Jurnal Teknologi

Full Paper

DENGUE DISEASE MAPPING IN BANDUNG, INDONESIA: AN ANALYSIS BASED ON POISSON-GAMMA, LOG-NORMAL, BYM AND MIXTURE MODELS

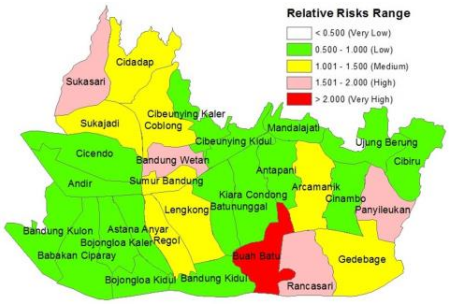
Farah Kristiani[a,b*], Nor Azah Samat[b], Sazelli Ab Ghani[b]

[a]Mathematics Department, Faculty of Information Technology and Sciences, Parahyangan Catholic University, Jln. Ciumbuleuit 94, Bandung – 40141, Indonesia
[b]Mathematics Department, Faculty of Mathematics and Sciences, Sultan Idris Education University, Tanjong Malim, Perak Darul Ridzuan-35900, Malaysia

*Corresponding author
farah@unpar.ac.id

Article history
Received
28 August 2015
Received in revised form
15 December 2015
Accepted
12 January 2016

Graphical abstract



Abstract

Dengue is the most rapidly spreading mosquitoes-borne viral disease in the world, especially in Bandung, Indonesia. This disease can be controlled if detected early. Therefore, in order to prevent and control this disease before it occurs, government and society must be cooperative to eradicate this dangerous disease. The statistical model used in the study of disease mapping can be considered as an important contribution. In this paper, the relative risk estimations using the Poisson-gamma, Log-normal, Besag, York and Mollié (BYM) and Mixture models for Bandung municipality will be investigated. In this study, the aggregated data of observed dengue data from Bandung, Indonesia from the year 2013 will be analyzed. The estimated relative risk will be displayed in tables and maps to obtain the clearer depictions of disease risks distribution in each area.

Keywords: Bandung, dengue, mosquitoes-borne, relative risk

1.0 INTRODUCTION

Dengue is one of the dangerous diseases in the world. This disease causes not only death but also economic losses. The economic losses are imposed by treated dengue occurrence that causes substantial costs [1]. The economic impact is divided into two, namely the direct and indirect economic impact [2]. The direct economic impacts are costs of the treatment or cure, for ambulatory case or hospitalized case. The examples of indirect economic impact are lost productivity, school absenteeism, and unpaid time of caregivers. Other losses include social losses such as causing panic or mortality in the family [3].

Generally, the number of dengue cases seems to be increasing from year to year worldwide, and this includes Bandung, Indonesia, specifically. Nathan, Dayal-Drager, Guzman *et al*. [1] mentioned that 150,000 cases in Indonesia were reported in 2007 which is the highest on record with over 25,000 cases reported from both Jakarta and West Java. According to Health Department of Indonesia, the total number of this disease in 2012, reported to World Health Organization (WHO), was 90,245 cases. This number increased to 112,511 cases in 2013. Bandung as the most populous city in West Java, Indonesia also has significant number of dengue cases. This information was obtained from Bandung Department of Health [4-5].

It is said that prevention is better than cure. This guideline applies in dealing with dengue disease. The primary preventions that can be done are with vector control and keep the person does not get bitten by a mosquito that has been infected. In Indonesia, current prevention and control strategies for dengue disease

include three important aspects. They include the environmental, biological and chemical aspect and involve various concerned, starting from Health Department, Health Services Workers, to society. For public education, there is a famous slogan for educating them, that is 3M (*Menutup, Menguras, Menimbun*), meaning Covering, Cleaning, Piling. Fogging is also conducted in areas which are infected by this disease. These strategies are only implemented after the dengue cases have occurred.

Ma and Li [6] stated that to prevent and to control infectious diseases more effectively, it is very important to first fully understand the mechanism of the spread and the transmission dynamics of the diseases. Only after we have the knowledge about the mechanism of the spread and the transmission dynamics of the diseases, we can provide useful predictions and guidance so that better strategies can be established.
Disease mapping is a method to depict transmission of a particular disease geographically. A good maps of disease risk can be useful as an important tool for disease control. It can be used to identify the risk area that deserve deeper observation and more attention for prevention and accurate treatment. This mapping refers to the visual representation of the geographical distributions.

This research concerned with the study of the geographical distribution of dengue disease in Bandung, West Java, Indonesia. If this transmission can be mapped, then the particular areas that are needed serious treatment will be showed up, hence the prevention and treatment of this disease can be implemented more effective and efficient.
Bandung is divided into two different administrative areas which are Bandung municipality and Bandung residence. In this research, the critical area that will be chosen is Bandung municipality. Therefore, in this article, the term Bandung will refer to Bandung municipality.

The existing disease mapping researches in Bandung only analyze the deterministic model using the existed observed dengue data [7-8]. In the real condition, the random effects cannot be neglected, hence the stochastic factor also need to be counted. Some stochastic researches indicated that the stochastic models are more realistic than the deterministic ones [9-10].

The Health Department usually used the total numbers of dengue occurrences across the regions for depicting the high-low risk areas. It is common to conclude that high risks of dengue occurrences in certain areas are represented by the large numbers of dengue cases of the areas without considering the population or the land size. The relative risk should be considered to be mapped in term of disease mapping, because it is taken to measure the excess risk found in relation to that supported purely by the local population, which is 'at risk' [11]. Relative risk is the ratio of the exposed group that will develop disease to the unexposed group that will develop the same disease. In brief, relative risk is usually used to compare the risks of different groups. Reference [12] explained the interpretation of relative risk as the probability that a person within a specified region contracted by the disease which is divided by the probability that a person in the population contracted by the disease.

In this research, in order to obtain a better risk map in Bandung, the high-low areas will be estimated and predicted using the observed dengue data provided by the Bandung Health Department in 2013. The existing models which will be used in this study are the Poisson-gamma, Log-normal, BYM and Mixture models.
Recently, many researches focused on dengue disease mapping have been familiar with the GIS (Geographical Information System) implementation [13-14]. The advantages of this software are producing and providing clear depiction of relative risk estimations which analyze the spatial data. The GIS software can also be a useful presentation tool for disease maps.

However, though the GIS software is useful with its advantages, it still has the drawback in limitation of the statistical capabilities. Whereas, the statistical capabilities related to data processing cannot be neglected to provide the more informative disease maps. Therefore, this research will use the WinBUGS software. This software is specialized in BUGS (Bayesian inference Using Gibbs Sampling) and can be run under Windows. It assumes a Bayesian probability model, in which all unknown parameters are treated as random variables as explained above. The relative risk which are estimated, will be displayed in maps to depict the high and low risk areas. These maps can be produced using the ArcGIS software. Results of the analysis will be presented in tables and maps.

It is better to assess the disease transmission which takes the relative risk estimation into account because human populations influence the transmission of the disease. This section will explain the existing models used in the study of disease mapping. Actually there are many models that commonly been used which includes the non-spatial, spatial and space-time models. However, in this paper, there are only four models which discussed in [15], will be elaborated.

2.0 METHODS

It is better to assess the disease transmission which takes the relative risk estimation into account because human populations influence the transmission of the disease. This section will explain the existing models used in the study of disease mapping. Actually there are many models that commonly been used which includes the non-spatial, spatial and space-time models. However, in this paper, there are only four models which discussed in [15], will be elaborated.

2.1 Poisson-gamma Model

Basically, the study area to be mapped is divided into M mutually exclusive regions ($i=1,2,\ldots,M$). Using the Bayesian inference to estimate the relative risk, it can be explained that there are two sources of information about any problem. These sources are provided by the

prior distribution and likelihood. Likelihood provides the information about the parameter via the data whilst prior distribution provides the information via prior beliefs or assumptions. The product of the likelihood and the prior distribution is called the posterior distribution. This distribution describes the behavior of the parameters after the data are observed and prior assumptions are made. It can be specified as proportionality:

$$p(\theta|y) \propto L(y|\theta)g(\theta), \quad (1)$$

where $g(\theta)$ is the joint distribution of the θ vector, with θ is the estimation of the relative risk.

In this model, the number of morbidity in *i*th area are assumed to be mutually independent and follow Poisson distributions,

$$y_i \sim Poisson(e_i\theta_i), \quad \forall i.$$

If the prior distribution is a $Gamma(a,b)$, then the posterior distribution of the relative risk would be $Gamma(a+y_i, b+e_i)$, with $a,b \sim Exponential(0.1)$.

However, the Poisson-gamma model also has some drawbacks. One of them is its inability to cope with the spatial correlation. This drawback can be overcame by BYM models which consider the neighboring areas.

2.2 Log-normal Model

Again it is assumed that the numbers of morbidity in each area are to be mutually independent and follow Poisson distributions,

$$y_i \sim Poisson(e_i\theta_i), \quad \forall i,$$

with $\log\theta_i = \alpha + v_i$, where $v_i = N(0,\tau_v^2)$.

In this case, α represents the overall log relative risk of disease in the whole study region compared with the reference rate and v_i represents the residual log relative risk in area i compared with the average over the study region [16].

2.3 Besag, York and Molli (BYM) Model

In this model, the relative risk is modelled with additional consideration. BYM considered two different components which influence the area specific random effects. They reckoned the component of the varying effect in a structured manner in space (clustering or correlated heterogeneity) and in an unstructured way between areas (uncorrelated heterogeneity).

The model is formulated as follows,

$$y_i \sim Poisson(e_i\theta_i),$$
$$\log\theta_i = \alpha + u_i + v_i, \quad (2)$$

where α is an overall level of the relative risk, u_i is the correlated heterogeneity and v_i is the uncorrelated heterogeneity. The prior distribution model for the uncorrelated heterogeneity is $v_i \sim N(0,\tau_v^2)$.

It is explained clearly that a spatial correlation structure is necessary in the clustering component. In this case, the risk estimation in any area will depend on neighboring areas. Therefore, the conditional autoregressive (CAR) model proposed by [16] will be used to model the distribution of the correlated heterogeneity as $[u_i|u_j, i \neq j, \tau_u^2] \sim N(\bar{u}_i, \tau_i^2)$, where $\bar{u}_i = \frac{1}{\sum_j \omega_{ij}}\sum_j u_j\omega_{ij}$, $\tau_i^2 = \frac{\tau_u^2}{\sum_j \omega_{ij}}$, $\omega_{ij} = 1$. if *i*, *j* are adjacent (or 0 if they are not).

In a full Bayesian inference, prior distributions for the parameters, in this case are τ_u and τ_v, must be specified. These parameters control the variability of *u* and *v*. From this source, *u* and *v* are considered to have Gamma distribution.

2.4 Mixture Model

This model was proposed by Lawson and Clark [10], with the equation as follows,

$$\log\theta_i = \alpha + v_i + p_iu_i + (1-p_i)\varphi_i. \quad (3)$$

The log function of θ_i is influenced by three additive components. v is the fixed component which represents the unstructured heterogeneity and measure the overdispersion in a individual region. Other two mixing components are (u,φ), represent different aspects of spatial correlation.

Special cases of this formulation arise depending on the value of p_i. If all $p_i = 1, \forall i$, then Equation (3) become $\log\theta_i = \alpha + v_i + u_i$, which is same with the BYM model in Equation (2). The other case is called a pure jump model which will happen if all $p_i = 0, \forall i$.

3.0 THE DATA SET

In this analysis, the data were provided by the Health Department of Bandung. For the preliminary investigation, the dengue data which mentioned in this paper refer to Dengue Fever, Dengue Hemorrhagic Fever and Dengue Shock Syndrome number of cases. The data were collected from all hospitals by the Health Department of Bandung in 2013 and covered all districts in Bandung. From the obtained data, it is known that Buah Batu had the highest dengue cases among other districts in Bandung, which were 540 cases occurred.

4.0 RESULTS AND DISCUSSION

The relative risks estimation obtained from Poisson-gamma, Log-normal, BYM and Mixture model are displayed in Table 1.

Table 1 Relative Risks Estimation of Dengue Cases in Bandung, 2013

District	P-G	Log-normal	BYM	Mixture
Bojongloa Kaler	0.5277	0.5319	0.5268	0.5413
Bandung Kulon	0.5485	0.5515	0.5444	0.5617
Bojongloa Kidul	0.5618	0.5672	0.5615	0.5766
Mandalajati	0.6337	0.6365	0.6448	0.6548
Andir	0.6738	0.6765	0.6703	0.69
Babakan Ciparay	0.7121	0.7124	0.7008	0.7266
Batununggal	0.73	0.7298	0.7429	0.749
Cibeunying Kidul	0.801	0.8016	0.8023	0.8202
Ujung Berung	0.8089	0.8092	0.8122	0.8285
Cinambo	0.8292	0.8304	0.8521	0.854
Astana Anyar	0.8294	0.8296	0.8162	0.845
Cibiru	0.8315	0.8304	0.832	0.8507
Cicendo	0.8351	0.8342	0.8381	0.8551
Bandung Kidul	0.8844	0.882	0.8867	0.9011
Kiaracondong	0.8881	0.888	0.8875	0.9085
Antapani	0.9147	0.9143	0.9202	0.9363
Cibeunying Kaler	0.9201	0.9197	0.9225	0.9402
Sumur Bandung	1.075	1.07	1.066	1.091
Coblong	1.126	1.124	1.13	1.151
Regol	1.15	1.15	1.147	1.173
Gedebage	1.195	1.19	1.203	1.218
Arcamanik	1.331	1.329	1.324	1.357
Sukajadi	1.35	1.349	1.355	1.38
Lengkong	1.402	1.401	1.393	1.429
Cidadap	1.498	1.5	1.514	1.535
Panyileukan	1.509	1.507	1.494	1.539
Bandung Wetan	1.511	1.513	1.475	1.537
Rancasari	1.527	1.527	1.536	1.564
Sukasari	1.685	1.686	1.7	1.729
Buah Batu	2.196	2.206	2.188	2.255

Samat and Percy [17] explained the interpretation of relative risk estimation. A relative risk less than one means that susceptible people within particular district is generally less likely to contract dengue compared to people in the overall population. On the contrary, a relative risk more than one means that susceptible people within particular district is generally more likely to contract dengue compared to people in the overall population.

From Table 1, thirteen of thirty districts have relative risks estimation more than one, even Buahbatu has relative risk more than two. This condition can be concluded that susceptible people in these districts are more likely to contract dengue compared to the overall population in Bandung, whilst susceptible people within the other districts are less likely to contract dengue compared to people within the overall population.

The results showed in Table 1 can be displayed in maps as in Figure 1-4 to depict the high and low risk areas in Bandung.

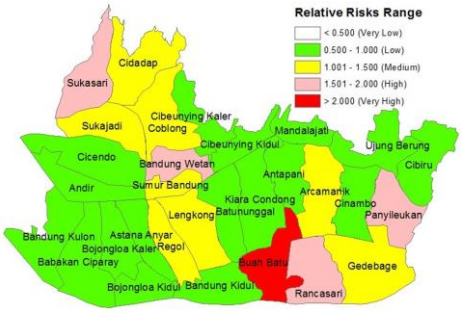


Figure 1 Disease Map of Estimated Relative Risk based on Poisson-gamma Model in Bandung, 2013

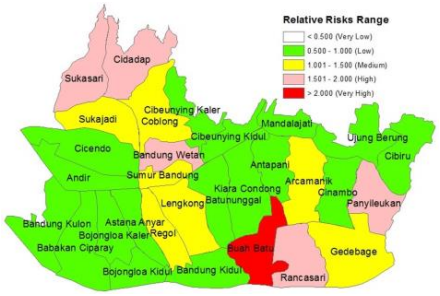


Figure 2 Disease Map of Estimated Relative Risk based on Log-normal Model in Bandung, 2013

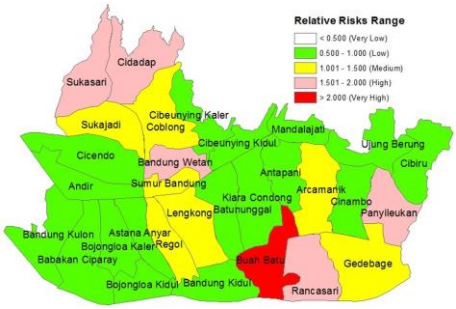


Figure 3 Disease Map of Estimated Relative Risk based on Mixture Model in Bandung, 2013

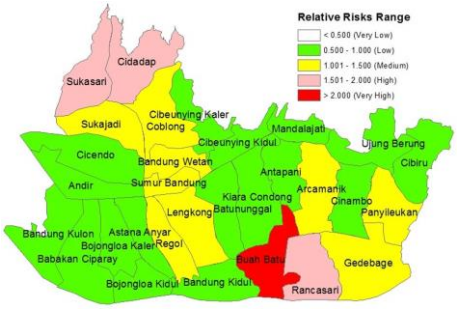


Figure 4 Disease Map of Estimated Relative Risk based on BYM Model in Bandung, 2013

These maps are used to represent the relative risk of each district that has been estimated by the models which have been explained previously. The purpose of using these maps are depicting and differing between the high and low risk areas of dengue cases for each district in Bandung, Indonesia in 2013. On these maps, there are five different areas classified depending on the relative risk estimation in each area. From the depiction above, it shows that there are no areas in very low level in 2013. All of those areas have low until very high risks level.

Furthermore, to identify the better model, the result of each model will be compared using Deviance Information Criterion (DIC) diagnostic as suggested by Spiegelhalter, *et al*. [18]. This method tool is intended to compare the Bayesian models using the criterion of deviance information. In this study, the DIC for each model calculated by WinBUGS software, is displayed in Table 2.

Table 2 Comparison of the Deviance Information Criterion (DIC) from Poisson-gamma, Log-normal, BYM and Mixture Model

Model	DIC
Poisson-gamma	268.453
Log-normal	268.632
BYM	268.586
Mixture	264.919

The best model which is chosen to fit the data is the model with the smallest DIC. Therefore, in this case, from the DIC comparison displayed in Table 2, the Mixture model is the best model that fit the data.

The results obtained above are the preliminary results of the dengue cases in Bandung for the year 2013. From the mapped estimated relative risks showed above, we can know the distribution of the high and low dengue risks areas and the critical areas can be identified. It is expected that the mapping of these relative risks would be better if the weekly data are used due to the characteristics of the disease.

For Poisson-gamma and Log-normal model, there are no possibility to take into account the spatial correlation. Hence, the BYM and Mixture model improve the previous models by taking into account the spatial correlation. These models consider the condition that estimate the relative risk in any area will vary because it is influenced by neighboring areas.

Mixture model accommodate two mixing components, which represent different aspects of spatial correlation. These two mixing components combine the spatial correlation component and discrete jumps model component. The DIC obtained from Mixture model is the smallest one since there are some possibilities of discrete jump areas in the critical areas. The causes of this condition will be investigated in the further research.

Dengue incubation period is about 3-8 days. Because of this fact, then getting the disease mapping that hopefully approaches the real condition must be based on the weekly real data. Unfortunately, in Bandung, one problem that is faced when obtaining the real observed data of dengue cases is that the researcher needs to collect the data separately from each hospital. It will be easier if there is cooperation between researchers and Health Department.

5.0 CONCLUSION

The relative risk estimations obtained by Poisson-gamma, Log-normal, BYM and Mixture models that are suggested in this paper provide an important approach to assess future risk. Some previous interpretations which had been explained are expected to be used as a reference in order to control and prevent dengue disease before it occurs. The estimated relative risk values can be displayed in a dengue risk map to show the high and low risk area of dengue risk. The higher the risk of a particular area, the more serious attention of government policy and financial support are needed.

It can be concluded that those models can be implemented using dengue data in Bandung municipality in 2013. The spatial model which is Mixture model, can be chosen as the better model that fit the data, since it has the smallest DIC. The other spatial model, BYM model, and the non-spatial models, Poisson-gamma and Log-normal, can also be applied to model the relative risks estimation, but are not as good as the Mixture model, since the BYM, Poisson-gamma and Log-normal model have greater DIC.

It is expected that the government can arrange better strategy in controlling and preventing the disease transmission. Thus, people at risk will be more aware and understand through scientific explanation that warns the probability of contracting dengue in the future.

Acknowledgement

The authors appreciate the valuable insights and comments provided by anonymous referees, leading to many improvements in the paper. The authors acknowledge support from the Health Department of Bandung and West Java Province for their data. The financial support provided by Parahyangan Catholic University and Sultan Idris Education University are gratefully acknowledged.

References

[1] Nathan, D., Dayal-Drager, D., Guzman, D., et al. 2009. Epidemiology, Burden of Disease and Transmission, Dengue, Guidelines for Diagnosis, Treatment, Prevention and Control. World Health Organization.

[2] Suaya, J., Shepard, D., Siquera, J., Martelli, C., Lum, L., Tan, L., et al. 2009. Cost of Dengue Cases in Eight Countries in The Americas and Asia: *A* Prospective Study. *American Journal of Tropical Medicine and Hygiene*. 80(05): 846-855.

[3] Widiyani, R. Empat Sekawan Penyebab DBD. April 2013. [Online]. Available: http://health.kompas.com/read/2013/04/03/18534298/Empat.Sekawan.Penyebab.DBD.

[4] Bandung, H. D. 2011. *Health Profile of Bandung Municipality in 2011*. Bandung: Bandung Government.

[5] Bandung, H. D. 2012. *Health Profile of Bandung Municipality in 2012*. Bandung: Bandung Government.

[6] Ma, Z. and Li, J. 2009. *Dynamical Modeling and Analysis of Epidemics.* World Scientific Publishing Co. Pte. Ltd.

[7] Supriatna, A., Soewono, E. and Gils, S. 2008, A two-age-classes Dengue Transmission Model. *Mathematical Biosciences.* 216: 114-121.

[8] Supriatna, A. 2009, Estimating the Basic Reproduction Number of Dengue Transmission during 2002-2007 Outbreaks in Bandung, Indonesia. Dengue Bulletin. 33: 21-22.

[9] Daley, D. and Gani, J. 1999. *Epidemiology Modelling: An Introduction*. Cambridge University Press.

[10] Wakefield, J. 2007. Disease Mapping and Spatial Regression with Count Data. *Biostatistics*. 8(2): 158-183.

[11] Lawson, A. 2006. *Statistical Methods in Spatial Epidemiology*. John Wiley and Sons Ltd.

[12] Samat, N. A. and Ma'arof, S. 2013. Dengue Disease Mapping with Standardized Morbidity Ratio and Poisson-Gamma Model: An Analysis of Dengue Disease in Perak, Malaysia. World Academy of Science, Engineering and Technology. *International Journal of Mathematical, Computational Science and Engineering*. 7(8): 640-644.

[13] Chang, A., Parrales, M., Jimenez, J., Sobieszczyk, M., Hammer, S., Copenhaver, D. and Rajan, P. 2009. Combining Google Earth and GIS Mapping Technologies in a Dengue Surveillance System for Developing Countries. *International Journal of Health Geographics.* 8: 49.

[14] Eisen, L. and Lozano-Fuentes, S. 2009. Use of Mapping and Spatial and Space-time Modelling Approaches in Operational Control of Aedes Aegypti and Dengue. *PLoS Neglected Tropical Diseases.* 3(4): 3.

[15] Lawson, A., Browne, W. and Rodeiro, C. 2003. *Disease Mapping with WinBUGS and MLWin*. John Wiley and Sons Ltd.

[16] Lawson, A. and Clark, A. 2002. Spatial Mixture Relative Risk Models Applied to Disease Mapping. *Statistics in Medicine*. 21: 359-370.

[17] Samat, N. and Percy, D., 2012. Dengue Disease Mapping in Malaysia based on Stochastic SIR Models in Human Populations. *International Conference of Statistics in Science, Business and Engineering* (ICSSBE).

[18] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. v. d., 2002. Bayesian Measures of Model Complexity and Fit. *Royal Statistical Society*. 64(4): 583-639.