RESEARCH ARTICLE

A SPATIAL ANALYSIS ON DENGUE OUTBREAKS IN TAMIL NADU, 2013-2018

*Jaisankar, R., Kesavan, J. and Ranjani, M.

Department of Statistics, Bharathiar University, Coimbatore, Tamil Nadu, India

ARTICLE INFO	ABSTRACT
	In 'epidemiology it is very important to develop surveillance systems that are being used to detect disease outbreaks. The requirements for such systems are the data related to cases and various risk factors associated with the disease over time. Meaningful information can be obtained from the collected data using statistical analysis which could be used for effective disease control. However, a geo-visualization approach by adding spatial parameters would be more informative through which one can identify the spatial patterns or space or space-time clusters of the disease. In this work, Space-time analysis using the Poisson model assumption is used for the identification of potential disease clusters of dengue in Tamil Nadu, using SaTScan and open-source Aeronautical Reconnaissance Coverage Geographic Information System (ArcGIS) software. Maps of the geographical location of disease clusters are provided to enhance the understanding of the results of statistical analysis.

INTRODUCTION

Disease surveillance is a continuous process of collecting information about disease incidences, organizing and analyzing the information collected and drawing productive interpretations so that the disease outbreaks can be determined and effectively controlled. This can be more meaningfully done by monitoring space-time trends of disease occurrences and changing patterns among the populations under risk which would be and helpful in identifying new risk factors. The essential requirement for this is the disease database daily, monthly or yearly, with the duration depending on the type of disease and over various geographical locations. In new spatial surveillance systems, one can statistically find out what are the significant areas of disease outbreaks and the geographical sizes of these outbreaks. One can also find disease clusters through which the expected number of cases of disease within study area can be determined. Through the visualization of spatial distribution of the disease over a defined area, the disease clusters may easily and efficiently analyzed and one may also detect the unusual patterns of disease outbreaks. The most common type of map for spatial distribution visualization is Choropleth maps over the geographical area under study using health indicators like disease occurrences, incidence rates or morbidity rate and mortality rate etc. Choropleth maps usually use color or pattern combinations to show different levels of disease risk associated with different geographical areas which are usually defined for administrative purposes, such as districts, taluks and cities. Dengue – hemorrhagic fever poses a challenge to the current generation. It is caused by an RNA virus of the family Flaviviridae and spread by Aedes mosquitoes. Every year a reasonable number of people get victims for this disease, particularly in tropic and sub-tropic nations.

According to World Health Organization (WHO) about half of the world's population are at risk with an estimate of 390 million infections per year. Dengue virus infection has diversified clinical presentations and such as Hemorrhagic episodes, high fever, Vasculopathy, deficiency and dysfunction of platelets, etc., and hence accurate diagnosis is difficult and relies on laboratory confirmation. The condition is usually self-limiting and antiviral therapy is not currently available. The present study uses the data on dengue infection in Tamil Nadu, provided by the department of health and family welfare, for finding the spatial heterogeneity at the district level and statistically significant disease clusters over the space and time using spatial scan statistic proposed by Kulldorff (1997). The results are visualized through ArcGIS software.

*****Corresponding author: Jaisankar, R.,**
Department of Statistics, Bharathiar University, Coimbatore, Tamil Nadu, India.

METHODOLOGY

Disease clustering can be classified as temporal clustering, spatial clustering or space-time clustering. Temporal clustering observes whether cases are located close to each other in time, spatial clustering observes whether cases are located close to each other in space and space-time clustering observes whether cases are close in space as well as in time.

Space-Time Scan Statistic: To apply the space-time scan statistics, it is necessary to collect data on the location of each case for a defined geographic area during a specified study period. Let n denotes the total number of observed cases and the study area is divided into z_S geographic regions like districts, taluks etc., and the study period are divided into z_T intervals.

Let n_{it} denotes the observed number of cases of the disease under study in region i at time t with $n = n_{11} + \dots + n_{z_S z_T}$ and e_{it} denotes the conditional null expected number of cases of the disease understudy in the region i at time t. For the kth category of the set confounders of the region i at time t, let the observed number of cases be n_{itk} and the population size is ξ_{itk}.

The centroid of the region i is denoted by (x_i, y_j) and the distance between the ith and j^{th} regions is defined as.

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}$$

Testing of Hypothesis: It is apt to assume a Poisson model for this problem. That is, if an entire study area is divided into m_S regions and the study period is divided into m_T intervals. The number of cases in the region i at time t is denoted by the Poisson random number nit with observed number nit

$$H_0 : N_{it} \sim \text{ Independent Poisson } \left(e'_{it}\right),$$
$$\left(e'_{it} \text{ is usually unknown}\right)$$

where $i = 1, 2, \dots, m_S$, $t = 1, 2, \dots, m_T$ and the time interval m_T includes the current time. On conditioning on the observed total number of cases, $n = \sum_{it} n_{it}$ the null hypothesis of on clustering can be

$$H_0 : E\left(N_{it}\right) = e_{it} \qquad \left(N_{11}, \dots, N_{z_s z_T}\right) \sim Multinomial\left(n, p\right)$$

where $p = \left(p_{11}, p_{12}, \dots, p_{z_S z_T}\right)$, $\qquad p_{it} = e_{it} / n$

where e_{it} denotes the conditional null expected number of cases in the region i at time t adjusted for the confounding factors such as sex and age and is given by

$$e_{it} = \sum_{k=1}^{K} \xi_{itk} \left(\frac{\displaystyle\sum_{j=1}^{z_S}\sum_{t=1}^{z_T} n_{jtk}}{\displaystyle\sum_{j=1}^{z_S}\sum_{t=1}^{z_T} \xi_{jtk}} \right), \qquad \left(i = 1, \dots z_S ; t = 1, \dots, z_T\right)$$

$$\sum_{i=1}^{z_S}\sum_{t=1}^{z_T}\sum_{k=1}^{K} n_{itk} = \sum_{i=1}^{z_S}\sum_{t=1}^{z_T} e_{it} = n$$

When the covariate-adjusted expected number of cases e_{it}^R using regression models such as a generalized linear mixed model is obtained then the conditional expectation is

calculated as $e_{it} = \dfrac{n e_{it}^R}{\displaystyle\sum_{i=1}^{z_S}\sum^{z_T} e_{it}^R}$, so that, $\displaystyle\sum_{i=1}^{z_S}\sum^{z_T} e_{it} = n$. In

general one may write $E\left(N_{it}\right) = \theta_{it} e_{it} \text{ and } \left(N_{11}, \dots, N_{z_s z_T}\right)$ follows multinomial. θ_{it} denotes the relative risk due to the region i and the time t, In order to test the significance of the most likely cluster of cases, a significance test using a Monte Carlo simulation is run. For each potential cluster, a large number of replications of datasets (usually 999 or 9999) using the restricted parameters is generated. The Likelihood Ratio is computed for these regions and the corresponding distribution is obtained. If the observed cluster of cases has a Likelihood Ratio higher than, say for example, 95% of the datasets generated under the hypothesis H_0, then the cluster is said to be significant at the 0.05 level of confidence.

Relative Risks

This is the estimated risk within the cluster divided by the estimated risk outside the cluster. It is calculated as the observed divided by the expected within the cluster divided by the observed divided by the expected outside the cluster. It is given by,

$$RR = \frac{c / E\left[c\right]}{\left(C - c\right) / \left(E\left[C\right] - E\left[c\right]\right)} = \frac{c / E\left[c\right]}{\left(C - c\right) / \left(C - E\left[c\right]\right)}$$

where c is the number of observed cases within the cluster and C is the total number of cases in the data set. Note that since the analysis is conditioned on the total number of cases observed, $E[C]=C$. Relative risks are calculated individually for every district of Tamil Nadu for each year. The results are shown in the figure 2(B).

DATA COLLECTION AND PROCESSING

Data collection: Data for the analysis was collected from Public Health and Preventive Medicine, Tamil Nadu, Health Department. The data covering details on dengue cases in Tamil Nadu, District-wise which are available for the years 2013 to 2018.

Data Pre-processing: For the analysis coordinate information about the geographical location are collected where the cases have occurred, which includes the details about location name, latitude and longitude of the location of the cases.
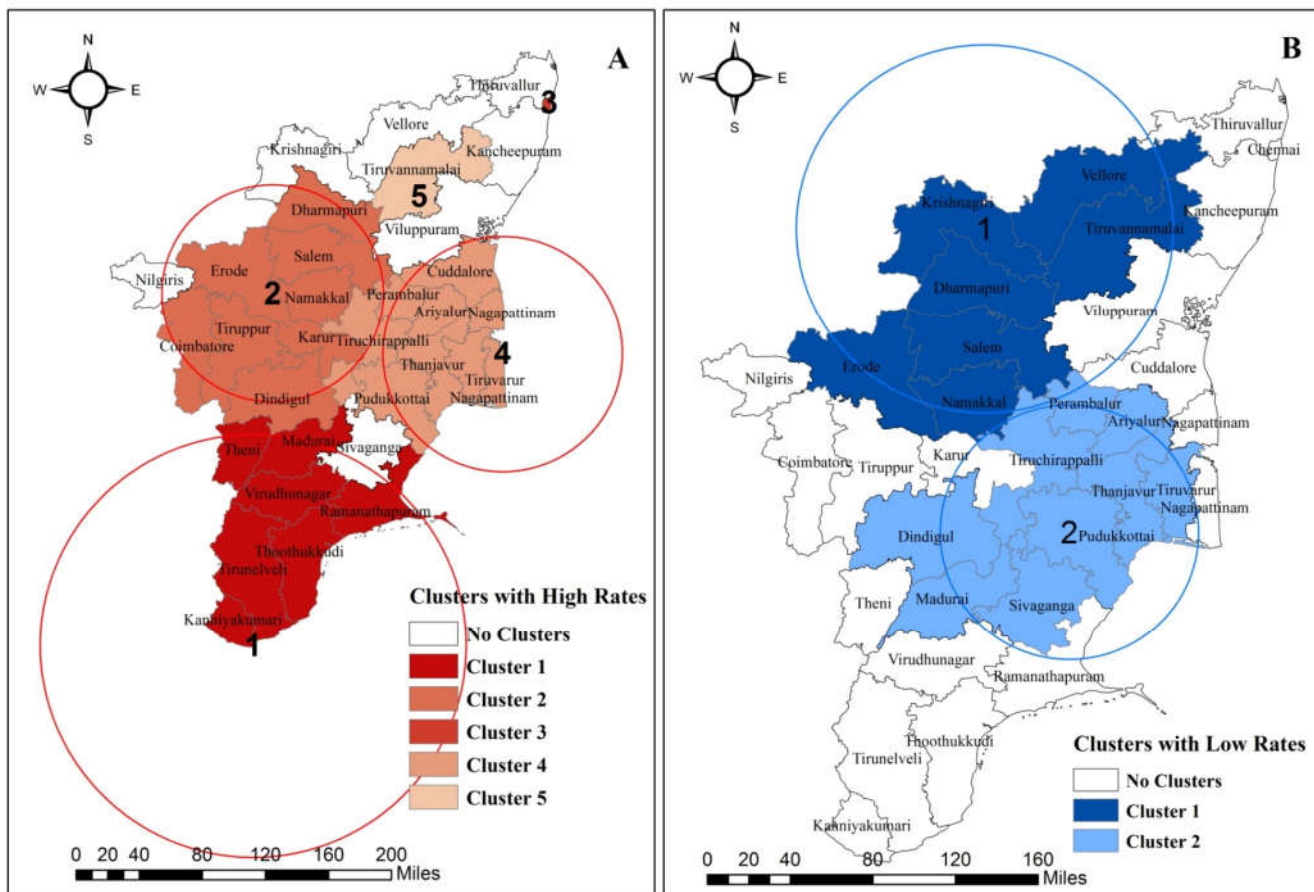
Figure 1: A) Most likely Clusters with High Rates of Dengue Incidences, B) Most likely Clusters with Low Rates of Dengue Incidences
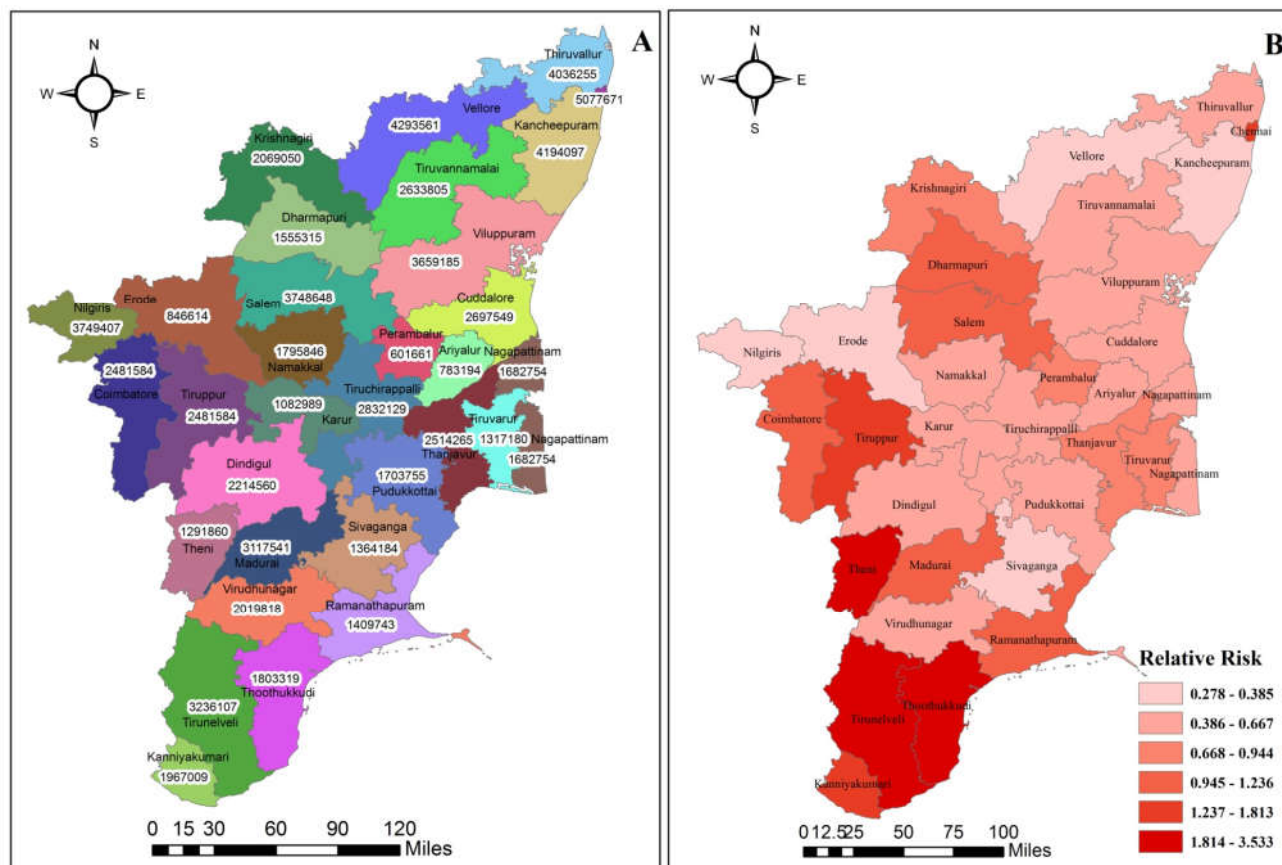

Figure 2: A) Average Population in Tamil Nadu State (2013 - 2018), B) Relative Risk of Dengue Incidences in Tamil Nadu State

Map Digitization: Geo-referencing process is used to assign real-world coordinates to each pixel of the raster using ArcGIS. In the presented work, scanned map of Tamil Nadu with district wise is digitized by obtaining coordinates from the markings on the map image itself. Using these GCPs (Ground Control Points), the image is warped and it is made to fit within the chosen coordinate system.

RESULTS AND DISCUSSION

High Rate Clusters: The total number of locations taken up is 32 districts of Tamil Nadu and the total number of cases reported is 41157 from January 2013 to October 2018.

Thanjavur, Sivaganga, Madurai, Ariyalur, Dindigul, Perambalur, Tiruvarur, Karur with an observed and expected number of cases as 303 and 1624.28 respectively. On the basis data collected from 2013 to 2018 the infected cases are lowest for the year 2014, almost in all the districts.

The detailed information is presented in the table 1 and figure 1(B). It is observed that there is a significant difference between the proportion of males and females affected by dengue. Males are affected more than that of females. Higher proportion of dengue cases are observed for the patients above 15 years of age. This behavior seen both for males and females.

Table 1. Cluster Information

	Location with Total Number of Clusters	Radius	Period Start	End	Observed	Expected	Population	O/E	RR	LLR	p
High Rate Clusters	Kanniyakumari (7)	215.416	2017	2017	8866	1367.425	14991686	6.484	7.989	9830.332	<0.05
	Erode (8)	110.960	2017	2017	5745	1737.824	18940229	3.306	3.680	3073.066	<0.05
	Chennai (1)	0.000	2017	2017	2740	464.793	5077860	5.895	6.244	2650.713	<0.05
	Nagapattinam (8)	120.170	2017	2017	2770	1290.368	14158530	2.147	2.229	664.232	<0.05
	Tiruvannamalai (1)	0.000	2017	2017	490	235.547	2583187	2.080	2.093	105.263	<0.05
Low Rate Clusters	Krishnagiri (7)	143.311	2014	2014	1428	4952.780	18447242	0.288	0.263	1915.097	<0.05
	Pudukkottai (9)	98.942	2014	2014	1441	4453.416	16582536	0.324	0.299	1506.867	<0.05

The Population averaged over time 76215400, exactly 9 individuals were affected by dengue among every 100000 people per annum. The analysis of the detection of clusters is done by SaTScan software with the assumption that the number of events in a geographical location is Poisson distributed. The largest admissible size of the cluster was fixed as 25%. The shapefile of the clusters is generated and saved by the package is easily imported in Google Earth so that interpretations can be made easily. The following are the details of detected clusters: The locations included in the most likely cluster are Kanniyakumari, Tirunelveli, Thoothukkudi, Virudhunagr, Ramanathapuram, Theni, Madurai with number of cases observed to be 8866 and expected number of cases as 1367.43, highest among detected clusters, which is found to be statistically significant and hence the occurrences are not by chance. The next significant cluster is found in Erode, Tiruppur, Namakkal, Karur, Salem, Coimbatore, Dharmapuri, Dindigul with observed and expected number of cases as 5745 and 1737.82 respectively. The third significant cluster is found in Chennai with observed and expected number of cases as 2740 and 464.79 respectively. The fourth significant cluster is found in Nagapattinam, Tiruvarur, Thanjavur, Ariyalur, Cuddalore, Tiruchirappalli, Perambalur, Pudukkottai with observed and expected number of cases as 2770 and 1290.37 respectively. The fifth significant cluster is found in Tiruvannamalai with observed and expected number of cases as 490 and 235.55 respectively. On the basis data collected from 2013 to 2018 the infected cases are highest for the year 2017, almost in all the districts. While making comparisons among the districts Tirunelveli is with highest incidence and Chennai is placed as second. The proportion of males affected by dengue is higher than that of females. The detailed information is presented in table 1 and figure 1(A).

Low Rate Clusters: The locations included in the most likely cluster are Krishnagiri, Dharmapuri, Salem, Tiruvannamalai, Vellore, Erode, Namakkal with number of cases observed to be 245 and expected number of cases as 1676.70, highest among detected clusters, which is found to be statistically significant and hence the occurrences are not by chance. The next significant cluster is found in Pudukkottai, Tiruchirappalli,

CONCLUSION

The spatial and Geographical visualization approaches are developed in this paper using the space-time cluster detection method which provides an efficient representation of the results of statistical analysis in geographical space. Space-time analysis is performed with the assumption of the Poisson model using SaTScan software. The disease clusters are detected with the cluster radius as 1 kilometer. With the proposed visualization method maps are generated using ArcGIS software. These maps show density of cases within each district, Clusters with High Rates of disease and statistically significant clusters. Proactive actions can be taken to prevent disease outbreaks, using these results. The present study is helpful in identifying the hot spot districts of an epidemic. The health department of the Government of Tamil Nadu may be advised to take extra measures for the prevention and control of the dengue in high-risk areas. For the present analysis, the dengue fever outbreaks in Tamil Nadu for past five years were taken as a sample case. On performing similar kinds of analysis for every district, taluk, etc, one may get productive information based on which disease surveillance can be done national wide for early detection of disease outbreaks and effective actions can be taken for the control of outbreaks.

REFERENCES

Gail, M., Krickeberg, K., Samet, J., Tsiatis, A., Wong, W. and Tango, T. 2009. *Statistics for biology and health*, Springer.

Kuldorff, M. 1997. A spatial scan statistic. *Communication Statistics Theory and Methods 26(6):1481-1495.*

Kulldorff M., 2011, SaTScan user guide, *pp. 4-107.*

Kulldorff, M. 2001. Prospective time periodic geographical disease surveillance using a scan statistic. *Royal Statistical Society 164(1): 61-72.*

Kulldorff, M., Athas F. William, Feuer J. Eric, Miller A. Barry and Key R. Charles. 1998. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico, *American Journal of Public Health 88(9):1377-1380.*

Kulldorff, M., Feuer J. Eric, Miller A. Barry and Freedman S. Laurence. 1997. Breast cancer clusters in the northeast united states: a geographic analysis. *American Journal of Epidemiology 146(2): 161-170.*

Kulldorff, M., Heffernan Richard, Hartman Jessica, Assunc Renato and Mostashari Farzad. 2005. A space time permutation scan statistic for disease outbreak detection. *plos journals 2(3):216-224.*

Kulldorff, M., Huang, L., Pickle, L. and Duczmal, L. 2006. An elliptic spatial scan statistics. *Wiley InterScience 25:3929-3943.*

Shuchi Mala and Raja Sengupta. 2013. Geo-visual approach for spatial scan statistics: an analysis of dengue fever outbreaks in Delhi. *International Journal of Advanced Computer Science and Applications 4(10):127-137.*

****