

# International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452  
Maths 2023; 8(3): 16-21  
© 2023 Stats & Maths  
<https://www.mathsjournal.com>  
Received: 16-02-2023  
Accepted: 26-03-2023

**Jaisankar R**  
Professor, Department of  
Statistics, Bharathiar  
University, Coimbatore,  
Tamil Nadu, India

**Ranjani M**  
Assistant Professor, Department  
of Statistics & Analytics, Mount  
Carmel College Autonomous,  
Bengaluru, Tamil Nadu, India

**Corresponding Author:**  
**Jaisankar R**  
Professor, Department of  
Statistics, Bharathiar  
University, Coimbatore,  
Tamil Nadu, India

## Bayesian cluster detection of dengue incidences in Tamil Nadu

**Jaisankar R and Ranjani M**

DOI: <https://doi.org/10.22271/math.2023.v8.i3a.973>

### Abstract

In epidemiology, the local properties of the relative risk surface are important rather than modelling the relative risk by itself. These surface properties may consist of boundaries between areas of risk, peaks of risk, local heterogeneities in risk etc. These features do not directly relate to a value at a specific location. The main concern of the Relative risk estimation is the 'global' smoothing of risk and estimating the true underlying risk level. But the cluster detection is mainly focused on the local features of the risk surface where some elevations or depressions of risk are found. In this article, analysis for disease clustering has been carried out and the high and low levels of disease clusters in dengue incidences are identified in the Coimbatore district, and also for overall Tamil Nadu.

**Keywords:** Relative risk, disease clustering, bayesian disease clustering, dengue incidences

### 1. Introduction

A spatial cluster may generally be defined as any spatially-bounded area of significantly elevated risk. This is very general and a lot of points are to be included further. The term 'spatially bounded' refers that the cluster must have some sort of spatial integrity. This may be a neighborhood criterion like that the 'areas must be adjoining' or in other words, at least two adjoining areas must meet a criterion, or the regions could be defined to have some type of 'external boundary'. A simple criterion which is called hot-spot clustering is often assumed. In such clustering, any area or region can be observed as a cluster. This is due to the assumption that 'no insistence on adjacency of regions within clusters' is called the zero-neighbourhood criterion. This is a simple criterion often assumed and commonly used in epidemiology (Rischarson *et al.*, 2004) [1]. This criterion is appealing when there is no prior knowledge of the behaviour of the disease is available and useful for such as the preliminary screening of data.

This definition of hot spot clusters ignores any contiguity that may be inherent in relevant clusters. For instance, the investigation of clusters of a given threshold size may be important. The size of the threshold is the minimum number of contiguous areas. Therefore, only groups of contiguous or adjacent regions of 'unusual' risk can be considered as clusters. In the case of infectious diseases, for understanding disease spread, a certain shape and size of the cluster may be are important. Apart from the above several definitions of clusters and clustering are available which leads to differences in the ability of detection methods. Sometimes the correlated heterogeneity term in relative risk models is considered a clustering term. Consideration of this term captures aggregation in the risk and which lead to an effect where neighbouring areas with similar risk levels which is a global feature of the risk, hence called the global clustering. This induces a smoothing of risk also.

The assumption of the global clustering is that the risk surface is clustered or has areas of elevated (reduced) risk. The localized behaviour or the location of clusters per se are not addressed by the global clustering and it is also termed general clustering (Besag and Newell 1991) [4]. But, an uncorrelated surface displays random changes in risk with location changes and so much more variable in both risk level and have few contiguous areas of similar risk. In essence, there are three major scenarios for clustering. The first one involves single region hot spot relative risk detection. The next one is that clusters may also be considered as objects or groupings with a specific criterion which includes a neighbourhood or proximity condition.

The third one is that clusters are defined as residuals. The Bayesian methodology may easily be adopted when clusters are defined as residuals.

## 2. The Dengue fever

Dengue Fever is an acute febrile disease caused by the dengue virus which is transmitted by the *Aedes Aegypti* mosquito. It is a serious public health problem in the world causing death within a short time. Every year a considerable proportion of people get victims of this disease around the world. The proportion of people infected with dengue is identified more particularly in tropic and sub-tropic regions. According to the report of the World Health Organization (WHO) in 2019, about half of the world's population are at risk with an estimate of 400 million infections per year. Tamil Nadu is one among five Dengue worst-hit states in South India, placed as the second among all the states of India. Hence, the controlling measures for dengue should be seriously implemented by the health authorities.

Dengue fever is an epidemiological disease with spatial characteristics that may be used to determine its distribution. Monitoring the spatial spread of disease, particularly for a disease like Dengue with a very high diffusion rate, is required to identify areas that have great potential to become endemic. Hence, mapping the spatial distribution of dengue fever cases may act as a guide for the reduction of further incidence.

## 3. Methodology

### 3.1 Clusters Defined as Residuals

A convenient method for clustering is based on the residual feature of the data. Assuming that  $Y_i$  is the count of disease within the  $i^{\text{th}}$  study region, the basic model for the average count  $\mu_i$

(i. e,  $E(Y_i) = \mu_i$ ) is, given by,  $\log \mu_i = \alpha_i + u_i$

Here,  $\alpha$  may be as a function of covariates which act as a linear or non-linear predictor and may also consist of different kinds of random effects. For simplicity, consider  $\alpha_i$  as the "smooth" part and  $i$  as the rough residual part. If  $\alpha_i$  consists of all pertinent non-clustering confounder effects then residual clustering information will be in the residual component. Hence, the estimated value of  $u_i$  will contain the information on any clusters unaccounted for in  $\alpha_i$ , which does not account for any pure noise in  $u_i$ . Hence, there would be at least two components an estimate of  $u_i$ , namely, clustered and unclustered (or frailty) components. Some additional components may be included which depends on whether the confounding in  $\alpha_i$  was sufficiently specified or estimated.

There are several approaches that exist in isolating residual clustering. First, a pure noise term within  $\alpha_i$  and  $u_i$  as a cluster term may be included. For instance, the assumption that  $\alpha_i = (v_i; \text{covariates})$ , a function of uncorrelated noise, in which  $v_i$  is the frailty or random effect term, and a function of covariates can be taken up. Second, a smoothed version of  $u_i$ , say,  $(u_i)$  where the pure noise is smoothed out is examined. The choice of the component that should be included in clustering depends on the following: If the clustering is likely to be irregular and that no clustering confounding effects are to be found, then a residual or smoothed residual is chosen. But, if there exists any prior information of the form of clustering, then including some of that information within the model itself is more important. The real issue behind here is

the ability of models and estimation procedures to differentiate spatial scales of clustering.

### 3.2. Cluster Detection using Residuals

In the case of count data, the suitable likelihood possibilities are either Poisson or a Binomial. For Poisson likelihood, it is assumed that  $Y_i, i = 1, 2, n$  are counts of disease cases in the  $i^{\text{th}}$  region and  $E_i, i = 1, 2, \dots, n$  are the corresponding expected counts of the disease in  $n$  small areas. Here,  $i \sim \text{Pois}(E_i\theta_i)$  given relative risk  $\theta_i$ . The log relative risk is usually modelled and hence the main focus of the model is  $\log \theta_i$ . The standardized form of Bayesian residuals for this likelihood is  $r_i = (Y_i - E_i\hat{\theta}_i) / \sqrt{E_i\hat{\theta}_i}$ , where  $\hat{\theta}_i$  is the converged posterior sample average value of the  $\theta_i$ .

Here, the Poisson likelihood was taken to examine the Dengue disease data and the model assumed is  $\log \theta_i = \alpha_0 + v_i$ , where the priors are  $\alpha_0 \sim (a, b)$  and  $v_i \sim N(0, \tau_v)$ . The term  $\tau_v$  set large and a large negative to positive range  $(a, b)$ . As the methodology is residual clustering, no correlated random effect is included in the analysis.

It is also possible to examine the predictive residuals for any given model. The predictive residual for each observation is  $r_i^{\text{pred}} = Y_i - \hat{Y}_i^{\text{pred}}$  where  $\hat{Y}_i^{\text{pred}} = \frac{1}{G} \sum_{g=1}^G f(Y_i | \theta^g)$ .  $f(Y_i | \theta^g)$  is the likelihood given the value of  $\theta^g$ , which would usually be small when compared to the standard Bayesian residual. An alternative approach to residual analysis could be based on the construction of a residual envelope, based on the comparison of the Bayesian residual:  $r_i = Y_i - \hat{Y}_i$  with  $r_i^* = \hat{Y}_i^{\text{pred}} - \hat{Y}_i$ . Unusual residuals could be assessed by assessing the ranking of  $r_i$  among the series of  $B$  simulated  $\{r_{ib}^*\} = b = 1, \dots, B$ . Further,  $p$ -values surface can be computed from a tally of exceedances:

$$P_{v_i} = \Pr(|r_i| > r_i^*) = \frac{1}{B} \sum_{b=1}^B I(|r_i| > r_{ib}^*)$$

The areas of unusually elevated values can be found using the mapped surface of  $P_{v_i}$  for hot spot detection.

In this Chapter residual clustering analysis has been performed for dengue disease counts data observed from the state of Tamil Nadu using the Poisson likelihood model. As the district of Coimbatore consists of many zero counts of dengue disease cases, the zero-inflated Poisson likelihood is considered for residual clustering in the Coimbatore district.

## 4. Results of Residual clustering in Tamil Nadu

### 4.1 High Rate Clusters Tamil Nadu

The total number of locations taken up in 32 districts in the state of Tamil Nadu. The analysis of the cluster detection is done by Win BUGS software with the assumption of events in a geographical location is Poisson distributed. The shapefile of the clusters is generated and saved by the package is imported by ArcGIS software so that interpretations can be made. The following are the details of cluster points.

The locations included in the high rate clusters are, Coimbatore, Tirunelveli, Madurai, Thoothukudi, Coimbatore, with the number of observed cases is 8866 and the expected number of cases 1367.43, highest among clusters which is found to be statistically significant and hence the occurrences are not by chance. The next significant cluster is found in Tiruppur, Theni, Salem, Kanniyakumari, Thanjavur, Tiruvallur, Tiruchirappalli, with the observed and expected

number of cases is 5745 and 1797.82. The third significant cluster is found in Ramanathapuram, Villupuram, Dharmapuri, The fourth significant cluster is Cuddalore, Tiruvannamalai, Krishnagiri, Dindigul, Kancheepuram. The fifth cluster is found in Virudhunagar, Thiruvavarur, Erode, Pudukkottai. On the basis of data collected from 2007 to 2018, the infected cases are highest for the year 2017, almost in all the districts. While making comparisons among the districts, Coimbatore and Tirunelveli with highest incidences of dengue disease and Chennai is placed second.

**4.2 Low Rate Clusters in Tamil Nadu**

The locations included in the most likely cluster are Krishnagiri, Dharmapuri, Salem, Tiruvannamalai, Vellore,

Erode, Namakkal, with the number of cases observed to be 245 and the expected number of cases as 1676.70, highest among the detected clusters with low rates, which is found to be significant and hence the occurrences are not by chance. The next significant cluster is found in Nagapattinam, Karur, Sivaganga, Perambalur, Ariyalur, The Nilgiris with the observed and expected number of cases as 2007 to 2018, the infected cases are lowest for the year in 2014, almost in all the districts. The detailed information is presented in Table. It is observed that there is a significant difference between the proportion of males and females affected by dengue. Males are affected more than females. A higher proportion of dengue cases are observed for patients above 15 years of age. This behavior is also seen both in males and females.

**Table 1 a:** First High Rate Cluster Areas

| Location     | Latitude | Longitude | Mean | Std.Dev | Median  | 2.50% | 50%  | 97.5% |
|--------------|----------|-----------|------|---------|---------|-------|------|-------|
| Theni        | 10.00832 | 77.47338  | 2772 | 1.128   | 0.0528  | 2771  | 2772 | 2773  |
| Tiruppur     | 11.10795 | 77.34068  | 2781 | 0.5827  | 0.01811 | 2780  | 2781 | 2782  |
| Coimbatore   | 11.01601 | 76.970306 | 3412 | 0.5166  | 0.01215 | 3411  | 3412 | 3413  |
| Thoothukkudi | 8.79299  | 78.14242  | 3844 | 1.005   | 0.044   | 3843  | 3844 | 3845  |
| Madurai      | 9.94472  | 78.13078  | 4565 | 0.811   | 0.02784 | 4564  | 4565 | 4566  |
| Tirunelveli  | 8.72879  | 77.70458  | 6389 | 1.518   | 0.07917 | 6388  | 6389 | 6390  |
| Chennai      | 13.08268 | 80.270721 | 7729 | 0.9249  | 0.03307 | 7728  | 7729 | 7730  |

**Table 2b:** Second High Rate Cluster Areas

| Location        | Latitude  | Longitude | Mean | Std.Dev | Median  | 2.50% | 50%  | 97.5% |
|-----------------|-----------|-----------|------|---------|---------|-------|------|-------|
| Dharmapuri      | 12.106527 | 78.136139 | 1200 | 0.4971  | 0.01094 | 1199  | 1200 | 1201  |
| Viluppuram      | 11.94551  | 79.4903   | 1210 | 0.7769  | 0.02878 | 1209  | 1210 | 1211  |
| Ramanathapuram  | 9.35886   | 78.83747  | 1292 | 0.51    | 0.0128  | 1291  | 1292 | 1293  |
| Vellore         | 12.916517 | 79.1325   | 1600 | 1.103   | 0.06091 | 1599  | 1600 | 1601  |
| Tiruchirappalli | 10.75961  | 78.78736  | 1871 | 0.5172  | 0.01299 | 1870  | 1871 | 1872  |
| Thiruvallur     | 11.59089  | 75.672707 | 1896 | 0.6496  | 0.02149 | 1895  | 1896 | 1897  |
| Thanjavur       | 10.7836   | 79.1336   | 2126 | 0.5146  | 0.01057 | 2125  | 2126 | 2127  |
| Kanniyakumari   | 8.0933    | 77.54909  | 2267 | 0.5985  | 0.0211  | 2266  | 2267 | 2268  |
| Salem           | 11.65212  | 78.157982 | 2625 | 0.5169  | 0.01324 | 2624  | 2625 | 2626  |

**Table 2a:** Low Rate Cluster Areas

| Location     | Latitude  | Longitude | Mean  | Std.Dev | Median  | 2.50% | 50%   | 97.5% |
|--------------|-----------|-----------|-------|---------|---------|-------|-------|-------|
| Karur        | 10.95958  | 78.08195  | 522.6 | 0.5326  | 0.01363 | 521.6 | 522.6 | 523.6 |
| Namakkal     | 11.219439 | 78.167725 | 754   | 0.5972  | 0.01378 | 753   | 754   | 755   |
| Nagapattinam | 10.77318  | 79.84145  | 791.3 | 0.5954  | 0.01704 | 790.4 | 791.4 | 792.3 |
| Pudukkottai  | 10.3829   | 78.8155   | 854.3 | 0.5995  | 0.02053 | 853.2 | 854.2 | 855.2 |
| Erode        | 11.329926 | 77.727757 | 870   | 0.7557  | 0.02947 | 868.9 | 870   | 870.9 |
| Thiruvavarur | 10.76508  | 79.63356  | 920.2 | 0.5052  | 0.01126 | 919.2 | 920.2 | 921.2 |
| Virudunagar  | 9.59375   | 77.9575   | 995.9 | 0.6058  | 0.01742 | 995   | 995.9 | 996.9 |

**Table 2b:** Second Low Rate Cluster Areas

| Location     | Latitude | Longitude | Mean  | Std.Dev | Median  | 2.50% | 50%   | 97.5% |
|--------------|----------|-----------|-------|---------|---------|-------|-------|-------|
| The Nilgiris | 11.41131 | 76.6927   | 147.2 | 0.7497  | 0.03197 | 146.3 | 147.2 | 148.2 |
| Ariyalur     | 11.1521  | 79.0694   | 222.5 | 0.6276  | 0.02165 | 221.5 | 222.5 | 223.4 |
| Perambalur   | 11.2341  | 78.8805   | 318.6 | 0.5112  | 0.01193 | 317.6 | 318.7 | 319.6 |
| Sivaganga    | 9.85498  | 78.5005   | 439.5 | 0.6538  | 0.02659 | 438.5 | 439.5 | 440.5 |

**5. Results of Residual clustering in Coimbatore District**

**5.1 High Rate Clusters in Coimbatore**

In Coimbatore, there is a considerable number of Dengue cases has been reported in Health and Welfare Department portals. It has 503 geographical locations based on the shapefile was taken up in the district of Coimbatore in Tamil Nadu. The analysis of the cluster detection is done by WinBUGS software with the assumption of events in a geographical location is Poisson distributed. The shapefile of the clusters is generated and saved by the package is imported

by ArcGIS software so that interpretations can be made. The following are the details of cluster points.

The locations included in the high rate clusters are Irugur village. The next significant clusters are Achipatti, Muthugoundenpudur, Kuniyamuthur, Perur villages and the third clusters are Saravanampatti, Sular, Ashokapuram, Vellanaipatti villages. Highest among clusters which is found to be statistically significant and hence the occurrences are not by chance.

**5.2 Low Rate Clusters in Coimbatore District**

The locations included in the low rate clusters are Palangarai A/b, Thekkampatti, Chinnathadagam, Idigarai villages, the next significant low rate clusters are Jadayampalayam, Vedasithur A/b, Poravipalayam, Alagumalai villages, and the

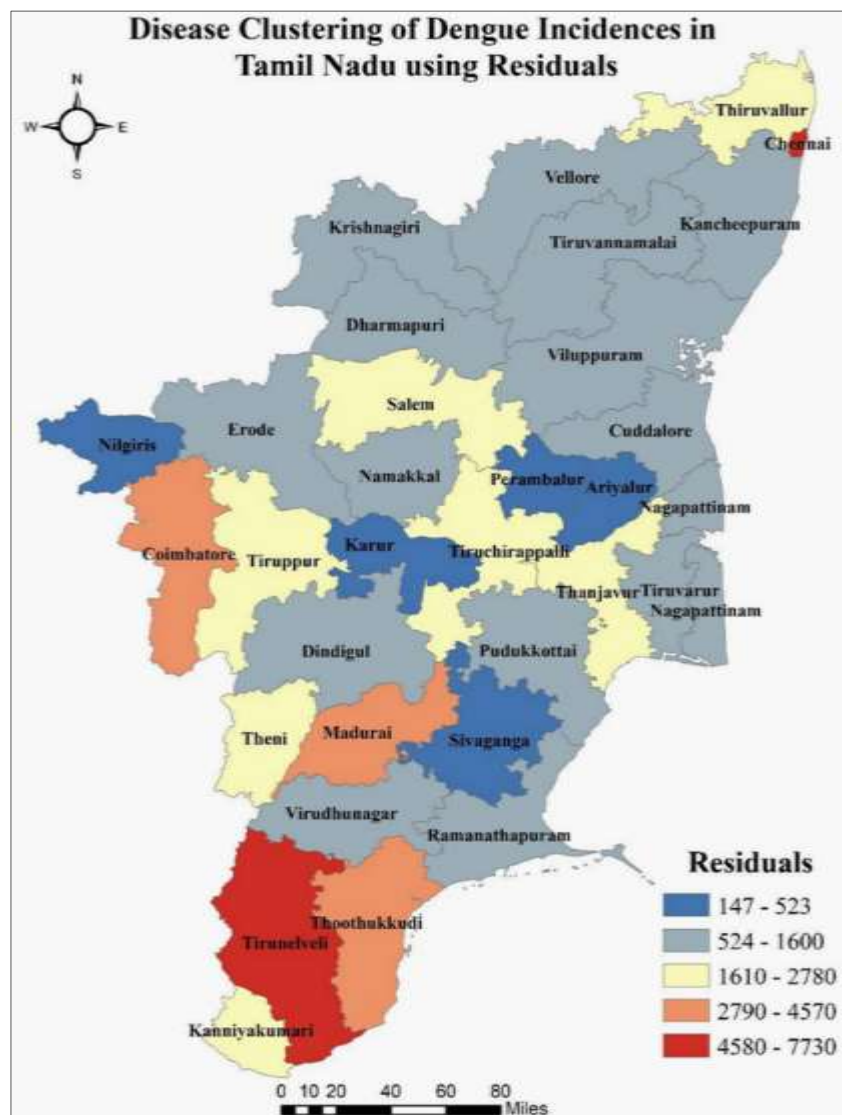
third clusters are Malaipalayam, Chinnakallipatty, Pachapalayam, Panappatti, Vadavalli, Jilobanaickenpalayam, Solapalayam villages. Lowest among clusters which is found to be statistically significant and hence the occurrences are not by chance.

**Table 3: High Rate Cluster Areas in Coimbatore**

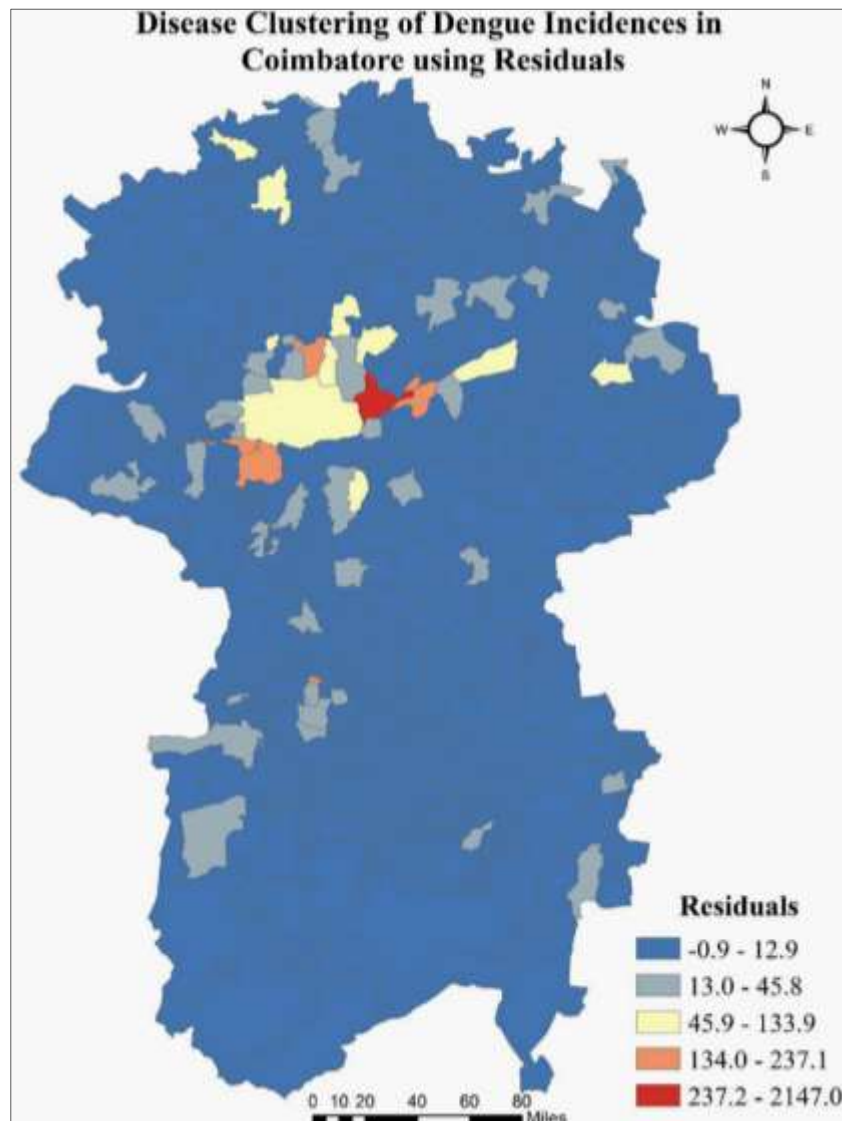
| S. No | Location                     | Mean  | Std. dev | Median  | 2.50% | Median | 97.50% |
|-------|------------------------------|-------|----------|---------|-------|--------|--------|
| 1     | Irugur A/B                   | 2147  | 0.7802   | 0.04208 | 2146  | 2147   | 2148   |
| 2     | Achipatty                    | 237.1 | 0.5475   | 0.0202  | 236.1 | 237.2  | 238    |
| 3     | Muthugoundenpudur Rly. Clny. | 237.1 | 0.5085   | 0.01943 | 236.1 | 237.1  | 238.1  |
| 4     | Kuniamuttur                  | 236.9 | 10.2     | 0.2066  | 236.1 | 237.1  | 238.1  |
| 5     | Perur                        | 218.7 | 0.5284   | 0.01954 | 217.7 | 218.8  | 219.7  |
| 6     | Saravanampatti               | 201.3 | 1.342    | 0.04713 | 200.3 | 201.3  | 202.4  |
| 7     | Sulur                        | 155.8 | 5.351    | 0.1982  | 155   | 156    | 156.9  |
| 8     | Ashokapuram                  | 133.9 | 0.5339   | 0.01695 | 133   | 133.9  | 134.8  |
| 9     | Vellanaipatti                | 110   | 0.5051   | 0.01801 | 109.1 | 110    | 111    |

**Table 4: Low Rate Cluster Areas in Coimbatore**

| S. No | Location             | Mean   | Std.dev | Median   | 2.50%   | Median | 97.50% |
|-------|----------------------|--------|---------|----------|---------|--------|--------|
| 1     | Solapalayam          | 0.6994 | 0.4786  | 0.01115  | -0.366  | 0.7275 | 1.5    |
| 2     | Jilobanaickenpalayam | 0.6922 | 0.4613  | 0.008892 | -0.3045 | 0.721  | 1.518  |
| 3     | Vadavalli            | 0.6825 | 0.4989  | 0.01136  | -0.4291 | 0.7175 | 1.513  |
| 4     | Panappatti           | 0.6796 | 0.4857  | 0.0108   | -0.3621 | 0.7259 | 1.498  |
| 5     | Pachapalayam         | 0.6759 | 0.475   | 0.01091  | -0.3301 | 0.7329 | 1.453  |
| 6     | Chinnakallipatty     | 0.6744 | 0.4808  | 0.01127  | -0.3895 | 0.7126 | 1.496  |
| 7     | Malaipalayam         | 0.6547 | 0.4757  | 0.01079  | -0.3644 | 0.6766 | 1.482  |



**Fig 1: Disease Clustering of Dengue Incidences in Tamil Nadu using Residuals**



**Fig 2:** Disease Clustering of Dengue Incidences in Coimbatore using Residuals

## 6. Conclusion

The spatial and geographical visualizations are developed in this Chapter using the Bayesian Spatial Disease clustering method, which provides an efficient representation of the results of statistical analysis in geographical space. The disease clusters are detected using WinBUGS software and maps which are generated using ArcGIS software. These maps show the density of cases within each region. Proactive actions can be taken to prevent disease outbreaks, using these results. The health department of the Government of Tamil Nadu may be advised to take extra measures for the prevention and control of dengue in high-risk areas.

For the present analysis, the dengue fever outbreaks in Tamil Nadu for the last twelve years were taken as a sample case. On performing similar kinds of analysis for every district, taluk, etc. According to that, the dengue incidences of the Coimbatore, districts of Tamil Nadu were analyzed over the period of 2012 to 2020. One may get productive information based on which disease surveillance can be taken for the control of outbreaks.

## 7. References

1. Anderson C, Lee D, Dean N. Identifying clusters in Bayesian Disease Mapping. *Biostatistics*. 2014;15(3):457-469.
2. Lawson AB, Rolejanaprasert C. Childhood Brain Cancer in Florida A Bayesian Clustering Approach. *Statistics and Public Policy*. 2014;1(1):99-107
3. Bernadinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian Analysis of Space-Time Variation in Disease Risk. *Statistics in Medicine*. 1995;14(21-22):2433-2443.
4. Besag J, York JC, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 1991;43(1):1-20.
5. Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society, Series-B*. 1974;36(2):192-236.
6. Clayton D, Kaldor J. Empirical Bayes Estimates of Age-Standardised Relative Risks for use in Disease Mapping. *Biometrics*. 1987;43(3):671-691.
7. Cressie N. *Statistics for spatial data*. John Wiley & Sons; New York; c1993.
8. Kulldorff M, Nagarwalla N. Spatial Disease Clusters: Detection and Inference. *Statistics in Medicine*. 1995;14(8):799-810.
9. Nivedha R, Duraisamy MR, Patil Santhosh Ganapathi, Monomani S. Genetic Diversity analysis of Sesame-A Bayesian Clustering Approach. *Electronic Journal of Plant Breeding*. 2019;10(2):743-753.

10. Gangon RE, Clayton M. Bayesian Detection and Modelling of Spatial Disease Clustering. *Biometrics*. 2000;56(3):922-935.
11. Richardson S, Thomson A, Best N, Elliott P. Interpreting Posterior Risk Estimators in Disease Mapping Studies. *Environmental Health Perspectives*. 2004;112(9):1016-1025.
12. Tiefelsdorf M. Modelling Spatial Processes – The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran’s I. *Lecture Notes in Earth Sciences*, Springer-Verlag, Berlin; c2000. p. 87.
13. Wakefield J, Albert KIM. A Bayesian Model for Cluster Detection. *Biometrics*; 2013;14(4):752-765.
14. WHO, Dengue in World Health Organization in India; c2019.  
[<http://www.searo.who.int/india/topics/dengue/en/>]