ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 15, Nomor 2, Tahun 2026, Halaman 471 - 482
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

# PERBANDINGAN ALGORITMA MCMC DAN INLA DALAM ESTIMASI MODEL GAMM BAYESIAN UNTUK KASUS DEMAM BERDARAH *DENGUE* DI JAWA BARAT

**Abdullah Zeith Alfarizy [1], Indah Dzulkharifah[1], Afiq Dzakwan Anasti[1], Nida Khairunnissa[1], Luthfia Aiman Hanin[1], Laily Nissa Atul Mualifah[1*], Kusman Sadik[1], Anik Djuraidah[1], Anang Kurnia[1]**
[1] Program Studi Statistika dan Sains Data, IPB University
***e-mail**: lailyatul@apps.ipb.ac.id

**DOI: 10.14710/j.gauss.15.2.471-482**

**Article Info:**
Received: 2026-06-12
Accepted: 2026-11-12
Available Online: 2026-11-25

**Keywords:**
Bayesian GAMM; Dengue Hemorragic Fever; INLA; MCMC; Negative Binomial

**Abstract:** Dengue hemorrhagic fever (DHF) remains a major public health challenge in West Java, with climate and sociodemographic factors potentially exhibiting nonlinear associations with disease incidence. Bayesian Generalized Additive Mixed Models (GAMM) accommodate nonlinear effects and regional heterogeneity, but Markov Chain Monte Carlo (MCMC) estimation is computationally demanding. This study aimed to identify factors associated with DHF incidence and compare Integrated Nested Laplace Approximation (INLA) and MCMC in estimating a Bayesian Negative Binomial GAMM. The model used data from 27 districts and cities in West Java during 2019–2025. INLA produced posterior estimates highly consistent with MCMC, with a posterior mean correlation of 0.997 and RMSE of 0.041, while being 24–40 times faster. No convincing difference in predictive performance was found based on WAIC. Population density was positively associated with DHF incidence, with increased risks in 2024 and 2025. Meanwhile, nonlinear components showed no statistically meaningful associations. These findings support INLA as a computationally efficient alternative for Bayesian GAMM inference.

## 1. PENDAHULUAN

Penyakit Demam Berdarah Dengue (DBD) masih menjadi tantangan kesehatan masyarakat di Indonesia, khususnya di Jawa Barat, dengan penyebaran yang dipengaruhi oleh faktor iklim dan sosiodemografi. Pada skala nasional, Mamenun *et al.* (2024) menemukan bahwa curah hujan, kelembapan relatif, dan suhu berkorelasi signifikan dengan insiden DBD. Pada level provinsi, penelitian Nugraha *et al*. (2021) mengidentifikasi pengaruh curah hujan dan angin terhadap kasus DBD di DKI Jakarta, sementara Giofandi *et al*. (2023) melalui pendekatan *Geographically Weighted Regression*, menemukan bahwa kepadatan penduduk menjadi faktor penting dalam penyebaran penyakit DBD. Fauzi *et al.* (2022) mengkaji tren temporal dan klaster spasial DBD di Provinsi Jawa Barat dan menemukan heterogenitas spasial yang signifikan dan tren temporal yang bervariasi antardaerah. Temuan tersebut menegaskan perlunya pemodelan yang mampu mengakomodasi variasi kejadian DBD antarwilayah. Dengan heterogenitas geografis dan demografis yang tinggi, identifikasi faktor risiko dan wilayah rawan menjadi penting untuk mendukung perumusan kebijakan intervensi kesehatan yang tepat sasaran.

Faktor iklim, seperti suhu dan curah hujan, sering menunjukkan hubungan yang nonlinier dengan risiko penyakit. Dalam kondisi tersebut, *Generalized Additive Mixed Model* (GAMM) mengakomodasi pola tersebut melalui fungsi penghalus (*smoothing functions*), serta heterogenitas antarwilayah dan dinamika temporal melalui efek acak wilayah dan waktu. Penelitian Wang *et al*. (2026) mengungkapkan bahwa Bayesian GAMM mampu menangkap

hubungan nonlinier antara faktor iklim dan variabel respons secara efektif. Oleh karena itu, pendekatan tersebut dinilai relevan untuk diterapkan pada data epidemiologi.

Pendekatan Bayesian GAMM menghadapi kendala komputasi, terutama pada *Markov Chain Monte Carlo* (MCMC) yang membutuhkan waktu konvergensi yang lama pada model dengan banyak parameter. Sebagai alternatif *Integrated Nested Laplace Approximations* (INLA) menggunakan aproksimasi deterministik yang lebih efisien. Salim, Satoto and Danardono (2025) menerapkan INLA untuk memodelkan dinamika spasiotemporal DBD di Daerah Istimewa Yogyakarta dengan prediktor iklim dan sosiodemografi, serta menunjukkan kelayakan metode tersebut pada konteks epidemiologi DBD di Indonesia. Andryan, Irawan and Yong (2026) membandingkan INLA dan MCMC dalam pemodelan spasial dan spasio-temporal DBD di Bandung menggunakan BYM, dengan hasil estimasi sebanding dan komputasi INLA lebih singkat. Namun, studi tersebut belum mempertimbangkan kovariat iklim dan sosiodemografi, overdispersi, serta hubungan nonlinier. Penelitian ini mengisi kesenjangan tersebut melalui perbandingan INLA dan MCMC pada GAMM Bayesian Negative Binomial dengan offset populasi, fungsi smooth, dan efek acak kabupaten/kota. Evaluasi mencakup performa prediktif, waktu komputasi, serta kedekatan estimasi dan distribusi posterior. Analisis faktor iklim dan sosial ekonomi menjadi kontribusi empiris pendukung.

Berdasarkan uraian tersebut, penelitian ini bertujuan membandingkan kesesuaian estimasi posterior, performa prediktif, dan efisiensi komputasi INLA dan MCMC dalam pemodelan Bayesian GAMM *Negative Binomial* untuk mengidentifikasi faktor-faktor yang berhubungan dengan kejadian DBD di Jawa Barat. Model mengintegrasikan kovariat iklim dan sosiodemografi serta efek acak dan temporal untuk mengakomodasi heterogenitas wilayah dan dinamika waktu. Hasil penelitian diharapkan memberikan kontribusi metodologis sekaligus rekomendasi pendekatan komputasi yang efisien untuk pemodelan epidemiologi.

## 2. TINJAUAN PUSTAKA

Demam Berdarah Dengue (DBD) merupakan penyakit menular yang ditularkan oleh nyamuk *Aedes aegypti* dan *Aedes albopictus* (Andryan, Irawan and Yong, 2026). Data jumlah kasus DBD merupakan data cacah (*count data*) yang sering mengalami overdispersi, yaitu kondisi varians data jauh melebihi rata-rata. Pada regresi Poisson dengan $Y_i \sim \text{Poisson}(\mu_i)$ berlaku seperti persamaan (1):

$$E(Y_i) = \text{Var}(Y_i) = \mu_i \tag{1}$$

Pelanggaran asumsi *equidispersion* pada regresi Poisson menyebabkan galat baku menjadi *underestimate* dan inferensi menjadi tidak valid sehingga distribusi *Negative Binomial* (NB) digunakan sebagai alternatif (Abdul Wahid, Suhaila and Rahman, 2021). *Negative Binomial* merupakan campuran (*mixture*) dari distribusi Poisson dan distribusi Gamma (Tiara *et al.*, 2023). Variabel respons $Y_i$ yang berdistribusi *Negative Binomial* dengan rata-rata $\mu_i > 0$ dan parameter dispersi $\theta > 0$ memiliki fungsi peluang didefinisikan pada persamaan (2):

$$f(y_i; \mu_i, \theta) = \left(\Gamma\left(y_i + \frac{1}{\theta}\right) \Big/ \Gamma\left(\frac{1}{\theta}\right) y_i!\right) \left(\frac{1}{1+\theta\mu_i}\right)^{\frac{1}{\theta}} \left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i}, y_i = 0,1,2,\ldots \tag{2}$$

Kondisi overdispersi pada data cacah ditandai dengan $\text{Var}(Y_i) > E(Y_i)$ sehingga parameter dispersi $\theta$ menjadi komponen yang menampung variabilitas tambahan di luar variabilitas Poisson (Tiara *et al.*, 2023).

Penaksiran parameter regresi *Negative Binomial* dilakukan menggunakan metode *Maximum Likelihood Estimation* (MLE). Fungsi *likelihood* untuk $n$ pengamatan independen didefinisikan pada persamaan (3):

$$L(\boldsymbol{\beta},\theta)=\prod_{i=1}^{n}\left(\Gamma\left(y_i+\frac{1}{\theta}\right)/\Gamma\left(\frac{1}{\theta}\right)\,y_i!\right)\left(\frac{1}{1+\theta\mu_i}\right)^{\frac{1}{\theta}}\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i} \quad (3)$$

dengan $\ln L(\boldsymbol{\beta},\theta)=\sum_{i=1}^{n}\ln f(y_i;\mu_i,\theta)$ dimaksimumkan melalui iterasi *Newton-Raphson* untuk memperoleh taksiran parameter (Tiara *et al.*, 2023).

Hubungan antara variabel prediktor dan jumlah kasus penyakit menular sering bersifat nonlinier, disertai korelasi temporal dan heterogenitas antarwilayah (Abdul Wahid, Suhaila and Rahman, 2021). *Generalized Linear Model* (GLM) kurang sesuai karena mengasumsikan hubungan linier dan pengamatan independen. *Generalized Additive Model* (GAM) mengakomodasi hubungan nonlinier melalui fungsi penghalus (*smooth function*), sedangkan *Generalized Additive Mixed Model* (GAMM) menambahkan efek acak (random effect) untuk menangkap heterogenitas antarunit pengamatan (Abdul Wahid, Suhaila and Rahman, 2021). Pada penelitian ini, variabel respons $Y_{ij}$ merupakan jumlah kasus DBD pada wilayah ke-$i$ dan tahun ke-$j$ yang diasumsikan $Y_{ij}\sim \mathrm{NB}(\mu_{ij},\theta)$. Model GAMM Negative Binomial dengan fungsi tautan logaritma dan *offset* populasi dirumuskan pada persamaan (4):

$$\ln(\mu_{ij})=\beta_0+\gamma_j+\sum_{k=1}^{p}\beta_k\;x_{kij}+\sum_{l=1}^{q}f_l\left(z_{lij}\right)+u_i+\ln(E_{ij}) \quad (4)$$

dengan $\beta_0$ adalah intersep, $\gamma_j$ efek tetap tahun ke-$j$, $\beta_k$ koefisien untuk variabel prediktor linier $x_{kij}$, $f_l(\cdot)$ fungsi penghalus untuk variabel prediktor $z_{lij}$, dan $\ln(E_{ij})$ merupakan *offset* populasi (Abdul Wahid, Suhaila and Rahman, 2021). Model probabilistik untuk unit wilayah ke-$i$ dan tahun ke-$j$ dispesifikasikan sebagai variabel respons $Y_{ij}\sim \mathrm{NB}(\mu_{ij},\theta)$ dengan $\mathrm{Var}(Y_{ij})=\mu_{ij}+\theta\mu_{ij}^2$. Efek acak wilayah ke-$i$ diasumsikan sebagai berikut:

$$u_i\sim N(0,\sigma_u^2), i=1,2,\dots,m \quad (5)$$

dengan $\sigma_u^2$ menyatakan ragam efek acak yang menangkap heterogenitas antarwilayah (Abdul Wahid, Suhaila and Rahman, 2021). Fungsi penghalus $f_l(\cdot)$ pada persamaan (4) direpresentasikan menggunakan kombinasi linier basis B-*spline* didefinisikan pada persamaan (6):

$$f_l(z)=\sum_{m=1}^{K}\alpha_{lm}\;B_m(z) \quad (6)$$

dengan $B_m(\cdot)$ merupakan fungsi basis B-*spline* ke-$m$ dan $\alpha_{lm}$ koefisien yang akan ditaksir (Gómez-Rubio, 2020). Spesifikasi prior pada kedua pendekatan dirancang setara secara konseptual. Implementasi MCMC menggunakan packages brms dengan pendekatan P-*spline* berbasis B-*spline* (K=15, bs=''ps'') yang menerapkan penalti pada selisih orde dua koefisien basis. Pada implementasi INLA, penalti tersebut disetarakan dengan prior *Random Walk* orde dua (RW2) pada koefisien basis berikut:

$$\alpha_{lm}-2\,\alpha_{l,m-1}+\alpha_{l,m-2}\sim N\left(0,\frac{1}{\tau_l}\right) \quad (7)$$

dengan $\tau_l$ hiperparameter presisi yang mengendalikan tingkat penghalusan untuk prediktor ke-$l$ (van Niekerk *et al.*, 2021). Kedua pendekatan secara konseptual ekuivalen karena keduanya menerapkan penalti pada selisih orde dua sehingga menghasilkan penghalusan yang setara pada skala fungsi.

Prior yang digunakan pada *brms* adalah *weekly informative* untuk koefisien efek tetap $\beta_k\sim N(0,2.5)$ dan intersep $\beta_0\sim N(0,5)$, sementara hiperparameter presisi *smooth* $\tau_l$ dan ragam efek acak $\sigma_u^2$ menggunakan prior *half normal* yang merupakan *default brms*. Pada R-INLA, prior yang setara digunakan melalui *panalised complexity prior* untuk $\tau_l$ dan $\sigma_u^{-2}$ yang secara konseptual mendorong model menuju struktur yang lebih sederhana apabila data

tidak mendukung kompleksitas tambahan (van Niekerk *et al.*, 2021). Penaksiran parameter pada model GAMM *Negative Binomial* dilakukan menggunakan kerangka inferensi bayes, yaitu distribusi prior, fungsi *likelihood*, dan distribusi posterior (Andryan, Irawan and Yong, 2026). Diberikan vektor parameter $\boldsymbol{\psi}$ dan vektor data $\boldsymbol{y} = (y_1, \dots, y_n)$, distribusi posterior diperoleh melalui Teorema Bayes didefinisikan pada persamaan (8):

$$f(\boldsymbol{\psi} \mid \boldsymbol{y}) = \frac{l(\boldsymbol{y} \mid \boldsymbol{\psi})\, g(\boldsymbol{\psi})}{k(\boldsymbol{y})} \tag{8}$$

dengan fungsi *likelihood* bersama $l(\boldsymbol{y} \mid \boldsymbol{\psi}) = \prod_{i=1}^{n} l(y_i \mid \boldsymbol{\psi})$, $g(\boldsymbol{\psi})$ merupakan distribusi prior, dan $k(\boldsymbol{y})$ merupakan distribusi marginal data yang berperan sebagai konstanta normalisasi (Andryan, Irawan and Yong, 2026). Posterior umumnya tidak berbentuk tertutup sehingga diperlukan metode aproksimasi. Sebagai ilustrasi, taksiran posterior untuk koefisien kepadatan penduduk $\beta_1$ diperoleh dari distribusi posterior marginal berikut:

$$\tilde{\pi}\big((\beta_1|\boldsymbol{y}\big) = \int \pi(\beta_1|\boldsymbol{x}_{-\beta_1}, \tau, \boldsymbol{y})\tilde{\pi}(\tau|\, \boldsymbol{y})d\tau \tag{9}$$

dengan $\boldsymbol{x}_{-\beta_1}$ menyatakan seluruh komponen laten selain $\beta_1$. Taksiran titik diperoleh sebagai rata-rata posterior $E(\beta_1|\boldsymbol{y})$ dan ketidakpastian parameter diringkas melalui interval kredibel 95%. Prosedur yang sama berlaku untuk seluruh koefisien efek linear lainnya ($\beta_2$ air minum bersih, $\beta_3$ suhu, $\beta_4$ kecepatan angin), intersep $\beta_0$, efek tetap tahun $\gamma_j$, dan efek acak wilayah $u_i$. Penaksiran posterior untuk koefisien fungsi penghalus $\alpha_{lm}$ berbeda karena adanya struktur penalti RW2 pada persamaan (7). Koefisien $\alpha_{lm}$ tidak ditaksir secara independen melainkan secara bersama untuk seluruh $m = 1, \dots, K$ dalam satu prediktor *smooth* ke-$l$, dengan distribusi prior bersama yang diinduksi oleh penalti RW2 sebagai berikut:

$$\pi(\boldsymbol{\alpha_l}|\tau_l) \propto \tau_l^{\frac{(K-2)}{2}} \exp\left\{-\frac{\tau_l}{2}\boldsymbol{\alpha}_l^T \boldsymbol{P}\boldsymbol{\alpha_l}\right\} \tag{10}$$

dengan $\boldsymbol{P}$ adalah matriks penalti orde dua yang elemennya mencerminkan selisih $\alpha_{lm} - 2\alpha_{l,m-1} + \alpha_{l,m-2}$ pada persamaan (7) dan $\tau_l$ tidak diketahui, distribusi posterior marginal fungsi penghalus diperoleh dengan mengintegrasikan $\tau_l$ keluar dari distribusi bersama sehingga ketidakpastian pada tingkat penghalus juga diperhitungkan dalam inferensi. Hal ini yang membedakan penaksiran parameter fungsi penghalus dari efek linear yang priornya tidak bergantung pada hiperparameter penalti (Andryan, Irawan and Yong, 2026).

Metode *Markov Chain Monte Carlo* (MCMC) membangkitkan sampel posterior melalui rantai Markov yang distribusi stasionernya adalah posterior (Andryan, Irawan and Yong, 2026). Penelitian ini menggunakan algoritma *Hamiltonian Monte Carlo* (HMC) yang memanfaatkan informasi gradien log-posterior dan diimplementasikan melalui paket *brms* dengan algoritma *No-U-Turn Sampler* (NUTS) yang menentukan parameter sampler secara adaptif (McElreath, 2020). MCMC bersifat fleksibel untuk berbagai struktur model Bayes, namun memerlukan waktu komputasi yang panjang dan rentan terhadap masalah konvergensi rantai serta autokorelasi antar-sampel posterior (Gómez-Rubio, 2020). Metode *Integrated Nested Laplace Approximation* (INLA) adalah alternatif yang lebih efisien secara komputasi untuk inferensi pada kelas *Latent Gaussian Models* (LGM) (van Niekerk *et al.*, 2021). Model GAMM *Negative Binomial* termasuk LGM karena prediktor aditif pada persamaan (4) merupakan medan Gauss laten yang dikondisikan oleh hiperparameter presisi $\tau$. Vektor laten $\boldsymbol{w}$ pada LGM berdistribusi Gauss multivariat didefinisikan pada persamaan (11):

$$\pi(\boldsymbol{w} \mid \tau) \propto |\, \boldsymbol{Q}(\tau)\,|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\, \boldsymbol{w}^{\top}\boldsymbol{Q}(\tau)\, \boldsymbol{w}\right\} \tag{11}$$

dengan $\boldsymbol{Q}(\tau)$ adalah matriks presisi yang struktur kerenggangannya ditentukan oleh prior pada setiap komponen efek (van Niekerk *et al.*, 2021). Prediktor aditif pada LGM ditulis

sebagai:

$$\eta_{ij} = \beta_0 + \gamma_j + \sum_{k=1}^{p} \beta_k \; x_{kij} + \sum_{l=1}^{q} f_l\left(z_{lij}\right) + u_i + \ln(E_{ij}) \tag{12}$$

dengan $\eta_{ij} = \ln\left(\mu_{ij}\right)$, $\beta_0$ intersep, $\gamma_j$ efek tetap tahun, $\beta_k$ koefisien efek tetap untuk variabel $x_{kij}$, dan $f_l(\cdot)$ fungsi penghalus (Andryan, Irawan and Yong, 2026). Aproksimasi posterior pada INLA dilakukan melalui tiga tahap berurutan. Pertama, distribusi posterior marginal untuk hiperparameter diaproksimasi dengan aproksimasi *Laplace*.

$$\tilde{\pi}(\tau \mid \boldsymbol{y}) \propto \frac{\pi(\boldsymbol{w},\tau,\boldsymbol{y})}{\pi_G(\boldsymbol{w}|\tau,\boldsymbol{y})} \mid_{\boldsymbol{w}=\boldsymbol{w}^*(\tau)} \tag{13}$$

dengan $\pi_G(\boldsymbol{w} \mid \tau, \boldsymbol{y})$ merupakan aproksimasi Gauss terhadap distribusi penuh kondisional dan $\boldsymbol{w}^*(\tau)$ modus dari distribusi tersebut (van Niekerk *et al.*, 2021). Kedua, distribusi posterior marginal untuk setiap komponen laten diaproksimasi melalui aproksimasi *Laplace* tersarang. Ketiga, marginal posterior untuk setiap parameter laten diperoleh melalui integrasi numerik didefinisikan pada persamaan (14):

$$\tilde{\pi}(x_i \mid \boldsymbol{y}) = \sum_k \tilde{\pi}(w_i \mid \tau_k, \boldsymbol{y})\, \tilde{\pi}(\tau_k \mid \boldsymbol{y})\, \Delta_k \tag{14}$$

dengan $\tau_k$ titik kuadratur dan $\Delta_k$ bobot integrasi (van Niekerk *et al.*, 2021). INLA bersifat deterministik dan tidak memerlukan simulasi berulang seperti MCMC sehingga umumnya jauh lebih cepat. Andryan, Irawan and Yong (2026) membandingkan INLA dan MCMC dalam pemodelan spasial dan spasio-temporal DBD di Bandung. Hasilnya menunjukkan bahwa INLA menghasilkan estimasi yang sebanding dengan MCMC dengan waktu komputasi 50–70 kali lebih singkat. INLA juga cenderung menghasilkan interval kredibel yang lebih konservatif, sedangkan MCMC unggul dalam keakuratan asimtotik. Dengan demikian, pemilihan metode bergantung pada keseimbangan antara akurasi inferensi dan efisiensi komputasi. Pemilihan model menggunakan Watanabe-Akaike Information Criterion (WAIC), yang mempertimbangkan kesesuaian dan kompleksitas model berdasarkan distribusi posterior log-*likelihood*.

$$\text{WAIC} = -2\,(\text{lpd} - p_{\text{WAIC}}) \tag{15}$$

dengan $\text{lpd} = \sum_{ij} \log\{ E_{\boldsymbol{\psi}_*|\boldsymbol{y}}[\, f(y_{ij} \mid \boldsymbol{\psi}_*)\, ]\}$ merupakan *log pointwise predictive density* dan $p_{\text{WAIC}} = \sum_{ij} \text{Var}_{\boldsymbol{\psi}_*|\boldsymbol{y}}\{\log f(\, y_{ij} \mid \boldsymbol{\psi}_*\,)\}$ adalah penalti kompleksitas yang dihitung berbasis varians posterior log-*likelihood* (Gómez-Rubio, 2020). Model dengan WAIC yang lebih kecil mengindikasikan kesesuaian model yang lebih baik.

## 3. METODE PENELITIAN

Data sekunder diperoleh dari BPS, Dinas Kesehatan Provinsi Jawa Barat, dan Google Earth Engine (GEE), mencakup 27 kabupaten/kota selama 2019–2025 (189 observasi). Data kasus DBD bersumber dari Dinas Kesehatan, sedangkan variabel sosial ekonomi dan jumlah penduduk (offset) dari BPS. Seluruh variabel lengkap tanpa imputasi maupun penghapusan observasi. Data iklim meliputi curah hujan CHIRPS Daily v2.0 (0,05°) serta suhu, titik embun, dan komponen angin ERA5-Land Hourly (0,1°). Curah hujan diakumulasikan tahunan, suhu dan titik embun dikonversi ke Celsius, serta kecepatan angin dihitung dari komponen (u) dan (v). Selanjutnya, variabel iklim diagregasi menggunakan reduceRegions berdasarkan rata-rata piksel pada poligon GADM level 2, kemudian dipadankan dengan data DBD dan sosial ekonomi berdasarkan kabupaten/kota dan tahun. Variabel-variabel yang digunakan dalam penelitian ini ditampilkan pada Tabel 1. Tahapan analisis yang dilakukan dalam penelitian ini meliputi:

1. Melakukan eksplorasi data untuk mengidentifikasi karakteristik distribusi variabel respons, termasuk pemeriksaan overdispersi melalui rasio varians terhadap rata-rata, *skewness* dan kurtosis dengan mengacu pada Persamaan (1) – (3). Eksplorasi juga mencakup pemeriksaan multikolinearitas antar-variabel penjelas menggunakan

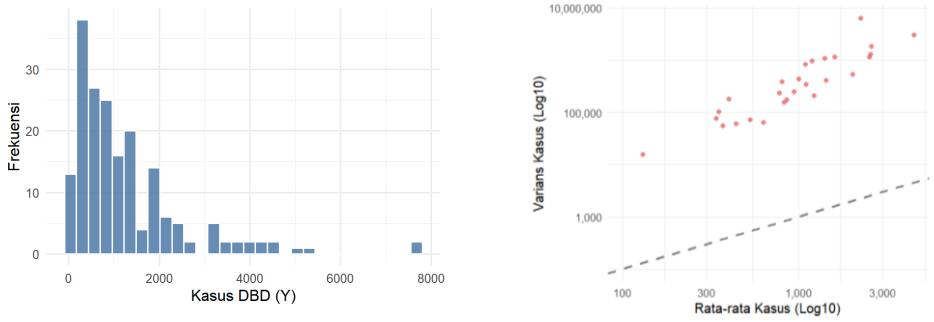
*Generalized Variance Inflation Factor* (GVIF) serta deteksi indikasi hubungan nonlinier melalui visualisasi awal.
2. Melakukan standarisasi menggunakan Z-*score* untuk mengubah skala semua variabel sehingga memiliki nilai tengah 0 dan simpangan baku 1 guna membantu stabilitas numerik saat estimasi model (Gelman, Hill and Vehtari, 2020).
3. Menentukan spesifikasi model GAMM Negative Binomial dengan offset logaritma populasi, efek tetap tahun, prediktor linier, efek acak wilayah berdistribusi normal (Persamaan 5), serta fungsi penghalus berbasis spline (Persamaan 6) dengan penalti perbedaan orde kedua (Persamaan 7), sebagaimana dirumuskan pada Persamaan (4).
4. Mengestimasi parameter model GAMM *Negative Binomial* menggunakan kerangka inferensi Bayes pada Persamaan (8), melalui dua pendekatan, yaitu INLA yang diimplementasikan melalui paket R-INLA, dan MCMC dengan algoritma HMC yang diimplementasikan melalui paket *brms.*
5. Memeriksa diagnostik konvergensi MCMC menggunakan statistik *Gelman-Rubin* ($\hat{R}$), *Effective Sample Size* (ESS), dan inspeksi keberadaan *divergent transitions*, sementara INLA menggunakan *Conditional Predictive Ordinate* (CPO).
6. Mengevaluasi kecocokan model dengan data melalui *posterior mean correlation*, khususnya melalui *density overlay*.
7. Membandingkan kedua pendekatan berdasarkan estimasi posterior parameter (titik taksiran dan interval kredibel), kinerja prediksi menggunakan WAIC (Persamaan 15), serta efisiensi komputasi berdasarkan waktu eksekusi algoritma.

Table 1. Variabel-variabel yang digunakan dalam analisis

| Variabel | Keterangan | Satuan | Sumber |
|---|---|---|---|
| Y | Jumlah kasus DBD | Jumlah | Dinkes Jawa Barat |
| X1 | Kepadatan penduduk | /km2 | BPS |
| X2 | Persentase penduduk miskin | Persen | BPS |
| X3 | Persentase rumah tangga yang memiliki Akses terhadap sanitasi layak | Persen | BPS |
| X4 | Persentase rumah tangga yang memiliki Akses terhadap sumber air minum layak | Persen | BPS |
| X5 | Persentase rumah tangga yang menempati Rumah layak huni | Persen | BPS |
| X6 | Curah hujan | mm | Google Earth Engine |
| X7 | Kecepatan angin | m/s | Google Earth Engine |
| X8 | Suhu rata-rata | °C | Google Earth Engine |
| X9 | Titik embun | °C | Google Earth Engine |
| X10 | Tahun | - | BPS |
| Offset | Jumlah penduduk per kabupaten/kota | Jumlah | BPS |

## 4. HASIL DAN PEMBAHASAN

Eksplorasi variabel respons menunjukkan bahwa jumlah kasus DBD memiliki variasi yang sangat lebar, dengan rata-rata 1250.51 kasus dan simpangan baku 1262.01 kasus, serta rentang dari 38 hingga 7680 kasus. Nilai *skewness* sebesar 2.234 dan kurtosis sebesar 9.544 mengindikasikan overdispersi dengan distribusi sangat menceng ke kanan dan berekor panjang yang dapat dilihat pada Gambar 1(a). Pola serupa terlihat pada kepadatan penduduk yang juga sangat beragam antarwilayah.




(a) (b)

Gambar 1. (a) Sebaran frekuensi dan (b) perbandingan rerata dengan varians kasus DBD Jawa Barat

Hubungan antara rata-rata dan varians kasus DBD menunjukkan bahwa hampir seluruh titik berada jauh di atas garis *equidispersion* (varians = rata-rata) yang menjadi karakteristik distribusi Poisson. Temuan tersebut menunjukkan adanya overdispersi yang kuat pada variabel respons sehingga asumsi *equidispersion* pada distribusi Poisson tidak terpenuhi. Oleh karena itu, distribusi *Negative Binomial* dipilih sebagai distribusi respons dalam pemodelan karena mampu mengakomodasi varians yang lebih besar daripada rata-rata melalui parameter dispersi tambahan (Payne *et al.*, 2018). Selain itu, besarnya variasi kasus antarwilayah mengindikasikan adanya heterogenitas yang tidak sepenuhnya dapat dijelaskan oleh kovariat yang diamati sehingga penggunaan efek acak wilayah dalam kerangka GAMM menjadi relevan (Pedersen *et al.*, 2019).

Tabel 2. Nilai GVIF untuk menentukan multikolinearitas

| Prediktor | GVIF | df | $GVIF^{\left(\frac{1}{2df}\right)}$ |
|---|---|---|---|
| Tahun | 7.375 | 6 | 1.181 |
| Curah hujan | 5.749 | 1 | 2.397 |
| Kecepatan angin | 2.252 | 1 | 1.500 |
| Suhu rata-rata | 21.656* | 1 | 4.653 |
| Titik embun | 15.762* | 1 | 3.970 |
| Kepadatan penduduk | 2.376 | 1 | 1.541 |
| Persentase penduduk miskin | 2.937 | 1 | 1.713 |
| Persentase rumah tangga yang memiliki Akses terhadap sanitasi layak | 7.386 | 1 | 2.717 |
| Persentase rumah tangga yang memiliki Akses terhadap sumber air minum layak | 2.042 | 1 | 1.715 |
| Persentase rumah tangga yang menempati Rumah layak huni | 8.191 | 1 | 2.862 |

* prediktor dengan multikolinearitas (GVIF > 10)

Pemeriksaan multikolinearitas menggunakan GVIF pada Tabel 2 menunjukkan suhu dan titik embun memiliki multikolinearitas, sehingga berpotensi meningkatkan ketidakstabilan estimasi parameter model (James *et al.*, 2021). Berdasarkan pertimbangan interpretabilitas epidemiologis, suhu rata-rata dipertahankan karena lebih mudah dikomunikasikan, sedangkan titik embun dikeluarkan dari pemodelan lanjutan. Variabel lainnya memiliki nilai GVIF yang berada dalam batas wajar sehingga tetap dipertahankan.

Eksplorasi nonlinieritas melalui GAM frekuentis menghasilkan nilai *Effective Degrees of Freedom* (EDF) sebagaimana disajikan pada Tabel 3. Curah hujan, persentase sanitasi layak, persentase rumah layak huni, dan persentase penduduk miskin menunjukkan nilai EDF yang lebih besar dari satu, yang mengindikasikan adanya pola hubungan nonlinier. Sebaliknya, kepadatan penduduk, persentase air minum layak, suhu rata-rata, dan kecepatan

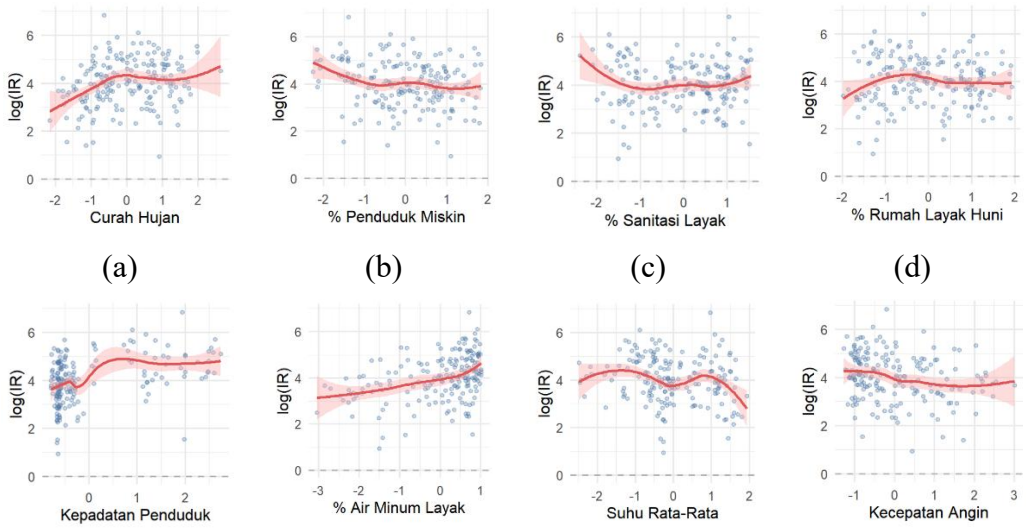
angin memiliki nilai EDF mendekati satu sehingga lebih sesuai dimodelkan sebagai efek linier. Nilai EDF yang semakin besar menunjukkan kompleksitas fungsi *smooth* yang semakin tinggi. Pada model GAM, EDF sebesar 1 ekuivalen dengan hubungan linier, sedangkan EDF lebih besar dari 1 menunjukkan bahwa bentuk hubungan memerlukan derajat kelengkungan tambahan untuk menjelaskan pola data (Wood, 2017).

Tabel 3. Nilai EDF dari GAM Awal

| Komponen *Smooth* | EDF |
|---|---|
| s(curah hujan) | 2.731* |
| s(% sanitasi layak) | 3.072* |
| s(% rumah layak huni) | 2.747* |
| s(% penduduk miskin) | 3.276* |
| s(kepadatan penduduk | 1.000 |
| s(% air minum layak) | 1.000 |
| s(suhu rata-rata) | 1.015 |
| s(kecepatan angin) | 1.001 |

*nilai EDF >1

Berdasarkan bukti gabungan dari nilai EDF, pola visual kurva *Locally Estimated Scatterplot Smoothing* (LOESS) pada Gambar 2 dan signifikansi komponen *smooth*, ditetapkan empat prediktor sebagai komponen nonlinier (*smooth*), yaitu curah hujan, persentase penduduk miskin, persentase sanitasi layak, dan persentase rumah layak huni yang ditunjukkan secara berurutan pada gambar 2a hingga 2d. Empat prediktor lainnya, yaitu kepadatan penduduk, persentase air minum layak, suhu rata-rata, dan kecepatan angin, dimodelkan sebagai efek linier (gambar 2e hingga 2h). Klasifikasi ini menjadi dasar penyusunan struktur model GAMM Bayesian.




Gambar 2. Kurva hubungan nonlinier log(IR) dan prediktor
(a) curah hujan, (b) penduduk miskin, (c) sanitasi layak, (d) rumah layak huni,
(e) kepadatan penduduk, (f) air minum layak, (g) sanitasi layak, dan (h) kecepatan angin

Hasil diagnostik menunjukkan bahwa kedua pendekatan memiliki kelayakan inferensi yang memadai. Pada model MCMC, empat *chain* dijalankan dengan masing-masing 4000 iterasi, termasuk 2000 iterasi *warm-up*, menggunakan algoritma NUTS dengan target penerimaan 0.95 dan *maximum tree depth* sebesar 12. Nilai R-hat maksimum sebesar 1.002,

ESS bulk minimum 2583, dan ESS tail minimum 3682 menunjukkan konvergensi yang baik dengan jumlah sampel efektif yang memadai. Terdapat dua *divergent transitions* dari 8000 *posterior draws* (0.025%), namun tidak terdapat iterasi yang mencapai batas *tree depth* dan nilai E-BFMI minimum sebesar 0.797 menunjukkan eksplorasi distribusi energi yang memadai. Hasil sampling MCMC dapat digunakan untuk inferensi posterior, dengan tetap mencatat *divergent transitions*. Sementara itu, model INLA menggunakan *simplified Laplace approximation* dengan strategi integrasi otomatis. Satu dari 189 observasi (0.53%), yaitu Kota Tasikmalaya tahun 2025 mengalami CPO failure dan dihitung ulang yang mengubah nilai CPO dari $3.22 \times 10^{-5}$ menjadi $5.16 \times 10^{-5}$ serta nilai *Probability Integral Transform* (PIT) dari 0.901 menjadi 0.842. Evaluasi kalibrasi pada 188 observasi tanpa *failure* menghasilkan statistik Kolmogorov–Smirnov sebesar 0.090 ($p = 0.098$) sehingga tidak ditemukan penyimpangan distribusi PIT yang signifikan dari Uniform(0,1) pada taraf 5%.

Tabel 4. Ringkasan konsistensi *posterior fixed effects* antara INLA dan MCMC

| Metrik | Nilai |
|---|---|
| Korelasi posterior mean | 0.9967 |
| RMSE posterior mean | 0.0408 |
| Korelasi *Credible Interval* (CI) | 0.9997 |
| Rata-rata *standardized difference* | 0.1600 |
| Maksimum *standardized differenc* | 0.2660 |

Hasil pada Tabel 4 menunjukkan kesesuaian estimasi posterior yang tinggi antara INLA dan MCMC. Pada 10 koefisien efek tetap, korelasi posterior mean sebesar 0.9967 dan RMSE 0.0408 menunjukkan kedekatan estimasi titik. Namun, karakteristik distribusi posterior masih berbeda. Selisih posterior mean terstandar maksimum mencapai 0.266, sementara SD posterior INLA lebih kecil 2–6% pada efek tetap dan 1.5% pada parameter dispersi. Jarak Wasserstein-1 terstandar berkisar 0.028–0.264 pada efek tetap dan 0.148 pada dispersi. Proporsi posterior draws MCMC dalam interval kredibel 95% INLA sebesar 0.910–0.944, dengan perbedaan batas interval pada efek tahun 2021 dan 2025 di sekitar nol. Pada 27 efek acak kabupaten/kota, korelasi mencapai 0.9971 dan RMSE 0.0345, meskipun standar deviasi INLA 16% lebih kecil (0.089 vs 0.106) dan jarak Wasserstein-1 terstandar 0.179. Temuan ini sejalan dengan Wang *et al.*, (2026) bahwa INLA menghasilkan inferensi sebanding dengan MCMC pada pemodelan penyakit berbasis areal.

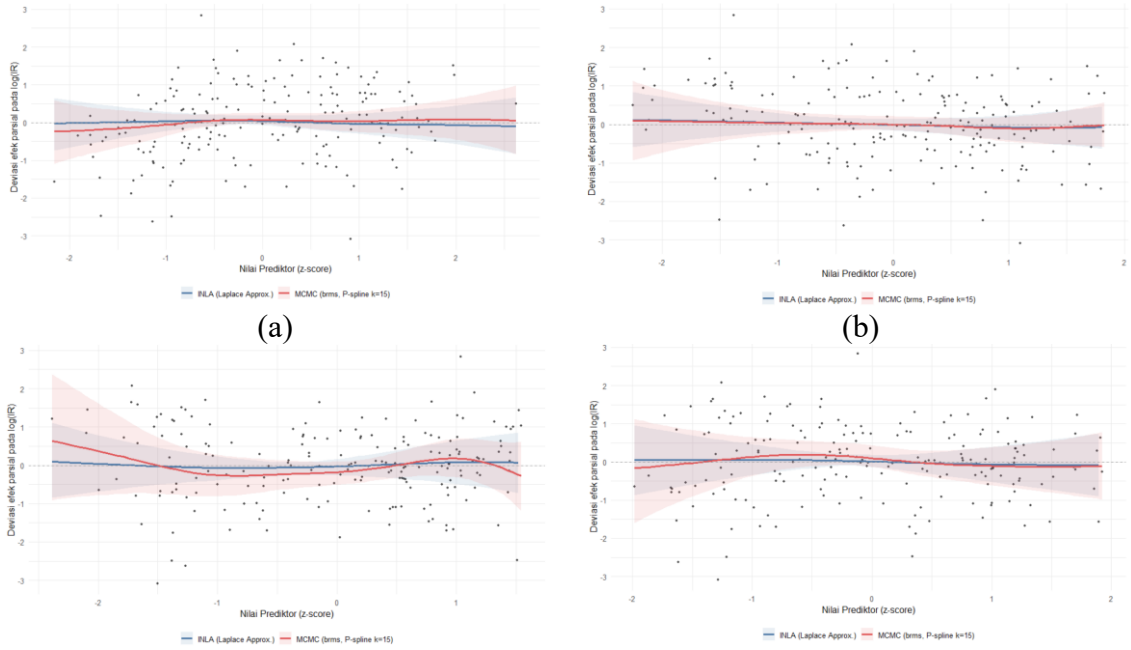
Tabel 5. Perbandingan waktu komputasi INLA dan MCMC

| Metode | Waktu (detik) | Waktu (menit) | Rasio terhadap INLA |
|---|---|---|---|
| INLA | 4.20 | 0.07 | 1 |
| MCMC | 168.30 | 2.80 | 40 |

Perbandingan waktu komputasi dilakukan pada Windows 11 x64, R versi 4.4.1, dengan 8 logical cores (4 physical cores) dan RAM 8 GB. Analisis menggunakan INLA versi 24.12.11 dengan 4 threads, sedangkan MCMC menggunakan brms versi 2.23.0 dan rstan versi 2.32.7. Pada pengukuran awal, INLA memerlukan 4.2 detik, sedangkan MCMC 168.3 detik atau sekitar 40 kali lebih lama, termasuk 113.3 detik untuk kompilasi Stan dan overhead. Tanpa waktu kompilasi, benchmark berulang menghasilkan median waktu INLA sebesar 3.26 detik (rentang 2.99–4.50 detik pada 10 pengulangan) dan MCMC sebesar 76.9 detik (rentang 75.1–81.4 detik pada 5 pengulangan), sehingga INLA sekitar 24 kali lebih cepat. Temuan ini sejalan dengan Chen *et al*. (2024) mengenai efisiensi INLA dibandingkan MCMC berbasis HMC. Kedua model menggunakan distribusi Negative Binomial, kovariat,

offset populasi, dan efek acak kabupaten/kota yang sama. Perbedaan terdapat pada fungsi nonlinier (RW2 dengan 50 interval pada INLA dan P-spline dengan 15 basis pada MCMC) serta spesifikasi prior. Dengan demikian, keunggulan komputasi INLA sebesar 24–40 kali berlaku pada 189 observasi dan konfigurasi penelitian ini.

Perbandingan WAIC menunjukkan nilai INLA sebesar 2825.92 lebih rendah dibandingkan MCMC sebesar 2828.92 , dengan korelasi log-*likelihood pointwise* sebesar 0.998 pada 189 observasi. Namun, selisih WAIC sebesar 3.00 hanya mencapai 1.42 kali *standard error* (2.11) sehingga belum menunjukkan perbedaan prediktif yang meyakinkan. Selain itu, diagnostik WAIC MCMC menunjukkan 25 observasi (13.2%) memiliki pWAIC lebih dari 0.4, mengindikasikan potensi ketidakstabilan estimasi prediktif *out-of-sample* akibat tingginya variasi log-l*ikelihood* posterior. Dengan demikian, belum terdapat bukti cukup untuk menyimpulkan INLA lebih unggul maupun membuktikan kesetaraan performa prediktif kedua metode.




(a) (b)

(c) (d)

Gambar 3. Kurva efek parsial *smooth* INLA vs MCMC: (a) curah hujan, (b) % penduduk miskin, (c) % sanitasi layak, (d) % rumah layak huni

Perbandingan fungsi *smooth* pada Gambar 3 menunjukkan pola kurva yang serupa antara INLA dan MCMC pada keempat prediktor nonlinier. Kesesuaian tertinggi terlihat pada persentase penduduk miskin, sedangkan perbedaan lokal pada prediktor lainnya terutama muncul di bagian ekor distribusi. Secara umum, kedua metode menangkap pola nonlinier yang konsisten. Oleh karena itu, dengan mempertimbangkan kesesuaian estimasi posterior dan efisiensi komputasi, interpretasi selanjutnya didasarkan pada hasil model INLA.

Tabel 6. Prediktor linier yang signifikan terhadap kejadian DBD pada model INLA

| Prediktor | Posterior Mean | 95% Credible Interval | IRR |
|---|---|---|---|
| Tahun 2024 | 0.871 | (0.493;1.243) | 2.390 |
| Tahun 2025 | 0.503 | (0.016;0.987) | 1.654 |
| Kepadatan penduduk | 0.377 | (0.165;0.587) | 1.458 |

Tabel 6 menunjukkan bahwa kepadatan penduduk berpengaruh positif terhadap kejadian DBD. Nilai Incidence Rate Ratio (IRR) sebesar 1.46 mengindikasikan bahwa peningkatan satu simpangan baku kepadatan penduduk berkaitan dengan peningkatan risiko DBD sebesar 45.8%. Temuan ini sejalan dengan Amelinda, Wulandari and Asyary (2022) yang menunjukkan bahwa kepadatan penduduk lebih tinggi cenderung meningkatkan incidence rate DBD akibat meningkatnya kontak antara vektor dan manusia. Selain itu, variasi temporal juga menunjukkan pengaruh yang signifikan. Dibandingkan tahun 2019 sebagai tahun referensi, risiko DBD pada tahun 2024 dan 2025 masing-masing meningkat sebesar 139.0% dan 65.4%, dengan peningkatan terbesar terjadi pada tahun 2024. Batas bawah CI tahun 2025 sangat dekat dengan nol (0.016), sehingga efek tahun tersebut perlu ditafsirkan secara hati-hati. Sementara itu, akses air minum layak, suhu rata-rata, dan kecepatan angin tidak menunjukkan pengaruh yang signifikan karena Credible Interval (CI) 95% yang masih mencakup nol.

Kurva smooth menunjukkan bahwa keempat prediktor memiliki pola hubungan yang relatif landai, dengan pita CI 95% yang mencakup nol di seluruh rentang pengamatan. Persentase penduduk miskin cenderung meningkat, curah hujan meningkat kemudian mendatar, sanitasi layak sedikit menurun lalu meningkat, sedangkan rumah layak huni relatif konstan. Hasil ini menunjukkan bahwa keempat prediktor belum memberikan bukti pengaruh nonlinier yang bermakna terhadap kejadian DBD setelah variabel lain dikendalikan. Pola kurva yang dihasilkan INLA dan MCMC juga relatif konsisten.

## 5. KESIMPULAN

Evaluasi terhadap GAMM Negative Binomial untuk permodelan kejadian DBD pada 27 kabupaten/kota di Jawa Barat selama 2019-2025 memperlihatkan bahwa INLA mampu menghasilkan estimasi posterior yang sangat dekat dengan MCMC, dengan korelasi posterior mean sebesar 0.9967 dan RMSE sebesar 0.0408, meskipun tidak identik dan memiliki *standard error* posterior sedikit lebih kecil. Kedua metode menghasilkan performa prediktif yang relatif sebanding, sedangkan INLA membutuhkan waktu komputasi 24 hingga 40 kali lebih singkat. Berdasarkan kesesuaian hasil inferensi dan efisiensi komputasinya, INLA selanjutnya digunakan sebagai dasar interpretasi faktor-faktor yang berhubungan dengan kejadian DBD. Hasil pemodelan menunjukkan bahwa kepadatan penduduk berhubungan positif dengan kejadian DBD, disertai peningkatan risiko pada tahun 2024 dan 2025. Sementara itu, komponen nonlinier belum menunjukkan hubungan yang bermakna. Temuan ini perlu diinterpretasikan dengan mempertimbangkan keterbatasan penggunaan satu data empiris dan konfigurasi model yang digunakan.

## DAFTAR PUSTAKA


Abdul Wahid, N.A., Suhaila, J. and Rahman, H.A. (2021) 'Effect of climate factors on the incidence of hand, foot, and mouth disease in Malaysia: A generalized additive mixed model', *Infectious Disease Modelling*, 6, pp. 997–1008. Available at: https://doi.org/10.1016/j.idm.2021.08.003.

Amelinda, Y.S., Wulandari, R.A. and Asyary, A. (2022) '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 Biomedica*, 93(6), p. e2022323. Available at: https://doi.org/10.23750/abm.v93i6.13503.

Andryan, D., Irawan, R. and Yong, B. (2026) 'Bayesian spatial and spatio-temporal modeling of Dengue transmission in Bandung: a comparative study of INLA and MCMC approaches', *Journal of Spatial Science*, pp. 1–18. Available at:

https://doi.org/10.1080/14498596.2026.2629378.
Chen, Z. *et al.* (2024) 'A comparison of computational algorithms for the Bayesian analysis of clinical trials', *Clinical Trials*, 21(6), pp. 689–700. Available at: https://doi.org/10.1177/17407745241247334.
Fauzi, I.S. *et al.* (2022) 'Temporal trend and spatial clustering of the dengue fever prevalence in West Java, Indonesia', *Heliyon*, 8, p. e10350. Available at: https://doi.org/10.1016/j.heliyon.2022.e10350.
Gelman, A., Hill, J. and Vehtari, A. (2020) *Regression and other stories*. Cambridge University Press.
Giofandi, E.A. *et al.* (2023) 'Analisis faktor spasial terhadap kejadian demam berdarah dengue menggunakan pendekatan Geographically Weighted Regression di Kota Pekanbaru, Provinsi Riau', *Jurnal Ilmu Lingkungan*, 22(1), pp. 50–59. Available at: https://doi.org/10.14710/jil.22.1.50-59.
Gómez-Rubio, V. (2020) *Bayesian inference with INLA*. Taylor & Francis
James, G. *et al.* (2021) *An introduction to statistical learning: With applications in R*. 2nd edn. Springer.
Mamenun *et al.* (2024) 'Spatiotemporal characterization of dengue incidence and its correlation to climate parameters in Indonesia', *Insects*, 15. Available at: https://doi.org/10.3390/insects15050366.
McElreath, R. (2020) *Statistical rethinking: A bayesian course with examples in r and stan*. 2nd edn. Chapman and Hall/CRC.
van Niekerk, J. *et al.* (2021) 'New frontiers in bayesian modeling using the INLA package in R', *Journal of Statistical Software*, 100(2), pp. 1–28. Available at: https://doi.org/10.18637/JSS.V100.I02.
Nugraha, F. *et al.* (2021) 'Studi ekologi hubungan kejadian Demam Berdarah Dengue (DBD) dengan faktor iklim di Kota Administrasi Jakarta Pusat, Indonesia tahun 1999-2018', *Jurnal Ilmu Kesehatan*, 10(3), pp. 142–148. Available at: https://doi.org/10.33221/jikm.v10i03.923.
Payne, E.H. *et al.* (2018) 'An empirical approach to determine a threshold for assessing overdispersion in poisson and negative binomial models for count data', 47(6), pp. 1722–1738. Available at: https://doi.org/10.1080/03610918.2017.1323223.
Pedersen, E.J. *et al.* (2019) 'Hierarchical generalized additive models in ecology: An introduction with mgcv', *PeerJ*, 7, p. e6876. Available at: https://doi.org/10.7717/peerj.6876.
Salim, M.F., Satoto, T.B.T. and Danardono (2025) 'Predicting spatio-temporal dynamics of dengue using INLA (integrated nested laplace approximation) in Yogyakarta, Indonesia', *BMC Public Health*, 25, p. 1321. Available at: https://doi.org/10.1186/s12889-025-22545-2.
Tiara, Y. *et al.* (2023) 'Overdispersion handling in poisson regression model by applying negative binomial regression', *Barekeng*, 17(1), pp. 417–426. Available at: https://doi.org/10.30598/barekengvol17iss1pp0417-0426.
Wang, Y. *et al.* (2026) 'How does climate influence mangrove litterfall production across different species? A case study in Zhanjiang, China', *Trees, Forests and People*, 23(December 2025), p. 101137. Available at: https://doi.org/10.1016/j.tfp.2025.101137.
Wood, S.N. (2017) *Generalized additive models: An introduction with R*. 2nd edn. Chapman and Hall/CRC.