of districts of Sri Lanka and dengue incidence data were selected as input data for the study. The outcome of this study has a practical importance for authorities dealing with dengue epidemic.

MATERIALS AND METHODS

Study Area

Sri Lanka is an island with an area of 65,000 km^2 which is located between latitudes 5° and 10°N and longitudes 79° and 82°E. The maximum length and the width of the island measured as 432 km and 224 km respectively. Population of Sri Lanka is 20,359,439 as at the population census conducted in year 2012. The average annual rainfall is 2500 mm. There are two major rainy seasons in Sri Lanka. Southwest Monsoon is in effect from May to August and Northeast Monsoon is in effect from November to January. Northeast Monsoon brings less rainfall compared to Southwest Monsoon. The average temperature ranges between 25 °C and 30 °C. The district which is the second level of administrative division of Sri Lanka is used as the base spatial unit in this study. All 25 districts were used in modeling the dengue epidemic.

Data Collection

For this study, rainfall data, population data, and dengue cases data were collected. The method of collection for each data type is given in the following sections. Shape files of administrative boundaries of Sri Lanka was obtained from GADM; the Database of Global Administrative Areas (GDAM 2016).

Remote Sensing Data

Rainfall data (2011-2015) was obtained using the Global Satellite Mapping of Precipitation (GSMaP) (RESTEC 2016). Temporal resolution of the data set is one hour and spatial resolution is 0.1 degrees (Longitude-Latitude). Preprocessing on acquired data is required to convert hourly basis data into monthly basis rainfall data. We took the cumulative sum of the hourly data set for a given month to generate the monthly rainfall data. This is mainly due to the fact that the case data is obtained on monthly basis and we need to match the time resolution of data sets. Each data set is loaded into a PostgreSQL database with PostGIS extension. Each district includes multiple rainfall measuring points. To obtain the rainfall figure for each administrative region, a spatial intersection operation is applied between loaded rainfall data points and administrative region boundary. Then the average monthly rainfall was calculated for the district as there is multiple rainfall measuring points encapsulated within a single district.

Population and Dengue Case Data

Population data is obtained from Department of Census of Sri Lanka for each district. District-level dengue case data is obtained from Epidemiology Unit of Sri Lanka. Both data sets were converted into a format usable in PostgreSQL database. Dengue incidence R_d was calculated for each region for every year as R_d = cases/ population*100,000. Monthly data for each year was used as repeated observations. For example, five observations of case data reported for the month of January for each year from 2011 to 2015 was recorded. Monthly dengue incidence from year 2011 to 2015 is given in Figure 1. According to the (Yang *et al.*, 2009, Fouque *et al.*, 2006, Tran *et al.*, 2004), it is clear that the dengue incidence closely follow the rainfall pattern. Hence, it is very important to study the rainfall pattern of the country. Just after every rainy season, there is a hike in dengue case data reported.

Temporal and Spatial Analysis of Data

In order to find out the relationship of dengue incidence with rainfall and population density Ordinary Least Squares (OLS) regression and Geographically Weighted Regression (GWR) analyses were conducted on the data set. OLS is a regression method that analyzes the impact of the independent variables on the dependent variable for the entire study region. However, local variations of influencing factors for each district are not taken into consideration when computing OLS. Since rainfall and other independent variables are spatially correlated (non-stationary, GWR is the most suitable modeling process for this situation.
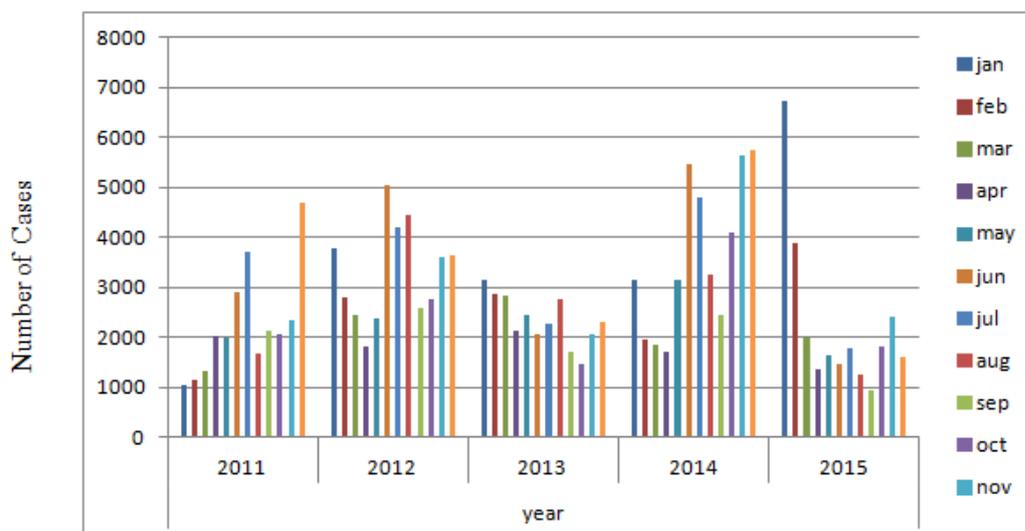

Figure 1: Dengue monthly incidence from 2011 to 2015.

To handle spatially varying properties, it is required to employ a technique which takes spatiality into account in regression analysis. GWR is the successful technique used over many years (Brunsdon *et al.*, 1998). We use GWR to analyze the spatial relationship of the dengue cases with rainfall and population data as given in equation 1. Let R_{di}, $i = 1,...,n$, be the response observations (dengue incidence) collected from location i in space. The corresponding covariate vector is $X_i = (1, X_{i1}, X_{i2},…, X_{ip})^t$ of dimension $(p+1)$ including the constant 1 for intercept.

$$R_{di}= \beta_0(u_i,v_i)+X_{ip}\beta_1(u_i,v_i)+X_{ir}\beta_2(u_i,v_i)+\varepsilon_i \quad (1)$$

Where $\beta(u_i,v_i)$ indicates the vector of the location-specific parameter estimates, (u_i,v_i) represents the geographic coordinates of location i in space, X_{ip} is covariate vector for population, X_{ir} is covariate vector for rainfall, and ε_i is the error term with mean zero and common variance σ^2.

Dengue monthly incidence (R_d) was chosen as the dependent variable that is calculated as per the equation given in the section 2.2.2. Population density (p) and the rainfall (r) in each district from 2011 to 2015 were selected as independent variables. The global model was executed first to determine which variables are important. The dengue incidence R_d was calculated according to the OLS equation given in the equation 2.

$$R_{di}= \beta_0+\beta_1 p_i+ \beta_2 r_i+\varepsilon \quad\quad (2)$$

where p_i is the population in ith region and r_i is the rainfall for ith region.

For some certain areas, rainfall and population density might be important predictors of dengue cases whereas it gives a weak prediction in other locations. Therefore, it is concluded that the study parameters are not suitable for OLS global model.

The standard residual map that GWR produces is an indicator of the model performance. Residuals are the portion of the total variability of the observed data that is unexplained by the model or the part of the model under and over predictions. Spatial Autocorrelation is also used to check whether the model residuals have a random spatial pattern.

RESULTS AND DISCUSSION

Spatial Analysis of Dengue Case Data in Accordance with Population and Rainfall Data

Spatial distribution of dengue incidence shows that incidence in Colombo district is continuously being high from 2011 to 2015. High population density in Colombo is accounted for having high incidence for five continuous years. The maps of dengue incidence from 2011 to 2015 are shown in Figure 2. Areas in green color represent the lowest incidence while the areas in red color represent the highest incidence.

Ordinary Least Square Analysis of Data

In OLS results VIF values shows whether the predictor variables are multicollinear. VIF < 10 means the variables are not multicollinear. In this study, rainfall and population density were used as explanatory variables. VIF value determined these two factors are not correlated and hence can be used in regression analysis together. Every explanatory variable used is unique and contributing to the variation in dengue incidence. OLS regression result also shows that the Adjusted R-Squared value is 0.332054 for the year 2014. This indicates the model built with a combination of population density and rainfall data explains 33.2% of the variation in dengue incidences. According to the OLS regression results, all explanatory variables (rainfall and population density) are statistically significant but the value for Jarque-Bera statistics is also significant. Significance in Jarque-Bera statistics indicates the model is biased and hence undesirable. Also the Koenker test is statistically significant (P value < 0.01 for both rainfall and population). This implies non-stationary relationship between the dependent and some or all of the explanatory variables. That reveals the explanatory variables (rainfall and population density) behave differently in different spatial regions.
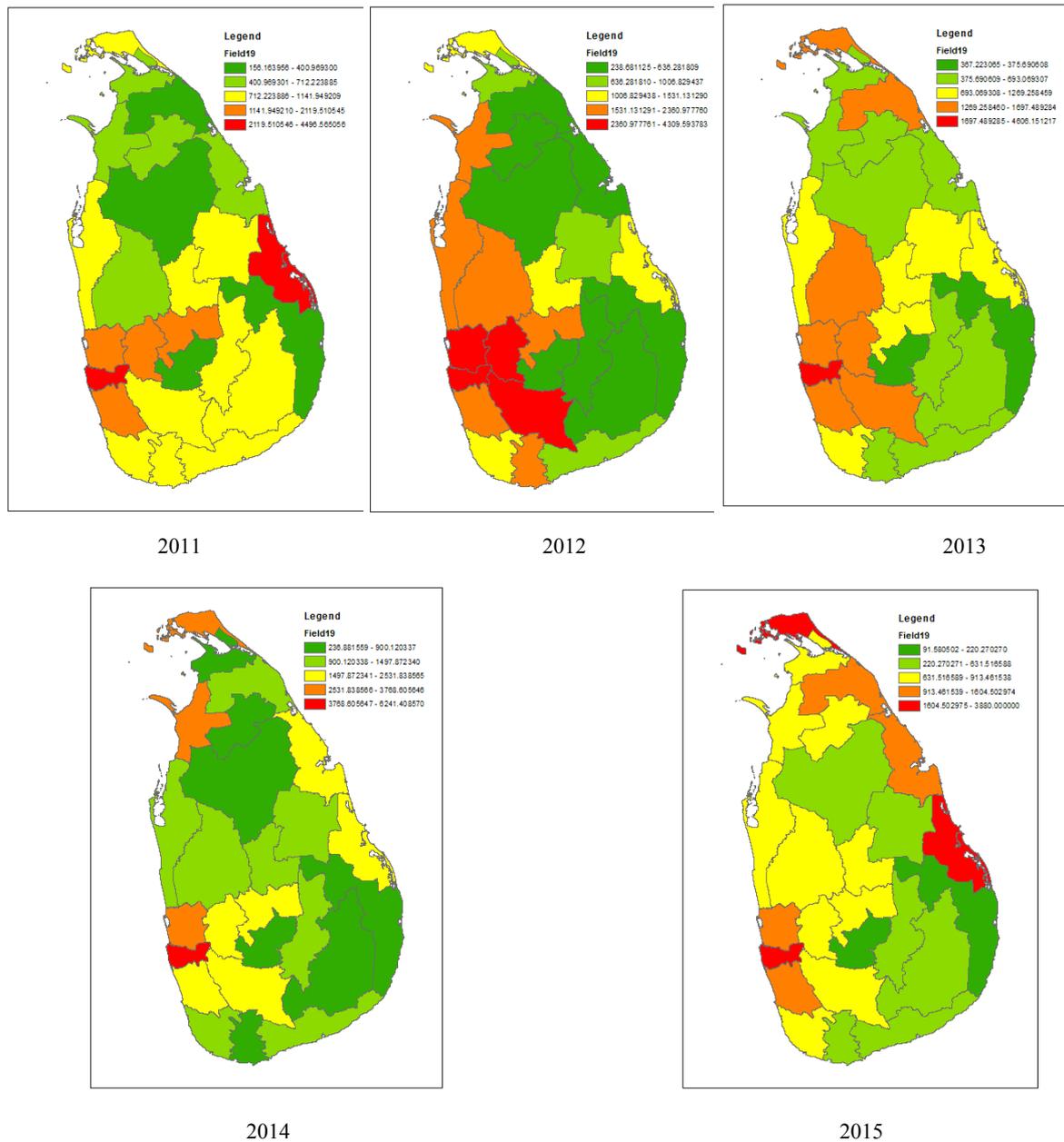
Figure 2: Spatial distribution of dengue incidence from 2011 to 2015 in Sri Lanka.

Geographically Weighted Regression Analysis

For GWR, adaptive kernel was used as kernel type and AIC was selected as bandwidth. In GWR AIC value determines the performance of the model. AIC can be used to compare two different models generated with regression analyses. AIC value for OLS is greater than in GWR. Hence GWR is a better analysis tool for dengue incidence with rainfall and population density as explanatory variables.

The GWR model results show that the Adjusted R-Squared values is 0.5632 (R2= 0.621). This indicates the model generated with population density and rainfall as explanatory variables can explain 56.3% of the variance in dengue incidences in 2014. These results also reveal that there are other variables besides population density and rainfall data that has stronger relationships with dengue incidence. These variables are not included in the model. But the model cannot provide a clue of those variables so that they cannot be included in the model. They have to be identified by experimenting with various candidate explanatory variables.
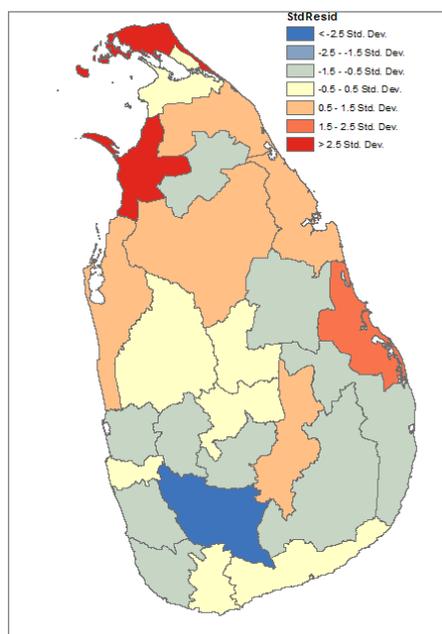
Figure3: GWR standard residual map for dengue incidence with rainfall and population density for the year 2014.

The standard residual map for the model developed for dengue incidence is shown in the Figure 3. The red areas indicate under predictions where the actual number of dengue cases is higher than the model predicted values. The blue areas indicate over predictions where actual dengue cases are lower than predicted values. Random locations of red or blue areas indicate the model performs fairly well. Red or blue clustered areas indicate under/over prediction of the model and hence the performance of the model is poor. Spatial clustering of over/under prediction indicates missing one or more key explanatory variables in the model. The standard residual map in Figure 3 shows clustered over and under predicted areas.

It is required to find how well each explanatory variable predicts the dengue incidence for each administrative region. It is revealed from previous sections that there are no global explanatory variables that hold consistent relationship across administrative regions. An analysis was conducted to reveal the variation in strength of explanatory variables in each administrative region in explaining the relationship between the variable and dengue incidence. Results of the analysis are shown in Figure 4 (a) and Figure 4 (b). Figure 4 (a) provides the spatial distribution of regression coefficients for rainfall and Figure 4 (b) provides the same for population density. Lighter colors represent lower coefficients and darker colors represent higher coefficients.

Mapping these coefficients shows the relationship between each explanatory variable and the dependent variable that how they change across the study area. The darker areas in figures indicate the explanatory variables, rainfall and population density, are strong predictors of the dengue incidence, whereas, the lighter areas are locations where they are comparatively weak.

GWR regression results show that relationship of incidence with rainfall and population density is spatially varying across districts of Sri Lanka. Figure 4(a) shows that spatial distribution of regression coefficient of population density is a strong predictor in eastern coastal areas mainly in Trincomalee district, and a weak predictor in Mannar. Figure 4(b) also shows that spatial distribution of regression coefficient of rainfall is a strong predictor in northern areas including Mannar and in eastern coast it is a weak predictor. There is an inverse effect of rainfall and population density on dengue incidence. When rain becomes a strong predictor in some areas population density is a weak predictor and vise versa. It is very important to understand this variation for making local policies to mitigate dengue. GWR model can also be used to predict values of dependent variables for locations within the study area with unseen explanatory variables values. This will give an estimate of the dependent variable (dengue incidence) using the regression model generated.
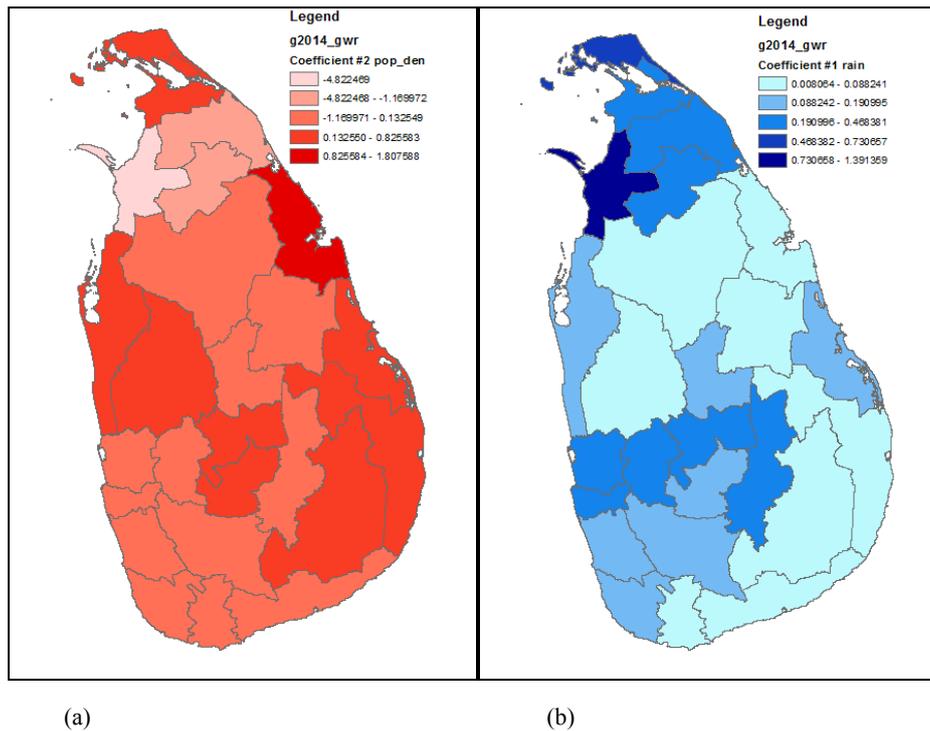
(a) (b)

Figure 4: spatial distribution of regression coefficients for (a) population density (b) rainfall.

CONCLUSIONS

We identified the relationship of dengue incidence with rainfall and population density that significantly influence dengue outbreaks. We analyzed the impact of these parameters on the distribution pattern of dengue outbreaks in Sri Lanka. Based on study results, it is concluded that study parameters are not suitable for OLS global model because of spatial heterogeneity in independent variables. GWR analysis is a form of linear regression that can model spatially varying relationships between variables. The GWR model showed that using population density and rainfall as explanatory variables, the model can explain 56.3% of the variance in dengue incidence. Densely populated areas are the most vulnerable ones that may provide suitable breeding grounds for virus to grow. There is a higher chance that these areas may experience severe dengue outbreaks if effective measures are not taken in order to control the disease or to minimize its risk. This outcome demands further study of the variables that were left out in the model as they are responsible of the dengue behavior. And also it stressed out the importance of having a spatially catered mitigation plan to deal with dengue epidemic as the impact in each area is spatially dependent. The geo-statistical dengue risk model developed in this study can be used to predict risk areas in districts of Sri Lanka that need special attention in order to effectively and efficiently manage and mitigate dengue outbreaks in future.

REFERENCES

Brunsdon, C., Fotheringham, S. and Charlton, M. (1998). Geographically Weighted Regression-Modeling Spatial Non-Stationarity, Journal of the Royal Statistical Society. *Series D (The Statistician)*. **47**(3):431-443.

Fouque, F., Carinci, R., Gaborit, P., Issaly, J. and Bicout, D.J. (2006). Aedes aegypti survival and dengue transmission patterns in French Guiana. *Journal of Vector Ecology*. **31**: 390-399.

GADM; the Database of Global Administrative Areas, 2016. Available :http://www.gadm.org/country. Accessed Date July 2016.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews*. **11**: 480-496.

Muhammad, N.A., Salleh, S., Neoh, H., Syed Z S. and Jamal, R. (2011). Dengue epidemic in Malaysia: Not a predominantly urban disease anymore. *BMC Research Notes*. **4**: 216.

Pathirana, S., Kawabata, M. and Gunatilake, R. (2009). Study of potential risk of dengue disease outbreak in Sri Lanka using GIS and statistical modeling. *Journal of Rural and Tropical Public Health.* **8**:8-17.

Picardal, J.P., Elnar, A.R.R. (2012). Rainfall, temperature and the incidence of dengue in Central Visayas, Philippines are not correlated. *CNU Journal of Higher Education*, 6.

RESTEC, 2016. Global Satellite Mapping of Precipitation (GSMaP). Available: https://www.restec.or.jp/en. Accessed Date June 2016.

Thammapalo, S., Chongsuwiwatwong, V., McNeil, D. and Geater, A. (2005). The climatic factors influencing the occurrence of dengue hemorrhagic fever in Thailand. *The Southeast Asian Journal of Tropical Medicine and Public Health.* **36**(1):191-196.

Tran, A, Deparis, X, Dussart, P, Morvan, J, Rabarison, P. (2004). Dengue spatial and temporal patterns, French Guiana, 2001. *Emerging Infectious Diseases.* **10**: 615-621.

WHO/TDR (2009). Dengue: guidelines for diagnosis, treatment, prevention and control - New edition. Geneva: World Health Organization, 160.

Available from http://www.who.int/tdr/publications /documents/dengue-diagnosis.pdf (Accessed Date 2016 June 01).

WHO (2012). Dengue and dengue hemorrhagic fever, Fact sheet 117, *World Health Organization.* Available from http://www.who.int/mediacentre/factsheets/fs117/ en/. (Accessed Date 2016 June 05).

WHO (1997). Obesity: preventing and managing the global epidemic. Report of a WHO Consultation presented at: the World Health Organization; June 3-5, 1997; Geneva, Switzerland.

WHO/WPRO (2013) Dengue in the Western Pacific Region. Emerging disease surveillance and response WHO.

Wiwanitkit, V. (2005). Strong correlation between rainfall and the prevalence of dengue in central region of Thailand in 2004. *Journal of Rural and Tropical Public Health.* **4**:41-42.

Yang, H.M, Macoris, M.L.G., Galvani, K.C., Andrighetti, M.T.M. and Wanderley, D.M.V. (2009). Assessing the effects of temperature on the population of Aedes aegypti, the vector of dengue. *Epidemiology and Infection.* **137**: 1188-1202.