Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2022, 2022:102

https://doi.org/10.28919/cmbn/7709

ISSN: 2052-2541

A BAYESIAN CONDITIONAL AUTOREGRESSIVE WITH INLA: A CASE STUDY OF TUBERCULOSIS IN JAVA, INDONESIA

ANIK DJURAIDAH[1,*], ZAKIYAH MAR'AH[2], RAHMA ANISA[1]

[1]Department of Statistics, IPB University, Bogor, Indonesia

[2]Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

Abstract: Most disease risk mapping models are modeled by using the Bayesian approach. In this approach, the posterior distribution is typically estimated by using Markov Chain Monte Carlo (MCMC) method. However, it is known that the MCMC method suffers from convergence issues and requires a high demand for computational resources. This study aims to identify potential covariates in tuberculosis reduction and to map the relative risk of tuberculosis disease in the Java region by using Bayesian Conditional Autoregressive (CAR) with Integrated Nested Laplace Approximation (INLA). The posterior distribution of Bayes was estimated analytically using INLA that resulting in a more efficient, fast, and accurate inference. The results showed that Besag York Mollié (BYM) has the smallest Deviance Information Criterion (DIC) and Mean Absolute Deviance (MAD), so this model gives the best prediction. The significant covariates in reducing the number of TB cases in Java are percentages of healthy homes, household clean and healthy behavior, non-smokers, and complete tuberculosis treatment. In addition, 52.1% of districts/cities in Java were found to have a relative risk greater than one, and most of them were located in West Java.

Keywords: Bayesian conditional autoregressive; integrated nested Laplace approximation; mapping disease;

*Corresponding author

E-mail address: anikdjuraidah@apps.ipb.ac.id

Received September 01, 2022

tuberculosis.

2010 AMS Subject Classification: 62M30, 62H11.

1. INTRODUCTION

Several endemic diseases are present in Indonesia [1], and tuberculosis (TB) is one of the diseases that has been reported with high infection cases. According to the 2019 Global Report on Tuberculosis (TB), Indonesia ranks the third-highest of TB burden throughout the globe with a TB incidence rate of 391 per 100,000 population [2]. Three provinces of West Java, East Java, and Central Java were reported to have the highest cases, which accounts for 44 percent of all TB cases in Indonesia [3]. WHO reported that the treatment of TB requires stringent adherence and relies on support from the local health authority. The prevention and control efforts of TB cases also play important role in reducing the spread of the diseases.

Disease modeling and mapping can assist in disease prevention through the early detection of high-risk locations [4]. A statistical approach to modeling and mapping can be utilized by policymakers in formulating actions that can be taken to reduce the spread of disease [5]. Disease modeling of several endemic diseases in Indonesia has been carried out, such as malaria [6], [7], and [8], and dengue fever [9]. Aisyah *et al*. [10] developed the spatial model for TB in East Java with the percentage of proper sanitation and healthy food processing places as significant covariates. Although spatial models can be used to determine the significant covariates, the model is limited to generating the estimate of the relative risk of an area.

The Bayesian approach can be used to model the relative risk of disease in a given area. Lee [11] and Lawson [12] proposed the Bayesian hierarchical model. In this model, the relative risk of an area was expressed as a set of spatial random effects and potential covariates. The outcome of the Bayesian model includes random effects that could capture unobserved heterogeneity or spatial autocorrelation. The conditional autoregressive (CAR) is mostly used to model these random effects in order to obtain the relative risk of disease [11]. Variations of model formulations for random effects have been proposed in existing studies [13, 14, 15]. Besag *et al*.

[13] proposed an intrinsic conditional autoregressive model (ICAR) and Besag York Mollié (BYM) model. In the ICAR model, the spatial effect of a particular region depends on the effects of all adjacent areas. Whereas, the BYM model includes an additional unstructured spatial random effect to account for region-specific noise.

Markov Chain Monte Carlo (MCMC) algorithms, such as the Gibbs sampling or Metropolis-Hastings, are commonly used to sample the posterior distribution of CAR [16]. According to the MCMC inference theory, the parameter simulation values will converge to the posterior distribution if the number of iterations is carried out an infinite number of times. In other words, it is not known how fast the resulting sample converges to its posterior distribution [17, 18]. Another method for determining the posterior distribution using integrated nested Laplace approximation (INLA) was recently proposed by Rue *et.al*. [17]. INLA is based on the Gauss latent model, which produces more accurate estimators and a shorter computation time than MCMC [19, 20]. Ferkingstad *et al*. [21] reported that the computation time of INLA is 250 times faster than that of MCMC.

The goal of this study is to develop a risk mapping of TB diseases in district/city Java. We used the Bayesian CAR model with its posterior computation using INLA. The objectives of this study are to (1) identify the potential covariates that have a significant effect and (2) to map the relative risk of TB disease in Java.

2. BAYESIAN CONDITIONAL AUTOREGRESSIVE

Suppose the study area is expressed as $\boldsymbol{S}$ which consists of n non-overlapping unit areas $\boldsymbol{S} = \{S_1, \cdots, S_n\}$ and the corresponding response value vector is expressed as $\boldsymbol{y} = (y_1, \cdots, y_n)^T$. The spatial pattern of the responses was modeled by the covariate matrix $\boldsymbol{X} = (\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_n^T)$ and the random effect vector $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_n)$. The covariate vector at the location-k is expressed as $\boldsymbol{x}_k^T = (1, x_{1k}, \cdots, x_{pk})$. The first term of this expression corresponds to an intercept term. The regression parameters are expressed as $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)$, and O_k is offset [21]. The general formulation of bayesian CAR is an extension of a generalized linear model and is given by

$$Y_k|\mu_k \sim f(y_k|\mu_k, \upsilon^2) \quad \text{for } k = 1, \cdots, n$$
$$g(\mu_k) = \boldsymbol{X}_k^T \boldsymbol{\beta} + \phi_k + O_k \tag{1}$$

The response Y_k has exponential family distribution $f(y_k|\mu_k, \upsilon^2)$, with expected value $E(Y_k) = \mu_k$ and scale parameter υ^2. The expected values of the responses are related to the linear covariate via an invertible link function $g(.)$ [11, 21].

The prior distribution of each regression parameter β_i is $\beta_i \sim Normal\,(m_j\,,\, v_j)$ for $j = 0, \cdots, p$. The scale parameter υ^2 is assigned a uniform prior distribution, that is $\upsilon^2 \sim Uniform((0, M_\upsilon)$. The simplest prior of random effects is

$$\phi_k \sim Normal\,(0, \sigma^2) \text{ and } \sigma^2 \sim Uniform((0, M_\sigma) \tag{2}$$

This specification is appropriate if the covariates included in equation (1) have removed all of the spatial structure in the response.

CAR models commonly used for spatial autocorrelation modeling are Intrinsic CAR (ICAR) and BesagYork Mollié (BYM), both proposed by Besag *et. al.* [13], the Leroux model proposed by Leroux *et. al.* [14], and the Stern-Cressie model proposed by Stern and Cressie [15]. The prior of random effect is assumed to have a Gaussian Markov random field (GMRF) and can be written as $\boldsymbol{\phi} \sim Normal(\mathbf{0}, \tau^2 \boldsymbol{Q}^{-1})$ where $\boldsymbol{Q}$ is a precision matrix that may be singular (intrinsic model). This matrix controls the spatial autocorrelation structure of the random-effects and is based on a non-negative symmetric $n \times n$ neighborhood or weight matrix $\boldsymbol{W}$. This specification forces (ϕ_j, ϕ_k) relating to geographically adjacent areas (that is $w_{kj} = 1$) to be correlated, whereas random-effects relating to non-contiguous areal units are conditionally independent given the values of the remaining random-effects[21]. CAR priors are commonly specified as a set of n univariate full conditional distributions $f(\phi_k|\,\boldsymbol{\phi}_{-k})$ for $k = 1, \cdots, n$, where $\boldsymbol{\phi}_{-k} = (\phi_1, \cdots, \phi_{k-1}, \phi_{k+1}, \cdots, \phi_n)$.

The prior for random-effect of ICAR is given by

$$\phi_k|\boldsymbol{\phi}_{-k} \sim \text{Normal}\left(\frac{\sum_{j=1}^n w_{kj}\phi_j}{\sum_{j=1}^n w_{kj}}, \frac{\tau^2}{\sum_{j=1}^n w_{kj}},\right) \tag{3}$$

The conditional expectation is the average of the random effects in neighboring areas, while the

conditional variance is inversely proportional to the number of neighbors. The prior for variance parameters is uniform distribution, $\tau^2 \sim Uniform(0, M_\tau)$. The limitation of ICAR is that it can only represent strong spatial autocorrelation, and produce overly smooth of random [21].

BYM or convolution model is a combination of ICAR with one independent random component (unstructured spatial). In BYM, ϕ_k in (1) is replacing with $\theta_k + \phi_k$, which θ_k, is an unstructured random component with priors given by

$$\theta_k \sim \text{Normal}\,(0, \sigma_\theta^2) \text{ and } \sigma_\theta^2 \sim Uniform((0, M_\sigma) \quad (4)$$

Leroux *et al*. [14] and Stern and Cressie [15] proposed alternative CAR priors for modeling varying strengths of spatial autocorrelation, using only a single set of random effects. The model by [14] is given by

$$\phi_k | \boldsymbol{\phi}_{-k} \sim \text{Normal}\left(\frac{\rho \sum_{j=1}^{n} w_{kj}\phi_j}{\rho \sum_{j=1}^{n} w_{kj} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{j=1}^{n} w_{kj} + 1 - \rho}\right) \quad (5)$$

while model by [15] is

$$\phi_k | \boldsymbol{\phi}_{-k} \sim \text{Normal}\left(\frac{\rho \sum_{j=1}^{n} w_{kj}\phi_j}{\sum_{j=1}^{n} w_{kj} +}, \frac{\tau^2}{\sum_{j=1}^{n} w_{kj}}\right) \quad (6)$$

In Equation (5) and (6), ρ is a spatial autocorrelation parameter, with ρ = 0 corresponding to independence, while ρ = 1 corresponds to strong spatial autocorrelation. The prior for ρ and τ^2 are expressed as

$$\rho \sim Uniform\,(0,1) \quad \text{and} \quad \tau^2 \sim Uniform\,(0, M_\tau).$$

In both cases when $\rho = 1$ is obtained ICAR, while when $\rho = 0$ the only difference is the denominator in the conditional variance.

3. Application Of Bayesian Conditional Autoregressive With INLA

3.1. Data Source

The data used in this study is from the publication of Statistics Indonesia and the Health Office of Java in Indonesia. The number of TB cases treated and reported in 100,000 population in districts/cities in Java in 2019 (TBP) was set as the response variable. Five covariates were set, and

they are: i) percentage of households with access to proper sanitation (APS), (ii) percentage of healthy homes (HH), (iii) percentage of household clean and healthy behavior (CHB), (iv) percentage of non-smokers (NS), (v) percentage of complete treatment of Tuberculosis (CTT), (vi) number of public health centers per 100,000 population (PHC).

The motivation behind covariates of APS, HH, CHB, and NS was based on the study by [22]. Their study showed that a contaminated household environment, such as smoke inside the house, sharing of toilets with other households, and non-use of potable water, increased the risk of tuberculosis. The covariates of CTT and PHC were chosen following the research by [23]. Their results showed that in endemic countries, diagnosis and treatment of smear-positive cases remain the key to tuberculosis control by reducing transmission from infectious cases.

3.2. Analysis Method

The data analysis procedure in this study are listed below:

1. Exploring the data.
2. Calculating the Moran's Index for response variable [24].
3. Calculating the spatial weighting matrix ($\boldsymbol{W}$) using queen contiguity, exponential weight, and inverse distance weight [25].
    a. For the queen contiguity, the weighting is performed using the following formula:

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous} \\ 0 & \text{if } i \text{ and } j \text{ are not contiguous} \end{cases}$$

    The value of w_{ij} is the value of the element in the adjacency matrix corresponding for the i-th area and j-th area. The value of one is given if the i-th area is adjacent to the j-th area, while a value of null is given if the i-th area is not adjacent to the j-th area.

    b. Formula for the exponential weight is $w_{ij} = exp(-d_{ij})$, where d_{ij} is distance between the i-th area and the j-th area.

    c. Formula for the inverse distance weight is $w_{ij} = {d_{ij}}^{-1}$.

4. Estimating of the Bayesian CAR model using INLA with following steps:

i. Determinating the following model:
   a. ICAR model without covariates (M1): $g(\mu_k) = \beta_0 + \phi_k$
   b. ICAR model with covariates (M2): $g(\mu_k) = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \beta_3 X_{3k} + \beta_4 X_{4k} + \phi_k$
   c. BYM model without covariates (M3): $g(\mu_k) = \beta_0 + \phi_k + \theta_k$
   d. BYM model with covariates (M4): $g(\mu_k) = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \beta_3 X_{3k} + \beta_4 X_{4k} + \phi_k + \theta_k$
   e. Leroux model without covariates (M5): $g(\mu_k) = \beta_0 + \phi_k$
   f. Leroux model with covariates (M6): $g(\mu_k) = \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \beta_3 X_{3k} + \beta_4 X_{4k} + \phi_k$

ii. Determinating the prior for hyperparamer of M1and M2:
   a. Random effect: $\phi_k|\boldsymbol{\phi}_{-k}$~Normal $\left(\frac{\sum_{j=1}^{n} w_{kj}\phi_j}{\sum_{j=1}^{n} w_{kj}}, \frac{\tau^2}{\sum_{j=1}^{n} w_{kj}}, \right)$
   b. Variance ϕ: $\log \tau_\phi \sim \log gamma\ (1, 0.0005)$

iii. Determinating the prior for hyperparamer of M3 and M4:
   a. Random effect: $\phi_k|\boldsymbol{\phi}_{-k}$~Normal $\left(\frac{\sum_{j=1}^{n} w_{kj}\phi_j}{\sum_{j=1}^{n} w_{kj}}, \frac{\tau^2}{\sum_{j=1}^{n} w_{kj}}, \right)$
   b. Variance ϕ: $\log \tau_\phi \sim \log gamma\ (1, 0.0005)$
   c. Unstructured random-effects: θ_k~Normal $(0, \sigma_\theta^2)$
   d. Variance θ: $\log \tau_\theta \sim \log gamma\ (1, 0.0005)$

iv. Determinating the prior for hyperparamer of M5 and M6:
   a. Random effect: $\phi_k|\boldsymbol{\phi}_{-k}$~~Normal $\left(\frac{\rho \sum_{j=1}^{n} w_{kj}\phi_j}{\rho \sum_{j=1}^{n} w_{kj}+1-\rho}, \frac{\tau^2}{\rho \sum_{j=1}^{n} w_{kj}+1-\rho}\right)$
   b. Variance ϕ: $\log \tau_\phi \sim \log gamma\ (1, 0.0005)$
   c. spatial autocorrelation: $\rho \sim Uniform\ (0,1)$

v. Estimating the parameter of M1-M6 using INLA

vi. Identifying the best model based on the following criteria:
   a. Deviance Information Criterion/DIC calculated by the following formula:

$$DIC = 2\bar{D}(\boldsymbol{y},\boldsymbol{\theta}) - D(\boldsymbol{y},\bar{\boldsymbol{\theta}}) \text{ or } DIC = D(\boldsymbol{y},\bar{\boldsymbol{\theta}}) + 2pD$$

where $D(\boldsymbol{y},\boldsymbol{\theta}) = -2\log p(\boldsymbol{y}|\boldsymbol{\theta})$ denotes the deviance for data $\boldsymbol{y}$ and model parameters $\boldsymbol{\theta}$, $D_{\bar{\boldsymbol{\theta}}}(\boldsymbol{y}) = D(\boldsymbol{y},\bar{\boldsymbol{\theta}})$ denotes the deviance for posterior mean $\bar{\boldsymbol{\theta}}$, $\bar{D}(\boldsymbol{y},\boldsymbol{\theta}) = E(D(\boldsymbol{y},\boldsymbol{\theta})|\boldsymbol{y})$ denotes the posterior mean deviance, and $pD = \bar{D}(\boldsymbol{y},\boldsymbol{\theta}) - D(\boldsymbol{y},\bar{\boldsymbol{\theta}})$ denotes the estimate of effective number of parameters [26]

b. Mean Absolute Deviance (MAD) calculated by the following formula:

$$\text{MAD} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}$$

where y_i denotes the response data at i-th area, $\hat{y}_i$ denotes the prediction at i-th area and n is the number of the observation area.

6. Mapping the relative risk obtained the best model at districts/cities in Java.

3.3. Data Exploration

The distribution of TBP in the Java region is presented in Figure 1. It was observed that the number of TBP in Central Java is higher than that in East Java. For example, in Central Java, the districts of Banyumas and Cilacap have a high number of TBP, with 253 and 272 TBP per 100,000 population, respectively. Pacitan and Trenggalek in East Java, on the other hand, have lower TBP of 65 and 79 per 100,000 population, respectively.

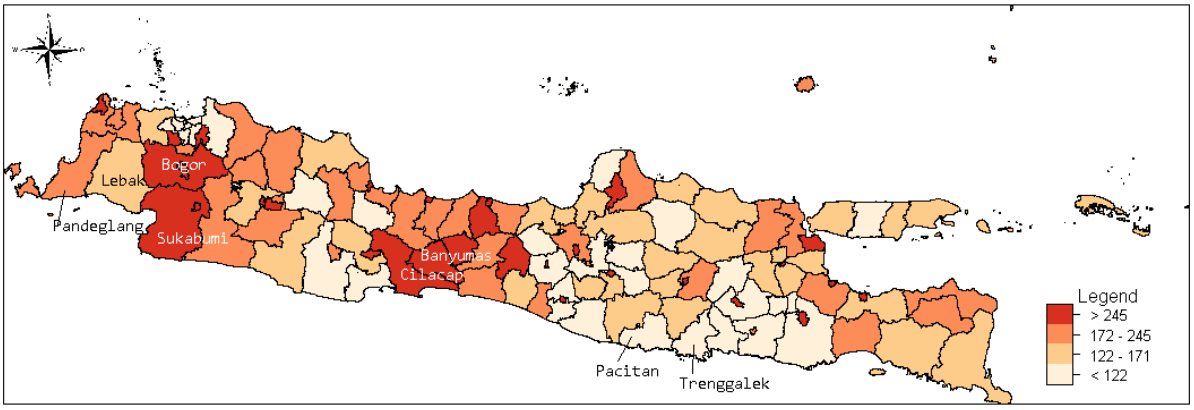


FIGURE 1. Number of TBP in 100,000 residents in district/city of Java in 2019 (non-scale map) indicated by the colormap. The darker region indicates high TBP and vice versa.

The result in Figure 1 also indicates a color grouping. The neighboring cities of a location that has a high number of TBP tend to have high TBP as well. For example, Bogor with 345 cases in 100,000 population, and its neighbor, Sukabumi, with 554 cases in 100,000 population. This pattern is also seen in groups of cities with a low number of TBP. Pacitan, for example, has 66 cases per 100,000 people, while Trenggalek has 80 cases per 100,000 people. This observation implies that there may be a spatial dependence in the data of TBP in Java.

A spatial dependence can be described by calculating Moran's index. The values of Moran's Index of TBP are summarized in Table 1. Moran's Index with Queen Contiguity is significant at a 5% significant level, thus the application is relevant for the weight in the CAR models (M1-M6).

TABLE 1 Moran's Index of TBP in Java in 2019

Weighted matrix	Moran's Index	E(I)	Var(I)	p-value
Queen Contiguity	0.18786	-0.00855	0.00362	0.0005
Exponential Distance	-0.00181	-0.00847	0.00018	0.3
Invers Distance	0.03442	-0.00847	0.0022	0.2

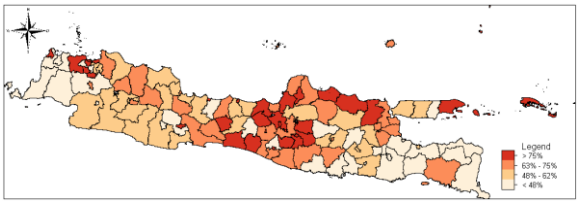
In order to improve the model, the observed variable of TBP was transformed using logarithmic transformation. Table 2 shows the correlation coefficient between variables. The result shows that covariates CHB and CTT have the highest correlation with ln(TBP), with correlation coefficients of 0.278 and 0.245, respectively. APS and PHC, on the other hand, have the lowest correlation coefficients of 0.004 and 0.076, respectively. Therefore, HH, CHB, NS, and CTT were selected as the covariates used in the CAR models. Furthermore, Table 2 indicates that there was no multicollinearity among covariates because the correlations between covariates are small.

TABLE 2. Correlation coefficient of variables used

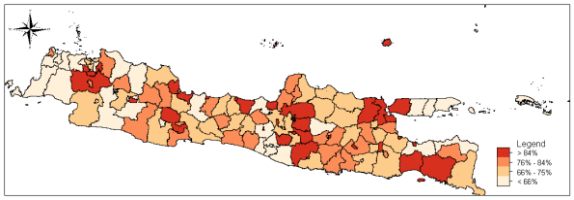
	ln(TBP)	APS	HH	CHB	NS	CTT	PHC
ln(TBP)	1.000	0.004	0.194	0.278	0.131	0.245	0.076
APS	0.004	1.000	0.212	0.330	0.393	0.147	-0.098
HH	0.194	0.212	1.000	0.300	0.033	-0.162	0.208
CHB	0.278	0.330	0.300	1.000	0.261	0.048	-0.060
NS	0.131	0.393	0.033	0.261	1.000	0.026	-0.214
CTT	0.245	0.147	-0.162	0.048	0.026	1.000	0.039
PHC	0.076	-0.098	0.208	-0.060	-0.214	0.039	1.000

Figure 2 shows the maps of the covariate data used in this study. The map of HH in Figure 2(a) shows that the majority of districts/cities have an HH of less than 62 %. There were only a few cities with HH greater than 75 %. The CHB map in Figure 2(b) indicates that only a few districts/cities have a CHB greater than 80 %. It could also be seen that the CHB values in West Java are low, which is the opposite of the observation in East Java.

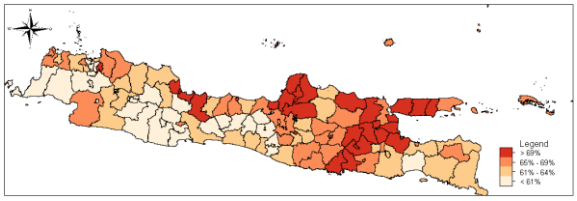
The NS map in Figure 2(c) shows that East Java has the highest percentage of NS. The lowest percentage, on the other hand, was found in half of West Java and half of Central Java. When we compare the TB map (Figure 1) to the NS map, we can see that the trends are opposite. The CTT map in Figure 2(d) shows that districts/cities in West Java have the highest CTT. This map is similar to the Java TBP map (Figure 1), which shows that West Java has a higher proportion of TBP than the rest of Java.



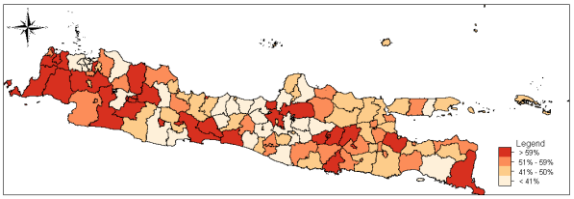
(a) HH



(b) CHB



(c) NS



(d) CTT

FIGURE 2. Map of covariates ((a) HH, (b) CHB, (c) NS, (d) CTT) used in the study (non-scale map), indicated by the colormap. The darker region indicates the high value of covariate and vice versa.

3.4. The Estimating of Bayes CAR

CAR models without covariates (M1, M3, M5) and with covariates (M2, M4, M6) were used in this study. CAR with covariates was used to measure the effect of including a covariate. Table 3 and Table 4 shows parameter estimates and credible intervals for CAR without and with covariates, respectively. The intercept estimates in the ICAR (M1) and BYM (M3), are similar because both models have the same basic model. For CAR models with covariates, it was observed

that covariant of HH, CHB, and CT are the significant parameters for ICAR (M2) and CAR-Leroux (M6). Whereas, the most significant parameters for BYM (M4) are HH, CHB, NS, and CTT.

TABLE 3. Mean and 95% credible interval of intercept of CAR without covariates

Notation	Model	Intercept
M1	ICAR	5.12 (5.11,5.14)
M3	BYM	5.12 (5.04,5.21)
M5	CAR Leroux	2.56 (1.56,3.56)

TABLE 4. Mean and 95% credible interval of CAR with covariates

Notation	Model	Intercept	HH	CHB	NS	CTT
M2	ICAR	2.829 (1.252,4.408)	**0.007** (0,0.007)	**0.01** (0.004,0.016)	0.012 (-0.006,0.029)	**0.007** (0.001,0.013)
M4	BYM	1.757 (-0.209,3.712)	**0.008** (0.001,0.016)	**0.008** (0.001,0.015)	**0.028** (0,0.057)	**0.009** (0.003,0.016)
M6	CAR Leroux	1.354 (0.216,2.491)	**0.008** (0.002,0.014)	**0.01** (0.004,0.017)	0.013 (-0.002,0.029)	**0.006** (0.001,0.012)

The DIC and MAD values for each model in Table 5 are nearly identical, ranging from 1067-1087 and 1.16-2.41, respectively. Among the models with covariates, CAR BYM (M4) has the lowest DIC and MAD values of 1068.88 and 1.20, respectively. Similar results occurred in the model without covariates, with CAR BYM (M3) having the lowest DIC of 1067.56 and MAD of 1.16. This finding implies that the CAR BYM (M4) model is the best model for TB disease in Java in 2019. The spatially structured random effect proportion was at 34%, indicating that the spatially structured random effect improves prediction accuracy.

TABLE 5. The DIC and MAD of Bayes CAR

Notation	Model	DIC	MAD
M1	ICAR without covariates	1087.96	2.41
M2	ICAR with covariates	1069.89	1.45
M3	CAR BYM without covariates	1067.56	1.16
M4	CAR BYM with covariates	1068.03	1.20
M5	CAR Leroux without covariates	1068.88	1.51
M6	CAR Leroux with covariates	1069.19	1.54

3.5. Risk Mapping of TB

Disease mapping is used to determine an area's relative risk (RR). An RR value higher than one indicates a higher risk of disease, whereas an RR value less than one indicates a healthy area. Figure 3 shows the BYM (M4) model's relative risk in districts/cities in Java. There is 52.1 percent of the areas are considered high-risk. West Java has the most districts/cities with high RR values when compared to other provinces in Java. West Java has 21.85 percent high-risk districts/cities, while East and Center Java has 14.29 percent and 15.97 percent high-risk districts/cities, respectively. Figure 1 shows that Lebak and Pandeglang, with 144 cases and 224 cases in 10.000 population, were classified as low and medium TBP areas, respectively, but Figure 3 shows that Lebak and Pandeglang have high RR values of 1.86 and 1.85, respectively. This occurred due to the closeness of those areas to Bogor and Sukabumi, both of which are classified as high-risk cities. While Pacitan and Trenggalek are considered low-risk districts.

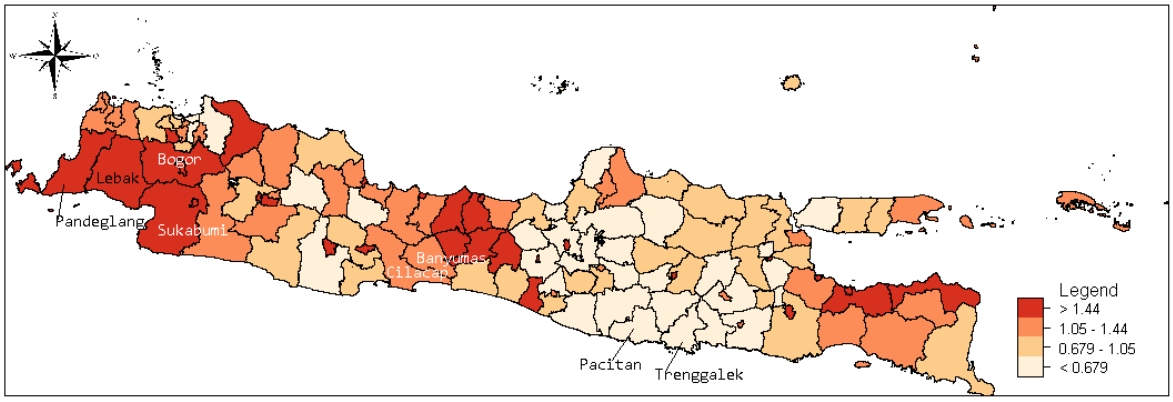


FIGURE 3. Relative risk map of TBP in Java in 2019 (non scale map) with colormap. The darker colors indicating a higher risk, and vice versa

4. CONCLUSION

The Besag York Mollié (BYM) with covariate model provides the most accurate prediction of TBP in Java in 2019. The percentage of non-smokers as a covariate was an important factor in the reduction of tuberculosis in Java. The percentage of districts/cities in Java classified as high-risk was 52.1 percent, with the majority of them being in West Java. This relative risk map of TBP can be used as a reference for residents as well as by local governments to handle the diseases, especially in high-risk areas.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] M. Nareza, Macam-macam Penyakit Endemik di Indonesia, 2020. https://www.alodokter.com/macam-macam-penyakit-endemik-di-indonesia.

[2] World Health Organization, Global Tuberculosis Report 2019, World Health Organization, Geneva, France, 2019. https://www.who.int/publications/i/item/9789241565714.

[3] Kementrian Kesehatan, Data dan Informasi: Profil Kesehatan Indonesia 2016, Jakarta, Kementrian Kesehatan RI. 2017.

[4] Y. Liu, S.C. Watson, J.R. Gettings, et al. A Bayesian spatio-temporal model for forecasting Anaplasma species seroprevalence in domestic dogs within the contiguous United States, PLoS ONE. 12 (2017), e0182028. https://doi.org/10.1371/journal.pone.0182028.

[5] D. Lee, A. Lawson, Cluster detection and risk estimation for spatio-temporal health data, arXiv:1408.1191. (2014). https://doi.org/10.48550/ARXIV.1408.1191.

[6] M.N. Siswanto, A. Djuraidah, Conditional autoregressive (CAR) modeling uses weighted matrix to first and second order (Case study: Malaria disease in Papua Province), Int. J. Eng. Manage. Res. 7 (2018), 297-301.

[7] S. Sukmawati, A. Djuraidah, A.H. Wigena, Spatial clustered regression analysis of 2017 Getis score Indonesian malaria prevalence data, J. Phys.: Conf. Ser. 1863 (2021) 012043. https://doi.org/10.1088/1742-6596/1863/1/012043.

[8] A. Djuraidah, P. Silvianti, B. Djaafara, et al. Modeling annual parasite incidence of malaria in Indonesia of 2017 using spatial regime, Indones. J. Geogr. 53 (2021), 185-191.

[9] D.R. Zulfanita, A. Saefuddin, A. Djuraidah, Modelling the number of cases of dengue hemorragic fever with mixed geographically negative binomial regression in West Java Province, Int. J. Sci. Res. Sci. Eng. Technol. 6 (2019), 71-77.

[10] F. Aisyah, A. Djuraidah, A. Alamudi, Modeling the number of tuberculosis patients in East Java with geographically weighted negative binomial regression, Theses, IPB University, Bogor, 2021. https://repository.ipb.ac.id/handle/123456789/108163.

[11] D. Lee, A comparison of conditional autoregressive models used in Bayesian disease mapping, Spat. Spatio-temporal Epidemiol. 2 (2011), 79–89. https://doi.org/10.1016/j.sste.2011.03.001.

[12] A.B. Lawson. Bayesian disease mapping: hierarchical modeling in spatial epidemiology, 3nd Edition, Chapman and Hall/CRC Press, Boca Raton, 2018.

[13] J. Besag, J. York, A. Mollie, Bayesian image restoration, with two applications in spatial statistics, Ann. Inst. Stat. Math. 43 (1991), 1–20. https://doi.org/10.1007/bf00116466.

[14] B.G. Leroux, X. Lei, N. Breslow, Estimation of disease rates in small areas: a new mixed model for spatial dependence, in: M.E. Halloran, D. Berry (Eds.), Statistical Models in Epidemiology, the Environment, and Clinical Trials, Springer New York, New York, NY, 2000: pp. 179–191. https://doi.org/10.1007/978-1-4612-1284-3_4.

[15] H. Stern, N. Cressie, Inference for extremes in disease mapping, In: A. Lawson, A. Biggeri, D. Böhning, et al. (eds.), Disease Mapping and Risk Assessment for Public Health, John Wiley & Sons, Chichester. (1999), 63–84.

[16] G. Onicescu, A.B. Lawson, Bayesian modeling and inference, In: A.B. Lawson, S.B. Robert, P. Haining, et al. (eds.), Handbook of Spatial Epidemiology, Chapman & Hall, Boca Raton, (2016), 133-158.

[17] H. Rue, S. Martino, N. Chopin, Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, J. R. Stat. Soc.: Series B (Stat. Methodol.) 71 (2009), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

[18] A. Riebler, S.H. Sørbye, D. Simpson, et al. An intuitive Bayesian spatial model for disease mapping that accounts for scaling, Stat. Methods Med. Res. 25 (2016), 1145–1165. https://doi.org/10.1177/0962280216660421.

[19] T.G. Martins, D. Simpson, F. Lindgren, et al. Bayesian computing with INLA: New features, Comput. Stat. Data Anal. 67 (2013), 68–83. https://doi.org/10.1016/j.csda.2013.04.014.

[20] F. Lindgren, H. Rue, Bayesian spatial modelling with R - INLA, J. Stat. Soft. 63 (2015), 1–25. https://doi.org/10.18637/jss.v063.i19.

[21] E. Ferkingstad, L. Held, H. Rue. Fast and accurate Bayesian model criticism and conflict diagnostics using R-INLA. Stat. 6 (2017), 331-334.

[21] D. Lee, CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors, J. Stat. Soft. 55 (2013), 1-24. https://doi.org/10.18637/jss.v055.i13.

[22] S.K. Singh, G.C. Kashyap, P. Puri, Potential effect of household environment on prevalence of tuberculosis in India: evidence from the recent round of a cross-sectional survey, BMC Pulm. Med. 18 (2018), 66. https://doi.org/10.1186/s12890-018-0627-3.

[23] P. Narasimhan, J. Wood, C.R. MacIntyre, D. Mathai, Risk Factors for Tuberculosis, Pulm. Med. 2013 (2013), 828939. https://doi.org/10.1155/2013/828939.

[24] L. Anselin, Spatial econometrics: Methods and models, Kluwer Academic Publisher, Dordrecht, 1988.

[25] S. Shekhar, H. Xiong, eds., Encyclopedia of GIS, Springer, Boston, MA, 2008. https://doi.org/10.1007/978-0-387-35973-1.

[26] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, et al. Bayesian measures of model complexity and fit, J. R. Stat. Soc. B. 64 (2002), 583–639. https://doi.org/10.1111/1467-9868.00353.