Contents lists available at GrowingScience

Journal of Future Sustainability

homepage: www.GrowingScience.com/jfs

# Spatial disease mapping using the Poisson-Gamma model

**R. Jainsankar[a] and M. Ranjani[b*]**

[a]*Professor, Department of Statistics, Bharathair University, Coimbatore, Tamil Nadu, India*
[b]*Assistant Professor, Department of Statistics & Analytics, Mount Carmel College, Autonomous, Bengaluru, Karnataka, India*

**C H R O N I C L E**

*Article history*:
Received: March 4, 2023
Received in revised format: March 28, 2023
Accepted: May 21, 2023
Available online: May 21, 2023

*Keywords:*
*Spatial Disease Mapping*
*Poisson-Gamma Model*
*Relative Risk Estimation*
*Dengue Incidences*

**A B S T R A C T**

In disease mapping, it is preferable to estimate the risk rather than the significance in general, but the variation in estimation precision across the geographical map of the study region must also be taken into consideration. In such a situation the conventional methods would not yield the best estimates. Heterogeneity is an important aspect to be considered as significant in Disease Mapping and relative risk estimation. The simple regression models often do not capture the extent of the variation exhibited in the spatial count data. This is the case when the spatial data is over-dispersed or there is spatial correlation due to unobserved confounders. In such situations, it is appropriate to include some additional terms, which may be in the form of the prior distribution. In this paper, a Poisson model with Gamma prior is used to model and map the dengue incidences in Tamil Nadu to explain the patterns of variations.



## 1. Introduction

Disease mapping involves the cartographical representation of the spatial distribution of disease risks (Clayton & Kaldor, 1987). In spatial epidemiology modelling for disease mapping and relative risk estimation consists of functions of fixed, unobserved covariates (Knorr-Held & Becker, 1999). In disease mapping, it is preferable to estimate the risk rather than the significance in general, but the variation in estimation precision across the geographical map of the study region must also be taken into consideration (Lawson et al., 2000). In such a situation the conventional methods would not yield the best estimates (Bernardinelli et al., 1995). Also, there may be high heterogeneity or substantial extra variability involved in the case of the geo-referenced data. To overcome this, Bayesian modelling would be highly helpful which consists of treating the parameters as random variables, and hence relying on shrinkage of estimators to obtain stabilized risk estimates (Eckert et al., 2007). Bayesian hierarchical models are found to be highly useful to obtain smooth risk estimates in disease mapping (Millar, 2009; Richardson et al., 2004). The Poisson-Gamma is the one which is potentially useful in many applications. This model considers the spread of the disease in human populations and enables covariate adjustments and spatial correlation between adjacent areas. The present work aims at the application of the Poisson-Gamma model for the data on dengue incidences in Tamil Nadu, India.In disease map re-construction and relative risk estimation for case event and count likelihoods, the simplest model is the Poisson Process model which is characterized by a point process in space (Cressie, 1993). Defining a study area say, $T$ and within that area $m$ disease incidences occur and define $\{s_i\}, i = 1, \dots, m$ this is called the realization of the disease process. Here it is assumed that all cases within the study area are recorded. The basic point process model assumed for such data within disease mapping is the heterogeneous Poisson process with first-order intensity with the basic assumptions that points (case events) are independently spatially-distributed and governed by the first-order intensity.

## 2. The Poisson Model for Count Data

The Poisson model is appropriate when there is a relatively lesser count of disease, and the population is large in each small area (Jaisankar & Kesavan, 2019; Jaisankar et al., 2019). The disease count $y_i$ is assumed to have a mean $\mu_i$ and is independently distributed as

* Corresponding author.
E-mail address: ranjanim@mccblr.edu.in (M. Ranjani)

ISSN 2816-8151 (Online) - ISSN 2816-8143 (Print)

doi: 10.5267/j.jfs.2024.5.004

$$y_i \sim Poisson(\mu_i)$$

The likelihood is given by

$$L(\boldsymbol{y}|\boldsymbol{\mu}) = \prod_{i=1}^{m} {\mu_i}^{y_i} \exp(-\mu_i) / y_i!.$$

The mean function is usually considered to consist of two components, (*i*) a component representing the background population effect, and (*ii*) a component representing the excess risk within an area. This second component is often termed the *relative risk*. The first component is commonly estimated or computed by comparison to rates of the disease in a standard population and a local expected rate is obtained. This is often termed *standardization* (Inskip et al., 1983). The usual assumption here is that the data is independently distributed with expectations,

$$E(y_i) = \mu_i = E_i \theta_i$$

where $E_i$ is the expected rate for the $i^{\text{th}}$ area and $\theta_i$ is the relative risk for the $i^{\text{th}}$ area. For the Bayesian hierarchical model, consider $\{y_i\}$ to be conditionally independent given knowledge of $\{\theta_i\}$. The expected rate is usually assumed to be fixed for the time period considered in the spatial realm (Ripley, 1988; Snow, 1854).

**3. The Poisson-Gamma Model**

The simple regression models do not capture the extent of variation present in count data. That is, overdispersion or spatial correlation due to unobserved confounders will usually not be captured by simple covariate models and often it is appropriate to include some additional terms or terms in a model which can capture such effects. In Bayesian model formulation, all parameters are stochastic and so the extension to the addition of random effects is a relatively straightforward procedure.

The overdispersion or extra-variation can be accommodated by either (*i*) inclusion of a prior distribution for the relative risk, such as the Poisson-Gamma model or (*ii*) by extension of linear or non-linear predictor term to include an extra random effect such as Log-Normal model (Besag & Newell, 1991; Besag et al., 1001).

The simplest extension to the Poisson likelihood model that accommodates extra variations is one in which the parameters of interest in the likelihood are given a prior distribution. The Poisson model is the most used model for counting data. This model is appropriate when there is a low count of disease in the population that is relatively large. The Bayesian hierarchical model that is familiar in disease mapping is the Poisson likelihood with the risk parameter under single Gamma prior distribution (Venkatesan & Srinivasan, 2010; Srinivasan & Venkatesan, 2014).

Let, $y_i, i = 1, \dots, n$ be counts of disease in arbitrary small areas. Also define, for the same areas, expected rates $\{E_i\}$ and relative risks$\{\theta_i\}$. It is assumed that $y_i \sim Pois(E_i\theta_i)$ given $\theta_i$ are *iid.* Assuming $\theta_i = \theta$, for all *i* and that the prior distribution of $\theta$, $p(\theta)$, is $\theta \sim Gamma(\alpha, \beta)$ where $E(\theta) = \alpha/\beta$, and $Var(\theta) = \alpha/\beta^2$. The posterior distribution of $\theta$ is given by,

$$[\theta|\boldsymbol{y}, \alpha, \beta] = \frac{\beta^* \alpha^*}{\Gamma(\alpha^*)} \theta^{\alpha^* - 1} \exp(-\theta\beta^*)$$

where $\alpha^* = \sum y_i + \alpha, \beta^* = \sum E_i + \beta$. It follows that the predictive distribution is

$$[y^*|y, \alpha, \beta] = \int f(y^*|\theta) f(\theta|\alpha, \beta) d\theta$$

$$= \prod_{i=1}^{m} \left[ \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(y_i^* + \alpha)}{(E_i + \beta)^{(y_i^* + \alpha)}} \right]$$

**4. Data Description**

The number of Dengue Disease cases reported from 2007 to 2018 have been obtained from the Department of Public Health and Prevention Medicine, Tamil Nadu. The observed cases were aggregated for each district of Tamil Nadu and expected numbers of cases were calculated using the indirect standardization method from the population of each district. The data covered the details on Dengue cases in Tamil Nadu, District-wise from 2007 to 2018. The coordinate information about the geographical location for the analysis is collected where the cases happened and the details about location name, latitude, and longitude of the location of the cases were assimilated.

The real-world coordinates to each pixel of the raster are assigned by the process of Geo-referencing using ArcGIS. In the present work, a scanned map of Tamil Nadu which is marked district wise is digitized by using the coordinates from the markings on the map image. With the aid of GCPs (Ground Control Points), the image is warped and is made to configure within the coordinate system that is chosen.

## 5. Results

The details of the results of Bayesian Disease mapping analysis through the Poisson Gamma model are presented below in the form of tables and map. Table 1 shows the posterior point estimate of relative risk for each of the study areas using the Poisson Gamma model.

**Table 1**
Posterior statistics for the parameter in the Poisson Gamma model

| Parameter | Mean | Std.dev | MC_error | val2.5pc | Median | val97.5pc |
|---|---|---|---|---|---|---|
| Alpha | -0.2213 | 0.1298 | 0.01067 | -0.4559 | -0.2417 | 0.05762 |
| Mean | 0.8083 | 0.1077 | 0.008852 | 0.6339 | 0.7853 | 1.059 |
| Sigma | 0.5335 | 0.5797 | 0.05705 | 0.3374 | 0.4252 | 3.578 |
| Tau | 5.519 | 1.787 | 0.1415 | 0.07812 | 5.531 | 8.783 |

**Table 2**
The Posterior relative risk for all the regions of Tamil Nadu using Poisson-Gamma model

| Area | Mean | Sd | MC_error | val2.5pc | Median | val97.5pc |
|---|---|---|---|---|---|---|
| RR[1] | 0.3760 | 0.0244 | 0.0002 | 0.3291 | 0.3756 | 0.4249 |
| RR[2] | 1.9020 | 0.0215 | 0.0001 | 1.8600 | 1.9020 | 1.9450 |
| RR[3] | 1.1960 | 0.0203 | 0.0001 | 1.1570 | 1.1960 | 1.2360 |
| RR[4] | 0.5594 | 0.0160 | 0.0001 | 0.5286 | 0.5592 | 0.5913 |
| RR[5] | 0.9803 | 0.0277 | 0.0002 | 0.9269 | 0.9800 | 1.0360 |
| RR[6] | 0.6064 | 0.0182 | 0.0001 | 0.5714 | 0.6062 | 0.6429 |
| RR[7] | 0.4659 | 0.0155 | 0.0001 | 0.4358 | 0.4658 | 0.4967 |
| RR[8] | 0.3236 | 0.0100 | 0.0001 | 0.3042 | 0.3235 | 0.3435 |
| RR[9] | 1.4470 | 0.0301 | 0.0002 | 1.3880 | 1.4460 | 1.5070 |
| RR[10] | 0.6220 | 0.0265 | 0.0002 | 0.5709 | 0.6216 | 0.6748 |
| RR[11] | 0.7026 | 0.0206 | 0.0001 | 0.6629 | 0.7023 | 0.7440 |
| RR[12] | 1.7940 | 0.0263 | 0.0002 | 1.7440 | 1.7940 | 1.8460 |
| RR[13] | 0.5926 | 0.0208 | 0.0001 | 0.5523 | 0.5922 | 0.6337 |
| RR[14] | 0.5406 | 0.0194 | 0.0002 | 0.5033 | 0.5402 | 0.5793 |
| RR[15] | 0.2375 | 0.0186 | 0.0002 | 0.2026 | 0.2370 | 0.2748 |
| RR[16] | 0.6969 | 0.0379 | 0.0002 | 0.6245 | 0.6963 | 0.7731 |
| RR[17] | 0.6410 | 0.0215 | 0.0002 | 0.5994 | 0.6407 | 0.6834 |
| RR[18] | 1.1580 | 0.0319 | 0.0002 | 1.0970 | 1.1580 | 1.2220 |
| RR[19] | 0.8900 | 0.0171 | 0.0001 | 0.8568 | 0.8897 | 0.9244 |
| RR[20] | 0.4165 | 0.0193 | 0.0001 | 0.3795 | 0.4161 | 0.4551 |
| RR[21] | 1.0610 | 0.0228 | 0.0002 | 1.0170 | 1.0610 | 1.1060 |
| RR[22] | 2.6820 | 0.0502 | 0.0004 | 2.5840 | 2.6820 | 2.7800 |
| RR[23] | 0.6175 | 0.0141 | 0.0001 | 0.5900 | 0.6174 | 0.6456 |
| RR[24] | 2.6530 | 0.0427 | 0.0003 | 2.5700 | 2.6530 | 2.7370 |
| RR[25] | 0.8338 | 0.0190 | 0.0001 | 0.7972 | 0.8336 | 0.8718 |
| RR[26] | 2.4770 | 0.0310 | 0.0002 | 2.4160 | 2.4760 | 2.5370 |
| RR[27] | 1.4350 | 0.0269 | 0.0002 | 1.3820 | 1.4350 | 1.4870 |
| RR[28] | 0.5681 | 0.0165 | 0.0001 | 0.5363 | 0.5680 | 0.6006 |
| RR[29] | 0.8842 | 0.0286 | 0.0002 | 0.8302 | 0.8836 | 0.9418 |
| RR[30] | 0.4784 | 0.0118 | 0.0001 | 0.4560 | 0.4783 | 0.5018 |
| RR[31] | 0.4232 | 0.0120 | 0.0001 | 0.4001 | 0.4230 | 0.4469 |
| RR[32] | 0.6171 | 0.0192 | 0.0001 | 0.5798 | 0.6169 | 0.6551 |

From Table 2, it is observed that the range of the relative risk estimates is calculated using the Poisson-Gamma model on average. Tables 2 shows the posterior summary of various parameters estimated by the MCMC which consists of mean, standard deviation, Monte Carlo error and selected the percentiles (2.5%, 50%, and 97.5%). MC error was estimated using the batch mean method. The Bayesian estimators of the parameter's alpha, tau and the mean, sigma of the Poisson Gamma model were observed as -0.1607, 0.8592, 2.636, and 5.519 respectively. The respective confidence intervals of these parameters were (-0.1618, 0.0887), (0.6737,1.093), (1.499, 4.117) and (0.3609, 0.9793), respectively.

The iteration trace of alpha, tau, and mean trace plots about the sigma parameter were also analyzed. The results of iteration show that the Markov chain-based Gibbs sampler of model parameters offered convergence. For each area and parameter (alpha, tau, and sigma), the correlogram plots of autocorrelations of the chain are developed to understand the convergence of the parameters in detail. The densities of sections of the chains for the study area and parameters alpha, mean, tau and sigma and was produced an approximate visual kernel estimate of the posterior densities of the two models which are shown. The relative risks of dengue disease are estimated by two models. The Poisson Gamma and used to estimate the relative risks refer to the above equations respectively. On the contrary, the estimates of the Bayesian Poisson-Gamma not only on $y_i$ and $e_i$ but also on the probability distributions and other parameters.

Using the results obtained from the Poisson Gamma model a map indicating the relative risk of dengue at each district of Tamil Nadu was constructed and presented in Fig. 1. It is seen that about 18 districts have a relative risk ranging between 0.238 to 0.703, 7 districts come under the relative risk of 0.703 to 1.196, 4 districts come under relative risk of 1.196 to 1.902. Also, it is observed that 3 districts have a higher relative risk of 1.902 to 2.282.

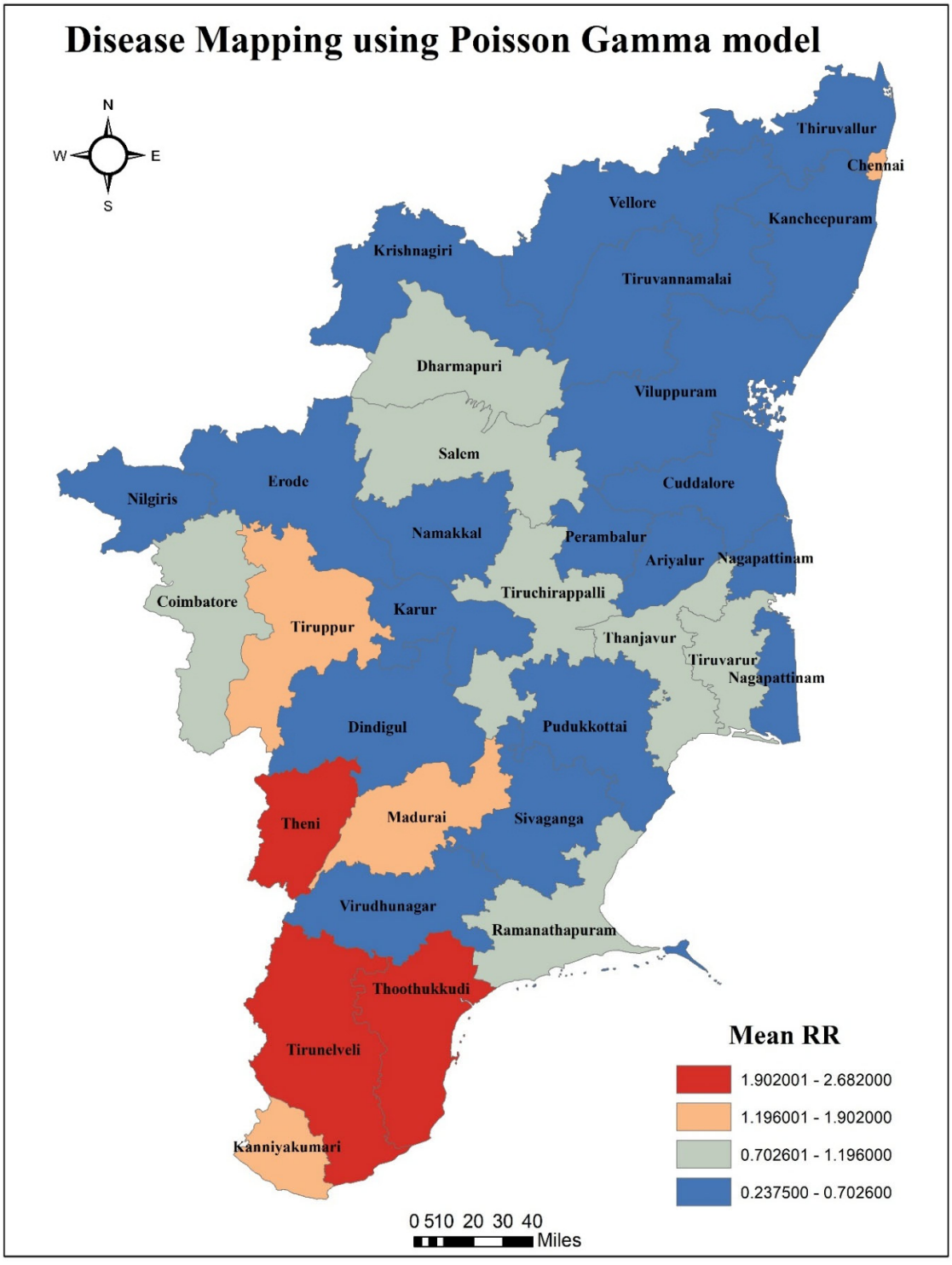


**Fig. 1.** The map indicating the relative risk of dengue at each district of Tamil Nadu

## 6. Conclusion

In this study, based on the results of the estimations, the study areas that have very high relative risk claims through the Poisson-Gamma model are Theni, Tirunelveli and Thoothukudi. The Poisson Gamma model provides smoother estimates. It is expected that the results of this research can be used by the health authorities as a reference for estimating the risk of dengue disease in all districts of Tamil Nadu. It can also be used to estimate the high and very high relative risk of the study region. The more accurate results may be obtained from further research. Apart from that, other types of Bayesian models, especially the spatial one (Cressie, 1993), can be applied to the estimation.

## References

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in Medicine*, *14*(21-22), 2433-2443

Besag, J., & Newell, J. (1991). The detection of clusters in rare diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 154(1), 143-155.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1), 1-20.

Clayton, D., & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 671-681.

Cressie N. (1993). Statistics for spatial data. *John Wiley & Sons* ; New York.

Eckert, N., Parent, E., Bélanger, L., & Garcia, S. (2007). Hierarchical Bayesian modelling for spatial analysis of the number of avalanche occurrences at the scale of the township. *Cold Regions Science and Technology*, *50*(1-3), 97-112.

Inskip, H., Beral, V., Fraser, P., & Haskey, J. (1983). Methods for age-adjustment of rates. *Statistics in medicine*, *2*(4), 455-466.

Jaisankar, R., & Kesavan, J. (2019). A study on spatial variations in temporal trends of dengue incidences in Tamil Nadu, India. *International Journal of Scientific & Technology Research, 8*(9), 788-92.

Jaisankar, R., Kesavan, J., & Ranjani M. (2019). A Spatial Analysis on Dengue Outbreaks in Tamil Nadu, 2013-2018. *International Journal of Current Research,* Vol. 11, Issue, 11, pp.8116-8120.

Knorr-Held, L., & Becker, N. (1999). Bayesian modelling of spatial heterogeneity in disease maps with application to German cancer mortality data.

Lawson, A. B., Biggeri, A. B., Boehning, D., Lesaffre, E., Viel, J. F., Clark, A., & Divino, F. (2000). Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group. *Statistics in Medicine*, 19(17), 2217-41.

Millar, R. B. (2009). Comparison of hierarchical Bayesian models for over dispersed count data using DIC and Bayes' factors. *Biometrics*, *65*(3), 962-969.

Richardson, S., Thomson, A., Best, N., & Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental health perspectives*, *112*(9), 1016-1025.

Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge university press.

Snow, J. (1854). On the communication of cholera by impure Thames water. *Medical Times and Gazette*, *9*, 365-366

Srinivasan, R., & Venkatesan, P. (2014). Bayesian random effects model for disease mapping of relative risks. *Ann Biol Res*, *5*(1), 23-31.

Venkatesan P, & Srinivasan R. (2010). Bayesian model of HIV/AIDS in India: A spatial analysis. *Applied Bayesian Statistical Analysis,* 51-56.

WHO, Dengue in World Health Organization in India. (Accessed July 2019). [http://www.searo.who.int/india/topics/dengue/en/]