PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 05 No. 1, April 2026
Pages 139-152

Bayesian Hierarchical Lognormal Modeling of Dengue Incidence with Area-Specific Temporal Effects

Erwan Setiawan[1]*, Anang Kurnia[2], Kusman Sadik[3]

[1]Mathematics Education Study Program FKIP Universitas Suryakancana
Pasirgede Raya Street, Bojongherang, Cianjur District, Cianjur Regency, 43216, West Java, Indonesia
[2,3]The School of Data Science, Mathematics, and Informatics IPB University
Meranti Street IPB Dramaga Campus, Bogor, 16680, West Java, Indonesia

E-mail Correspondence Author: *erwan@unsur.ac.id*

Abstract

This study applies and evaluates a Bayesian hierarchical lognormal model for dengue fever incidence rates across 27 regencies/cities in West Java, Indonesia, from 2014 to 2022, while accounting for area-level heterogeneity. Six predictors were considered: population density, population growth rate, net enrollment rate at junior secondary education level, access to adequate sanitation, poverty rate, and GRDP growth rate at constant prices. The model included area-specific random intercepts and random slopes for year to represent differences in baseline incidence and temporal patterns. Prior sensitivity was assessed using three scenarios for the predictor coefficients. The results showed that population density had the clearest positive association with dengue incidence, whereas the other covariates did not show clear independent effects based on posterior credible intervals. Random-effect estimates indicated stronger area-level heterogeneity in baseline incidence than in temporal trends. Evaluation on the 2022 testing data yielded MAE and RMSE values of 32.38 and 51.57, respectively, indicating moderate predictive performance. These findings suggest that Bayesian hierarchical modeling is useful for describing regional heterogeneity, although additional spatial, climatic, or temporal components may improve predictive accuracy.

Keywords: *Bayesian hierarchical model, dengue incidence rate, lognormal model, random effects, west java.*

: https://doi.org/10.30598/parameterv5i1pp139-152

Research Aritticle** · **Open Acces
Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter

Submitted: July 2025
Accepted: April 2026

1. INTRODUCTION

Dengue fever remains a recurring public health problem in Indonesia, particularly in West Java. Previous studies have shown that dengue incidence in West Java varies across regencies/cities and over time, with evidence of temporal trends, spatial clustering, and high-risk areas in densely populated regions [1], [2]. These patterns indicate that dengue incidence should not only be analyzed through average population-level associations, but also through a modeling framework that can accommodate regional heterogeneity and temporal variation.

Several studies on dengue incidence in Indonesia have applied classical regression, mixed models, time-series analysis, clustering, or spatial-temporal approaches to identify risk patterns and associated factors [1], [3], [4], [5]. These studies provide important empirical evidence that dengue incidence is related to regional conditions such as population density, sanitation, education-related indicators, poverty, and other local socioeconomic or environmental factors [2], [3]. However, many existing studies focus mainly on point estimates or local association patterns, whereas uncertainty in area-level effects, prior sensitivity, and predictive validation are not always discussed in detail.

Bayesian hierarchical models offer a relevant framework for this problem because they allow fixed effects and random effects to be estimated jointly, quantify posterior uncertainty, and represent unobserved heterogeneity across areas [6]. In epidemiological modeling, Bayesian spatial and spatio-temporal models have been widely used to incorporate spatial or temporal random effects, assess predictive uncertainty, and improve risk mapping [7], [8]. A systematic review of Bayesian spatial and spatio-temporal models for dengue highlighted the importance of model structure, prior specification, covariate selection, and opportunities for further methodological development in dengue modeling [7].

In the Indonesian context, recent Bayesian dengue studies have increasingly considered spatial and temporal structures, including Bayesian spatio-temporal models for dengue prediction and risk mapping [8], [9]. Nevertheless, applications focusing on regency/city-level dengue incidence in West Java using a Bayesian hierarchical lognormal model with area-specific random intercepts and area-specific random slopes for year remain limited. Such a structure is useful because the random intercept captures differences in baseline dengue incidence among regencies/cities, while the random slope for year allows temporal patterns to vary across areas.

Therefore, this study applies and evaluates a Bayesian hierarchical lognormal model for dengue incidence rates across 27 regencies/cities in West Java from 2014 to 2022. The model includes six area-level predictors: population density, population growth rate, net enrollment rate at junior secondary education level, access to adequate sanitation, poverty rate, and GRDP growth rate at constant prices. Model evaluation is conducted through prior sensitivity analysis, MCMC diagnostics, LOOIC, Pareto-k diagnostics, and predictive accuracy on the 2022 testing data using MAE and RMSE [10].

2. METHOD

2.1. Data

This study used secondary panel data for 27 regencies/cities in West Java Province from 2014 to 2022, yielding 243 area-year observations. The response variable was the annual dengue incidence rate, measured as the number of dengue cases per 100,000 population in each regency/city. The explanatory variables consisted of population density, population growth rate, net enrollment rate at junior secondary education level, household access to adequate sanitation,

poverty rate, and gross regional domestic product (GRDP) growth rate at constant prices [11], [12], [13], [14]. The complete list of variables used in this study is presented in **Table 1**.

Table 1. List of variables used in the study

Types of Variables	Variable in dataset	Description	Unit/Scale
Response	IR	Annual dengue incidence rate in each regency/city	Cases per 100,000 population
Fixed Effect	pop_density	Population density of each regency/city	1000 persons per km^2
	pop_growth	Annual population growth rate	Percent (%)
	ner_junior	Net enrollment rate at junior secondary education level	Percent (%)
	Sanitation	Percentage of households with access to adequate sanitation	Percent (%)
	Poverty	Percentage of population living below the poverty line	Percent (%)
	GRDP_growth	Growth rate of gross regional domestic product at constant prices	Percent (%)
Random Effect	Area	Administrative area in West Java Province	Categorical
	Year	Observation year	Time index, 2014 = 0

The variables sanitation, ner_junior, and pop_density were divided by 10 to improve numerical stability. All continuous predictors were then centered using the training-data means. Regency/city was included as a random intercept to capture differences in baseline dengue incidence across areas, while year was included as an area-specific random slope to capture heterogeneity in temporal trends.

2.2 Research Model

In this study, the incidence rate of dengue fever was modeled using a Bayesian hierarchical lognormal model. The model was specified based on a panel structure consisting of 27 regencies/cities observed annually from 2014 to 2022. The year variable was transformed into a time index by subtracting 2014, so that $T_t = 0$ corresponds to 2014 and $T_t = 8$ corresponds to 2022. The hierarchical model is written in **Equation (1)** [15].

$$\begin{aligned} \log(Y_{jt}) &= \eta_{jt} + \varepsilon_{jt} \\ \eta_{jt} &= \beta_0 + \sum_{k=1}^{6} \beta_k X_{kjt} + u_{0j} + u_{1j} T_t \\ \varepsilon_{jt} &\sim N(0, \sigma^2) \\ (u_{0j}, u_{1j})^T &\sim N_2(\mathbf{0}, \Sigma_u) \end{aligned} \tag{1}$$

where

$$\Sigma_u = \begin{pmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{pmatrix}$$

Here, Y_{jt} denotes the dengue incidence rate in regency/city j at year t, where $j = 1, \ldots, 27$ and $t = 0, \ldots, 8$. The term X_{kjt} denotes the k-th explanatory variable for regency/city j at year t. The parameter β_0 is the overall intercept, and β_k represents the fixed effect of the k-th covariate. The

term u_{0j} represents the regency/city-specific random intercept, while u_{1j} represents the regency/city-specific random slope for year. The residual error is represented by ε_{jt}, and the covariance matrix Σ_u allows the random intercept and random slope for year to be correlated. The Bayesian hierarchical model is completed by assigning prior distributions to the unknown parameters. Prior distributions represent initial assumptions about the model parameters before observing the data. Let

$$\theta = \{\beta_0, \dots, \beta_6, \mathbf{u}_1, \dots, \mathbf{u}_{27}, \Sigma_u, \sigma\}$$

denote the set of unknown parameters in the model. The joint prior distribution can be written as **Equation (2)**.

$$p(\theta) = p(\beta_0) \prod_{k=1}^{6} p(\beta_k) \prod_{j=1}^{27} p(\mathbf{u}_j|\Sigma_u)\, p(\Sigma_u) p(\sigma) \tag{2}$$

where $\mathbf{u}_j = \left(u_{0j}, u_{1j}\right)^T$.

Prior sensitivity was assessed using three scenarios that differed only in the prior scale assigned to the predictor coefficients, while the priors for the intercept, random-effect standard deviations, correlation matrix, and lognormal residual standard deviation were held constant. Model 1 imposed stronger shrinkage toward zero, Model 2 represented an intermediate prior specification, and Model 3 allowed more diffuse predictor effects. Because the lognormal model is defined on the logarithmic scale, the priors for the intercept and predictor coefficients were specified on the log-incidence-rate scale. The intercept prior was centered at 3.679, corresponding to the logarithm of the median incidence rate in the training data. The random-effect correlation matrix was assigned an $LKJ(2)$ prior, providing mild regularization toward zero correlation while allowing either positive or negative correlation. The prior specifications are summarized in **Table 2**, and their implications for $\exp(\beta_k)$ are illustrated in **Figure 1**.

Table 2. Prior specifications for each model scenario

Parameter	Notation	Model 1: Regularizing	Model 2: Main	Model 3: Diffuse
Intercept	β_0	$N(3.679\,, 1.00^2)$	$N(3.679\,, 1.00^2)$	$N(3.679\,, 1.00^2)$
Predictor coefficients	$\beta_k,\ k = 1, \dots, 6$	$N(0\,, 0.05^2)$	$N(0\,, 0.10^2)$	$N(0\,, 0.30^2)$
SD of random intercept	τ_0	$t_3^+(0\,, 0.75)$	$t_3^+(0\,, 0.75)$	$t_3^+(0\,, 0.75)$
SD of random slope for year	τ_1	$t_3^+(0\,, 0.15)$	$t_3^+(0\,, 0.15)$	$t_3^+(0\,, 0.15)$
Correlation matrix of random effects	R_u	$LKJ(2)$	$LKJ(2)$	$LKJ(2)$
Lognormal residual SD	σ	$t_3^+(0\,, 0.75)$	$t_3^+(0\,, 0.75)$	$t_3^+(0\,, 0.75)$

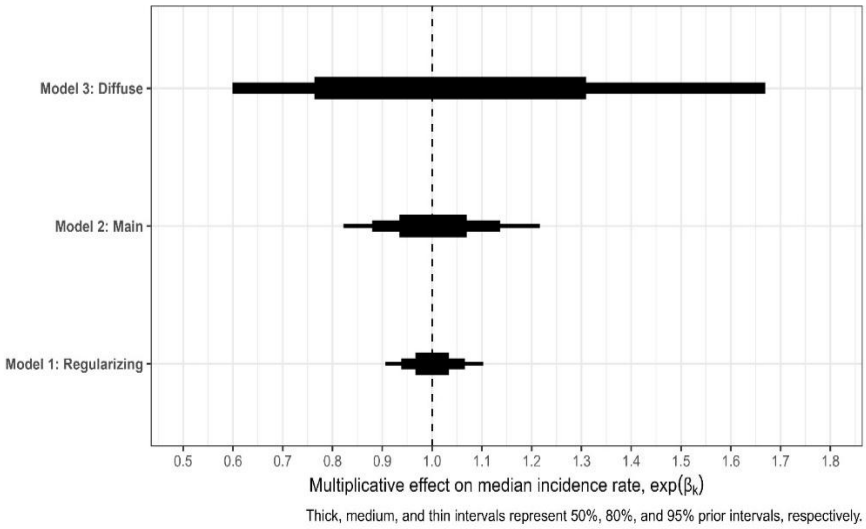

Figure 1. Prior implications for predictor coefficients

Figure 1 shows that all three scenarios are centered around $\exp(\beta_k) = 1$, which corresponds to no covariate effect. However, the scenarios differ in their dispersion. Model 1 has the narrowest interval, indicating stronger shrinkage toward no effect. Model 3 has the widest interval, allowing larger positive or negative covariate effects. Model 2 lies between these two scenarios, representing an intermediate prior specification.

The incidence rate (IR) of dengue fever is defined as **Equation (3)**.

$$\text{IR} = \frac{\text{the number of dengue fever cases}}{\text{total population}} \times 100000 \tag{3}$$

Because the incidence rate is a positive continuous measure, it was modeled using a lognormal distribution. The probability density function is given by **Equation (4)**.

$$f\left(Y_{jt}|\eta_{jt},\sigma\right) = \frac{1}{Y_{jt}\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(\log Y_{jt} - \eta_{jt}\right)^2}{2\sigma^2}\right), \qquad Y_{jt} > 0 \tag{4}$$

The likelihood function for the training data can be written as **Equation (5)**.

$$p(Y|\theta) = \prod_{(j,t)\in\mathcal{D}_{train}} \frac{1}{Y_{jt}\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(\log Y_{jt} - \eta_{jt}\right)^2}{2\sigma^2}\right) \tag{5}$$

where $\mathcal{D}_{train}$ denotes the set of observations used for model fitting. In this study, the training data consist of observations from 2014 to 2021, while the 2022 data were used for testing model predictions.

The posterior distribution represents the updated distribution of the model parameters after observing the data. Based on Bayes' theorem, the posterior distribution is proportional to the product of the likelihood and the prior distribution. Using the prior structure in **Equation (2)**, the posterior distribution can be written as **Equation (6)**.

$$p(\theta|Y) \propto p(Y|\theta)p(\beta_0)\prod_{k=1}^{6} p(\beta_k)\prod_{j=1}^{27} p\left(\mathbf{u}_j|\Sigma_u\right)p(\Sigma_u)p(\sigma) \tag{6}$$

2.3 Analysis Data

The analysis was conducted using R version 4.3.2. The Bayesian hierarchical lognormal models were fitted using the brms package [16], which interfaces with Stan through rstan. The models were fitted using four chains, 10,000 iterations per chain, and 7,000 warm-up iterations. The panelr [17] package was also used to support the management of panel data structures. The overall data analysis procedure is presented in **Figure 2**.

Figure 2 presents the research flow used in this study. The analysis began with data preparation and exploration, including checking and handling missing values when necessary. The dataset was then split into training data from 2014 to 2021 and testing data from 2022. The training data were used to specify and fit the Bayesian hierarchical lognormal model under several prior distribution scenarios. Model performance was evaluated using LOOIC-based criteria, and the selected model was then applied to testing data. Finally, predictive performance was assessed using MAE and RMSE, followed by interpretation of the model results.

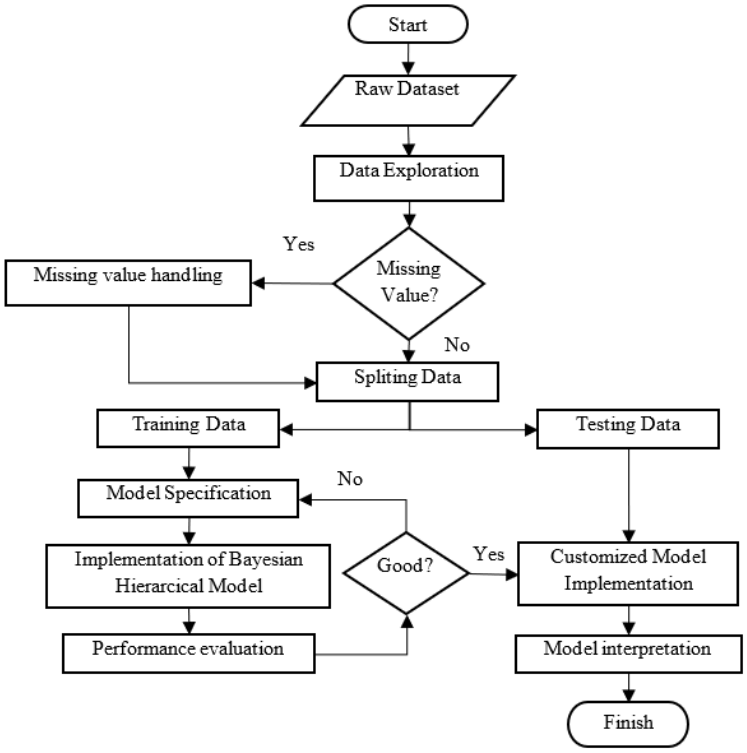

Figure 2. Research flow

3 RESULTS AND DISCUSSION

The research dataset contains information on dengue fever rates and several predictor variables from 27 regencies/cities in West Java during 2014 to 2022. **Figure 3** presents the annual incidence rate patterns for each regency/city. The figure shows that dengue incidence varied substantially across regions, both in terms of baseline levels and temporal fluctuations. Some regencies/cities showed relatively stable patterns, whereas others experienced sharp increases or decreases in certain years. These descriptive patterns provide an empirical motivation for using a hierarchical model with regency/city-specific random intercepts and regency/city-specific random slopes for year, allowing baseline incidence and temporal variation to differ across regions.

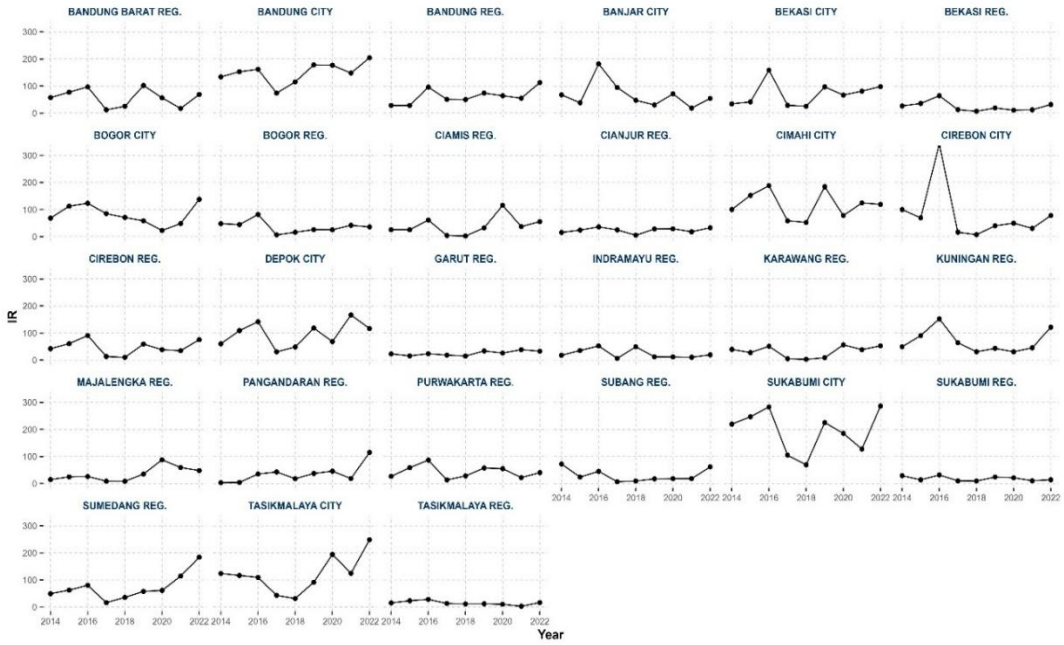

Figure 3. **The time series data of the incidence rate in each regency/city**

The MCMC diagnostics are summarized in **Table 3**. Overall, the posterior sampling showed satisfactory convergence across the three prior scenarios. The maximum $\hat{R}$values ranged from 1.001 to 1.004, which are very close to 1, indicating good convergence of the Markov chains. The effective sample sizes were also sufficiently large, with minimum bulk ESS values above 2,200 and minimum tail ESS values above 4,000 across all models. In addition, no divergent transitions were detected in any model. These results indicate that the posterior samples were stable and reliable for subsequent inference and model comparison.

Table 3. MCMC diagnostics for the three prior scenarios

Model	Prior scenario	Max $\hat{R}$	Min Bulk ESS	Min Tail ESS	Divergent transitions
Model 1	Regularizing prior	1.001	2239.81	4033.74	0
Model 2	Main prior	1.004	2284.88	4822.49	0
Model 3	Diffuse prior	1.002	2252.30	4277.51	0

After confirming MCMC convergence, model comparison was conducted using LOOIC and predictive accuracy on the held-out 2022 testing data. LOOIC was used to assess expected out-of-sample predictive performance, while Pareto-k diagnostic evaluated the reliability of the PSIS-LOO approximation. Testing MAE and RMSE were then used to assess prediction accuracy on unseen observations. Lower LOOIC, MAE, and RMSE values indicate better predictive performance.

Table 4. Model comparison based on information criteria and testing accuracy

Model	Prior Scenario	LOOIC		Testing		Pareto-$k > 0.7$
		Estimate	SE	MAE	RMSE	
Model 1	Regularizing prior	2080.00	29.78	32.66	51.90	0
Model 2	Main prior	2080.99	29.76	32.38	51.57	0
Model 3	Diffuse prior	2082.42	29.63	32.02	50.76	0

Table 4 shows that the three prior scenarios produced comparable predictive performance. Model 1 yielded the lowest LOOIC, whereas Model 3 produced the lowest testing MAE and RMSE. However, the differences were small. The LOOIC difference between Model 1 and Model 2 was only 0.99, which is negligible relative to the corresponding standard error of approximately 29.8. Similarly, the differences in testing MAE and RMSE among the three models were modest. All models had no observations with Pareto-k > 0.7, indicating that the PSIS-LOO approximation was stable across all prior scenarios. Therefore, there was no strong evidence that one prior scenario clearly outperformed the others. Model 2 was retained as the main inferential model because it represents the intermediate prior specification, balancing regularization and flexibility while maintaining competitive performance across both LOOIC and testing-error metrics. Accordingly, the posterior parameter estimates from Model 2 are reported in Table 5.

Table 5. Parameter estimation for Model 2

	Estimate	Est.Error	l-95%CI	u-95%CI
sd(Intercept) (τ_0)	0.45	0.09	0.29	0.66
sd(year) (τ_1)	0.06	0.03	0.00	0.13
Cor(Intercept, year) (ρ)	0.07	0.37	-0.67	0.76
Intercept (β_0)	3.64	0.10	3.44	3.84
pop_density (β_1)	0.13	0.03	0.08	0.18
pop_growth (β_2)	0.03	0.07	-0.11	0.17
ner_junior (β_3)	0.02	0.08	-0.14	0.19
sanitation (β_4)	-0.02	0.04	-0.09	0.04
poverty (β_5)	0.05	0.03	-0.02	0.11
GRDP_growth (β_6)	-0.02	0.02	-0.07	0.02
Sigma (σ)	0.74	0.04	0.66	0.82

The posterior estimate of the random intercept standard deviation was 0.45, with a 95% credible interval of [0.29, 0.66]. This result indicates meaningful variation in baseline dengue incidence across regencies/cities on the log scale. In contrast, the random slope standard deviation for year was smaller, with an estimate of 0.06 and a 95% credible interval of [0.00, 0.13], suggesting that temporal trends varied only weakly across regions. The correlation between the random intercept and random slope was uncertain, indicating no clear relationship between regional baseline incidence and annual temporal change. These findings suggest that regional differences in dengue incidence were more evident in baseline levels than in temporal trends, which is consistent with previous studies reporting spatial variability in dengue incidence rates [18].

The fixed-effect estimates indicate that population density had the clearest positive association with dengue incidence, as its 95% credible interval excluded zero. This finding is consistent with previous dengue studies showing that densely populated areas tend to be associated with higher dengue risk or exposure [19], [20]. Since the model is specified on the logarithmic scale, the coefficient is interpreted on the log incidence-rate scale. Exponentiating the estimate gives $\exp(0.13) = 1.14$, suggesting that a 1,000 persons/km² increase in population density from its centered value was associated with approximately 14% higher median dengue incidence rate, holding other covariates constant. The remaining covariates had credible intervals that included zero, indicating limited posterior evidence of independent associations after accounting for area-level random effects. The residual standard deviation of the lognormal model was estimated at 0.74, indicating remaining within-area and temporal variability not fully explained by the observed covariates and random effects. Overall, these results suggest that area-level heterogeneity in baseline dengue incidence was more pronounced than heterogeneity in

temporal trends, and that population density was the most consistently supported covariate in the model.

Table 6. Parameter estimation of random effect

Area	Random Intercept				Random Slope			
	Estimate	Est. Error	Q2.5	Q97.5	Estimate	Est. Error	Q2.5	Q97.5
BANDUNG BARAT REG.	0.2615	0.2446	-0.2130	0.7404	-0.0211	0.0564	-0.1563	0.0779
BANDUNG CITY	0.1140	0.3111	-0.4824	0.7378	-0.0008	0.0520	-0.1104	0.1089
BANDUNG REG.	0.4342	0.2438	-0.0351	0.9187	0.0228	0.0566	-0.0799	0.1534
BANJAR CITY	0.5285	0.2608	0.0242	1.0575	-0.0276	0.0606	-0.1721	0.0745
BEKASI CITY	-0.3327	0.2783	-0.8986	0.2046	0.0115	0.0549	-0.0958	0.1356
BEKASI REG.	-0.3027	0.2819	-0.8707	0.2380	-0.0433	0.0629	-0.1931	0.0523
BOGOR CITY	0.0118	0.2527	-0.4875	0.5144	-0.0360	0.0588	-0.1801	0.0554
BOGOR REG.	-0.0250	0.2460	-0.5094	0.4644	-0.0169	0.0545	-0.1446	0.0883
CIAMIS REG.	-0.1254	0.2520	-0.6172	0.3652	0.0207	0.0543	-0.0787	0.1484
CIANJUR REG.	-0.2495	0.2401	-0.7299	0.2098	-0.0018	0.0523	-0.1126	0.1098
CIMAHI CITY	-0.0767	0.2935	-0.6561	0.5086	-0.0094	0.0528	-0.1271	0.0981
CIREBON CITY	-0.3405	0.2748	-0.8903	0.1899	-0.0500	0.0660	-0.2088	0.0465
CIREBON REG.	-0.0559	0.2486	-0.5527	0.4216	-0.0127	0.0519	-0.1311	0.0878
DEPOK CITY	0.2713	0.2795	-0.2626	0.8334	0.0063	0.0533	-0.1031	0.1219
GARUT REG.	-0.1660	0.2399	-0.6392	0.2946	0.0173	0.0542	-0.0840	0.1443
INDRAMAYU REG.	-0.4432	0.2688	-0.9933	0.0687	-0.0303	0.0590	-0.1716	0.0698
KARAWANG REG.	-0.3547	0.2455	-0.8394	0.1232	-0.0113	0.0537	-0.1305	0.0969
KUNINGAN REG.	0.4174	0.2590	-0.0894	0.9395	-0.0267	0.0587	-0.1678	0.0743
MAJALENGKA REG.	-0.1638	0.2495	-0.6663	0.3144	0.0444	0.0639	-0.0478	0.2009
PANGANDARAN REG.	-0.2580	0.2467	-0.7431	0.2216	0.0591	0.0712	-0.0356	0.2298
PURWAKARTA REG.	0.2341	0.2488	-0.2441	0.7190	-0.0051	0.0530	-0.1224	0.1031
SUBANG REG.	-0.2314	0.2492	-0.7298	0.2455	-0.0345	0.0593	-0.1762	0.0615
SUKABUMI CITY	0.7480	0.2641	0.2412	1.2786	-0.0122	0.0590	-0.1470	0.0992
SUKABUMI REG.	-0.2812	0.2517	-0.7804	0.2060	-0.0149	0.0532	-0.1362	0.0875
SUMEDANG REG.	0.4927	0.2549	0.0151	1.0094	0.0163	0.0553	-0.0926	0.1420
TASIKMALAYA CITY	0.3929	0.2714	-0.1241	0.9423	0.0083	0.0544	-0.1033	0.1257
TASIKMALAYA REG.	-0.6369	0.2549	-1.1416	-0.1439	-0.0528	0.0693	-0.2191	0.0496

Table 6 presents the area-specific random effect estimates from Model 2. The random intercepts represent deviations of each regency/city from the overall baseline log incidence rate after accounting for the observed covariates. Most regencies/cities had random intercept credible intervals that included zero, indicating that their baseline incidence was not clearly different from the overall average. However, Sukabumi City, Banjar City, and Sumedang Regency showed positive random intercept estimates with 95% credible intervals above zero, suggesting higher baseline dengue incidence than the overall average. In contrast, Tasikmalaya Regency had a negative random intercept estimate with a 95% credible interval below zero, indicating lower baseline incidence. Because the model was specified on the logarithmic scale, the random intercept estimates can also be interpreted multiplicatively after exponentiation. For example, the random intercept of Sukabumi City was 0.7480, corresponding to $\exp(0.7480) \approx 2.11$, indicating that its baseline median dengue incidence was estimated to be about twice the overall baseline level, holding other covariates constant. Conversely, Tasikmalaya Regency had a random

intercept of -0.6369, corresponding to $\exp(-0.6369) \approx 0.53$, indicating a lower baseline incidence than the overall average. The estimated random slopes for year were generally small, and all corresponding 95% credible intervals included zero. This indicates limited posterior evidence of strong area-specific differences in temporal trends. Therefore, the random effect results suggest that heterogeneity in dengue incidence across West Java was more pronounced in baseline incidence levels than in year-to-year temporal changes.

After evaluating the overall predictive accuracy, the prediction results were examined at the regency/city level. **Figure 5** compares the observed dengue incidence rates with the posterior predictive mean estimates for each regency/city in the 2022 testing data. This comparison highlights areas where the model performed well and areas where larger prediction errors remained, providing additional insight into regional heterogeneity in dengue incidence.

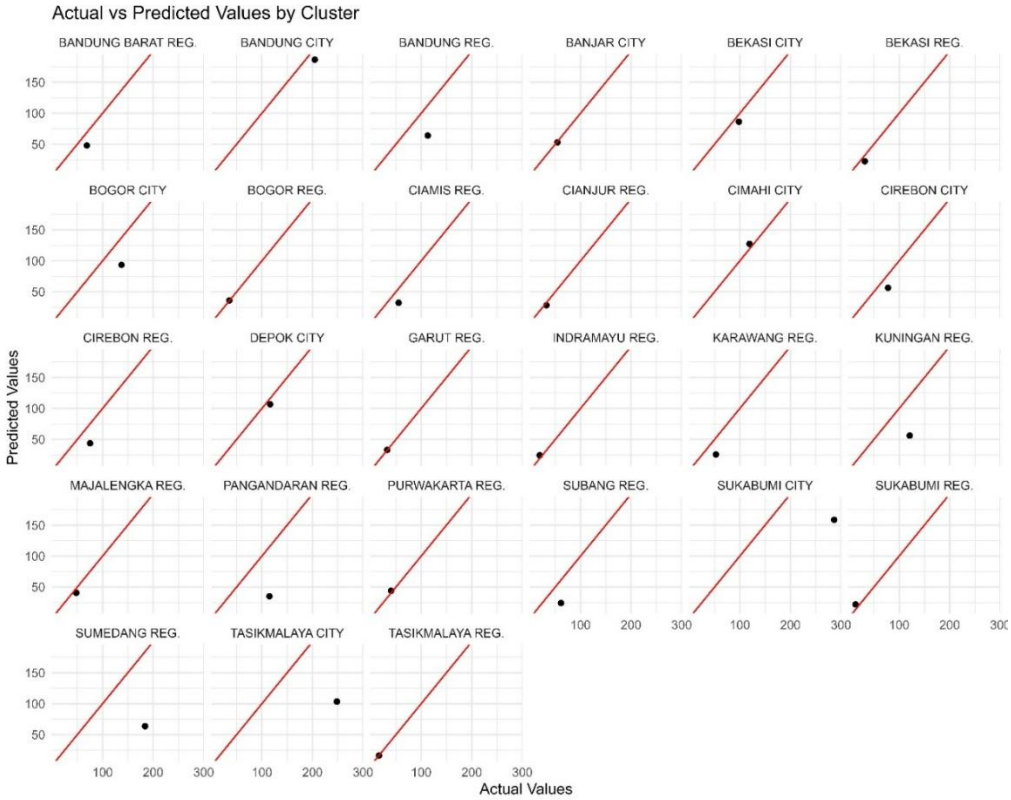

Figure 5. Predicted values for each regency/city

In **Figure 5**, the red line represents the ideal condition in which the predicted value equals the observed value. Points close to this line indicate more accurate predictions, whereas points far from the line indicate larger prediction errors. The results show that predictions were reasonably close for several regencies/cities, but some regions displayed substantial discrepancies. For example, in Tasikmalaya City, the predicted value was lower than the observed value. This pattern explains why RMSE was larger than MAE, since RMSE is more sensitive to large errors. These findings indicate that the model captures part of the regional pattern of dengue incidence, but additional explanatory variables or more flexible temporal and spatial structures may be required to improve prediction in areas with unusually high or low incidence. This is consistent with Bayesian dengue modeling literature, which emphasizes the importance of climatic variables, spatially structured random effects, and spatio-temporal dependence in improving dengue risk modeling and prediction [21], [22].

4 CONCLUSION

This study applied a Bayesian hierarchical lognormal model to analyze dengue fever incidence rates across 27 regencies/cities in West Java from 2014 to 2022. The model incorporated area-specific random intercepts and random slopes for year, allowing baseline incidence and temporal variation to differ across regencies/cities. The results showed substantial area-level heterogeneity in baseline dengue incidence, while the variation in year-specific slopes was relatively weak. Among the six covariates considered, population density showed the clearest positive association with dengue incidence, whereas population growth, junior secondary education enrollment, sanitation, poverty rate, and GRDP growth rate did not show clear independent effects based on their 95% credible intervals. The prior sensitivity analysis indicated that the three prior scenarios produced comparable predictive performance. Model 2 was retained as the main inferential model because it represented an intermediate prior specification while maintaining competitive performance in terms of LOOIC and testing-error metrics. The testing MAE and RMSE for Model 2 were 32.38 and 51.57, respectively, indicating moderate predictive performance on the held-out 2022 data. These remaining prediction errors suggest that although the model captured important regional heterogeneity, dengue incidence may also be influenced by additional factors not included in the current model, such as climatic variables, mosquito-vector indicators, spatial dependence, or more flexible temporal dynamics. Future research may extend this framework by incorporating spatially structured random effects, dynamic temporal components, and environmental covariates to improve predictive accuracy and support more targeted dengue surveillance.

Acknowledgments

The authors gratefully acknowledge the academic and institutional support from IPB University, particularly the School of Data Science, Mathematics and Informatics, during the development of this study.

Funding Information

The first author received doctoral scholarship and research support from Beasiswa Pendidikan Indonesia (BPI) under the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, as stated in the official scholarship award documents. The funder had no role in the study design, data analysis, interpretation of results, or preparation of the manuscript.

Author Contributions Statement

Erwan Setiawan: Conceptualization, methodology, formal analysis, software, visualization, writing-original draft, and editing. Anang Kurnia: Supervision, methodology, validation, and writing-review. Kusman Sadik: Supervision, validation, and writing-review. All authors discussed the results and approved the final manuscript.

Conflict of Interest Statement

No potential conflict of interest was reported by the authors.

Data Availability

The data supporting the findings of this study are publicly available from the West Java Government Open Data platform and BPS-Statistics of Jawa Barat Province. Derived datasets and analysis scripts are available from the corresponding author upon reasonable request.

REFERENCES

[1] Y. Yanti, S. Rahardiantoro, and G. A. Dito, "Spatio-temporal Clustering Analysis of Dengue Hemorrhagic Fever Cases in West Java 2016 – 2021," *Indones. J. Stat. Its Appl.*, vol. 7, no. 1, pp. 56–63, 2023, doi: 10.29244/ijsa.v7i1p56-63.

[2] I. S. Fauzi, N. Nuraini, R. W. S. Ayu, and B. W. Lestari, "Temporal trend and spatial clustering of the dengue fever prevalence in West Java, Indonesia," *Heliyon*, vol. 8, no. 8, p. e10350, 2022, doi: 10.1016/j.heliyon.2022.e10350.

[3] A. N. Hanif, G. P. D. Sohibien, and I. Y. Wulansari, "The Influences of Climate Change and Social Vulnerability on Dengue Fever Incidence Rate in West Java Province 2019–2023," in *Proceedings of 2025 International Conference on Data Science and Official Statistics (ICDSOS)*, Politeknik Statistika STIS, 2025. doi: 10.34123/icdsos.v2025i1.606.

[4] F. A. Siregar and T. Makmur, "Time Series Analysis of Dengue Hemorrhagic Fever Cases and Climate: a Model for Dengue Prediction," *J. Phys. Conf. Ser.*, vol. 1235, p. 012072, 2019, doi: 10.1088/1742-6596/1235/1/012072.

[5] E. Setiawan, K. A. Notodiputro, and B. Sartono, "Generalized Linear Mixed-Model Tree for Modeling Dengue Fever Cases," *Cogito smart J.*, vol. 10, no. 2, pp. 380–392, 2024, doi: 10.31154/cogito.v10i2.715.380-392.

[6] A. Gelman and J. L. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[7] A. Aswi, S. M. Cramb, P. Moraga, and K. Mengersen, "Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review," *Epidemiol. Infect.*, vol. 147, p. e33, 2019, doi: 10.1017/S0950268818002807.

[8] M. F. Salim, T. B. T. Satoto, and Danardono, "Predicting spatio-temporal dynamics of dengue using INLA (integrated nested laplace approximation) in Yogyakarta, Indonesia," *BMC Public Health*, vol. 25, no. 1321, 2025, doi: 10.1186/s12889-025-22545-2.

[9] I. G. N. M. Jaya, B. N. Ruchjana, A. S. Abdullah, and N. Sunengsih, "Multilevel model of dengue disease transmission in West Java province, Indonesia by means INLA," *Commun. Math. Biol. Neurosci.*, 2020.

[10] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Stat. Comput.*, vol. 27, pp. 1413–1432, 2017, doi: 10.1007/s11222-016-9696-4.

[11] A. Ruliansyah, Y. Yuliasih, and S. Hasbullah, "Pemanfaatan Citra ASTER Dalam Penentuan Dan Verifikasi Daerah Rawan Demam Berdarah Dengue (DBD) Di Kota Banjar Provinsi Jawa Barat," *ASPIRATOR - J. Vector-borne Dis. Stud.*, vol. 6, pp. 55–62, 2015, doi: 10.22435/ASPIRATOR.V6I2.3631.55-62.

[12] A. Apriyani, A. H. Sutomo, and S. R. Umniyati, "Sanitasi lingkungan dan keberadaan jentik Aedes sp dengan kejadian demam berdarah dengue di Banguntapan Bantul," *Ber. Kedokt. Masy.*, vol. 33, no. 2, 2017, doi: 10.22146/bkm.12704.

[13] J. A. Sugianto, C. Cecilia, and Sulistiawati, "Dengue Incidence in Relation to Environmental Factors in Sampang Regency: Analysis of Five Years Data," *Media Kesehat. Masy. Indones.*, vol. 17, no. 1, 2021, doi: 10.30597/mkmi.v17i1.10598.

[14] B. Haryanto, "Indonesia Dengue Fever: Status, Vulnerability, and Challenges," in *Current Topics in Tropical Emerging Diseases and Travel Medicine*, 2018. doi: 10.5772/intechopen.82290.

[15] T. A. B. Snijders and R. J. Bosker, "Multilevel analysis: an introduction to basic and advanced multilevel modeling," 1999, [Online]. Available: https://api.semanticscholar.org/CorpusID:63173084

[16] P.-C. Bürkner, "brms: An R Package for Bayesian Multilevel Models using Stan," *J. Stat. Softw.*, vol. 80, no. 1, pp. 1–28, 2017, doi: 10.18637/jss.v080.i01.

[17] J. A. Long, "panelr: Regression Models and Utilities for Repeated Measures and Panel Data," *R Packag. version 0.7.3*, 2020, [Online]. Available: https://cran.r-project.org/package=panelr

[18] D. J. Gubler, "Dengue and Dengue Hemorrhagic Fever," *Clin. Microbiol. Rev.*, vol. 11, pp. 480–496, 1998, doi: 10.1128/CMR.11.3.480.

[19] N. Norjanah and M. R. Ridha, "Spatial Autocorrelation of Dengue and Its Relationship with Population Density in South Kalimantan, Indonesia," *Althea Med. J.*, vol. 11, no. 2, pp. 113–119, 2024, doi: 10.15850/amj.v11n2.3190.

[20] Y. S. Amelinda, R. A. Wulandari, and A. Asyary, "The effects of climate factors, population density, and vector density on the incidence of dengue hemorrhagic fever in South Jakarta Administrative City 2016-2020: an ecological study," *Acta Biomed.*, vol. 93, no. 6, p. e2022323, 2022, doi: 10.23750/abm.v93i6.13503.

[21] A. C. C. Costa, C. T. Codeço, N. A. Honório, G. R. Pereira, C. F. N. Pinheiro, and A. A. Nobre, "Surveillance of dengue vectors using spatio-temporal Bayesian modeling," *BMC Med. Inform. Decis. Mak.*, vol. 15, no. 93, 2015, doi: 10.1186/s12911-015-0219-6.

[22] S. Yip, N. C. Him, N. I. Jamil, D. He, and S. K. Sahu, "Spatio-temporal detection for dengue outbreaks in the Central Region of Malaysia using climatic drivers at mesoscale and synoptic scale," *Clim. Risk Manag.*, vol. 36, no. 100429, 2022, doi: 10.1016/j.crm.2022.100429.