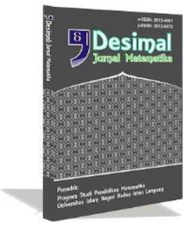
Contents lists available at DJM
DESIMAL: JURNAL MATEMATIKA
p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm
http://ejournal.radenintan.ac.id/index.php/desimal/index



Bayesian spatial data analysis: Application of pneumonia spread in west java

Muhammad Yusuf Al Habsy, Fulkan Kafilah Al Husein, Muhammad Harun Yahya, Ro'fah Nur Rachmawati*

Indonesia Defense University, Indonesia

ARTICLE INFO

Article History
Received : 11-01-2024
Revised : 14-02-2024
Accepted : 20-03-2024
Published : 30-04-2024

Keywords:
Besag York Mollie; Desease Mapping; INLA; Latent Gaussian; Spatial Statistics.

*Correspondence: E-mail: rofah.rachmawati@idu.ac.id

Doi: 10.24042/djm.v7i1.23154

ABSTRACT

Pneumonia has a notable influence on public health, especially among susceptible demographics like children and the elderly. This respiratory disease can be transmitted through human interaction. Analyzing the spread of the illness within a community requires assessing the characteristics of the community itself. The objective of this research is to describe the distribution of pneumonia cases and their causes in the West Java Province using RStudio software. The analytical method employed is the Integrated Nested Laplace Approximations (INLA) approach, a Bayesian statistical method used for estimation in complex Bayesian models, particularly in hierarchical or nested structure. The sample utilized comprises the entire population, totaling 27 Districts/Cities within West Java Province. The influence of differences in population size, number of people living in poverty, waste production, the quantity of primary healthcare facilities, total number of vehicles, and the count of HIV patients in Cities/Regencies in West Java on the spread of pneumonia will be analyzed. The result of analysis show that the population and number of health centers variables had a significant influence on the mapping of pneumonia disease in each location. And also, the Relative Risk (RR) and Standardized Incidence Ratio (SIR) maps show that some regions have a higher risk of pneumonia compared to other regions. These findings are expected to provide insights for public policies in addressing health issues, particularly in the efforts to prevent and control diseases like pneumonia. Moreover, these results serve as a foundation for further studies regarding other factors that might contribute to the spread of this disease at the local level.

http://ejournal.radenintan.ac.id/index.php/desimal/index

INTRODUCTION

Pneumonia has significant impact on public health, particularly among vulnerable groups such as children and the elderly (Sun, 2020). Pneumonia is a respiratory disease that can be spread through human contact (Chakraborty, 2020). To understand how the distribution of the disease in the community, it can be seen how the character of the community itself (Gajdács, 2020). The deeper insight of how pneumonia distribute among community will support more effective planning of public health (Cillóniz, 2022). In this case, the character of the community chosen as an explanatory variable, namely differences in population size, number of people living in poverty, waste production, the quantity of primary healthcare facilities (Puskesmas), total number of vehicles, and the count of HIV patients (Wang, 2022). Statistical analysis will be conducted to evaluate whether there are correlations between these variables and the incidence of pneumonia in each city or district, with the hope that this research will offer a comprehensive insight about how pneumonia distribute in West Java Province (Zhong, 2021). The statistical analysis used is a spatial Bayesian model with INLA (Satria, 2021). INLA is a statistical method used to estimate complex hierarchical or Bayesian models (Gómez, 2020). INLA approach is highly beneficial for understanding the spread of a disease, especially infectious diseases significantly influenced by their surrounding environments, such as pneumonia (Soh, 2022). Proximity to an affected area increases the risk of transmission compared to distant locations. Information about the geographical layout of an area and the pattern of disease spread becomes more apparent when utilizing spatial approaches (Murugesan, 2020). Therefore, data about pneumonia will be better monitored, focusing not only on who is affected but also on the geographical proximity between areas and the occurring cases (Wang, 2020).

The main objective of this paper is to describe the distribution of pneumonia cases and their causes in the West Java Province. Interventions in the future consideration of factors above is expected to provide a more holistic understanding of the dynamics of this disease spread in the West Java region (Purwanto, 2023). The novelty of this research lies in the selection of variables that influence the spread of pneumonia itself. By combining the influences of the aforementioned variables, the spread of pneumonia can be described in terms of its distribution patterns (Suaya, 2020). According to Yatnaningtyas, R., et al (2016), the conceptual framework used in the causal factors of pneumonia is a modified concept of The Epidemiological Triad or causal components proposed by John E. Gordon (1950), one of which is population and waste production. Multilevel modeling analysis of pneumonia showed that socioeconomics is a contributing factor (Machmud, R. 2009). Influential extrinsic risk factors are the number of health centers, and the percentage of the population that has access to proper sanitation facilities, as well as air quality, one of which comes from motor vehicles (Renika, I., & Amin, C. 2021). A person infected with pneumonia is certainly infected with HIV but not all those infected with HIV experience pneumonia infection (Carolin, B. T., et al. 2020).

From these modelling, we obtained satisfying results including pneumonia distribution mapping with relative risk and colour grading in it. In the remainder of this paper, we present the dataset, the brief methodology and INLA inference which are explained in Section 2. Results and discussions are reported in Section 3. Some concluding remarks and possible future development are summarized in Section 4.

METHOD

Dataset

The dataset in this study consists of the number of pneumonia cases in 27 cities/districts in West Java Province throughout 2021. The data was obtained from the official open data portal of the West Java Provincial Government, namely Open Data Jabar and the West Java Central Bureau of Statistics. Pneumonia cases are spread across 27 cities/districts of West Java Province in units of population. From the sources described in previous section, the variables selected are the population, the number of poor people in units of people, the amount of waste production in units of tons/day, the number of health center units, the number of vehicle units, and the number of HIV cases in 27 cities/districts in West Java Province throughout 2021. It is defined that variable population is x_{pop} , the number of poor people is x_{poor}, the amount of waste production is x_{waste}, the number of health center is x_{hc}, the number of vehicle $x_{vehicle}$, the number of HIV cases is x_{hiv}.

The West Java Province shapfile data format is also used to view the geographic information system obtained from geo-sai.my.id. The next process is mapping the existing dataset with SIR method. And after the data already plotted, next the data will be analyzed with Bayesian Spatial model as the next step.

Bayesian Spatial Model

We describe the spatial modeling of Pneumonia disease in West Java province in several steps below.

Stage 1. It's important to note that there is no need to include an intercept term as it is automatically included. In the random effect, we specify the index variable 'idarea' with the respective indices of the random effect. We establish a Penalized Compelxity prior for the marginal precision τ_b using $P\left(\left(\frac{1}{\sqrt{\tau_b}}\right) > U\right) = a$. Assuming a reasonable upper bound for the marginal standard deviation is approximately 0.5. This choice is conservative and assumes that the unstructured random effect contributes more to the variability than the spatially structured effect.

Stage 2. After the prior is obtained and entered into g. then the data frame g is ready to be modeled into the model formula 'bym2'. But before the data frame modeled into 'bym2' the equation is established to incorporates the dependent variable Y on the left side and the random effect 'bym2' on the right side. the observed counts $\boldsymbol{Y_i}$ in each area i are modeled using a Poisson distribution with a mean of $\boldsymbol{E_i\theta_i}$, where $\boldsymbol{E_i}$ represents the expected counts and θ_i represents the relative risk in that specific area. The logarithm of the relative risk θ_i is expressed as the sum of an intercept, which captures the overall disease risk level, and random effects that account for additional variability beyond what is expected from a Poisson distribution. The relative risk θ_i indicates whether area i has a higher (θ_i > 1) or lower (θ_i < 1) risk compared to the average risk in a standard population. For instance, if θ_i = 2, it means that the risk in area i is twice the average risk in the standard population**.**

The general equation model for spatial data can be represented as follows (Moraga, 2020):

$$\boldsymbol{Y_i \sim Po(E_i\theta_i), i = 1, .., n},$$

$$log\,\boldsymbol{log}\,(\boldsymbol{\theta_i}) = \boldsymbol{\alpha + \mu_i + v_i} \qquad \textbf{(1)}$$

Here**,** α represents the overall risk in the region of study, μ_i is a random effect specific to area **i** to model spatial dependence between the relative risks, and v_i is an unstructured exchangeable component that models uncorrelated noise, $v_i \sim N$ (0, σ^2_v) (Moraga, 2020).

After the general equation model is established above 'bym2' model will be used. From (Moraga, 2020) Besag-York-Mollié (BYM) is a popular spatial model in disease mapping applications (Besag et al., 1991). In this model, the spatial random effect μ_i is assigned a Conditional Autoregressive (CAR) distribution which smoothes the data according to a certain neighborhood structure that specifies that two areas are neighbors if they share a common boundary. In this case, 'bym2' is the development of BYM model above. Simpson et al. (2017) proposed a new parametrization of the BYM model called BYM2 which makes parameters interpretable and facilitates the assignment of meaningful Penalized Complexity (PC) priors. The BYM2 model uses a scaled spatially structured component **u*** and an unstructured component **v***

$$\boldsymbol{b} = \frac{1}{\sqrt{\tau_b}}(\sqrt{1-\varphi}\boldsymbol{v}_* + \sqrt{\varphi}\boldsymbol{u}_* \qquad \textbf{(2)}$$

Here, the precision parameter $\tau_b > 0$ controls the marginal variance contribution of the weighted sum of **u*** and **v***. The mixing parameter $0 \leq \varphi \leq 1$ measures the proportion of the marginal variance explained by the structured effect **u***. Thus, the BYM2 model is equal to an only spatial model when $\varphi = 1$, and an only unstructured spatial noise when $\varphi = 0$ (Riebler et al., 2016). After the 'bym2' model used, the analysis of Bayesian inference next will be done by INLA. In the next sub-chapter Bayesian Inference with INLA will be explained.

Bayesian inference with INLA

The Integrated Nested Laplace Approximation (INLA) is a method that enables approximate Bayesian inference in latent Gaussian models, including generalized linear mixed models and spatial/spatio-temporal models. These models can be represented in the following form:

$$y_i|x,\theta \sim \pi(x_i,\theta), i = 1,..,n,$$
$$x|\theta \sim N(\mu(\theta), Q(\theta)^{-1}), \theta \sim \pi(\theta), \qquad \textbf{(3)}$$

In the given model, y represents the observed data, x represents a Gaussian field, and θ represents the hyperparameters. The mean of the Gaussian field is denoted as $\mu(\theta)$, and the precision matrix (the inverse of the covariance matrix) is represented as $Q(\theta)$. It is important to note that both $y$and x can have high dimensions. However, to ensure efficient and fast inferences, it is preferable to have a small dimension for the hyperparameter vector θ. This is because approximations are computed using numerical integration over the hyperparameter space (Moraga, 2020)(Djuraidah, 2021).

RESULTS AND DISCUSSION

Table 1 presents the values of the intercept (β_{int}) and the coefficients (β_i) for the explanatory variables in Equation **(1)**. It is found that the intercept and all explanatory variables have a not significant effect on the estimation of pneumonia disease mapping in each location, except for population and number of health center. This can be seen from the credibility interval value which does not contain 0.

Table 1. Estimator of Fixed Effect in Equation 1

Fixed Effect	Mean	Standard Deviation	Credibility Interval
$\boldsymbol{\beta}_{\boldsymbol{int}}$	1.28	0.88	(-0.45, 2.98)
$\boldsymbol{\beta}_{\boldsymbol{pop}}$	-0.837	0.33	(-1.49, -0.2)
$\boldsymbol{\beta}_{\boldsymbol{poor}}$	-0.01	0.3	(-0.6, 0.59)
$\boldsymbol{\beta}_{\boldsymbol{waste}}$	-1.08	0.76	(-2.59, 0.42)
$\boldsymbol{\beta}_{\boldsymbol{hc}}$	0.34	0.14	(0.06, 0.62)
$\boldsymbol{\beta}_{\boldsymbol{vehicle}}$	0.05	0.06	(-0.07, 0.17)
$\boldsymbol{\beta}_{\boldsymbol{hiv}}$	-0.17	0.17	(-0.5, 0.16)

Can be seen in Table 1, it shows the value of the effect seen from the mean which shows the magnitude of the β value, there are two variables that show a significant effect. namely the population variable and the number of health centers variable with the coefficient value of the population variable is -0.837 and the number of health centers variable is 0.34. If the population increases by one million, the number of pneumonia cases will decrease by -0.837, while if the number of health centers increases by 10 health centers, the number of pneumonia cases will increase by 0.34. Next, the relative risk of each region will be estimated and plotted using the *library* (leaflet). We get the summary of fitted values which contains a summary of the relative risk including the posterior mean, lower limit limit, upper limit limit. Specifically the mean column is the posterior mean. From the posterior mean that has been obtained, the value will be adjusted to the location in the shp file to be plotted according to the SIR and RR values of each location.

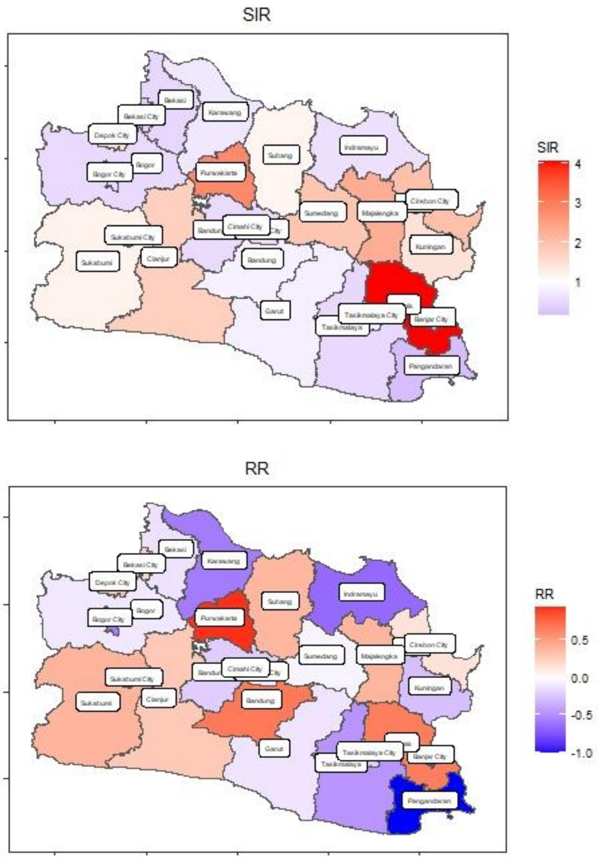
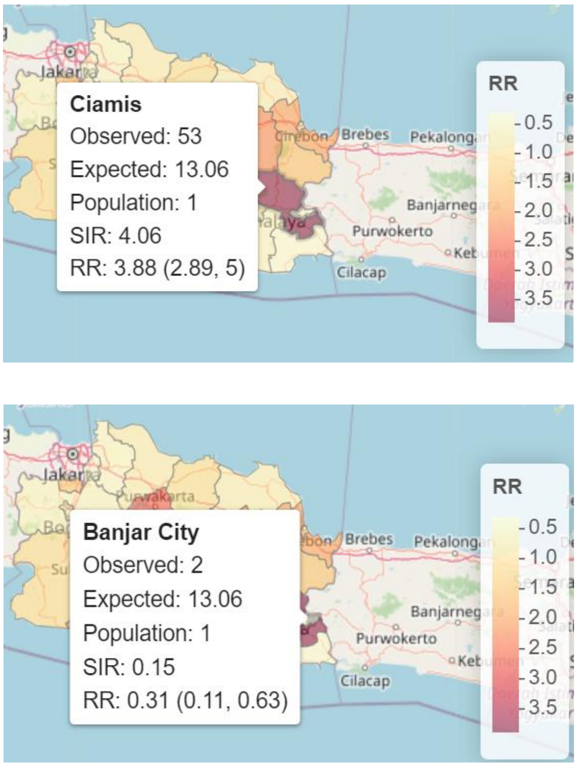


Figure 1. Plot RR and SIR

In general, the comparison of RR and SIR plots shows that the mapping of pneumonia in each region of West Java is quite similar, meaning that these two data support the hypothesis that certain areas have a much higher risk of being exposed to pneumonia than other areas. For more specific analysis, plots (*leaflet*) were used to provide more detailed information and data visualization of each region. The areas that show the most prominent distribution of pneumonia will be shown more easily in the plot image (*leaflet*) below.


Ciamis
Observed: 53
Expected: 13.06
Population: 1
SIR: 4.06
RR: 3.88 (2.89, 5)

Banjar City
Observed: 2
Expected: 13.06
Population: 1
SIR: 0.15
RR: 0.31 (0.11, 0.63)

Figure 2. Top: Ciamis leaflet Plot; Bottom: Banjar City leaflet Plot

For example, the Ciamis region has the highest risk of exposure to pneumonia among other regions at 3.88, meaning that the Ciamis region has a risk of exposure to pneumonia 3.88 times greater than other regions. In contrast, the city of Banjar has the smallest relative risk among other regions, which is 0.31, meaning that the city of Banjar has a 0.31 greater risk of pneumonia than other regions. It can be show in the figure 3. Dataset like this can be taken from each region and can be compared with each other. in this study these results are only limited to data exposure. for the study of the causes of higher or lower disease exposure specifically can be developed in other papers further with the data already in this paper.

CONCLUSIONS AND SUGGESTIONS

From this research, it can be concluded that this study aims to analyze the effect of differences in population size, the number of poor people, waste production, the number of health centers (Puskesmas), total motor vehicles, and the number of HIV patients in Municipalities/Districts in West Java on the spread of pneumonia. The analysis method used is the Integrated Nested Laplace Approximations (INLA) approach, a complex Bayesian statistical method used for estimation in hierarchical or nested Bayesian models. The results showed that the population and number of health centers variables had a significant influence on the mapping of pneumonia disease in each location. This can be seen from the confidence intervals that do not contain the value 0, indicating that an increase of one million population will reduce pneumonia cases by -0.837, while an increase of 10 health centers will increase pneumonia cases by 0.34. The Relative Risk (RR) and Standardized Incidence Ratio (SIR) maps show that some regions have a higher risk of pneumonia compared to other regions. For example, the Ciamis region has a 3.88 times higher risk of pneumonia than other regions, while Banjar City has a lower risk of 0.31.

These results are expected to provide insights for public policies in addressing health issues, especially in prevention and control efforts of diseases such as pneumonia. In addition, these results also serve as a basis for further research on other factors that may contribute to the spread of this disease at the local level. This study is expected to make a significant contribution to the understanding of the factors that influence the spread of pneumonia, helping in the formation of more effective health policies for the welfare of the community.

REFERENCES

Badan Pusat Statistik Jawa Barat (2021). Diakses 15 Desember 2023 dari https://jabar.bps.go.id/

Besag, J., York, J., and Mollié, A. (1991). *Bayesian image restoration with applications in spatial statistics (with discussion)*. Annals of the Institute of Statistical Mathematics, 43:1–59

Carolin, B. T., Suprihatin, A. M. P., & PK, A. M. (2020). Analisis *faktor risiko kejadian human immunodeficiency virus* (HIV) *pada lelaki seks lelaki* (LSL). *Jurnal Kebidanan*, *6*(2), 141-147.

Chakraborty, C., Sharma, A. R., Sharma, G., Bhattacharya, M., & Lee, S. S. (2020). *SARS-CoV-2 causing pneumonia-associated respiratory disorder (COVID-19)*: *diagnostic and proposed therapeutic options*. European Review for Medical & Pharmacological Sciences, 24(7).

Cillóniz, C., Menéndez, R., García-Vidal, C., Péricas, J. M., & Torres, A. (2020). *Defining community-acquired pneumonia as a public health threat: arguments in favor from Spanish investigators*. Medical Sciences, 8(1), 6. https://doi.org/10.3390/medsci8010006

Djuraidah, A., Rachmawati, R. N., Wigena, A. H., & Mangku, I. W. (2021). *Extreme data analysis using spatio-temporal bayes regression with INLA in statistical downscaling model*. International Journal of Innovative Computing, Information and Control, 17(1).

https://doi.org/10.24507/ijicic.17.01.259

Gajdács, M., Paulik, E., & Szabó, A. (2020). *Knowledge, attitude and practice of community pharmacists regarding antibiotic use and infectious diseases: a cross-sectional survey in Hungary (KAPPhA-HU)*. Antibiotics, 9(2), 41. https://doi.org/10.3390/antibiotics9020041

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. CRC Press.

Machmud, R. (2009). *Pengaruh kemiskinan keluarga pada kejadian pneumonia balita di Indonesia*. Kesmas: Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal), *4*(1), 36-41.

Moraga, P. (2020). *Geospatial Health Data Modeling and Visualization with R-INLA and Shiny*. New York: CRC Press.

Murugesan, B., Karuppannan, S., Mengistie, A. T., Ranganathan, M., & Gopalakrishnan, G. (2020). *Distribution and trend analysis of COVID-19 in India: geospatial approach.* Journal of Geographical Studies, 4(1), 1-9. http://dx.doi.org/10.21523/gcj5

Open Data Jabar (2021). Jumlah Kasus Pneomonia di Provinsi Jawa Barat Diakses 15 Desember 2023 dari https://opendata.jabarprov.go.id/id.

Purwanto, P., Astuti, I. S., Tanjung, A., Rohman, F., & Utomo, K. S. B. (2023). *Mapping the spatial transmission risk and public spatial awareness in the use of personal protective equipment: COVID-19 pandemic in East Java, Indonesia*. International Journal of Disaster Risk Reduction, 97, 104018. https://doi.org/10.1016/j.ijdrr.2023.104018

Renika, I., & Amin, C. (2021). *Identifikasi faktor risiko ekstrinsik pneumonia pada balita di pulau Jawa tahun 2018 menggunakan geographically weighted regression* (Doctoral dissertation, Universitas Muhammadiyah Surakarta).

Riebler, A., Sørbye, S. H., Simpson, D., and Rue, H. (2016*). An intuitive Bayesian spatial model for disease mapping that accounts for scaling*. Statistical Methods in Medical Research, 25(4):1145–1165

Satria, R., Aguero-Valverde, J., & Castro, M. (2021). *Spatial analysis of road crash frequency using Bayesian models with Integrated Nested Laplace Approximation (INLA)*. Journal of Transportation Safety & Security, 13(11), 1240-1262. https://doi.org/10.1080/19439962.2020.1726542

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). *Penalising model component complexity: A principled, practical approach to constructing priors*. Statistical Science, 32:1–28.

Suaya, J. A., Mendes, R. E., Sings, H. L., Arguedas, A., Reinert, R. R., Jodar, L., ... & Gessner, B. D. (2020). *Streptococcus pneumoniae serotype distribution and antimicrobial nonsusceptibility trends among adults with pneumonia in the United States, 2009–2017*. Journal of infection, 81(4), 557-566. https://doi.org/10.1016/j.jinf.2020.07.035

Sun, Y., Li, H., Pei, Z., Wang, S., Feng, J., Xu, L., ... & Zhan, S. (2020). *Incidence of community-acquired pneumonia in urban China: a national population-based study*. Vaccine, 38(52), 8362-

8370. https://doi.org/10.1016/j.vaccine.2020.11.004

Soh, S., Ho, S. H., Seah, A., Ong, J., Richards, D. R., Gaw, L. Y. F., ... & Lim, J. T. (2022). *Spatial Methods for Inferring Extremes in Dengue Outbreak Risk in Singapore*. Viruses, 14(11), 2450. https://doi.org/10.3390/v14112450

Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). *Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing*. Jama, 323(14), 1341-1342. doi:10.1001/jama.2020.3151

Wang, Y., Sun, K., Li, L., Lei, Y., Wu, S., Jiang, Y., ... & Yang, J. (2022). *The impacts of economic level and air pollution on public health at the micro and macro level.* Journal of Cleaner Production, 366, 132932. https://doi.org/10.1016/j.jclepro.2022.132932

Yatnaningtyas, R., Latra, I. N., & Andari, S. (2016). *Identifikasi faktor-faktor yang mempengaruhi pneumonia pada balita di surabaya menggunakan geographically weighted negative binomial regression.* Jurnal Sains dan Seni ITS, *5*(2). DOI: 10.12962/j23373520.v5i2.16597

Zhong, Q., & Peng, J. (2021). *Mean platelet volume/platelet count ratio predicts severe pneumonia of COVID-19*. Journal of clinical laboratory analysis, 35(1), e23607. https://doi.org/10.1002/jcla.23607