Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2020, 2020:89

https://doi.org/10.28919/cmbn/5061

ISSN: 2052-2541

ELEVATION AS A PROXY FOR DENGUE DISEASE TRANSMISSION IN INDONESIA

I. GEDE NYOMAN MINDRA JAYA[1,2,*], YUDHIE ANDRIYANA[1], BERTHO TANTULAR[1]

[1]Department of Statistics, Padjadjaran University, Bandung, Indonesia

[2]Faculty of Spatial Science, Groningen University, Groningen, Netherlands

Abstract: Dengue is a severe health problem. In recent decades, the number of infections has increased significantly worldwide, and approximately 70% of cases were found in Asia. To predict local mosquito distribution and estimate the local risk, a combination of ecological factors has been widely used. However, it is not easy to control the dengue transmission because the transmission is influenced by several ecological factors that might have complex interactions among the factors. In addition, ecological information is commonly limited. The elevation is an important weather proxy for *the Aedes Aegypti* life cycle because it is correlated to a variety of vital ecological factors, especially temperature. This study evaluated the effect of the elevation and identified high-low dengue risk provinces in Indonesia using the clustering Bayesian spatial hierarchical model. We found that the elevation's effect decreased for the provinces with an elevation higher than 289.44m.

Keyword: Bayesian; clustering; dengue disease; elevation; Indonesia.

2010 AMS Subject Classification: 91D20.

1. INTRODUCTION

Dengue is a mosquito-borne viral infection. In recent decades, the number of infections has increased significantly cause serious health problems. There were 390 million dengue virus

*Corresponding author

E-mail address: mindra@unpad.ac.id

Received September 26, 2020

infections per year in 129 countries, with 70% of cases found in Asia [1, 2]. The South-East Asia and Western Pacific regions are the most seriously affected. Indonesia is one of the countries in South-East Asia, with high dengue cases every year [1].

The dengue virus is transmitted through the bites of infected female *Aedes Aegypti* mosquitoes. It is found in many urban areas and mostly in human-made containers. It feeds mainly early in the morning and in the evening before sunset [3].

Dengue disease control is very difficult because there is no accurate information about the mosquito vector distribution [4, 5]. To predict local mosquito distribution and estimate the local risk, a combination of ecological factors have been used often. However, the life cycle of being influenced by the complex interaction between environmental factors and the weather variables' lack of information makes it challenging to develop a prediction model [5, 6].

The elevation is a vital weather proxy for the *Aedes Aegypti* life cycle because it is correlated to some crucial ecological factors, especially temperature [7]. Although elevation does not directly affect dengue virus transmission, it can be an essential proxy in assessing the environmental variable's effect with complex interactions in describing the *Aedes Aegypti* mosquito breeding [5]. Elevation has been used as a proxy for mosquito-borne Zika virus transmission in the Americas [5]. In this study, we use elevation as a proxy for mosquito-borne dengue virus transmission in Indonesia.

Besides the complexity of environmental factors' interactions, spatial dependency, and heterogeneity need more attention to disease modeling to obtain more reliable risk prediction [8-11]. In this study, we develop a clustering Bayesian hierarchical model based on Poisson log-linear model. The model is developed by considering spatial dependency, heterogeneity, and elevation variation across areas. The model is applied to model dengue incidences in Indonesia with 34 provinces.

The paper is structured as follows section 2 describes the method, section 3 figure out the application and discussion, and section 4 focuses on the conclusion and discusses the conclusion.

2. MATERIAL AND METHOD

2.1. Global Moran's I

Global Moran's I is widely used to provide information on the spatial dependency between neighboring areas [12]. Let we assume the regions consist of non-overlapping areas (e.g., provinces) with y_i is the number of dengue cases and N_i as the population at risk at spatial unit i for i=1,…,n. Moran's I statistic measures the spatial autocorrelation globally [12]:

$$I = \frac{1}{s_y^2} \frac{\sum_{i=1}^{n} \sum_{\{j:i \neq j\}} w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} \sum_{\{j:i \neq j\}} w_{ij}} \quad (1)$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, i and j were the region indexes and w_{ij} indicated the adjacency between district i and district j.

$$w_{ij} = \begin{cases} 1 & \text{if } i, j \text{ are adjacent neighbors} \\ 0 & \text{otherwise} \end{cases}$$

The adjacent neighbors in this study are defined by mean k-nearest neighbors. Moran's I has values between 0 – 1. If the Moran's I close to zero, it is indicated there is no spatial autocorrelation. Conversely, if the value close to one, it shows strong spatial autocorrelation.

2.2. Standardized morbidity ratio (SMR)

In spatial epidemiology, SMR is known as a crude risk because it is an unreliable estimate, especially for a small area. The areas with a small population are commonly categorized as high-risk areas; even the number of cases is small [8-11]. The SMR is expressed as the ratio between the number of cases (y_i) on the expected count (E_i) as follows [13]:

$$SMR_i = \frac{y_i}{E_i} \quad (2)$$

The expected number of cases E_i in each area is commonly calculated by considering the structure in population at risk. However, if the information was not provided, then $E_i = N_i(\sum_{i=1}^{n} y_i / \sum_{i=1}^{n} N_i)$ where n denotes the number of areas and N_i is population at risk in area ith. Although the SMR is not reliable for small areas, it helps understand the distribution patterns of various types of diseases. However, the inability of SMR to model covariates causes SMR to be fully applied in this study.

2.3. Poisson Loglinear model

To overcome the unreliable SMR and the inability to include the covariate in the model. We propose to use Poisson log linear model. The Poisson model is given by:

$$y_i|E_i,\theta_i \sim Poisson(E_i\theta_i);\ i=1,\dots,n \tag{3}$$

Poisson regression model fits for count data. Log linear models are widely used to explain the effects of the *K*-risk factors on the mean function. The model for *n* spatial units is given by:

$$\log \mathbb{E}(y_i) = \text{offset}(\log E_i) + \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i;\ i=1,\dots,n \tag{4}$$

where $\text{offset}(.)$ that describes the regression coefficient of $\log E_i$ is fixed to one, $\boldsymbol{x}_i = (1, \boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iK})'$is a $(K+1)\times 1$ is vector of area-level risk factors, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)'$ is $(K+1)\times 1$ is vector of regression parameters, and $\epsilon_i = \omega_i + v_i$ denotes the random effect components used to accommodate the spatial dependency (ω_i) and heterogeneity (v_i). In this study we consider elevation as the covariate and intrinsic conditional autoregressive (iCAR) and exchangeable priors to account spatial dependency (ω) and heterogeneity (v), respectively. iCAR model is defined follows Besag York and Molie model [14]:

$$\omega_i|\boldsymbol{\omega}_{-i},\tau_\omega,\mathbf{W} \sim \mathcal{N}\left(\frac{\sum_{j=1}^n w_{ij}\omega_j}{\sum_{i=1}^n w_{ij}}, \frac{\sigma_\omega^2}{\sum_{i=1}^n w_{ij}}\right) \tag{5}$$

where $\mathbf{W} = \left(w_{ij}\right)$ is the 'adjacency' matrix with $w_{ij}=1$ if i and j are adjacent (i.e. are first-order contiguous) and $w_{ij}=0$ otherwise. The spatial heterogeneity (v_i) is commonly assigned by an exchangeable prior, that is:

$$v_i|\sigma_v^2 \sim \mathcal{N}\left(0, \frac{1}{\sigma_v^2}\right) \tag{6}$$

where σ_v^2 is the variance parameter of v_i.

2.3. Clustering loglinear model approach

The ecological data are commonly high variation. Elevation data has high variation from 0m to 2163.19m, and the outlier problems potentially occur in disease modeling. To deal with the group outlier, we develop a clustering log-linear model by means dummy variable approach. We divide the elevation data based on quantile statistics and use this information as clusters. There are five candidate model that we consider.

Table 1. Candidate models

No	Model	Specification
M1	Based line model	$\eta_i = \alpha + \beta_1 x_i$
M2	Fixed 1	$\eta_i = \alpha + \beta_1 x_i + \sum_{j=1}^{J-1} \gamma_j D_{ji} + \sum_{j=1}^{J-1} \beta_j x_i D_{ji}$
M3	Fixed 2	$\eta_i = \alpha + \beta_1 x_i + \omega_i + v_i + \sum_{j=1}^{J-1} \gamma_j D_{ji} + \sum_{j=1}^{J-1} \delta_j x_i D_{ji}$
M4	Random 1	$\eta_i = \alpha + \beta_1 x_i + \gamma_j + \delta_j x_{i(j)}$
M5	Random 2	$\eta_i = \alpha + \beta_1 x_i + \omega_i + v_i + \gamma_j + \delta_j x_{i(j)}$

To evaluate which model is better for our data, we use two different model selection criteria those are deviance information criterion (DIC) [15] and Watanabe Akaike information criterion (WAIC) [16]. The statistical model presented in Table 1 is fitted using the Integrated Nested Laplace Approximation (INLA). To identify the hotspot clusters, we apply Bayesian exceedance probability [11]. The models were estimated using R-INLA packages.

3. RESULT

Study area and data. Indonesia is located in southeast Asia. It has a total land area of 1,904,569 km^2 [17]. Indonesia country can be seen on the Figure 1. The provinces names are corresponding to the id in the map and Table 1.

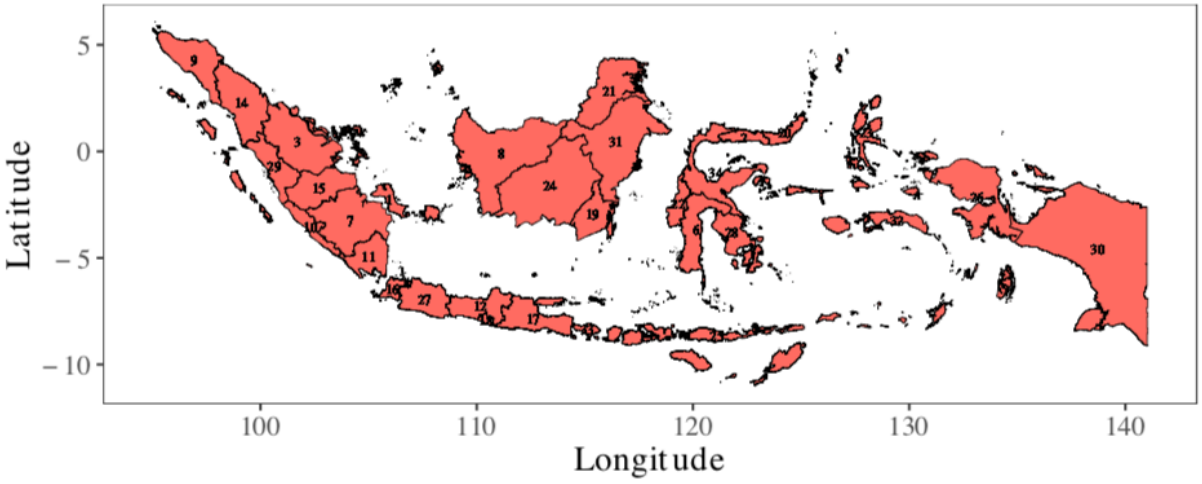


Figure 1. Indonesia country with 34 provinces.

Population projection by the Central Bureau of Statistics in 2019 puts Indonesia's population at 268,026,566 [18], which spread in 34 provinces. The dengue disease data used for this study recorded in 2019. Dengue disease Cases were accessed from https://www.kemkes.go.id/. The elevation data we extracted from https://www.worldclim.org/.

Table 2. Raw data for the analysis

id	Province	Population	Cases	Elevation (m)
1	Bangka Beliting	1,488,792	1,012	38.37
2	Gorontalo	1,202,631	1,221	155.05
3	Riau	6,971,745	4,126	37.15
4	DKI Jakarta	10,557,810	8,705	10.93
5	Kepulauan Riau	2,189,653	1,865	38.76
6	Sulawesi Selatan	8,851,240	3,265	502.28
7	Sumater Selatan	8,470,683	2,799	13.40
8	Kalimantan Barat	5,069,127	2,798	42.35
9	Aceh	5,371,532	2,386	2,163.19
10	Bengkulu	1,991,838	1,479	356.23
11	Lampung	8,447,737	5,611	49.08
12	Jawa Tengah	34,718,204	9,124	673.10
13	Nusa Tenggara Barat	5,070,386	2,971	0
14	Sumatera Utara	14,562,549	7,731	1,202.55
15	Jambi	3,624,579	2,158	49.81
16	Banten	12,927,316	2,915	61.61
17	Jawa Timur	39,698,631	18,031	1,299.16
18	Yogyakarta	3,842,932	3,301	268.03
19	Kalimantan Selatan	4,244,096	2,381	310.84
20	Sulawesi Utara	2,506,981	2,381	767.24
21	Kalimantan Utara	724,245	1,774	488.58
22	Sulawesi Barat	1,380,256	796	256.36
23	Maluku Utara	1,255,771	1,195	662.94
24	Kalimantan Tengah	2,714,859	1,616	75.36
25	Nusa Tenggara Timur	5,456,203	4,059	490.66
26	Papua Barat	959,617	354	22.68
27	Jawa Barat	49,316,712	23,483	706.30
28	Sulawesi Tenggara	2,704,737	1,479	541.10
29	Sumater Barat	5,411,197	2,263	622.05
30	Papua	3,379,302	597	619.36
31	Kalimantan Timur	3,721,389	6,723	65.13
32	Maluku	1,802,870	236	1,915.12
33	Bali	4,336,923	4,979	1,201.87
34	Sulawesi Tengah	3,054,023	1,947	0

Figure 3 shows the raster data from the worldclim database and Table 2 shows the statistics of the number of Cases, population size, population density, and the aggregated of elevation (m).

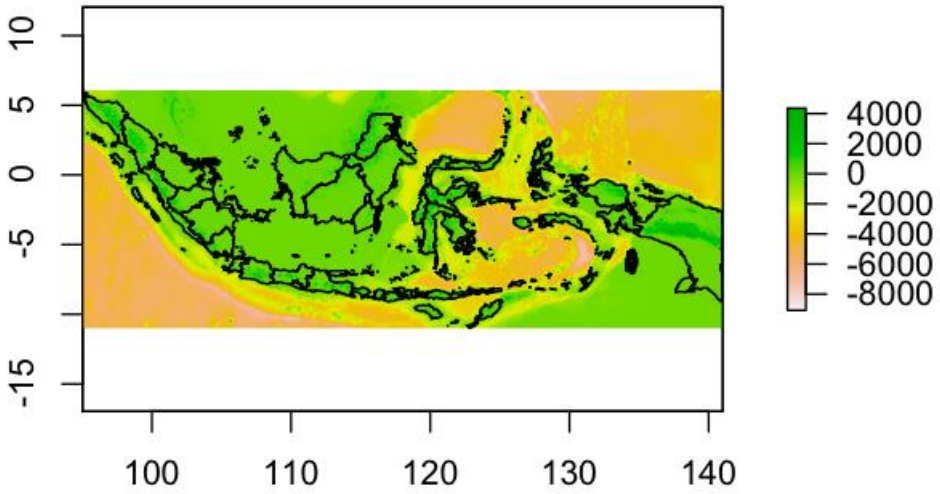


Figure 2. The raster of elevation data of 34 provinces in Indonesia

Table 3. The statistics of the number of Cases, population size, population density, and the aggregated of elevation

Variable	Minimum	Q(0.25)	Q(0.50)	Q(0.75)	Maximum
Cases	236	1513	2384	4109	23483
Population	724,245	2,268,985	4,043,514	8,078,739	49,316,712
Elevation (m)	0	44.03	289.44	652.71	2163.19

The statistics of each variable shows that the data have high variability and indicates an outlier. The spatial distribution of all variables is presented below.

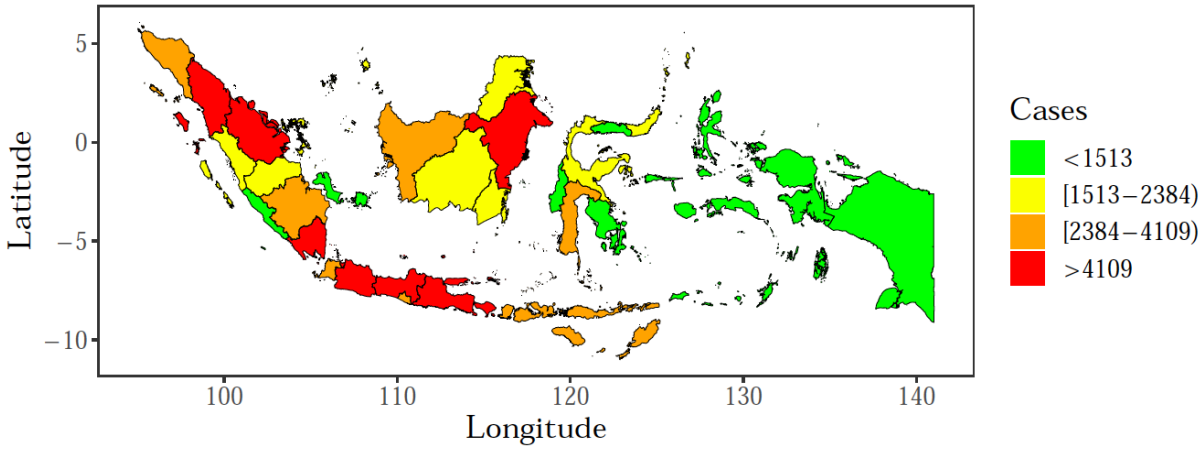


Figure 3. Spatial distribution of the number of dengue disease cases

Figure 3 shows the spatial distribution of dengue disease over 34 provinces in 2019. The high cases were found in most provinces in Java and Sumatera island and north of Kalimantan. High cases were related to high population areas, as presented in Figure 3. The number of cases highly correlates to the population size with r-Pearson=0.864, and it has a similar spatial pattern. It indicates that the areas with a high number of cases do not describe the high-risk areas. The high cases were recorded caused by the high population. In this study, we focus on estimating the high-risk areas as a function of elevation.

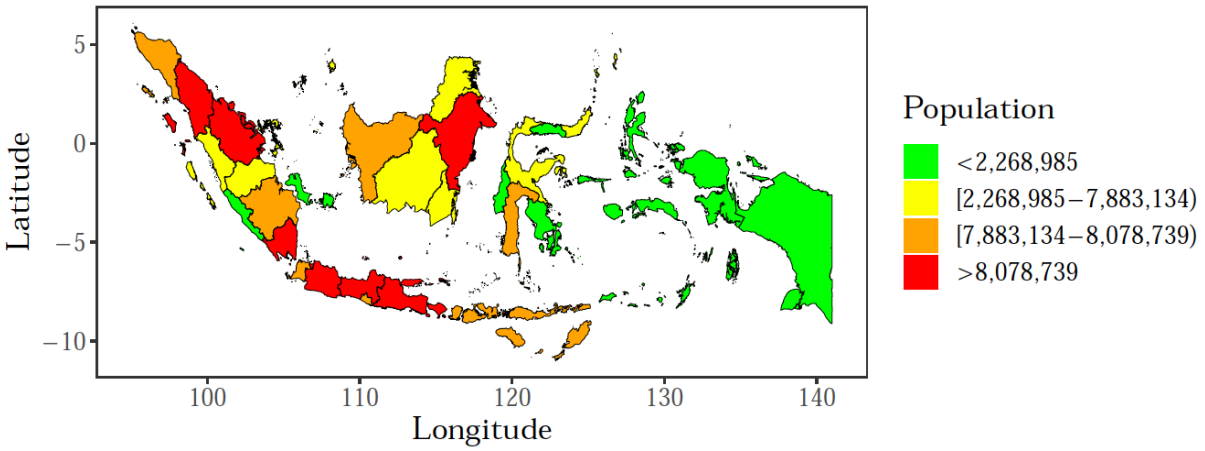


Figure 4. Spatial distribution of population

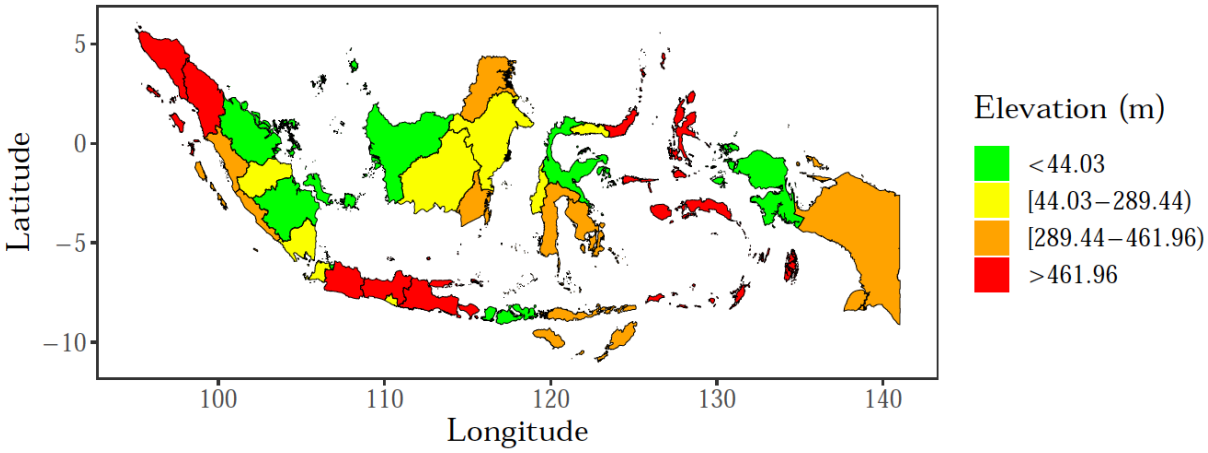
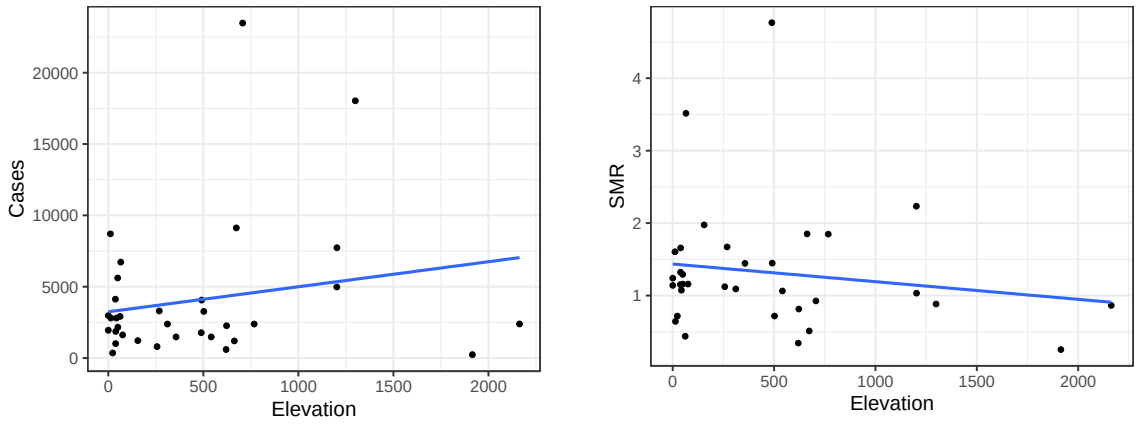


Figure 5. Spatial distribution of elevation (m)

Elevation data were extracted from the worldclim database using a raster package in R. The elevations were extracted corresponding to the centroid coordinates for each area. The elevation data were high variation (see Table 3). To model the dengue disease, we have to control the population size because it can be a confounding factor. Figure 6(a) shows there is a positive correlation between the number of cases with elevation. It is caused by the population at risk does

not consider in computing the correlation index. Figure 6(b) shows the correlation between the elevation with crude risk. It shows that there is a negative correlation between both variables.



(a) Elevation versus Cases (b) Elevation versus crude risk (SMR)

Figure 6 (a) Elevation versus dengue cases and (b) elevation versus crud risk

As we mentioned above, the elevation has a high variation with some outliers. There are several solutions to overcome this high variation. First, we can use a nonlinear regression model, and second, we can develop clustering regression. Clustering regression can be extended by developing regression dummy variables. Here we focus on clustering regression. The clusters are constructed based on the quartile of elevation data. The exploration of the relationship between crude risk and elevation is shown in Figure 7.

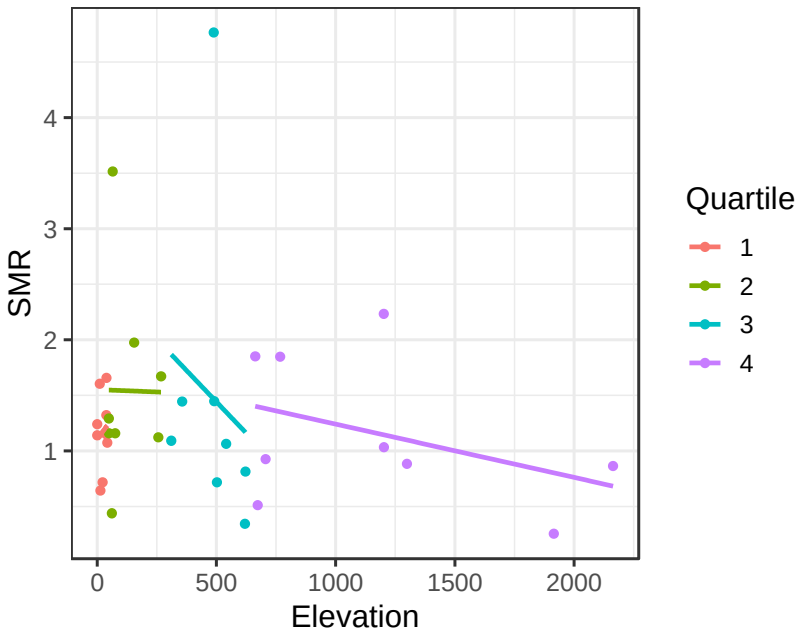


Figure 7. Elevation versus crude risk grouped quartile of elevation data

Figure 7 shows there are varying effects across the group of the elevation on crude risk. However, the variation of dengue risk is influenced by spatial dependency and heterogeneity. Hence, it is essential to consider the spatial dependence and heterogeneity in the model to obtain a more reliable disease risk prediction. To include the spatial dependency into the model, we have to define the spatial weight matrix that describes the dependencies between areas. We develop a spatial weight matrix using the k-nearest neighbor approach. We do not use spatial contiguity because some areas are not connected due to archipelago country. Using the number of cases, we found the optimal weight matrix is the matrix constructed by four (k=4) neighboring. See Figure 8 for Moran's I for $k = 1, \ldots, 8$ and Figure 9 for the connecting map.

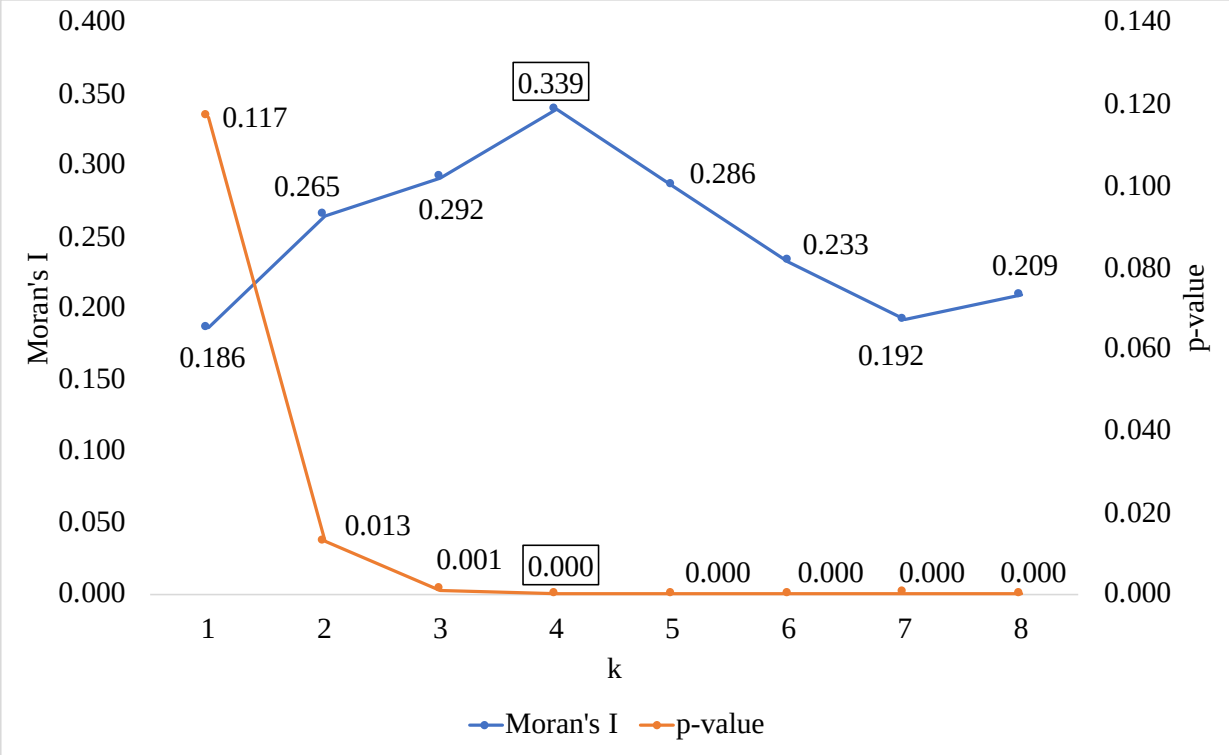


Figure 8. Moran's I for $k = 1, \ldots, 8$

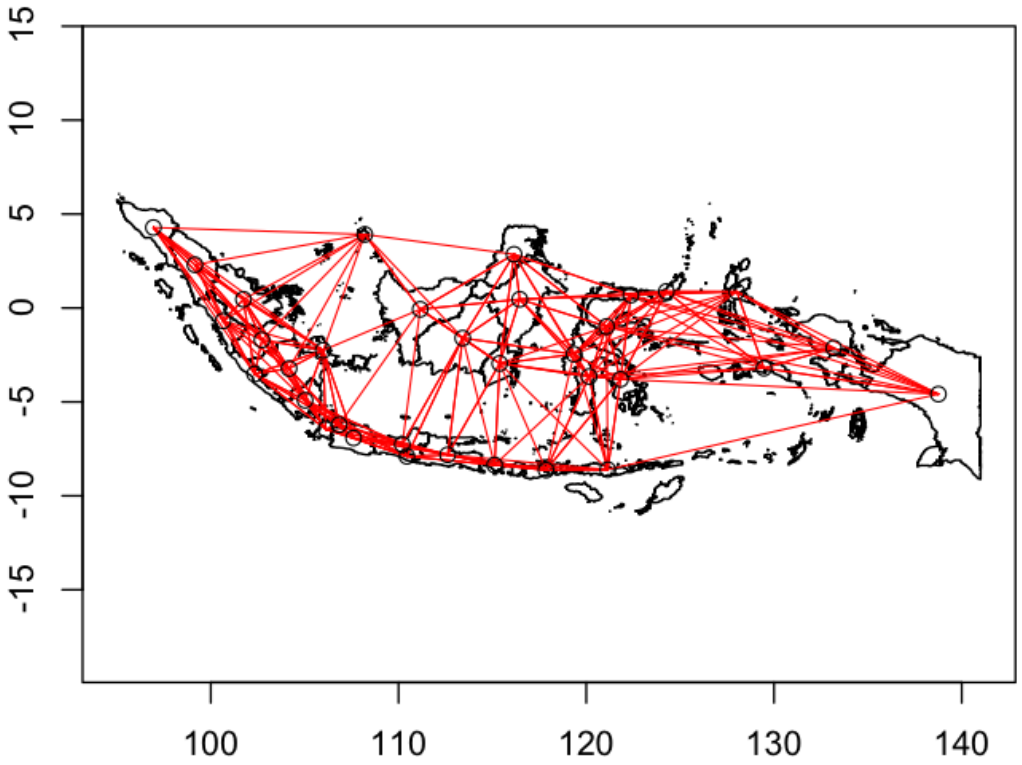


Figure 9 Connecting map

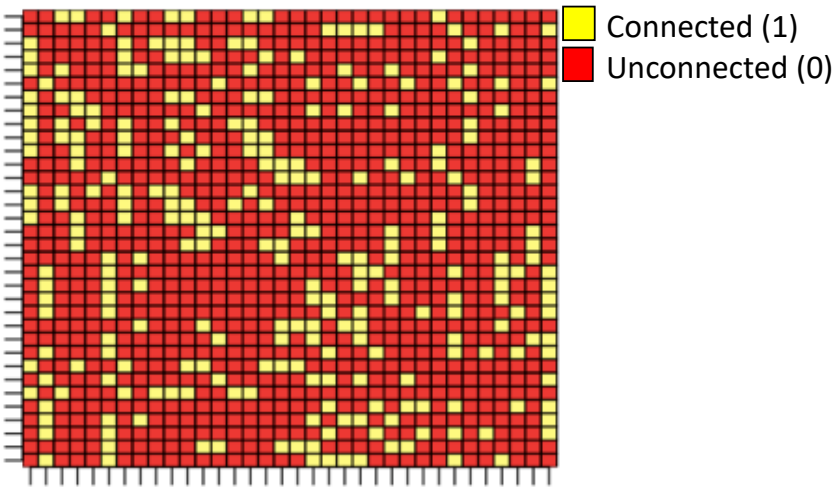


Figure 10. Sparse spatial weight matrix

Figure 10 shows the sparse spatial matrix W used to accommodate the spatial dependency into the spatial model.

Next, we compare five different models of dengue disease risk. We estimate the model using integrated nested Laplace approximation (INLA). Table 4 shows the model comparison based on deviance information criteria (DIC) and Watanabe Akaike information criteria (WAIC).

Table 4. Models comparisons

No	Model	DIC	WAIC
M1	Based line model	26829.24	18683.27
M2	Fixed 1	24445.28	22814.46
M3	Fixed 2	396.91	386.63
M4	Random 1	24445.13	22801.12
M5	Random 2	396.92	386.68

Based on the DIC and WAIC criteria we chose the M3 as the best model to explain the dengue disease in Indonesia in 2018. The M3 is specified as:

$$\hat{\eta}_{i(1)} = 0.1746 - 0.0003x_{i(1)} + \hat{\omega}_{i(1)}$$

$$\hat{\eta}_{i(2)} = 0.1201 + 0.0012x_{i(2)} + \hat{\omega}_{i(2)}$$

$$\hat{\eta}_{i(3)} = 0.5088 - 0.0010x_{i(3)} + \hat{\omega}_{i(3)}$$

$$\hat{\eta}_{i(4)} = 0.5357 - 0.0005x_{i(4)} + \hat{\omega}_{i(4)}$$

The detail parameters estimates of the intercept, slope, the relative risk are presented in Table 5.

Table 5. Parameter estimates of M3

Quartile	Intercept (α)	Slope (β_1)	Relative Risk		
			Mean	Min	Max
Q1 (very low)	0.175	-0.0003	1.173	0.643	1.657
Q2 (low)	0.120	0.0012	1.542	0.439	3.514
Q3 (high)	0.509	-0.0010	1.460	0.345	4.755
Q4 (very high)	0.536	-0.0005	1.156	0.258	2.233

Table 5 shows the very low, high and very high elevation areas have negative effect on the disease risk. It indicates that for very low, high and very high areas the risk of dengue disease tend to be low. See the Figure 11 for detail spatial distribution.

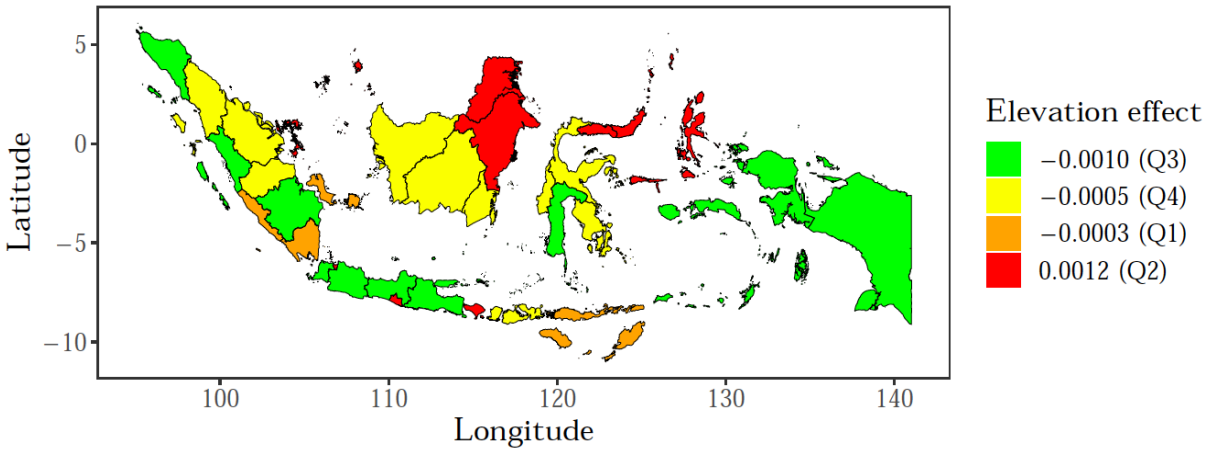


Figure 11. Elevation effects for all quartiles

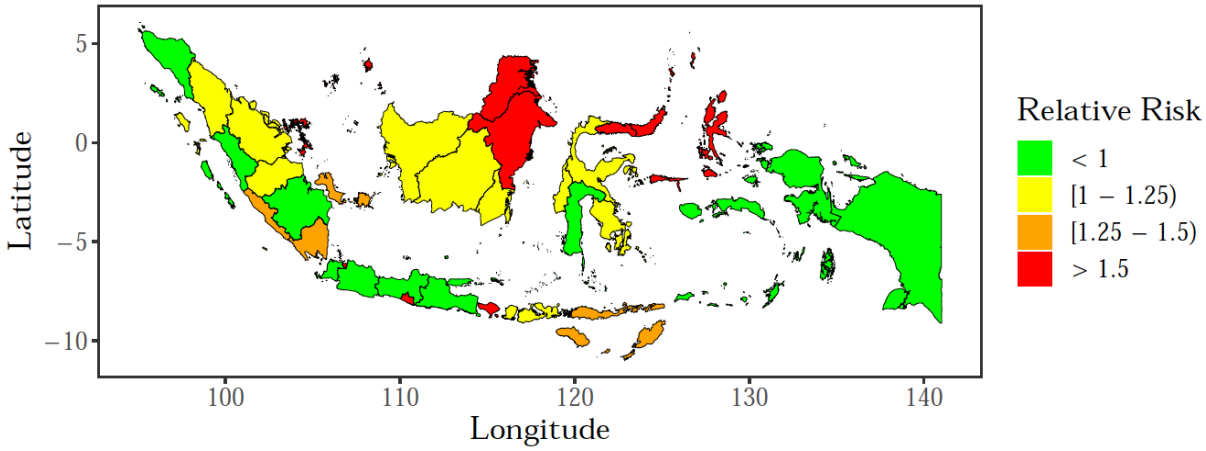


Figure 12. Spatial distribution of the relative risk

Figure 12 shows the high-risk estimates of dengue disease across 34 provinces in Indonesia in 2019. The spatial distribution of dengue disease risk seems to correspond to the elevation effects in Figure 11.

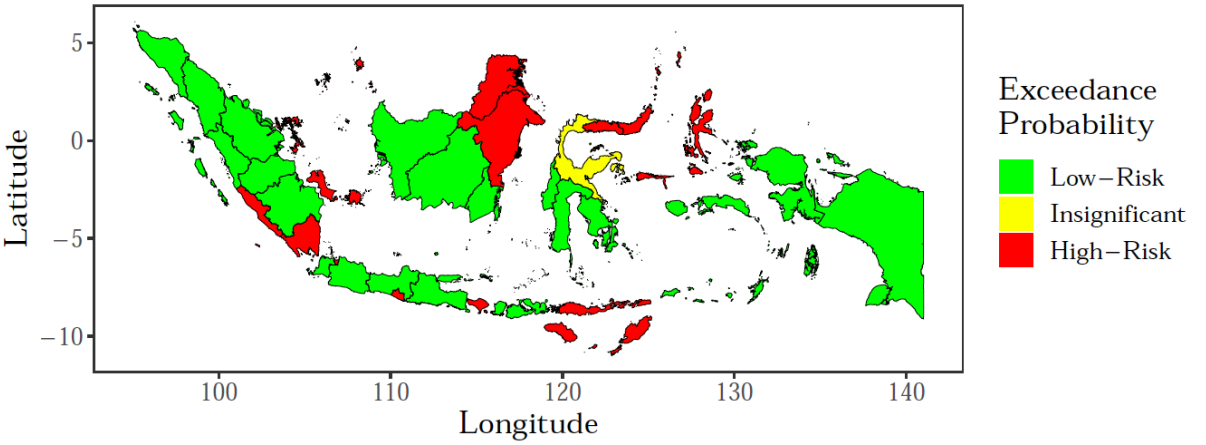


Figure 13. High, insignificant, and low risk areas

Figure 13 shows the High, insignificant, and low-risk areas based on the exceedance probability. The areas are categorized as high-risk if the exceedance probabilities are higher than 0.950. Low risk if the exceedance probabilities are lower than 0.05 and insignificant for otherwise. Our clustering Bayesian hierarchical model found that the two provinces in Java island are categorized as a high-risk area. The provinces are DKI Jakarta and Yogyakarta. In Sumatera island, we found more regions are categorized as high-risk areas. We also found that Bali, Maluku, Nusa Tenggara Timur provinces are also classified as high-risk areas. The result is much different from the plot of the number of cases in Figure 1. It indicates we have to be careful in modeling disease risk because some confounding variables might influence the disease risk variation, such as population at risk and elevation.

4. CONCLUSION

Our analysis described that elevation is a proxy for the dengue disease in Indonesia. We found that for quantiles Q3 (289.44 - 461.96) and quantile Q4 (> 461.96), the effect of the elevation on dengue risk disease decreased, respectively. We realized that a single risk factor such as elevation does not determine all spatial transmission processes of dengue disease. However, we believe it is a pragmatic proxy for the dengue disease range because it has correlation with ecological factors that is critical for mosquito development, such as weather variables. Our study has several limitations. In our model, we only considered the elevation variable, which is fixed over time. Our analysis does not account for seasonality variability in explaining dengue disease transmission. Indonesia has two seasons, rainy and dry seasons, that might be significantly contribute to mosquitoes development and disease transmission.

ACKNOWLEDGMENT

The authors gratefully acknowledge the funding from Universitas Padjadjaran through research grant No: 1427/UN6.3.1/LT/2020 and supported by South Initiatives Project 2020, SI-2019-01-82.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] WHO, Dengue and severe dengue, WHO, (2020). Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. [Accessed 26 September 2020].

[2] J. Brady, P. W. Gething, S. Bhatt, J. P. Messina, J. S. Brownstein, A. G. Hoen, C. L. Moyes, A. W. Farlow, T. W. Scott and S. I. Hay, Refining the global spatial limits of dengue virus transmission by evidence-based consensus, PLOS Negl. Trop. Dis. 6(8) (2012), e1760.

[3] M. Trips, G. McClelland, J. Gillet, C. Teesdale and T. Rao, Diel periodicity in the landing of Aedes aegypti on man, Bull World Health Organ. 48 (1973), 623-629.

[4] M. U. Kraemer, M. E. Sinka, K. A. Duda, A. Q. Mylne, F. M. Shearer, C. M. Barker, C. G. Moore, R. G. Carvalho and G. E. Coelho, The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus, eLife, 4 (2015), e08347.

[5] A. G. Watts, J. Miniota, H. A. Joseph, O. J. Brady, M. U. G. Kraemer and A. W. Grills, Elevation as a proxy for mosquito-borne Zika virus transmission in the Americas, PLoS ONE, 12(5) (2017), e0178211.

[6] K. L. Ebi and J. Nealon, Dengue in a changing climate, Environ. Res. 151(1) (2016), 115-123.

[7] O.J. Brady, N. Golding, D.M. Pigott, M.U.G. Kraemer, J.P. Messina, R.C. Reiner Jr, T.W. Scott, D.L. Smith, P.W. Gething, S.I. Hay, Global temperature constraints on Aedes aegypti and Ae. albopictus persistence and competence for dengue virus transmission, Parasit. Vectors. 7 (2014), 338.

[8] P. Yin, L. Mu, M. Madden and J. E. Vena, Hierarchical Bayesian modeling of spatio-temporal patterns of lung cancer incidence risk in Georgia, USA: 2000–2007, J. Geogr. Syst. 16(1) (2014), 387–407.

[9] I. G. N. M. Jaya and H. Folmer, Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia, J. Geogr. Syst. 22(1) (2020), 105-142.

[10] I.G.N.M. Jaya, H. Folmer, B.N. Ruchjana, F. Kristiani, Y. Andriyana, Modeling of Infectious Diseases: A Core Research Topic for the Next Hundred Years, in: R. Jackson, P. Schaeffer (Eds.), Regional Research Frontiers - Vol. 2, Springer International Publishing, Cham, 2017: pp. 239–255.

[11] A. Lawson, Bayesian Disease Mapping, Hierarchical Modelling in Spatial Epidemiology, 2nd ed., CRC Press Taylor and Francis Group, Boca Raton, (2018).

[12] P. A. P. Moran, Notes on Continuous Stochastic Phenomena, Biometrika, 37 (1950), 17-23.

[13] D. Clayton and J. Kaldor, Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping, Biometrics, 43(3) (1987), 671-681.

[14] J. Besag, J. York, A. Mollié, Bayesian image restoration, with two applications in spatial statistics, Ann. Inst. Stat. Math. 43 (1991) 1–20.

[15] A. Gelman, J. Hwang and A. Vehtari, Understanding predictive information criteria for Bayesian models, Stat. Comput. 24 (2014), 997–1016.

[16] S. Watanabe, Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory, J. Mach. Learn. Res. 11 (2010), 3571-3594.

[17] U. Nation, United Nation Statistics, (2015). Available: https://web.archive.org/web/20071031023924/http://unstats.un.org/unsd/demographic/products/dyb/DYB2004/Table03.pdf. [Accessed 25 September 2020].

[18] Kementrian Kesehatan RI, Data dan Informasi Profil Kesehatan Indonesia 2019, Jakarta: Kementrian Kesehatan RI, (2019).