Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bayesian disease mapping from incomplete hospital-based surveillance: accounting for catchment and missing case origins in dengue surveillance in Makassar, Indonesia

Bobby Poerwanto

bobby.poerwanto@postgrad.manchester.ac.uk

University of Manchester

Ian Hall

University of Manchester

Emma L Fairbanks

University of Manchester

Thomas House

University of Manchester

Olatunji Johnson

University of Manchester

Research Article

Keywords: Dengue, Bayesian disease mapping, Inhomogeneous Poisson process, Hospital catchment, Missing location data, Cross-validation

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11332410/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Bayesian disease mapping from incomplete hospital-based surveillance: accounting for catchment and missing case origins in dengue surveillance in Makassar, Indonesia

Bobby Poerwanto[1,2*], Ian Hall[1†], Emma L Fairbanks[1†], Thomas House[1†], Olatunji Johnson[1†]

[1*]Department of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, Greater Manchester, United Kingdom.
[2]Department of Statistics, Universitas Negeri Makassar, Dg. Tata Raya, Makassar, 90224, Sulawesi Selatan, Indonesia.

*Corresponding author(s). E-mail(s): bobby.poerwanto@postgrad.manchester.ac.uk;
Contributing authors: ian.hall@manchester.ac.uk; emma.fairbanks@manchester.ac.uk; thomas.house@manchester.ac.uk; olatunji.johnson@manchester.ac.uk;
†These authors contributed equally to this work.

Abstract

Dengue surveillance data are typically collected at healthcare facilities rather than at patients' residences, so reported case locations reflect healthcare-seeking behaviour as much as the underlying disease process. In practice, cases carry differing levels of location detail: some are geolocated, some are known only to the subdistrict, and some carry no residential information at all. This incompleteness complicates spatial risk estimation. We develop a hierarchical Bayesian point-process framework that maps dengue risk while jointly modelling hospital catchment behaviour and unobserved case origins. Subdistrict-level intensity is modelled on a per-person basis using a Besag–York–Mollié (BYM) structure with a population offset; a distance-based gravity (power-law) kernel represents hospital attraction; and cases with incomplete location are handled within a single likelihood that accommodates three levels of spatial resolution and marginalises over unknown origins. Applied to 1,842 dengue cases from eight hospitals in

Makassar, Indonesia (2021–2025), the framework recovers held-out case origins substantially better than a conventional exponential catchment kernel and than deterministic allocation baselines. The full model identifies Manggala as the subdistrict with the highest estimated relative risk (2.7), with Tamalate, Rappocini and Tallo also meeting the specified high-risk criterion. Rankings by model-estimated relative risk differ from rankings by recorded case count. The framework offers a principled approach to disease mapping from incomplete, facility-based surveillance data.

Keywords: Dengue, Bayesian disease mapping, Inhomogeneous Poisson process, Hospital catchment, Missing location data, Cross-validation

1 Introduction

Dengue fever is among the most significant vector-borne diseases globally, with its burden rising rapidly in tropical and subtropical regions. Despite longstanding recognition of the climatic, environmental, and demographic determinants of dengue transmission, its burden remains highly heterogeneous across space, particularly in rapidly urbanising settings in Southeast Asia [1, 2].

Globally, reported dengue cases increased from about half a million in the year 2000 to over 5 million in 2019. In 2024, the estimated burden reached over 14 million cases and 9,500 deaths [3]. Indonesia consistently ranks among the countries with the highest dengue burden in Southeast Asia. In fact, the World Health Organization identified Indonesia as one of the world's most dengue-endemic nations from 2004 to 2010 [4].

Dengue cases in Indonesia have remained persistently high in recent years. The Ministry of Health reported 114,720 cases and 894 deaths in 2023 [5]. In 2024, a sharp increase occurred: by April, there were 88,593 cases and 621 deaths recorded [6], rising to 210,644 cases and 1,239 deaths by epidemiological week 43 (21-27 October 2024) [5]. Several factors have intensified dengue transmission dynamics, including rapid urbanization, population growth, and environmental changes, resulting in increasingly heterogeneous spatial risk patterns within cities. In addition, climatic variation, rainfall patterns, and environmental conditions that facilitate Aedes aegypti proliferation have further contributed to this surge [7].

Indonesia has recurrent dengue outbreaks in both urban and peri-urban settings [8, 9]. The spatial distribution of dengue incidence in Indonesian cities is highly heterogeneous, reflecting differences in environmental conditions, population density, and infrastructure [4]. In Makassar City, dengue cases are consistently reported, but notable variation across subdistricts suggests localized transmission dynamics that aggregated statistics do not adequately capture [10]. These patterns highlight the need for spatially explicit modeling approaches to identify high-risk areas and potentially inform spatial targeting of interventions.

A significant challenge in dengue surveillance is that case records are often generated at healthcare facilities, where the residential location of the patient is often

recorded incompletely or not at all. As a result, observed case counts reflect not only the disease process but also healthcare-seeking behavior, accessibility, and hospital catchment dynamics [8]. This reliance on facility-based data can introduce bias into spatial risk estimates if used directly as indicators of disease distribution. Therefore, distinguishing latent disease intensity from the observation process is essential for reliable spatial inference.

A wide range of statistical and spatial modeling approaches has been applied to dengue data in previous studies. Conventional regression frameworks, such as generalized linear models, typically assume independence between observations and therefore fail to account for the spatial autocorrelation that is intrinsic to infectious disease processes [11]. This limitation can lead to biased inference and underestimation of uncertainty when disease risk exhibits spatial clustering.

To address spatial dependence, conditional autoregressive (CAR) models explicitly incorporate neighbourhood structure, allowing adjacent areas to share information through spatial smoothing and improving estimation of area-level risk [12]. However, CAR models primarily capture structured spatial dependence and do not explicitly account for unstructured area-specific heterogeneity arising from local factors unrelated to geographical proximity. Bayesian hierarchical disease mapping models, particularly the Besag–York–Mollié (BYM) model, extend the CAR formulation by combining spatially structured and unstructured random effects within a unified hierarchical framework [13]. Subsequently, the reparameterised BYM2 model was introduced to improve parameter identifiability and facilitate principled prior specification while retaining the advantages of the original BYM model [14]. Together, these models provide a coherent Bayesian framework for integrating fixed effects, structured and unstructured spatial variation, and prior information, enabling robust disease mapping with comprehensive uncertainty quantification [9, 15].

Despite these methodological advances, two challenges remain insufficiently addressed in the analysis of routinely collected dengue surveillance data. First, most spatial models assume that the location of each case is fully and uniformly observed. In practice, hospital-based surveillance records location at differing levels of resolution: some cases are geolocated to a residential coordinate, some are known only to the subdistrict of residence, and some carry no residential information at all, being recorded only as a facility total [16, 17]. This is a form of coarsened and missing spatial data, and ignoring it, or replacing the unlocated cases by ad hoc imputation, can produce substantial spatial misclassification and biased risk estimates [18, 19].

Second, the observation mechanism linking residential disease occurrence to hospital-based records is rarely modelled explicitly. Patients may seek care at hospitals outside their subdistrict of residence, so the reporting hospital does not necessarily indicate where a case originated [16, 20]. Travel distance is an important determinant of spatial accessibility to healthcare facilities and is commonly represented using distance-decay or gravity models [21–24]. This provides a natural basis for representing the spatial catchment of each reporting hospital when residential locations are incomplete.

To address these limitations, we propose a unified hierarchical Bayesian framework that explicitly distinguishes between the latent disease process and the hospital

observation process. The model represents the underlying spatial intensity of dengue as an inhomogeneous Poisson process (IPP) with a population-adjusted Besag–York–Molli'e (BYM2) log-intensity. The hospital observation process is represented by a distance-based gravity kernel, in which the expected contribution of cases from a given residential area to a reporting hospital decreases with travel distance, with hospital-specific parameters allowing catchment patterns to vary across facilities [22, 24]. A single likelihood accommodates the varying resolution of the recorded locations: geolocated cases contribute their exact distance to the reporting hospital, subdistrict-level cases contribute their region of residence, and unlocated cases have their origins treated as latent variables and inferred within the model, with weights determined jointly by the estimated risk and distance-based accessibility. Uncertainty in the unlocated cases is thereby propagated through the posterior distribution rather than fixed by imputation [18].

Using dengue surveillance data from Makassar City, Indonesia, this study develops and evaluates an IPP–BYM2 framework for jointly modelling per-person dengue risk and hospital catchment behaviour under incomplete origin information. We make two contributions. Methodologically, we show that the framework's ability to recover the residential origins of unlocated cases can be evaluated directly by out-of-sample cross-validation, and that under this criterion a gravity (power-law) catchment kernel substantially outperforms both the conventional exponential kernel and deterministic allocation rules. Substantively, the full model identifies Manggala as the area with the highest estimated relative risk and distinguishes rankings by relative risk from those by recorded case count, providing evidence to inform further assessment of spatial control priorities.

2 Methods

Study area

Makassar, the capital of South Sulawesi Province, Indonesia, is a rapidly urbanising metropolitan area of high population density. The study covers its 14 mainland subdistricts (*kecamatan*), which differ substantially in population, area, and environmental conditions (Fig. 1). Dengue is endemic in the city, while cases are reported year-round, they are concentrated in the rainy season [10, 25].

Dengue case and covariate data

Hospital-recorded dengue cases were obtained from the medical records of eight hospitals in Makassar over 2021–2025 (Fig. 1). The hospitals recorded patient location at three levels of resolution. Five hospitals recorded a residential coordinate for each patient, from which the travel distance to the hospital and the subdistrict of residence were obtained (849 cases). Two hospitals recorded only the subdistrict of residence (773 cases). The eighth recorded only its total number of cases, with no residential information (220 cases). In all, 1,842 cases were analysed, of which 1,622 had a known subdistrict of residence and 220 did not. The offshore island district of Sangkarrang

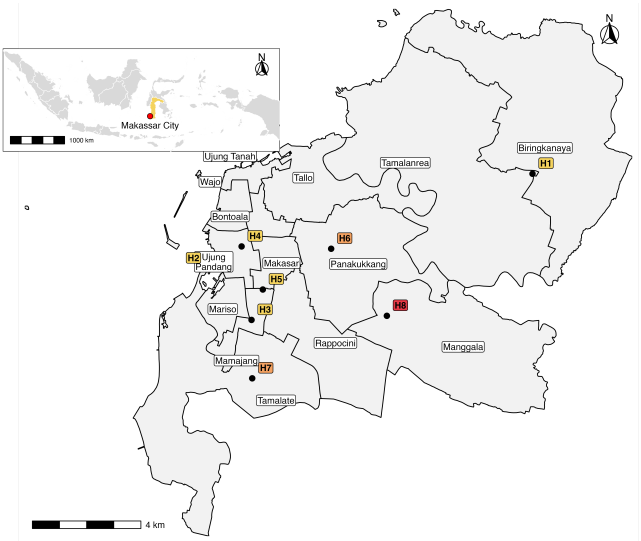


Fig. 1 Administrative boundary map of Makassar City showing subdistricts and hospital locations, with an inset indicating the location of Makassar City in South Sulawesi Province, Indonesia.

was excluded as it forms a disconnected component of the spatial adjacency graph and contributed no cases, leaving the 14 mainland subdistricts.

The resident population of each subdistrict, from the "Makassar in Figures" reports of the Central Bureau of Statistics (BPS) (https://makassarkota.bps.go.id/), was used as the population at risk. Two subdistrict-level covariates were considered: the mean Normalised Difference Vegetation Index (NDVI), from Sentinel-2 imagery, as an indicator of vegetation that provides resting sites and a favourable microclimate for *Aedes* mosquitoes [26, 27]; and health facility density (health facilities per 10,000 residents), included to adjust for spatial variation in case ascertainment. Candidate climatic covariates (precipitation and vapour pressure deficit, from CHIRPS and ERA5-Land) were examined but not retained, as their source resolution ($\geq 0.05°$) is too coarse to resolve meaningful variation across a city of this size (Section 3).

Statistical framework

We proceed in two stages. We first establish an area-level model for the expected number of dengue cases per subdistrict, using it to determine which combination of covariates and spatial structure best represents the areal risk surface. We then extend that structure to a point-process model that additionally represents the hospital observation process and the incomplete case locations. Throughout, index subdistricts by $j = 1, \ldots, J$ ($J = 14$) and hospitals by $h = 1, \ldots, H$ ($H = 8$). Covariates are standardised to zero mean and unit variance.

Area-level models

Let y_j denote the number of cases resident in subdistrict j and P_j its resident population. We model $y_j \sim \text{Poisson}(\lambda_j)$, where λ_j is the expected number of cases, with

$$\log \lambda_j = \log P_j + \eta_j, \tag{1}$$

so that η_j is the log rate per resident, with the population at risk accounted for through the offset $\log P_j$. We compare four nested specifications of the linear predictor η_j:

$$\text{M1 (null):} \quad \eta_j = \alpha, \tag{2}$$

$$\text{M2 (covariates):} \quad \eta_j = \alpha + \mathbf{X}_j^\top \boldsymbol{\beta}, \tag{3}$$

$$\text{M3 (spatial):} \quad \eta_j = \alpha + \phi_j, \tag{4}$$

$$\text{M4 (covariates + spatial):} \quad \eta_j = \alpha + \mathbf{X}_j^\top \boldsymbol{\beta} + \phi_j, \tag{5}$$

where α is an intercept, $\mathbf{X}_j$ is the covariate vector with coefficients $\boldsymbol{\beta}$, and ϕ_j is a Besag–York–Mollié (BYM2) spatial random effect [14]. The BYM2 effect combines a spatially structured and an unstructured component,

$$\phi_j = \sigma\Big(\sqrt{1-\rho}\,\epsilon_j + \sqrt{\rho/\kappa}\,\theta_j\Big), \tag{6}$$

in which $\sigma > 0$ is the marginal standard deviation, $\rho \in [0,1]$ is the proportion of variance that is spatially structured, and $\epsilon_j \sim \mathcal{N}(0,1)$ are independent effects. The structured component $\theta = (\theta_1, \ldots, \theta_J)$ follows an intrinsic conditional autoregressive (ICAR) prior [13, 28], with density

$$p(\theta) \propto \exp\Big(-\tfrac{1}{2}\sum_{i\sim j}(\theta_i - \theta_j)^2\Big), \tag{7}$$

where $i \sim j$ denotes pairs of subdistricts sharing a boundary (queen contiguity) and a sum-to-zero constraint is imposed for identifiability. The fixed scaling factor κ standardises the structured component so that σ has a consistent interpretation as the marginal standard deviation across different adjacency graphs [14]. This sequence isolates the contributions of covariates (M1 vs. M2) and of spatial dependence (M2 vs. M4), and identifies the intensity structure carried forward to the full model.

Proposed model: IPP–BYM2

The area-level model requires each case to be assigned to a subdistrict of residence. In practice, residence is recorded at three levels of resolution (exact coordinate, subdistrict, or nothing at all), while the location that is always known is the reporting hospital. We therefore embed the intensity in an inhomogeneous Poisson process (IPP) over cases, with an explicit hospital observation model.

Let the population-adjusted expected case intensity in subdistrict j be

$$\lambda_j = P_j \exp\big(\alpha + \mathbf{X}_j^\top \boldsymbol{\beta} + \phi_j\big), \tag{8}$$

using the M4 structure of Eq. (6). A case resident in subdistrict j is observed at hospital h with an attraction that declines with the travel distance d_{hj}. Following the gravity-model literature on spatial accessibility [21, 22, 24], the expected number of cases from j observed at h is

$$\mu_{hj} = \lambda_j \, k(d_{hj}), \qquad k(d) = (1 + \gamma_h d)^{-b}, \tag{9}$$

where the gravity (power-law) kernel $k(\cdot)$ has a hospital-specific scale $\gamma_h > 0$ and a shared shape exponent $b > 0$. The conventional exponential kernel $k(d) = \exp(-\gamma_h d)$ is a one-parameter alternative; the additional shape parameter of the power-law form lets the kernel combine a rapid initial decline with a heavier tail, and we compare the two forms by out-of-sample origin recovery (below).

Under the IPP, the log-likelihood is a sum of log-intensity terms over the observed cases minus a compensator equal to the total expected number of cases [29, 30],

$$\log L = \sum_{i=1}^{N} \log \Lambda_i \; - \; \sum_{h=1}^{H} \sum_{j=1}^{J} \mu_{hj}. \tag{10}$$

The intensity Λ_i contributed by case i, reported at hospital h_i, depends on how precisely its residence was recorded. For a case whose residence is observed (exactly, at travel distance d_i, or to subdistrict y_i), the contribution is the intensity of the process at that location, $\Lambda_i = \lambda_{y_i} \, k(d_i)$, giving

$$\log \Lambda_i = \log \lambda_{y_i} - b \log\big(1 + \gamma_{h_i} d_i\big), \tag{11}$$

where d_i is the recorded travel distance for a geolocated case and the hospital-to-subdistrict distance $d_{h_i y_i}$ otherwise. For a case with no residential information, the residence is unknown and is marginalised over all subdistricts,

$$\log \Lambda_i = \log\Big(\sum_{j=1}^{J} \lambda_j \, k(d_{h_i j}) \Big), \tag{12}$$

evaluated by the log-sum-exp identity for numerical stability [31]. This treats the residence as a latent variable and propagates its uncertainty into the posterior rather than fixing it by imputation [18]. The posterior distribution of the origin of such a case follows directly,

$$P(\text{origin}_i = j \mid h_i, \text{data}) \propto \lambda_j \, k(d_{h_i j}), \tag{13}$$

and can be reported as an inferred origin map. Because each hospital's total caseload enters both the log-intensity and the compensator, the hospitals jointly inform the intensity and the catchment kernel [32, 33].

The hospital scale parameters share a common prior, $\log \gamma_h \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma)$, which mildly regularises the eight estimates toward a common value; each γ_h is nonetheless well identified by the many cases attending each hospital. We also examined whether the variation in γ_h could be attributed to hospital characteristics (bed capacity, physician numbers, and institutional affiliation) by regressing $\log \gamma_h$ on these predictors. With only eight hospitals, however, the predictors are strongly collinear with hospital case volume, which is itself mechanically related to the estimated scale, and the resulting coefficients were confounded and could not be interpreted as effects of the characteristics themselves. We therefore retain the common prior and summarise catchment descriptively.

Priors

We used weakly informative priors. In the full IPP–BYM2 model, the intercept α was assigned a normal prior with mean -6 and standard deviation 3, reflecting its interpretation as a log rate per resident. The covariate coefficients were assigned $\beta_k \sim \mathcal{N}(0, 1)$ for standardised covariates; the spatial standard deviation $\sigma \sim \mathcal{N}^+(0, 1)$ (half-normal); the mixing parameter $\rho \sim \text{Beta}(0.5, 0.5)$; and $\epsilon_j \sim \mathcal{N}(0, 1)$ with θ as in Eq. (7). For the observation model, the kernel shape $b \sim \text{LogNormal}(\log 1.5, 0.5)$ and the hospital scales $\log \gamma_h \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma)$ with hyperpriors $\mu_\gamma \sim \mathcal{N}(0, 1)$ and $\sigma_\gamma \sim \mathcal{N}^+(0, 1)$.

Bayesian inference and computation

Inference was carried out using Hamiltonian Monte Carlo, specifically the No-U-Turn sampler, as implemented in Stan [34, 35]. Four chains were run with 1,000 warm-up iterations followed by 2,000 sampling iterations each, yielding 8,000 posterior draws; results were unchanged when the number of sampling iterations was increased further. Convergence was assessed using the rank-normalised potential scale reduction factor ($\widehat{R}$) and effective sample size [36]: all parameters had $\widehat{R} \leq 1.01$ and adequate effective sample sizes, with no divergent transitions.

Identifying high-risk areas

To identify high-risk subdistricts we use exceedance probabilities on the relative-risk scale. Let $r_j = \lambda_j / P_j = \exp(\alpha + \mathbf{X}_j^\top \boldsymbol{\beta} + \phi_j)$ denote the model-estimated dengue rate per resident in subdistrict j. For each posterior draw m, relative risk was calculated as

$$\text{RR}_j^{(m)} = \frac{r_j^{(m)}}{\frac{1}{J}\sum_{\ell=1}^{J} r_\ell^{(m)}}, \qquad J = 14. \tag{14}$$

This scales relative risks so that their average across the 14 subdistricts equals one, giving each subdistrict equal weight. Thus, $\text{RR}_j > 1$ indicates that the rate per resident

in subdistrict j exceeds the average across subdistricts. The exceedance probability is the posterior probability of this event,

$$EP_j = P(\mathrm{RR}_j > 1 \mid \text{data}) \approx \frac{1}{M} \sum_{m=1}^{M} \mathbb{I}\big(\mathrm{RR}_j^{(m)} > 1\big), \tag{15}$$

estimated from the M posterior draws. The threshold compares each subdistrict's rate per resident with the average across subdistricts, giving each subdistrict equal weight. Subdistricts with $EP_j \geq 0.80$ were classified as high-risk and those with $0.50 \leq EP_j < 0.80$ as moderate [37, 38].

Model comparison and validation

For the area-level models (M1–M4), in-sample fit was summarised by the Widely Applicable Information Criterion (WAIC) [39, 40] and out-of-sample predictive accuracy by exact leave-one-area-out cross-validation, reported as the expected log pointwise predictive density (elpd). We report differences in elpd between models (Δelpd) with their standard errors, where a larger elpd indicates better out-of-sample prediction [41]. Exact cross-validation (refitting with each area removed) was used in preference to approximate leave-one-out because the Pareto-$\hat{k}$ diagnostic exceeded 0.7, a consequence of the near-saturation of a 14-area model with an areal random effect.

The full IPP–BYM2 model was evaluated on its central task of recovering the residential origins of cases, using cross-validation. We used leave-one-hospital-out cross-validation across the seven hospitals with recorded case origins. In each fold, all likelihood contributions from one hospital, including its case terms and compensator, were excluded before refitting the model. Its case total and residential origins therefore did not inform the fit. The held-out hospital's location was retained for prediction, and its catchment scale was informed through the shared hierarchical distribution estimated from the remaining hospitals. The posterior mean origin probabilities across subdistricts (Eq. (13)) were then compared with the held-out hospital's observed origin distribution. Agreement was summarised by two standard measures of distance between probability distributions. The total variation distance, $\mathrm{TVD} = \frac{1}{2}\sum_j |\hat{p}_j - p_j|$ between the inferred ($\hat{p}$) and observed (p) origin proportions, measures the total misallocated probability (lower is better). The multinomial log predictive score, $\sum_j n_j \log \hat{p}_j$ where n_j is the observed number of cases from subdistrict j, is the log-probability the model assigns to the true origins (higher is better) [41]. Both were compared against two deterministic allocation baselines (uniform allocation and allocation proportional to population) and used to choose between the exponential and gravity kernels on held-out data. Residual spatial autocorrelation was assessed by Moran's I on the Pearson residuals of the fitted model, to confirm that the spatial effect absorbs the clustering present in the raw counts.

3 Results

Descriptive characteristics

A total of 1,842 dengue cases were analysed across the 14 subdistricts, of which 1,622 had a recorded subdistrict of residence and 220, all from a single hospital, did not (Fig. 2). Case counts, population, and the two retained covariates are summarised in Table 1. Raw counts were highest in Panakukkang (259), Tamalate (231) and Manggala (203); however, these are among the most populous subdistricts, and per-1,000 incidence varied more modestly and did not simply track the counts, for example, populous Biringkanaya had a low rate (0.5 per 1,000) while the small central subdistrict of Ujung Pandang had the highest (1.9 per 1,000). This discrepancy between counts and rates motivates the population-adjusted analysis.

Table 1 Descriptive summary by subdistrict, ordered by case count: resident population, dengue cases with known residence, cases per 1,000 residents, mean NDVI, and health facility density (facilities per 10,000 residents).

Subdistrict	Population	Cases	per 1,000	NDVI	Health dens.
Panakukkang	153,599	259	1.7	0.22	1.2
Tamalate	174,552	231	1.3	0.21	1.0
Manggala	168,720	203	1.2	0.30	1.0
Rappocini	148,277	177	1.2	0.16	1.6
Tallo	134,776	152	1.1	0.12	1.1
Mariso	56,006	103	1.8	0.16	1.9
Makassar	80,835	103	1.3	0.09	1.0
Biringkanaya	200,807	94	0.5	0.24	1.1
Mamajang	55,729	62	1.1	0.11	1.7
Tamalanrea	112,678	62	0.6	0.22	1.4
Bontoala	55,337	50	0.9	0.08	2.4
Ujung Pandang	25,394	47	1.9	0.14	6.4
Ujung Tanah	40,161	41	1.0	0.10	1.9
Wajo	29,461	38	1.3	0.06	3.7

The raw case counts showed statistically significant positive spatial autocorrelation (Moran's $I = 0.23$, $p = 0.04$, queen contiguity), indicating clustering across subdistricts and supporting a spatial random-effect structure. Among candidate covariates, precipitation and vapour pressure deficit showed negligible within-city variation (coefficients of variation below 2%), reflecting the coarse resolution of the source climate products relative to the size of the city, and were not retained. Population density and precipitation were also collinear with other predictors (variance inflation factors of 5.7 and 7.1); population was in any case represented through the offset. This left NDVI and health facility density as the covariates carried forward.

Among the five hospitals with coordinate data, travel distances were strongly right-skewed: most patients attended a nearby facility (median distances of 2–4.5 km) but a meaningful minority travelled several kilometres across subdistrict boundaries (Fig. 3). This concentration near the facility together with a non-negligible tail motivates

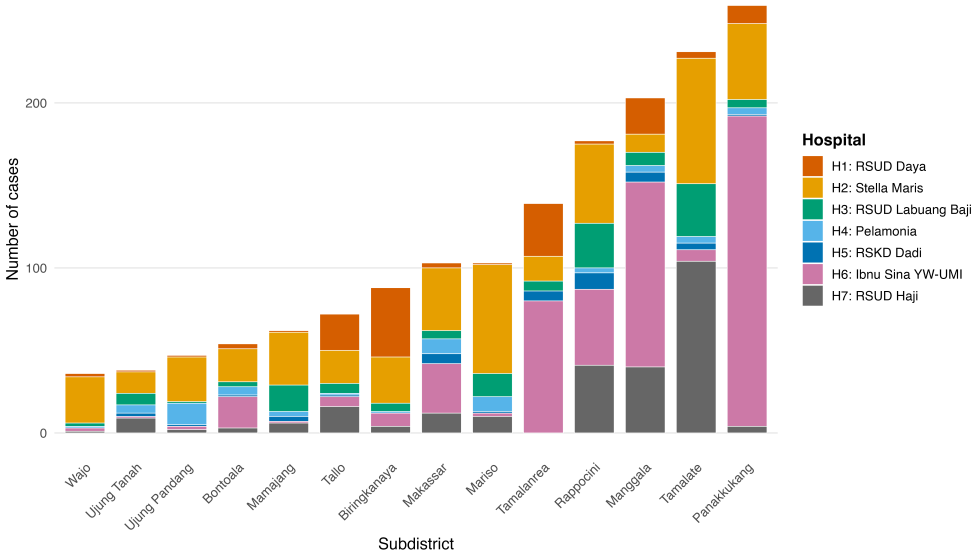


Fig. 2 Distribution of dengue cases by subdistrict and by reporting hospital.

a distance-decay catchment kernel with a heavier tail than a simple exponential, and the kernel parameters were estimated within the full model rather than fixed from these marginal distance distributions.

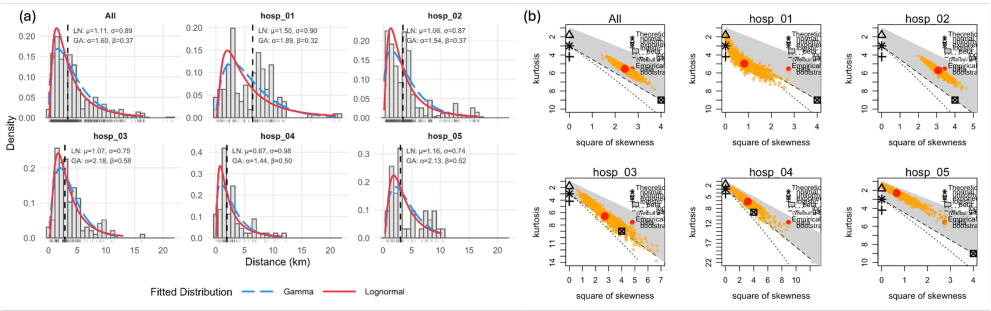


Fig. 3 Distribution of patient-to-hospital travel distances for the five hospitals with coordinate data: (a) histograms with fitted lognormal (red) and gamma (blue) densities and median (dashed); (b) Cullen and Frey skewness–kurtosis diagrams.

Area-level model selection

We fitted the four area-level models (M1–M4, Eqs. (1)–(6)), all with the population offset and with NDVI and health facility density as covariates, to quantify the separate

contributions of covariates and spatial dependence and to select the intensity structure for the full model. This model-selection stage used the aggregated counts of the 1,622 cases with a recorded subdistrict of residence (from the seven hospitals reporting location), since unlocated cases cannot be aggregated to subdistricts; all 1,842 cases are used by the full model below.

Table 2 Comparison of candidate area-level Bayesian models

Model	Structure	WAIC	SE(WAIC)	$\Delta\text{elpd}_{\text{CV}}$ vs M3
M1	Null Poisson	758	154	–
M2	Poisson + Covariates	430	92	–
M3	BYM2	113	2.7	0 (reference)
M4	BYM2 + Covariates	113	2.6	+1.65 (SE 1.57)

The non-spatial models (M1, M2) fitted markedly worse than the spatial models, supporting the inclusion of a spatial random effect. WAIC did not distinguish M3 and M4 (both 113); with 14 areas and an areal random effect, the models are near-saturated in-sample, and residual Pearson dispersion was far below its null expectation. Under exact leave-one-area-out cross-validation, M4 had a modestly higher estimated predictive accuracy than M3 ($\Delta\text{elpd} = 1.65$, SE 1.57), but the difference was small relative to its uncertainty and does not establish a clear predictive advantage. For a held-out area, M4 uses covariate information in addition to the spatial structure. The estimated spatial standard deviation was lower under M4 (0.55 versus 0.73), a reduction of approximately 25% in standard deviation, equivalent to approximately 43% in variance based on these point estimates. This reduction should not be interpreted as a causal proportion of variance explained. M4 was retained as the covariate-adjusted area-level structure for the full IPP model, while recognising the limited evidence distinguishing it from M3.

Catchment kernel and origin recovery

The central claim of the framework is that it can infer where cases originate from the reporting hospital and the spatial intensity. We evaluated this directly, and used it to choose the catchment kernel, by held-out cross-validation of origin recovery: each hospital with known origins was held out in turn, all of its likelihood contributions were excluded, and the model was refitted using the remaining hospitals. The predicted origin distribution for the held-out hospital was then compared with its observed distribution. This provides a like-for-like comparison of the exponential and gravity (power-law) kernels, and of the model against deterministic allocation baselines, on data the model did not see.

Table 3 shows that the proposed model recovered held-out case origins substantially better than every alternative. The gravity kernel reduced the mean total variation distance to 0.243, against 0.366 for a population-proportional baseline and 0.419 for uniform allocation, and improved the log-score by 348 nats over the best baseline.

Table 3 Held-out recovery of case origins (seven hospitals with known origins). The total variation distance (TVD, case-weighted mean; lower is better) and pooled multinomial log-score (higher is better) compare the inferred origin distributions with the truth.

Method	Mean TVD	Log-score
IPP–BYM2, gravity kernel (proposed)	**0.243**	**−3736**
IPP–BYM2, exponential kernel	0.370	−4436
Population-proportional allocation	0.366	−4084
Uniform allocation	0.419	−4281

Holding the intensity model fixed, replacing the gravity kernel with the conventional exponential kernel degraded recovery markedly (log-score −4436 versus −3736), and the exponential specification did not even outperform the trivial population-proportional baseline. The gravity kernel was therefore adopted; its shared shape parameter was estimated at $b = 1.38$ (90% CrI: 1.28–1.47), consistent with the range reported for gravity models of spatial interaction [24, 42].

Parameter estimates

Table 4 presents posterior estimates from the final IPP–BYM2 model. On the per-person scale, NDVI was positively associated with dengue risk ($\beta = 0.29$, 90% CrI: 0.04–0.53), its credible interval excluding zero, indicating that more vegetated subdistricts carried somewhat higher risk per resident. Health facility density showed a weak negative association whose interval included zero ($\beta = -0.22$, 90% CrI: −0.44 to 0.01), and we do not interpret it as a robust effect. The shared kernel shape was $b = 1.38$ (1.28–1.47); the spatial standard deviation was $\sigma = 0.47$ (0.30–0.71) and the mixing parameter $\rho = 0.54$ (0.01–0.99), the latter weakly identified as expected with only 14 areas.

Table 4 Posterior parameter estimates from the final per-person IPP–BYM2 model with a gravity catchment kernel.

Parameter	Mean	90% CrI
NDVI (per-person log-RR)	0.29	(0.04, 0.53)
Health facility density	−0.22	(−0.44, 0.01)
Kernel shape b	1.38	(1.28, 1.47)
Spatial SD σ	0.47	(0.30, 0.71)
Spatial mixing ρ	0.54	(0.01, 0.99)

Figure 4 and Table 5 present the relative-risk estimates from the final model, which combines a population offset, hospital catchment modelling and inference for missing origins. Manggala had the highest estimated relative risk, 2.73 (90% CrI: 2.41–3.06; $EP = 1.00$). Four subdistricts met the specified high-risk criterion ($EP \geq$

0.80): Manggala, Tamalate, Rappocini and Tallo, with the latter three having lower point estimates (RR 1.1–1.3). Panakukkang had the highest recorded case count (259), but its estimated RR was 0.95; Biringkanaya, the most populous subdistrict, had 94 recorded cases and an estimated RR of 0.91. These comparisons show that rankings by recorded case count and model-estimated relative risk differ. They do not isolate the contribution of population adjustment from that of the catchment model or missing-origin inference. The estimates can inform further assessment of spatial priorities, alongside absolute disease burden, uncertainty and intervention feasibility.

Table 5 Relative risk per person (RR) and exceedance probability ($EP = P(\text{RR} > 1)$) by subdistrict from the final model, ordered by RR. High: $EP \geq 0.80$; Moderate: $0.50 \leq EP < 0.80$.

Subdistrict	RR (90% CrI)	EP	Class
Manggala	2.73 (2.41, 3.06)	1.00	High
Tamalate	1.26 (1.12, 1.42)	0.999	High
Rappocini	1.24 (1.09, 1.39)	0.998	High
Tallo	1.12 (0.98, 1.27)	0.926	High
Ujung Tanah	1.11 (0.86, 1.38)	0.737	Moderate
Tamalanrea	1.00 (0.82, 1.20)	0.493	Low
Panakukkang	0.95 (0.84, 1.07)	0.247	Low
Biringkanaya	0.91 (0.76, 1.08)	0.182	Low
Mariso	0.85 (0.71, 1.00)	0.046	Low
Wajo	0.85 (0.64, 1.09)	0.137	Low
Makassar	0.63 (0.53, 0.74)	0.000	Low
Bontoala	0.59 (0.47, 0.73)	0.000	Low
Mamajang	0.39 (0.30, 0.49)	0.000	Low
Ujung Pandang	0.36 (0.27, 0.48)	0.000	Low

Model calibration and inferred origins

The observation model was well calibrated (Fig. 5b): predicted and observed hospital case totals lay close to the line of equality across the eight hospitals. Residual spatial autocorrelation was removed: Moran's I on the Pearson residuals was -0.03 ($p = 0.38$), compared with 0.23 ($p = 0.04$) for the raw counts, confirming that the spatial random effect absorbs the clustering that motivated the analysis. Hospital catchment behaviour, summarised as the mean distance of each hospital's inferred case origins, was similar across facilities (about 3–5 km; Fig. 5a), with 75–84% of a hospital's inferred cases originating within 5 km.

Finally, the framework yields a posterior distribution over the residential origins of the 220 unlocated cases (Fig. 5c). These are inferred to originate predominantly from the high-risk eastern subdistricts near the reporting hospital, with the largest expected contributions from Manggala and neighbouring areas, an inference that would be unavailable to any analysis restricted to cases with complete location information.

The substantive conclusions were robust to modelling choices: the estimated risk surface and the identification of Manggala as the dominant hotspot were essentially

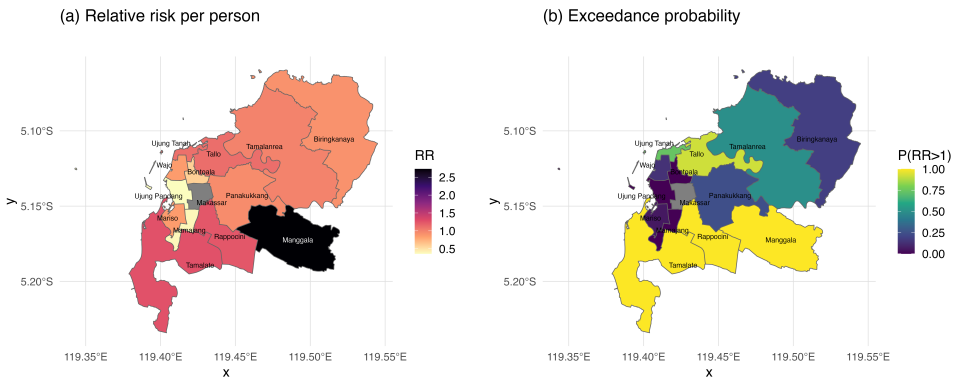


Fig. 4 (a) Posterior mean relative risk of dengue per person by subdistrict, and (b) exceedance probability, $P(\mathrm{RR} > 1)$, that a subdistrict's rate per resident exceeds the average across subdistricts. Risk is strongly concentrated in Manggala, in the east of the city.

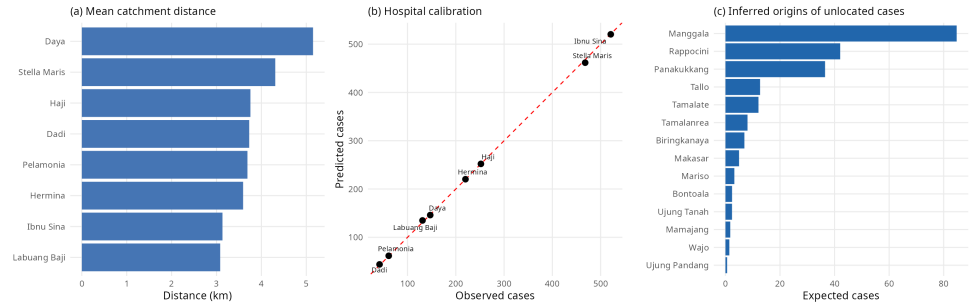


Fig. 5 Outputs of the final model: (a) estimated hospital catchment, as the mean distance of each hospital's inferred case origins; (b) observed versus predicted hospital case totals (dashed line: perfect agreement); (c) inferred residential origins of the 220 unlocated cases, as expected cases per subdistrict.

unchanged under alternative intercept and spatial priors and under omission of the covariates (rank correlation above 0.99 in each case).

4 Discussion

A Bayesian statistical framework was developed to map per-person dengue risk in Makassar while accounting for hospital catchment behaviour and cases with incomplete residential location. By combining an inhomogeneous Poisson process (IPP), a per-person BYM2 intensity, and a distance-based gravity kernel, the framework simultaneously estimated subdistrict-level relative risk, hospital catchment patterns, and the probable origins of unlocated cases within a single likelihood spanning three levels of location resolution.

The analysis produced two principal findings. Methodologically, the framework's central task, inferring where cases originate from the reporting hospital and the spatial intensity, was evaluated directly by out-of-sample cross-validation, under which a gravity catchment kernel clearly outperformed both the conventional exponential kernel and simple deterministic allocation rules. Substantively, the full model identified Manggala as the subdistrict with the highest estimated relative risk, while four subdistricts met the specified high-risk criterion. These estimates reflect the combined population adjustment, catchment model and treatment of missing origins.

Exceedance-probability analysis classified Manggala, Tamalate, Rappocini and Tallo as high risk under the specified threshold. Manggala had the highest estimated relative risk (2.7; $EP = 1.00$). Panakukkang had the highest recorded case count, whereas Biringkanaya had the largest resident population; neither met the model's high-risk criterion. The contrast between recorded counts and model-estimated relative risk reflects the combined population adjustment, catchment model and treatment of missing origins, rather than demonstrating the effect of population adjustment alone. These findings support closer investigation of Manggala and the other areas classified as high risk. Decisions about control resources should also consider absolute burden, uncertainty in the estimates, existing service coverage and the feasibility and expected benefit of interventions.

The pre-eminence of Manggala is consistent with its ecology. Manggala is a peripheral suburban district in eastern Makassar undergoing rapid demographic expansion alongside complex environmental challenges. Its geomorphology, dominated by the low-lying basin structures of the Antang area, subjects it to chronic seasonal inundation, compounded by the operations of the Tamangapa Landfill. Inadequate urban drainage fails to clear post-flood waters which, combined with widespread discarded open containers near the landfill, generate extensive stagnant pools that offer optimal micro-habitats for *Aedes aegypti* proliferation. These ecological dynamics plausibly elevate Manggala to a primary focus of dengue transmission in the city [4, 43, 44].

On the per-person scale, NDVI retained a positive association with dengue risk (0.29; 90% CrI: 0.04–0.53), consistent with vegetated environments providing favourable conditions for *Aedes* mosquitoes, although the effect was considerably weaker than on the raw-count scale and its context-dependence is well documented [4, 26, 27, 45]. We caution against over-interpretation: NDVI is spatially correlated with population and area, and much of its apparent strength in count-based analyses is attributable to this confounding. Health facility density showed only a weak, non-significant association after adjustment for population, and we do not interpret it as evidence that healthcare access reduces dengue risk; on the raw-count scale the same covariate appears protective, but this too is largely a population artefact.

Estimated mean catchment distances were broadly similar across the eight hospitals (approximately 3–5 km), with the majority of each hospital's inferred cases originating within 5 km. We deliberately refrain from attributing catchment breadth to specific institutional characteristics. With only eight hospitals, candidate predictors such as physician numbers and institutional affiliation are strongly collinear with hospital case volume (the three faith-based hospitals are also the three highest-volume

hospitals), and case volume itself is mechanically related to the estimated catchment scale. Any regression of catchment on such characteristics would therefore be confounded, and earlier suggestions in the literature that, for example, faith-based hospitals draw wider catchments cannot be substantiated with data at this resolution.

A methodological contribution of this study is the treatment of cases with missing origin information, a frequent problem in facility-based surveillance where patients seek care away from home or addresses are incompletely recorded. Here 220 of the 1,842 cases (11.9%) lacked a subdistrict of residence. Rather than assigning these to fixed locations by deterministic rules, which discard the uncertainty in the true origin before analysis begins, the framework treats each origin as a latent variable and infers a full posterior distribution over subdistricts (Eq. (13)), propagating that uncertainty into the risk estimates. This is consistent with the wider use of latent-variable and point-process methods for incomplete spatial data in epidemiology and ecology [18, 32, 33, 46].

Crucially, this capability was tested rather than asserted. Because the located cases provide a ground truth, the framework's ability to recover origins was evaluated out-of-sample by excluding each hospital's entire likelihood contribution in turn, refitting the model, and comparing its predicted origin distribution with the observed one. This assessed prediction for a held-out hospital whose case total and origins were both excluded from fitting. Under this criterion, the model outperformed both uniform and population-proportional allocation, as well as a version of the model using the conventional exponential catchment kernel (Table 3). These results provide support for the gravity kernel and the latent-origin approach used to account for cases with missing residential information.

Several limitations should be taken into account. First, the analysis is ecological: covariates and risk are measured at the subdistrict level, so the associations we report (for example, between NDVI and risk) describe area-level patterns and cannot be interpreted as individual-level effects, and are subject to the ecological fallacy. Second, the analysis is cross-sectional, pooling cases over 2021–2025; it does not capture seasonal or interannual variation in transmission, and a spatio-temporal extension is a natural next step. Third, with only 14 subdistricts and a spatial random effect per area, the intensity model is close to saturated in-sample: information criteria have little power to discriminate between models, which is why we relied on out-of-sample cross-validation throughout, and why the covariate and spatial-mixing parameters are only weakly identified. Fourth, the data derive from eight major hospitals and may not capture dengue cases presenting to primary healthcare or other facilities, or cases that do not enter the formal healthcare system. Previous evidence from Indonesia suggests substantial under-ascertainment of symptomatic dengue along the pathway from healthcare attendance to diagnosis and surveillance reporting [47]; the observation model also lacks a competing-destinations term, treating each hospital's attraction independently rather than as a choice among alternatives. Finally, the study is confined to one city, so the substantive findings may not transfer to settings with different populations, environments, or health systems.

These limitations point to natural extensions. A spatio-temporal formulation would capture seasonal and interannual variation and outbreak dynamics [15, 48]; incorporating human mobility would better represent how infection and healthcare-seeking connect areas beyond simple adjacency [49, 50]; and a competing-destinations observation model would represent hospital choice as a selection among alternatives. More immediately, the framework is directly applicable to other cities and diseases whose surveillance data share the same feature of incomplete, facility-based case locations.

5 Conclusion

This study introduced a per-person IPP–BYM2 framework for mapping dengue risk from facility-based surveillance data in which case locations are recorded at three levels of resolution and some are missing entirely. Treating unlocated origins as latent variables and marginalising over them, the framework jointly estimated per-person relative risk, hospital catchment behaviour, and the probable origins of unlocated cases. Its central capability, recovering case origins, was validated out-of-sample, under which a gravity catchment kernel outperformed both the conventional exponential kernel and deterministic allocation baselines. Substantively, the full model identified Manggala as the subdistrict with the highest estimated relative risk, with Tamalate, Rappocini and Tallo also meeting the specified high-risk criterion. Differences between rankings by recorded case count and estimated relative risk can inform spatial assessment, but control priorities should also consider absolute burden, uncertainty and intervention feasibility. More broadly, the framework offers a principled and testable approach to disease mapping from the incomplete, facility-based surveillance data that are common in low- and middle-income settings.

Acknowledgements

We would like to thank the Provincial Government of South Sulawesi and the Makassar City Government for granting permission and facilitating data collection at the eight participating hospitals: RSUD Daya, Stella Maris Hospital, RSUD Labuang Baji Sulsel, Pelamonia Makassar Hospital, RSKD Dadi, Ibnu Sina YW-UMI Hospital, RSUD Haji Makassar, and Hermina Hospital. We also extend our sincere appreciation to these hospitals for their support and for providing the dengue data used in this study.

Funding

BP gratefully acknowledges the financial support provided by the Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia, through the doctoral scholarship program. ELF acknowledges support from a University of Manchester Healthier Futures Fellowship and L'Oréal-UNESCO For Women in Science Award. TH and IH acknowledge support from the Wellcome Trust (Grant number 227438/Z/23/Z). TH acknowledges support from the Medical Research Council (Grant Number UKRI483) and UKRI (Grant Number MR/Z505328/1).

The funding bodies had no role in the design of the study, data analysis, interpretation of results, or preparation of the manuscript.

Competing interests

The authors declare that they have no competing interests relevant to this study.

Ethics approval and consent to participate

This study received ethical approval from the Health Research Ethics Committee, Faculty of Medicine, Universitas Hasanuddin, Indonesia (Approval No. 598/UN4.6.4.5.31/PP36/2026; Protocol No. UH26040522; approved on 30 April 2026). The study used routinely collected secondary surveillance and hospital data that were analysed in anonymised form. No identifiable personal information was used in this research.

Consent for publication

Not applicable.

Data availability

The dengue surveillance and hospital data used in this study were obtained from collaborating health institutions under ethical approval and data-sharing restrictions and are therefore not publicly available; derived, aggregated datasets may be made available from the corresponding author upon reasonable request and subject to approval from the data providers. Population and health facility data were obtained from the annual *Makassar in Figures* reports of the Central Bureau of Statistics (BPS) of Makassar City (https://makassarkota.bps.go.id/), and the NDVI covariate was derived from publicly available Sentinel-2 imagery (Copernicus Data Space Ecosystem, https://browser.dataspace.copernicus.eu/).

Materials availability

The spatial boundary data and derived datasets generated during the current study are available from the corresponding author upon reasonable request, subject to applicable data-sharing restrictions.

Code availability

The statistical models were implemented in `R` and `Stan`. The Stan model code and the R analysis scripts used to generate all results, tables and figures in this study are available at https://github.com/bobbybp/IPPBYM2. The individual patient data are not included in the repository owing to data-sharing restrictions (see Data availability).

Author contribution

BP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. TH: Conceptualization, Formal analysis, Methodology, Project administration, Supervision, Writing – review and editing. IH: Formal analysis, Methodology, Supervision, Writing – review and editing. OJ: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review and editing. ELF: Methodology, and Writing – review All authors read and approved the final manuscript.

References

[1] Ni, H., Cai, X., Ren, J., Dai, T., Zhou, J., Lin, J., Wang, L., Wang, L., Pei, S., Yao, Y., *et al.*: Epidemiological characteristics and transmission dynamics of dengue fever in china. Nature Communications **15**(1), 8060 (2024) https://doi.org/10.1038/s41467-024-52460-w

[2] Colón-González, F.J., Gibb, R., Khan, K., Watts, A., Lowe, R., Brady, O.J.: Projecting the future incidence and burden of dengue in southeast asia. nature communications **14**(1), 5439 (2023) https://doi.org/10.1038/s41467-023-42378-0

[3] Haider, N., Hasan, M.N., Onyango, J., Billah, M., Khan, S., Papakonstantinou, D., Paudyal, P., Asaduzzaman, M.: Global dengue epidemic worsens with record 14 million cases and 9000 deaths reported in 2024. International Journal of Infectious Diseases **158**, 107940 (2025) https://doi.org/10.1016/j.ijid.2025.107940

[4] Mamenun, Koesmaryono, Y., Sopaheluwakan, A., Hidayati, R., Dasanto, B.D., Aryati, R.: Spatiotemporal characterization of dengue incidence and its correlation to climate parameters in indonesia. Insects **15**(5), 366 (2024) https://doi.org/10.3390/insects15050366

[5] Ministry of Health Republic Indonesia: Waspada Penyakit di Musim Hujan (2024). https://kemkes.go.id/id/waspada-penyakit-di-musim-hujan

[6] World Health Organization: Disease Outbreak News; Dengue – Global Situation (2024). https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON518

[7] Sajib, A.H., Akter, S., Saha, G., Hossain, Z.: Demographic-environmental effect on dengue outbreaks in 11 countries. PLOS ONE **19**(9), 0305854 (2024) https://doi.org/10.1371/journal.pone.0305854

[8] Djaafara, B.A., Elyazar, I.R., Silalahi, F.S., Surya, A., Handito, A., Thohir, B., Aryani, D., Kamal, M., Ramadona, A.L., Gunawan, D., *et al.*: Dengue transmission heterogeneity across indonesia's archipelago: Climate-driven spatiotemporal patterns and policy implications. PLOS Neglected Tropical Diseases

20(3), 0014135 (2026) https://doi.org/10.1371/journal.pntd.0014135

[9] O'Driscoll, M., Imai, N., Ferguson, N.M., Hadinegoro, S.R., Satari, H.I., Tam, C.C., Dorigatti, I.: Spatiotemporal variability in dengue transmission intensity in jakarta, indonesia. PLoS neglected tropical diseases **14**(3), 0008102 (2020) https://doi.org/10.1371/journal.pntd.0008102

[10] Thamrin, S.A., Siswanto, S., Ansariadi, A., Jaya, A.K.: Spatial pattern analysis of dengue fever using conditional autoregressive in makassar city, indonesia. In: Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia. EAI, ??? (2020). https://doi.org/10.4108/eai.2-8-2019.2290482

[11] Lin, C.-H., Wen, T.-H.: How spatial epidemiology helps understand infectious human disease transmission. Tropical Medicine and Infectious Disease **7**(8), 164 (2022)

[12] Besag, J.: Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological) **36**(2), 192–225 (1974)

[13] Besag, J., York, J., Mollié, A.: Bayesian image restoration, with two applications in spatial statistics. Annals of the institute of statistical mathematics **43**(1), 1–20 (2011) https://doi.org/10.1007/BF00116466

[14] Riebler, A., Sørbye, S.H., Simpson, D., Rue, H.: An intuitive bayesian spatial model for disease mapping that accounts for scaling. Statistical methods in medical research **25**(4), 1145–1165 (2016) https://doi.org/10.48550/arXiv.1601.01180

[15] Aswi, A., Cramb, S., Moraga, P., Mengersen, K.: Bayesian spatial and spatiotemporal approaches to modelling dengue fever: a systematic review. Epidemiology & Infection **147**, 33 (2019) https://doi.org/10.1017/S0950268818002807

[16] Alegana, V.A., Khazenzi, C., Akech, S.O., Snow, R.W.: Estimating hospital catchments from in-patient admission records: a spatial statistical approach applied to malaria. Scientific reports **10**(1), 1324 (2020) https://doi.org/10.1038/s41598-020-58284-0

[17] Jacquez, G.M.: Current practices in the spatial analysis of cancer: flies in the ointment. International journal of health geographics **3**(1), 22 (2004) https://doi.org/10.1186/1476-072X-3-22

[18] Little, R.J., Rubin, D.B.: Statistical Analysis with Missing Data. John Wiley & Sons, Hoboken, NJ (2019). https://doi.org/10.1002/9781119482260

[19] Diggle, P.J., Giorgi, E.: Model-based geostatistics for prevalence mapping in low-resource settings. Journal of the American Statistical Association **111**(515), 1096–1120 (2016) https://doi.org/10.1080/01621459.2015.1123158

[20] Victoor, A., Delnoij, D.M., Friele, R.D., Rademakers, J.J.: Determinants of patient choice of healthcare providers: a scoping review. BMC health services research **12**(1), 272 (2012) https://doi.org/10.1186/1472-6963-12-272

[21] Luo, W., Wang, F.: Measures of spatial accessibility to health care in a gis environment: synthesis and a case study in the chicago region. Environment and planning B: planning and design **30**(6), 865–884 (2003) https://doi.org/10.1068/b29120

[22] Guagliardo, M.F.: Spatial accessibility of primary care: concepts, methods and challenges. International journal of health geographics **3**(1), 3 (2004) https://doi.org/10.1186/1476-072X-3-3

[23] Delamater, P.L., Messina, J.P., Shortridge, A.M., Grady, S.C.: Measuring geographic access to health care: raster and network-based methods. International journal of health geographics **11**(1), 15 (2012) https://doi.org/10.1186/1476-072X-11-1

[24] Stacherl, B., Sauzet, O.: Gravity models for potential spatial healthcare access measurement: a systematic methodological review. International Journal of Health Geographics **22**(1), 34 (2023) https://doi.org/10.1186/s12942-023-00358-z

[25] Sasmono, R.T., Wahid, I., Trimarsanto, H., Yohan, B., Wahyuni, S., Hertanto, M., Yusuf, I., Mubin, H., Ganda, I.J., Latief, R., *et al.*: Genomic analysis and growth characteristic of dengue viruses from makassar, indonesia. Infection, Genetics and Evolution **32**, 165–177 (2015) https://doi.org/10.1016/j.meegid.2015.03.006

[26] Mahato, R.K., Htike, K.M., Sornlorm, K., Koro, A.B., Yadav, R.K., Kafle, A., Sharma, V.: Spatial autocorrelation of environmental factors influencing dengue outbreaks using moran's i: A study from nepal (2020–2023). Plos one **20**(6), 0324798 (2025) https://doi.org/10.1371/journal.pone.0324798

[27] Murdock, C.C., Evans, M.V., McClanahan, T.D., Miazgowicz, K.L., Tesla, B.: Fine-scale variation in microclimate across an urban landscape shapes variation in mosquito population dynamics and the potential of aedes albopictus to transmit arboviral disease. PLoS neglected tropical diseases **11**(5), 0005640 (2017) https://doi.org/10.1371/journal.pntd.0005640

[28] Rue, H., Held, L.: Gaussian Markov Random Fields: Theory and Applications. Chapman and Hall/CRC, Boca Raton, FL (2005). https://doi.org/10.1201/9780203492024

[29] Diggle, P.J.: Statistical Analysis of Spatial and Spatio-temporal Point Patterns.

CRC press, Boca Raton, FL (2013). https://doi.org/10.1201/b15326

[30] Baddeley, A., Rubak, E., Turner, R., *et al.*: Spatial Point Patterns: Methodology and Applications with R vol. 1. Chapman and Hall/CRC, Boca Raton, FL (2016). https://doi.org/10.1201/b19708

[31] Moraga, P.: Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny. Chapman and Hall/CRC, Boca Raton, FL (2019). https://doi.org/10.1201/9780429341823

[32] Li, Y., Brown, P., Gesink, D.C., Rue, H.: Log gaussian cox processes and spatially aggregated disease incidence data. Statistical methods in medical research **21**(5), 479–507 (2012) https://doi.org/10.1177/0962280212446326

[33] Diggle, P.J., Moraga, P., Rowlingson, B., Taylor, B.M.: Spatial and spatio-temporal log-gaussian cox processes: Extending the geostatistical paradigm. Statistical Science **28**(4) (2013) https://doi.org/10.1214/13-sts441

[34] Hoffman, M.D., Gelman, A., *et al.*: The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. J. Mach. Learn. Res. **15**(1), 1593–1623 (2014)

[35] Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan: A probabilistic programming language. Journal of statistical software **76**, 1–32 (2017) https://doi.org/10.18637/jss.v076.i01

[36] Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.-C.: Rank-normalization, folding, and localization: An improved rhat for assessing convergence of mcmc (with discussion). Bayesian analysis **16**(2), 667–718 (2021) https://doi.org/10.1214/20-BA1221

[37] Blangiardo, M., Boulieri, A., Diggle, P., Piel, F.B., Shaddick, G., Elliott, P.: Advances in spatiotemporal models for non-communicable disease surveillance. International Journal of Epidemiology **49**(Supplement_1), 26–37 (2020) https://doi.org/10.1093/ije/dyz181

[38] Andrade-Pacheco, R., Rerolle, F., Lemoine, J., Hernandez, L., Meïté, A., Juziwelo, L., Bibaut, A.F., Laan, M.J., Arnold, B.F., Sturrock, H.J.: Finding hotspots: development of an adaptive spatial sampling approach. Scientific reports **10**(1), 10939 (2020) https://doi.org/10.1038/s41598-020-67666-3

[39] Watanabe, S., Opper, M.: Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. Journal of machine learning research **11**(12) (2010) https://doi.org/10.5555/1756006.1953045

[40] Gelman, A., Hwang, J., Vehtari, A.: Understanding predictive information criteria for bayesian models. Statistics and computing **24**(6), 997–1016 (2014) https://doi.org/10.1007/s11222-013-9416-2

[41] Vehtari, A., Gelman, A., Gabry, J.: Practical bayesian model evaluation using leave-one-out cross-validation and waic. Statistics and computing **27**(5), 1413–1432 (2017) https://doi.org/10.1007/s11222-016-9696-4

[42] Wang, F.: Measurement, optimization, and impact of health care accessibility: a methodological review. Annals of the Association of American Geographers **102**(5), 1104–1112 (2012) https://doi.org/10.1080/00045608.2012.657146

[43] Suryadi, I., Lestari, V.D., Widyatuti, N., Fitriani, N.: Edukasi dan pendampingan pembuatan ovitrap sebagai upaya pengendalian demam berdarah. Shihatuna: Jurnal Pengabdian Kesehatan Masyarakat **3**(1), 38–44 (2023) https://doi.org/10.30829/shihatuna.v3i1.12907

[44] Pasau, H.B., et al.: Preferensi oviposisi dan kelimpahan telur nyamuk aedes aegypti dengan menggunakan ovitrap di wilayah kerja puskesmas antang kecamatan manggala kota makassar. PhD thesis, Universitas Hasanuddin (2024)

[45] Aminuddin, M.A.M., Dom, N.C., Porusia, M., Yatim, S.R.M.: Urban microclimate impacts on aedes mosquitoes' life cycle: A concept paper. Environment-Behaviour Proceedings Journal **8**(25), 215–221 (2023) https://doi.org/10.21834/e-bpj.v8i25.4866

[46] Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., Rue, H.: Going off grid: Computationally efficient inference for log-gaussian cox processes. Biometrika **103**(1), 49–70 (2016) https://doi.org/10.1093/biomet/asv064

[47] Wahyono, T., Nealon, J., Beucher, S., Prayitno, A., Moureau, A., Nawawi, S., Thabrany, H., Nadjib, M.: Indonesian dengue burden estimates: review of evidence by an expert panel. Epidemiology & Infection **145**(11), 2324–2329 (2017) https://doi.org/10.1017/S0950268817001030

[48] Rotejanaprasert, C., Chinpong, K., Lawson, A.B., Maude, R.J.: Comparative evaluation of spatiotemporal methods for effective dengue cluster detection with a case study of national surveillance data in thailand. Scientific Reports **14**(1), 31064 (2024) https://doi.org/10.1038/s41598-024-82212-1

[49] Wesolowski, A., Qureshi, T., Boni, M.F., Sundsøy, P.R., Johansson, M.A., Rasheed, S.B.w., Engø-Monsen, K., Buckee, C.O.: Impact of human mobility on the emergence of dengue epidemics in pakistan. Proceedings of the national academy of sciences **112**(38), 11887–11892 (2015) https://doi.org/10.1073/pnas.1504964112

[50] Simini, F., González, M.C., Maritan, A., Barabási, A.-L.: A universal model for

mobility and migration patterns. Nature **484**(7392), 96–100 (2012) https://doi.org/10.1038/nature10856