Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Bayesian hierarchical spatiotemporal modeling for forecasting diarrhea risk among children under 5 in Bandung city, Indonesia

I Gede Nyoman Mindra Jaya[a*], Anna Chadidjah[a], Yudhie Andriyana[a], Farah Kristiani[b] and Anggi Nur Fauziah[a]

[a]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia*
[b]*Department of Mathematics, Faculty of Technology and Data Science, Parahyangan University, Bandung city 40141, Indonesia*

C H R O N I C L E

Article history:
Received: July 2, 2023
Received in revised format: July 25, 2023
Accepted: August 10, 2023
Available online: August 10, 2023

Keywords:
Bayesian
Diarrhea
Forecasting
Children under five years old
INLA

A B S T R A C T

The main objectives of this research are to identify significant spatial and temporal components associated with diarrhea and provide an accurate forecast. Using data from the Bandung city health surveillance system, the analysis reveals a decreasing trend in both the number of incidences and the estimated relative risks of diarrhea in most districts. Key factors contributing to diarrhea variation include temporally structured, spatially structured, and unstructured effects of space-time interaction Type I. No clear seasonal pattern is observed in diarrhea incidence among children under five, emphasizing the need for consistent vigilance and preventive measures. Spatial clustering was observed in the eastern and western parts of Bandung city. The forecasting model predicts a continued decline in diarrhea incidence and relative risk throughout 2022.

1. Introduction

Diarrhea is a significant health concern, especially for children younger than five (Jaya et al., 2019). Diarrhea in children is characterized by the presence of three or more liquid stools within a 24-hour period (Alebel et al., 2018; Li et al., 2020; Walker et al., 2013). The World Health Organization (WHO) reports that diarrhea is the most common cause of death among children. It is astounding that nearly 1.7 billion incidences of diarrhea among children under five years are reported annually worldwide. This results in the unexpected deaths of approximately 525,000 per year (Kotloff et al., 2017; WHO, 2017). Diarrhea remains a significant threat in Indonesia, especially among children under the age of five, despite a declining trend in incidence over the years. According to the Indonesian Ministry of Health, approximately four million cases of diarrhea in this age group were reported in 2018 (Ministry of Health RI, 2019). Bandung is one of the most populous cities in Indonesia, and has a concerning incidence of childhood diarrhea. According to Rahayu et al. (2023), the prevalence of diarrhea among toddlers in the city of Bandung in 2019 was 17,554 cases, accounting for 8.93% of the total toddler population in the same city. According to George et al. (2014) and Paul (2020), environmental variables and the adoption of hygienic and health-conscious behaviors are strongly correlated with the number of cases of diarrhea among children under 5. To control the spread of diarrhea, an effective and efficient early warning system (EWS) is necessary. It is essential, particularly for small cities like Bandung. Identifying hotspot areas is the first step in developing an early warning system (Fang et al., 2020; Jaya & Folmer, 2020, 2021, 2022). The spatiotemporal forecasting model has a chance of helping with the recognition of high-risk areas and facilitating the implementation of interventions that have been customized to reduce those risks. A key component of an early warning system is the utilization of spatiotemporal maps that monitor the spatial and temporal distribution of a disease. According to Ugarte et al. (2012), mapping disease risk provides significant insights into the spatiotemporal progression of the disease's incidence rate, enabling epidemiologists to enhance their understanding of disease outbreaks and develop relevant hypotheses.
Because of limited data availability, it is difficult to develop spatiotemporal forecasts for diarrhea among children under 5. Data on the number of cases and the population at risk is generally available. While covariate variables such as weather and

* Corresponding author.
E-mail address: mindra@unpad.ac.id (I. G. N. M. Jaya)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.8.008

socioeconomic variables are unavailable, especially for small areas. In this study, we consider developing a pure spatiotemporal model with two essential components: (i) the intercept component and (ii) the random effect component. The random effect component includes spatially and temporally structured and unstructured effects, seasonal components, and spatial and temporal interaction effects. The spatially structured effect captures the spatial clustering patterns observed in the data, while the spatially unstructured effect is responsible for spatial heterogeneity. Temporally structured and unstructured effects are included to account for temporal autocorrelation and temporal heteroscedasticity, respectively. In addition, the seasonal component is used to account for the regular pattern in diarrhea cases caused by seasonal variations. The spatial and temporal interaction effects are used to capture an extra variation that cannot be explained by the main spatial and temporal components (Jaya & Folmer, 2020).

The simplicity of the univariate modeling approach is one of its primary advantages over multivariate models with covariates. Univariate models are simpler to implement and interpret because they focus on a single variable and do not require the forecasting of additional covariables. Even without covariates, univariate models can still generate accurate forecasts, especially when there is a strong autocorrelation among the observations. Univariate models can effectively capture and forecast the dynamics of a variable by exploiting its temporal patterns and dependencies. When including covariates is not necessary or possible, univariate models are a good choice for forecasting because they are easy to use and accurate (Assad et al., 2023; Pena & SaNchez, 2007). Bayesian methods are often used to estimate pure models (Blangiardo & Cameletti, 2015). They have advantages when it comes to specifying the random parts by putting the parameters in a hierarchical order. When analytical solutions are not possible (Blangiardo & Cameletti, 2015), Markov Chain Monte Carlo (MCMC) is a popular algorithm in a Bayesian setting. However, for complex models with a huge number of parameters involved, the computational aspect of MCMC can be challenging. MCMC can result in significant Monte Carlo errors and require substantial computational resources. Rue et al. (2009) introduced the Integrated Nested Laplace Approximation (INLA) as an alternative to the MCMC. INLA has demonstrated the ability to reduce computation time while providing parameter estimates comparable to those obtained through MCMC techniques.

The remaining sections of the paper are as follows: The spatial-temporal model is presented in Section 2 using an INLA-based methodology. The third section describes the application of the model to diarrhea data for children under the age of five. The discussion section in Section 4. Section 5 is the conclusion of the study.

2. Method

2.1 Spatiotemporal autocorrelation

The spatiotemporal analysis starts by examining the possibility of spatiotemporal autocorrelation in the diarrhea cases data among children under 5 in Bandung city. Let the number of diarrhea cases in district i ($i = 1,2, \dots, n = 30$) and month t ($t = 1,2,\dots, T = 60$), denoted as y_{it}, the spatiotemporal Moran's formula can be expressed as follows (Jaya & Folmer, 2020):

$$\text{Moran's I} = \frac{nT \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{s=1}^{T} \widetilde{w}_{(it,js)}(y_{it}-\bar{y})(y_{js}-\bar{y})}{\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{n}\sum_{s=1}^{T} \widetilde{w}_{(it,js)} \sum_{i=1}^{n}\sum_{t}^{T}(y_{it}-\bar{y})^2} \tag{1}$$

where $\bar{y}$ is the average of the observed diarrhea cases and $\widetilde{w}_{(it,js)}$ is the spatiotemporal weight defined as:

$$\tilde{w}_{(it,js)} = \begin{cases} w_{ij} & \text{if } t = s \\ 1 & \text{if } i = j \text{ and } |t-s| = 1 \\ 0 & \text{otherwise} \end{cases}$$

w_{ij} takes a value of 1 if areas i and j are neighbors and 0 otherwise. A high value of Moran's I indicate substantial positive spatiotemporal correlation in diarrhea cases, indicating clustering or similarity in the spatial pattern. Conversely, a value close to zero suggests the absence of spatiotemporal autocorrelation.

2.2. Bayesian Spatiotemporal Hierarhical Model

We used a spatiotemporal model to model the diarrhea transmissions over space and time. The number of diarrhea cases for children under five years old y_{it}, were defined to follow a Poisson distribution such that:

$$p(y_{it}|E_{it},\theta_{it}) = \frac{\exp(-E_{it}\theta_{it})(E_{it}\theta_{it})^{y_{it}}}{y_{it}!} \; ; \; i = 1,\dots,n \text{ and } t = 1,\dots,T \tag{2}$$

where E_{it} represents the expected number of diarrhea counts and θ_{it} denotes the relative risk in area i and at month t. The expected number of counts E_{it} is defined as:

$$E_{it} = N_{it}\frac{\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}/nT}{\sum_{i=1}^{n}\sum_{t=1}^{T} N_{it}/nT} \;\; i = 1,\dots,n \text{ and } t = 1,\dots,T, \tag{3}$$

with N_{it} represents the number of children under five years old in area i at month t, with n is the number areas, and T the length of time periods. Count data often encounters the issue of overdispersion, which can arise from various factors such as

an abundance of zeros or the presence of outliers (Lee et al., 2012). This issue can be handled by employing the Negative Binomial (NB) distribution with the probability mass function defined as follows (Mohebbi et al., 2014):

$$p(y_{it}|E_{it},\theta_{it},\varsigma)=\frac{\Gamma(y_{it}+\varsigma)}{\Gamma(y_{it}+1)\Gamma(\varsigma)}\left(\frac{E_{it}\theta_{it}}{E_{it}\theta_{it}+\varsigma}\right)^{y_{it}}\left(\frac{\varsigma}{E_{it}\theta_{it}+\varsigma}\right)^{\varsigma}. \tag{4}$$

The NB distribution has an expectation or mean $E(y_{it})=E_{it}\theta_{it}$ and variance $\mathrm{Var}(y_{it})=E_{it}\theta_{it}+(E_{it}\theta_{it})^2/\varsigma$ with ς denote the overdispersion parameter. We employed the log linear models to describe the space-time variation in the relative risk θ_{it} over space, as expressed by the following equation:

$$\eta_{it}=\log(\theta_{it})=\alpha+\omega_i+\upsilon_i+\phi_t+\varphi_t+\gamma_t+\delta_{it} \text{ for } i=1,\dots,n \text{ and } t=1,\dots,T, \tag{5}$$

where α represents the intercept, which reflects the global relative risk. The spatially structured effects are denoted by ω_i, capturing the spatial patterns and clustering. The spatially unstructured effects are represented by υ_i accounting for spatial heterogeneity. The temporally structured effects are denoted by ϕ_t, capturing the temporal trends and patterns. The seasonal effects are expressed by ϕ_t, accounting for the regular seasonal variations. The temporally unstructured effects are denoted by γ_t, capturing the temporal heterogeneity. Lastly, the interaction effect is denoted by δ_{it}, representing the combined spatial and temporal influence. The conditional autoregressive model of Leroux (LCAR) (Leroux et al., 1999) was utilized to effectively capture the spatially correlated random effects such that:

$$\omega_i|\boldsymbol{\omega}_{-i},\boldsymbol{W}\sim N\left(\frac{\rho\sum_{j=1}^{n}w_{ij}\omega_j}{\rho\sum_{j=1}^{n}w_{ij}+1-\rho},\frac{\sigma_\omega^2}{\left(\rho\sum_{j=1}^{n}w_{ij}+1-\rho\right)}\right) \text{ for every } t \text{ and } i=1,\dots,n \tag{6}$$

where $\mathbf{W}=\left(w_{ij}\right)$ denotes the spatial weights matrix, ρ denotes the spatial autocorrelation coefficient, and σ_ω^2 the variance of $\boldsymbol{\omega}=(\omega_1,\dots,\omega_n)'$. We employed the queen contiguity weight matrix in our analysis. The weight matrix is designed to incorporate the adjacency relationship between areas, which is determined by the presence of a shared boundary or vertex. The exchangeable prior was used to capture the spatial heterogeneity such that:

$$\upsilon_i\sim N(0,\sigma_\upsilon^2) \text{ for every } t \text{ and } i=1,\dots,n \tag{7}$$

where σ_υ^2 denote the variance of $\upsilon=(\upsilon_1,\dots,\upsilon_n)'$. For temporally structured effects we utilized Random Walk of order 1 or 2 (RW1 or RW2) such that:

$$\text{RW1: } \phi_{t+1}-\phi_t|\sigma_\phi^2\sim N\left(0,\sigma_\phi^2\right) \text{ for every } i \text{ and } t=1,..,T \tag{8}$$

$$\text{RW2: } \phi_t-2\phi_{t+1}+\phi_{t+2}|\sigma_\phi^2\sim N\left(0,\sigma_\phi^2\right) \text{ for every } i \text{ and } t=1,\dots,T \tag{9}$$

with σ_ϕ^2 the variance of $\boldsymbol{\phi}=(\phi_1,\dots,\phi_T)'$.

The seasonal component (φ_t) for m seasonality period is defined as:

$$\varphi_t+\varphi_{t+1}+\cdots+\varphi_{t+m-1}|\tau_\kappa\sim N\left(0,\sigma_\varphi^2\right) \text{ for every } i \text{ with } t=1,\dots,T-m+1, \tag{10}$$

with σ_φ^2 denotes the variance of $\boldsymbol{\varphi}=(\varphi_1,\dots,\varphi_T)'$. Temporally unstructured effects γ_t is assumed to follow exchangeable prior:

$$\gamma_t|\sigma_\gamma^2\sim N\left(0,\sigma_\gamma^2\right) \text{ for every } i \text{ and } t=1,\dots,T, \tag{11}$$

with σ_γ^2 denotes the variance of $\boldsymbol{\gamma}=(\gamma_1,\dots,\gamma_T)'$. For the space-time interaction effects $\boldsymbol{\delta}=(\delta_{11},\dots,\delta_{nT})'$, we considered four types of interactions. These interactions correspond to different combinations of spatially unstructured and temporally unstructured effects (Type I), spatially unstructured effects and temporally structured effects (Type II), spatially structured effects and temporally unstructured effects (Type III), and spatially structured effects with temporally structured effects (Type IV), respectively. By considering these different interaction types, we aimed to capture the complex interplay between spatial and temporal factors in our modeling approach (see Knorr-Held (2000) for details).

We employed vague Gaussian distributions as prior for the model parameters. The parameters' prior distributions were specified as follows: $\alpha\sim\mathcal{N}(0,10^6)$. Following Simpson et al., (2017), the square root hyperparameters $\sigma_\omega,\sigma_\upsilon,\sigma_\phi,\sigma_\varphi,\sigma_\gamma,$ and σ_δ are assumed to follow Penalized Complexity (PC) prior. To account for the limited range of values for ρ (between 0 and 1), we employed a non-informative Gaussian prior with a large variance for the log transformation of ρ, such that $log\left(\frac{\rho}{1-\rho}\right)\sim\mathcal{N}(0,10)$ (Bivand et al., 2015).

The spatiotemporal model (5) was estimated using INLA. The marginal posterior distributions are utilized to compute posterior means, standard deviations, and forecast values (Jaya & Folmer, 2020). In addition, the INLA provide several model

selection criteria, such as the deviance information criterion (DIC) and the Watanabe–Akaike information criterion (WAIC) (Spiegelhalter et al., 2002; Watanabe, 2010). These criteria help evaluate the model's validity. In addition, the INLA method permits the computation of additional metrics for prediction performance, such as Conditional Predictive Oridinate (CPO), marginal predictive likelihood (MPL), root mean squared error (RMSE), mean absolute error (MAE), and correlation coefficient (r) (Blangiardo & Cameletti, 2015).

The marginal predictive posterior distribution, denoted as $p(\hat{\mathbf{y}}|\mathbf{y})$ (Wang et al., 2018), is utilized for Bayesian prediction.The INLA method facilitates the prediction process by constructing a model that combines historical observations from previous time intervals with missing values for the desired time intervals.

The exceedance probability criteria proposed by Lawson (2010) were utilized to identify spatiotemporal hotspots. The probability of spatiotemporal exceedance is computed using the marginal posterior distribution. It indicates the probability that the estimated of the relative risk exceeds a predetermined cut-off value c. This probability is denoted as $\widehat{\Pr}(\theta_{it} > c|\mathbf{y})$. The calculation of this probability is determined as follows:

$$\widehat{\Pr}(\theta_{it} > c|\mathbf{y}) = 1 - \int_{\theta_{it} \leq c} p(\theta_{it}|\mathbf{y})\, d\theta_{it} \tag{12}$$

The estimation can be performed using the Laplace method (Blangiardo & Cameletti, 2015). The identification of hotspots through the utilization of exceedance probability necessitates the establishment of two predetermined parameters. The initial parameter to consider is the threshold value, denoted as c, for θ_{it}. A value of 1 signifies that a county possesses an average level of relative risk, while values of 2 or 3 indicate a significantly heightened level of risk. The threshold γ of the probability represents the second parameter. According to (Lawson & Rotejanaprasert, 2014), typical values for the parameter γ include 0.90, 0.95, and 0.99.

3. Result

3.1. Study Area

Bandung is the capital of West Java province, located in the middle of the province and contains 30 districts (Fig. 1). Bandung is the most populous city, with an approximate population of 2,526,395 and a small area of 167.3 square kilometers.

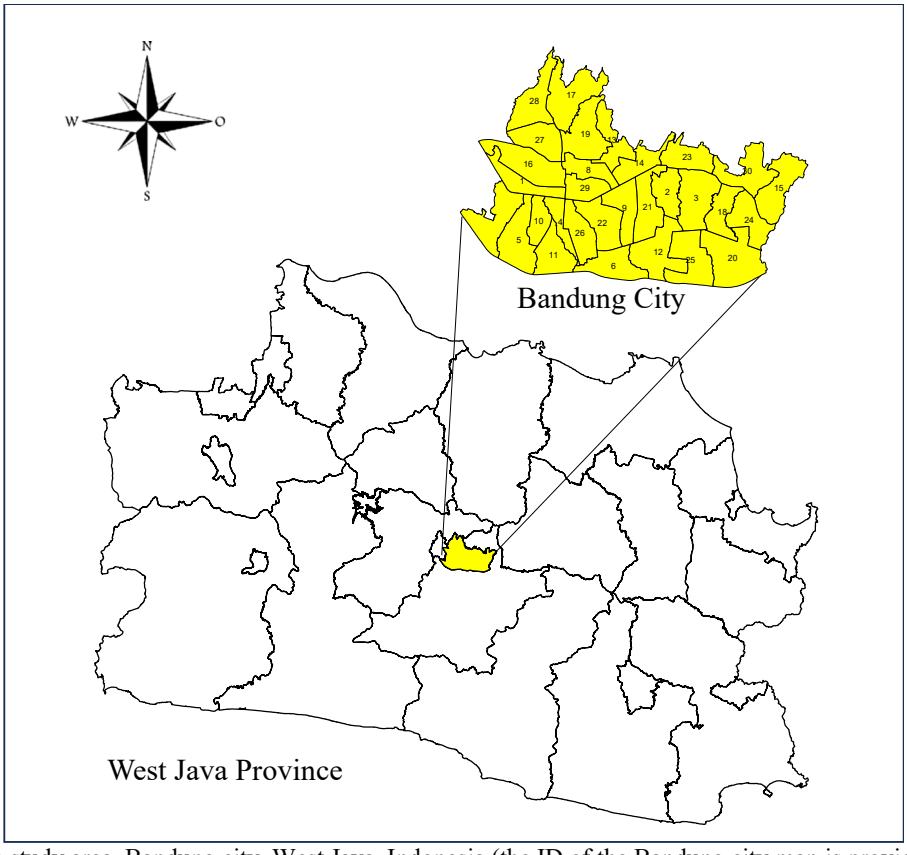
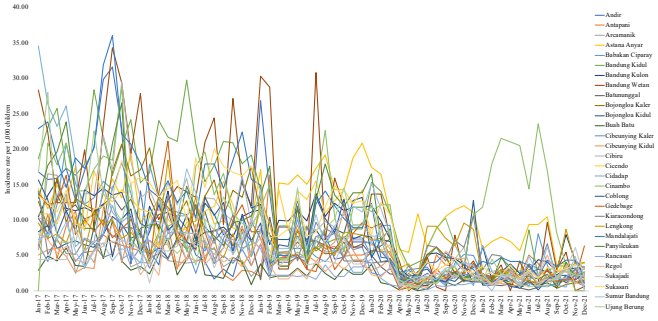


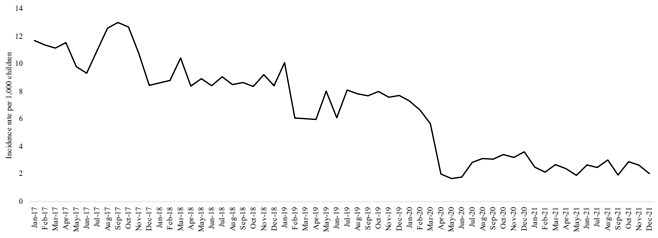
Fig. 1 Map of the study area, Bandung city, West Java, Indonesia (the ID of the Bandung city map is provided in Table 1)

3.2. Data description

Over a 5-year surveillance period, from 2017 to 2021, monthly diarrheal morbidity data were collected from each of the 30 study districts from Bandung Health Office (2017-2021). Throughout the duration of the study, not a single district reported missing data. There were 2,463 cases of diarrhea among children under 5 resulting in a monthly incidence rate of 7 cases per 1,000 children under 5. The incidence rate of diarrhea varied across areas and times, exhibiting an overall upward trend with temporal fluctuations. From 2017 to 2021, however, the number of occurrences observed decreased. During the period from 2017 to 2021, Bandung city Kidul district had the highest average incidence rate (11.66 per 1,000), while Buah Batu district had the lowest incidence rate (2.71 per 1,000) (see Table 1).



(A)



(B)

Fig. 1 (A) Monthly temporal trend for 30 districts; (B) Monthly temporal trend cumulative of 30 districts of incidence rate of diarrhea for Bandung city

Table 1
The annual incidence rate of diarrhea in Bandung city, 2017–2021

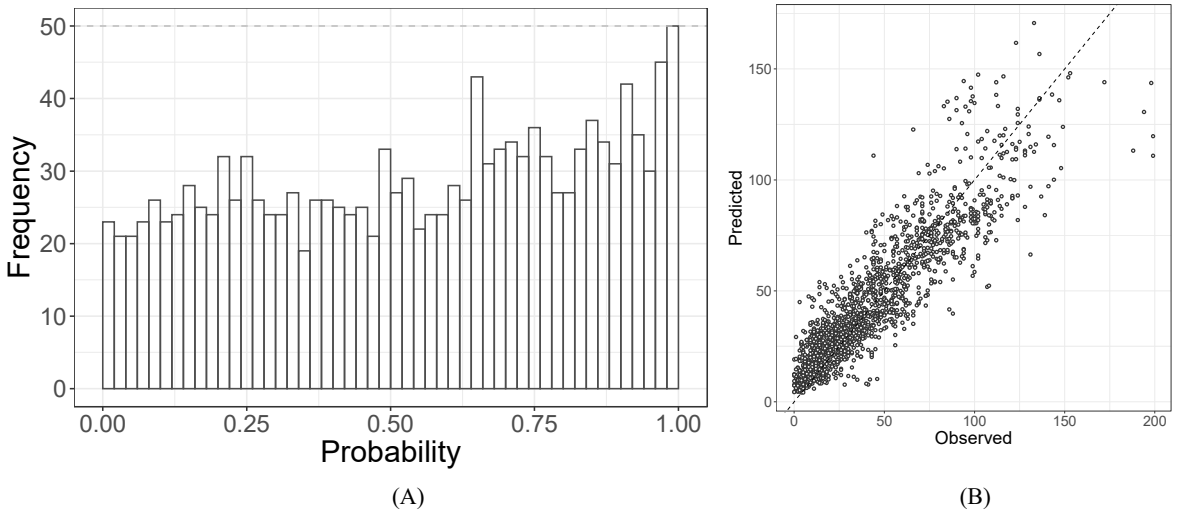
id	District	2017	2018	2019	2020	2021	Mean
1	Andir	18.180	5.926	8.006	2.975	1.345	7.287
2	Antapani	7.127	5.611	5.674	3.235	2.329	4.795
3	Arcamanik	7.860	4.417	6.048	2.772	2.121	4.644
4	Astana Anyar	13.487	9.863	15.451	10.566	6.440	11.161
5	Babakan Ciparay	13.646	9.462	6.056	4.225	2.650	7.208
6	Bandung city Kidul	20.434	20.517	8.091	6.349	2.929	11.664
7	Bandung city Kulon	10.770	9.586	10.878	5.101	3.762	8.019
8	Bandung city Wetan	20.947	15.581	13.183	5.000	2.722	11.487
9	Batununggal	11.284	8.342	6.393	1.992	1.246	5.851
10	Bojongloa Kaler	13.673	13.452	6.122	4.086	1.942	7.855
11	Bojongloa Kidul	7.776	8.696	9.013	3.713	2.255	6.291
12	Buah Batu	4.928	3.797	2.565	1.147	1.118	2.711
13	Cibeunying Kaler	6.752	5.805	6.418	2.253	1.831	4.612
14	Cibeunying Kidul	5.975	5.955	5.933	1.869	1.183	4.183
15	Cibiru	15.437	12.611	8.134	3.531	2.237	8.390
16	Cicendo	12.088	9.862	12.283	2.724	1.801	7.751
17	Cidadap	14.498	9.007	5.228	2.634	1.718	6.617
18	Cinambo	15.347	11.569	10.195	7.399	13.304	11.563
19	Coblong	11.804	9.273	8.619	2.745	1.403	6.769
20	Gedebage	10.482	9.630	5.357	3.318	2.251	6.207
21	Kiaracondong	7.259	6.512	6.473	4.836	3.182	5.652
22	Lengkong	8.399	7.871	7.201	6.576	2.768	6.563
23	Mandalajati	22.092	15.578	9.189	4.394	3.831	11.017
24	Panyileukan	15.140	9.524	11.452	3.998	2.392	8.501
25	Rancasari	9.354	7.021	2.734	1.135	1.946	4.438
26	Regol	6.023	5.520	3.505	2.564	2.249	3.972
27	Sukajadi	6.005	6.064	5.032	3.070	2.232	4.481
28	Sukasari	12.990	15.824	10.967	5.183	2.848	9.563
29	Sumur Bandung city	6.000	9.319	8.761	4.753	2.324	6.232
30	Ujung Berung	12.994	7.969	8.602	3.648	1.871	7.017
Mean		11.625	9.339	7.785	3.926	2.741	7.083

3.3 Bayesian hierarchical spatiotemporal modeling

We analyzed the data for spatiotemporal autocorrelation before performing spatiotemporal modeling. We used spatiotemporal Moran's I to measure spatiotemporal autocorrelation on spatiotemporal data. The calculation yields a large spatiotemporal Moran's I value of 0.6173 (p-value = 0.0196), indicating that there is evidence of spatiotemporal autocorrelation in diarrhea data for children under the age of five in Bandung. Next, the models are estimated utilizing two likelihood distributions, Poisson and Negative Binomial, and two temporal trends, random walk order 1 (RW1) and random walk order 2 (RW2). This method yields sixteen models. Multiple criteria, including DIC, WAIC, MAE, RMSE, R, MPL, and CPO are used to evaluate the performance of these models. The primary focus of our evaluation is CPO Failure, which indicates that the Poisson likelihood distribution is inadequate for representing the frequency of diarrhea cases in children under 5, as its CPO Failure value is greater than zero. Consequently, we evaluate the models using different criteria. In general, the models have comparable DIC and WAIC values. In contrast, the model that includes a NB likelihood and a temporal trend RW2, specifically interaction types I and III, exhibits a statistically significant reduction in both MAE and RMSE in comparison to the alternative models. We chose a model with a NB likelihood, a RW2 to capture temporal trends, and interaction type I because we needed to find a good balance between model complexity and performance. This specific model will be designated Model I.

Tabel 2
Model Selection

Model	Likelihood	Temporal Trend	DIC	WAIC	MAE	RMSE	R	CPO Failure	MPL
Type I	Poisson	RW1	10513.626	10275.573	10.565	13.280	0.461	1182	-6288.384
Type II	Poisson	RW1	10524.019	10545.248	30.422	36.735	0.544	377	-8306.206
Type III	Poisson	RW1	10514.812	10278.838	9.720	12.422	0.463	1188	-6299.213
Type IV	Poisson	RW1	10507.689	10496.642	23.941	42.382	0.509	448	-8138.201
Type I	NB	RW1	12141.771	12145.083	10.957	13.651	0.461	0	-6288.191
Type II	NB	RW1	11744.924	11755.624	25.491	31.597	0.544	0	-8188.070
Type III	NB	RW1	12255.183	12260.717	11.663	14.342	0.463	0	-6293.719
Type IV	NB	RW1	11734.374	11743.096	12.021	17.622	0.509	0	-8006.922
Type I	Poisson	RW2	10515.984	10281.183	6.680	10.542	0.461	1183	-6405.967
Type II	Poisson	RW2	10810.414	11209.093	46.252	59.582	0.544	126	-12271.217
Type III	Poisson	RW2	10515.445	10275.440	7.245	11.385	0.463	1195	-6414.339
Type IV	Poisson	RW2	10727.154	11019.627	1.540E+27	2.910E+28	0.509	209	-11869.394
Type I	NB	RW2	12154.207	12166.113	6.692	10.518	0.461	0	-6407.446
Type II	NB	RW2	11859.172	11866.325	136.497	399.434	0.544	0	-11511.499
Type III	NB	RW2	12259.836	12266.430	6.740	10.613	0.463	0	-6403.770
Type IV	NB	RW2	11847.509	11856.527	447.889	1667.593	0.509	0	-11313.034



(A) (B)

Fig. 2. (A) The probability integral transform (PIT) and (B) observed versus predicted cases

Fig. 2 (A) provides support for Model I, as the PIT histogram closely resembles the uniform distribution histogram. This indicates that the model is effectively capturing the underlying distribution of the data, as the observed values align well with the expected uniform distribution. Furthermore, Fig. 1 (B) demonstrates a strong correlation between the predicted and observed cases. The points on the scatter plot cluster closely along the diagonal line, indicating a close alignment between the predicted values and the actual observations. This correlation reinforces the accuracy and reliability of the model's predictions, as it successfully captures the patterns and trends present in the data.

3.4 Relative Risk and Exceedance Proability Estimation

Model I is considered to provide an estimation of the relative risk, exceedance probability and an accurate forecast for January to December 2022. Table 3 shows the fixed effects component exhibits an average relative risk of 0.655%. Furthermore, Table 4 shows the hyperparameter of the random effect components.

Table 3
Estimated overall relative risk (intercept)

Parameter	Mean	SD	Parameter	Mean	SD
Overall relative risk	0.655	1.103	0.541	0.654	0.793

Table 4
Estimated hyperparameter

Hyperparameter	Mean	SD	*q*(0.025)	*q*(0.50)	*q*(0.975)	Fractional Variance
Overdispersion	11.105	1.536	8.480	10.967	14.491	
Leroux coefficient	0.577	0.086	0.398	0.581	0.730	
SD for Spatially structured effect	0.237	0.027	0.188	0.236	0.295	6.223
SD for Spatially unstructured effect	0.337	0.015	0.310	0.336	0.367	12.557
SD for Temporally structured effect	0.771	0.068	0.648	0.766	0.917	65.628
SD for Temporally unstructured effect	0.202	0.008	0.185	0.202	0.218	4.500
SD for Seasonal effects	0.002	0.002	0.000	0.002	0.008	0.001
SD for interaction effect type I	0.317	0.020	0.280	0.316	0.359	11.092

Table 3 shows the six random effect components that contribute to the spatiotemporal transmission of diarrhea disease for children under five years old in Bandung, Indonesia. These include (i) spatially structured effects, (ii) spatially unstructured effects, (iii) temporally structured effects, (iv) temporally unstructured effects, (v) seasonal effects, and (vi) space-time interaction effects. The four main factors that critically explain diarrhea transmission are temporally structured effects, spatially structured effects, spatially unstructured effects, and space-time interaction effects. Notably, the seasonal factor does not contribute significantly to the explanation of diarrhea's temporal variation. The analysis reveals that temporally structured effects account for 65.628% of the spatiotemporal variability in diarrhea, as measured by the fractional variance. In addition, the average standard deviation of these effects is calculated to be 0.77. The contribution of spatially structured effects to the total variation is 6.223%, whereas the contribution of unstructured effects is 12.557%. The space-time interaction accounts for 11.092% of the observed variation in spatiotemporal phenomena. In contrast, the seasonal component accounts for a negligible 0.001% of the total variation, whereas the temporally unstructured effect component accounts for a substantial 4.500%. Through careful consideration of these factors, we can gain important insights into the complex spatiotemporal transmissions of diarrhea, thereby facilitating the development of targeted interventions and preventive measures. As indicated by a parameter value of 11.105, the results of the analysis reveal a significant degree of overdispersion. In addition, a correlation coefficient of 0.57 demonstrates that our research demonstrates a significant spatial correlation. Fig. 3 shows the relative contributions of the six components to the overall diarrhea risk in Bandung.

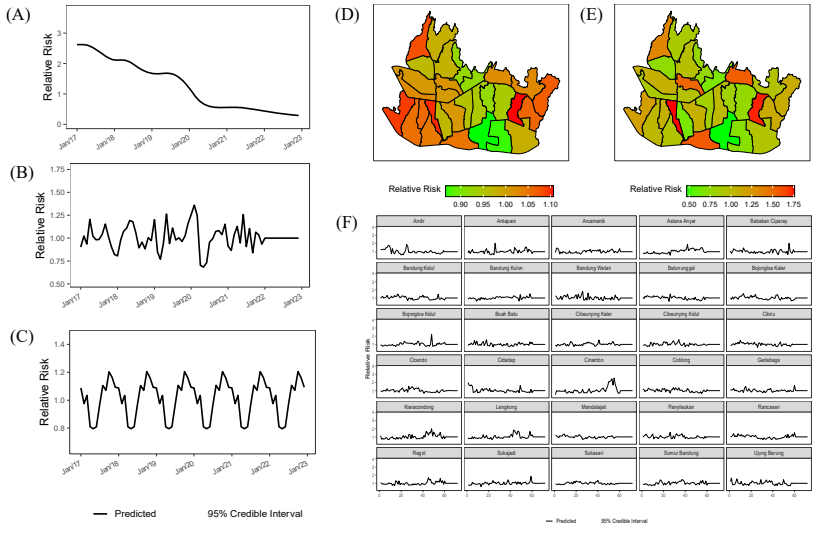


Fig. 3. (A) Temporally structured, (B) Temporally Unstructured, (C)) Seasonal, (D) Spatially structured, (E) Spatial Unstructured, and (F) Interaction effects

Fig. 3 (A) shows the temporally structured effect modeled with an RW2, which reveals an ongoing decrease in the relative risk of diarrhea from 2017 to 2022. Notably, the most significant decline occurred between 2019 and 2020, indicating an upward trend in the reduction of diarrhea cases over time. Fig. 3 (D) also illustrates the spatially structured effect that was modeled using the Leroux model. The results clearly identify clusters of high-risk areas for diarrhea on the western and eastern sides of the city of Bandung city. These clusters provide valuable insights into the disease's spatial distribution, allowing for targeted interventions and concentrated efforts to reduce the risk of diarrhea in these specific areas.

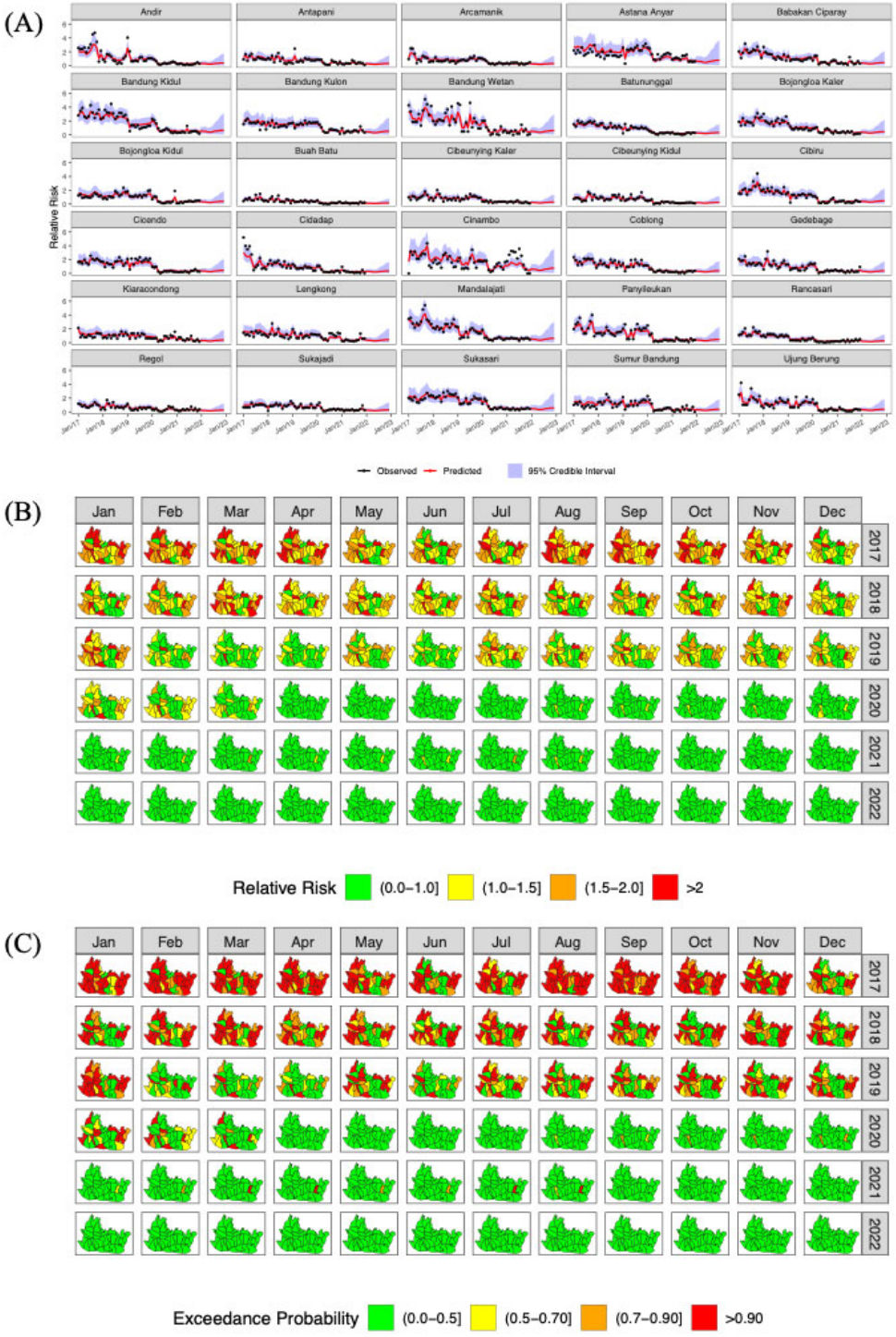
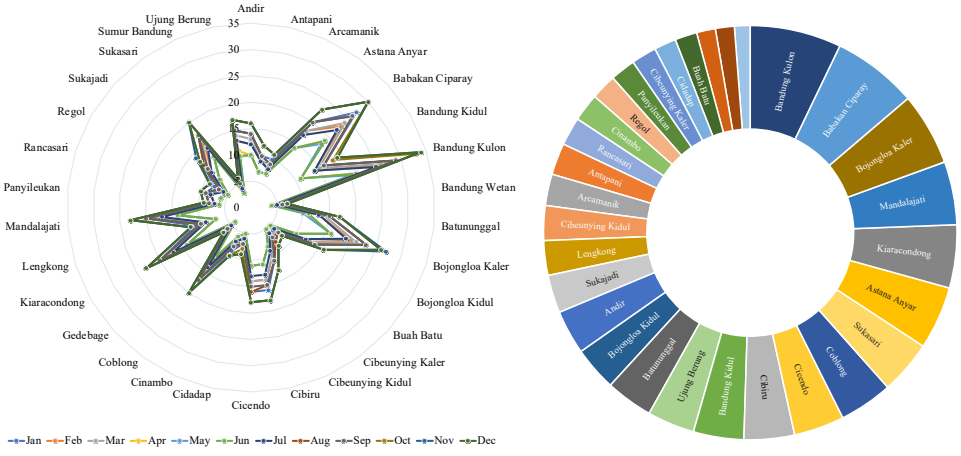


Fig. 4. (A) Temporal trend, (B) Spatiotemporal maps, and (C) Exceedance probaility of of the estimated relative risk over January 2017 – December 2022 in Bandung city

From 2017 to 2022, Fig. 4(A) depicts the temporal trend of the predicted relative risk of diarrhea in children under five years old in Bandung. In general, the relative risk decreases each year, with the greatest decline occurring at the end of 2019. Fig. 4(B) depicts the relative risk distribution map, whereas Fig. 4(C) depicts the identified hotspots based on the exceedance probability measure. Fig. 4(B-C) demonstrates that the western and southern regions of Bandung exhibited a higher and statistically significant relative risk of diarrhea during the period of 2017-2019. However, a substantial reduction in risk is observed beginning in early 2020. Notably, the risk continues to decline from 2020 to 2022, and the projected values for 2022 indicate the absence of hotspots. These findings indicate a positive trend in the reduction of relative risk, especially in 2022 as predicted. The analysis highlights the efficacy of interventions and preventive measures in reducing the risk of diarrhea in children under the age of five in Bandung, resulting in better public health outcomes.

3.5 Forecasted incidences of diarrhea

We predicted the number of diarrhea cases in 30 subdistricts of Bandung city from January to December 2022 using Bayesian spatiotemporal models with type I interactions. For the forecast, we utilized the infant birth rate for the same time period. We used the number of babies born in 2021 as a proxy for 2022 on the assumption that the birth rate is relatively stable over multiple years. Fig. 5 depicts the forecasting outcomes, and Table 5 provides a comprehensive breakdown. Forecasts indicate a continual decline in the overall number of diarrhea cases. Nonetheless, several districts continue to experience a heavy caseload. Bandung city Kulon, Babakan Ciparay, Bojonglua Kaler, Mandala Jati, and Kiara Condong are the five sub-districts with the highest monthly case counts throughout 2022. Each of these districts consistently reported a sizeable number of cases, which is noteworthy. In 2022, the total number of cases for each district exceeds 200. Fig. 5 and Table 5 provide a comprehensive visual and numerical representation of the forecasting results, highlighting the persisting high burden in specific districts while illustrating the decreasing trend in diarrhea cases. These results emphasize the need for targeted interventions and concentrated efforts to address the ongoing problems in these high-risk regions.



(A) (B)

Fig. 5. (A) Monthly Forecasted Diarrhea Incidences 2022 (B) Cummulative Forecasted Diarrhea Incidences 2022 in Bandung city, Indonesia

Table 5
Forecasted Diarrhea Incidences in Bandung city, Indonesia: January to December 2022

Id	District	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1	Andir	14	12	13	10	10	10	12	14	14	16	16	16	157
2	Antapani	10	9	9	7	7	7	9	10	10	12	12	12	114
3	Arcamanik	10	8	9	7	7	7	8	9	9	11	11	11	107
4	Astana Anyar	20	18	18	14	14	14	17	20	20	23	23	23	224
5	Babakan Ciparay	27	23	24	19	18	19	22	26	26	30	30	30	294
6	Bandung city Kidul	16	14	15	11	11	11	14	16	16	18	19	19	180
7	Bandung city Kulon	29	26	27	21	20	21	25	29	29	33	34	34	328
8	Bandung city Wetan	6	5	5	4	4	4	5	6	6	7	7	7	66
9	Batununggal	15	13	14	11	10	11	13	15	15	17	17	17	168
10	Bojongloa Kaler	23	20	21	16	16	16	19	23	23	26	27	26	256
11	Bojongloa Kidul	14	12	13	10	10	10	12	14	14	16	16	16	157
12	Buah Batu	7	6	6	5	5	5	6	7	7	8	8	8	78
13	Cibeunying Kaler	8	7	7	5	5	5	6	8	7	9	9	9	85
14	Cibeunying Kidul	11	10	10	8	8	8	9	11	11	13	13	13	125
15	Cibiru	16	14	14	11	11	11	13	15	15	18	18	18	174
16	Cicendo	16	14	14	11	11	11	13	16	15	18	18	18	175

Table 5
Forecasted Diarrhea Incidences in Bandung city, Indonesia: January to December 2022 (Continued)

Id	District	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
17	Cidadap	7	7	7	5	5	5	6	7	7	8	9	9	82
18	Cinambo	9	7	8	6	6	6	7	8	8	10	10	10	95
19	Coblong	17	15	16	12	12	12	14	17	17	20	20	20	192
20	Gedebage	6	6	6	4	4	4	5	6	6	7	7	7	68
21	Kiaracondong	20	18	19	14	14	14	17	20	20	23	23	23	225
22	Lengkong	11	9	10	7	7	7	9	10	10	12	12	12	116
23	Mandalajati	20	18	18	14	14	14	17	20	20	23	23	23	224
24	Panyileukan	8	7	7	6	6	6	7	8	8	9	9	9	90
25	Rancasari	9	8	8	6	6	6	8	9	9	10	10	10	99
26	Regol	8	7	7	5	5	5	6	7	7	9	9	9	84
27	Sukajadi	12	10	11	8	8	8	10	12	11	13	14	13	130
28	Sukasari	17	15	16	12	12	12	14	17	17	20	20	20	192
29	Sumur Bandung city	5	4	5	3	3	3	4	5	5	6	6	6	55
30	Ujung Berung	15	13	14	11	10	10	13	15	15	17	17	17	167
Total		406	355	371	283	279	282	340	400	397	462	467	465	4507

4. Discussion

We estimated the parameters of a spatiotemporal model using a hierarchical Bayesian framework. The main purpose of our study was to determine the spatial and temporal transmission of diarrhea among children under 5 in Bandung, Indonesia, and to obtain an accurate forecast from January to December 2022. We analyzed the data collected by Bandung city health office from January 2017 through December 2021.

We discovered that four primary factors contribute to the spatiotemporal variation of diarrhea transmission among children under 5 in the city of Bandung. These factors are (i) temporally structured effects, which explain the decline in diarrhea risk during the study period; (ii) spatially structured effects, which explain the spatial cluster in some close areas; (iii) spatially unstructured effects, which explain spatial heterogeneity; and (iv) spatiotemporal interaction type I. Additionally, the presence of spatially unstructured effects suggests that a variety of factors connected to local spatial heterogeneity influence the disease's occurrence within a region (Ghosh et al., 2023; Li et al., 2020; Ngesa et al., 2014; Raza et al., 2020). In addition, the analysis has successfully identified a Type I spatiotemporal interaction component, indicating that additional factors contribute to the observed spatiotemporal variation. This interaction is based on an unstructured space-time phenomenon, which means that there are independent variables without a temporal pattern that contribute to the spatiotemporal fluctuations in diarrhea cases (Amegbor & Addae, 2023; Jaya & Folmer, 2020; 2022). In addition, there is no seasonal pattern in the transmission of diarrhea risk. It is consistent with Barnes et al.'s (1998) previous publication. It suggests that diarrhea is not affected by seasonal variations and can manifest throughout the year without identifiable regular trends.

According to four significant factors, our prediction results indicate that the relative risk of diarrhea decreases during period 2017 to 2021. At the end of 2019, the decline was the greatest. In addition, we found strong spatial clustering in certain western and eastern districts of Bandung, indicating that diarrhea transmission typically takes place in close proximity. The result is consistent with Ntirampeba et al.'s (2018) study findings. This result demonstrates that diarrhea can be transmitted due to shared environmental characteristics such as sanitation condition (Sudasman et al., 2019; Otsuka et al., 2019).

Our forecasting model forecasts that the incidence and relative risk of diarrhea will continue to decrease through December 2022. However, it is noteworthy that certain subdistricts, namely Bandung City Kulon, Babakan Ciparay, and Bojongloa Kaler, persist in encountering a substantial volume of diarrhea cases. In order to effectively tackle this enduring issue, it is imperative to prioritize targeted interventions and allocate resources accordingly. By prioritizing the enhancement of sanitation practices, raising awareness about hygiene, and implementing suitable preventive measures, it is possible to diminish the incidence of cases in these regions with elevated risk and safeguard the welfare of the impacted communities.

The analysis conducted in our study has yielded a novel hypothesis indicating that enhanced sanitation practices have a protective effect against the occurrence of diarrhea. This result is consistent with previous research (Sudasman et al., 2019; Otsuka et al., 2019). Nevertheless, the issue of water quality continues to be a matter of concern, underscoring the necessity for forthcoming research to incorporate supplementary indicators in the assessment of drinking water quality. The decrease in the relative risk of diarrhea in Bandung city can be attributed to a range of factors, encompassing political advancements, national policies and strategies, disease control initiatives, and enhancements in sanitation and water infrastructure.

However, it is essential to recognize the limitations of our research. The data were collected by the Health Office of Bandung through surveillance efforts. Noting that these are secondary data that may be susceptible to underreporting is essential. In addition, risk factors were omitted from our study because its primary objective was to obtain accurate forecast values for the specified time periods.

5. Conclusions

Using Bayesian hierarchical spatiotemporal modeling, the spatiotemporal transmissions of diarrhea among children under five in Bandung, Indonesia have been successfully explained. Additionally, the model accurately predicted the relative risk and incidence of diarrhea from January to December 2022. The result of this study suggests that the incidence of diarrhea is not randomly distributed across space and time and tends to be concentrated in regions with poor sanitation, such as the western and eastern regions. The analysis also revealed that the seasonal factor does not adequately explain diarrhea transmission in the city of Bandung. Although there has been a general decrease in the risk of diarrhea, the anticipated results indicate that certain districts, such as Bandung Kulon, have a significant number of cases.

References

Alebel, A., Tesema, C., Temesgen, B., Gebrie, A., Petrucka, P., & Kibret, G. (2018). Prevalence and determinants of diarrhea among under-five children in Ethiopia: A systematic review and meta-analysis. *PLoS ONE, 13*(6), e019968.

Amegbor, P., & Addae, A. (2023). Spatiotemporal analysis of the effect of global development indicators on child mortality. *International Journal of Health Geographics*, 22(9), 1-15.

Assad, D., Cara, J., & Ortega-Mier , M. (2023). Comparing Short-Term Univariate and Multivariate Time-Series Forecasting Models in Infectious Disease Outbreak. *Bulletin of Mathematical Biology*, 85(9), 1-51.

Bandung HD (2018) *Health profile of Bandung municipality in 2017*. Bandung Government, Bandung

Bandung HD (2019) *Health profile of Bandung municipality in 2018*. Bandung Government, Bandung

Bandung HD (2020) *Health profile of Bandung municipality in 2019*. Bandung Government, Bandung

Bandung HD (2021) *Health profile of Bandung municipality in 2020*. Bandung Government, Bandung

Bandung HD (2022) *Health profile of Bandung municipality in 2021*. Bandung Government, Bandung

Barnes, G., Uren, E., Stevens, K., & Bishop, R. (1998). Etiology of Acute Gastroenteritis in Hospitalized Children in Melbourne, Australia, from April 1980 to March 1993. *Journal of Clinical Microbiology*, 36(1), 133–138.

Bivand, R., Gómez-Rubio, V., & Rue, H. (2015). Spatial data analysis with R-INLA with some extensions. *Journal of Statistical Software January*, 63(20), 1-31.

Blangiardo, M., & Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Chennai: John Wiley & Sons.

Fang, X., Liu, W., Ai, J., He, M., Wu, Y., Shi, Y., . . . Bao , C. (2020). Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC Infectious Diseases*, *20*(222), 1-8.

George, C., Perin, J., Calani, K., Norman, W., Perry, H., Davis Jr, T., & Lindquist, E. (2014). Risk Factors for Diarrhea in Children under Five Years of Age Residing in Peri-urban Communities in Cochabamba, Bolivia. *American Journal of Tropical Medicine and Hygiene*, *91*(6), 1190–1196.

Ghosh, K., Chakraborty, A., & SenGupta, S. (2023). Identifying spatial clustering of diarrhoea among children under 5 years across 707 districts in India: a cross sectional study. *BMC Pediatrics*, *23*(272), 1-15.

Jaya, I., & Folmer, H. (2020). Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia. *Journal of Geographical Systems*, *22*, 105–142.

Jaya, I., & Folmer, H. (2021). Bayesian spatiotemporal forecasting and mapping of COVID-19 risk with application to West Java Province, Indonesia. *Journal of Regional Science*, 64(1), 849-881.

Jaya, I., Folmer, H., & Lundberg, J. (2022). A joint Bayesian spatiotemporal risk prediction model of COVID-19 incidence, IC admission, and death with application to Sweden. *The Annals of Regional Science*, 1-45. https://doi.org/10.1007/s00168-022-01191-1

Jaya, I., Ruchjana, B., Andriyana, Y., & Agata, R. (2019). Clustering with spatial constraints: The case of diarrhea in Bandung city, Indonesia. *Journal of Physics: Conference Series*, 1397, 012068.

Knorr, H. L. (2000). Bayesian Modeling of inseparable space-time variation in disease risk. *Statistics in Medicine*, *19*(17-18), 2555-2567.

Kotloff, K., Platts-Mills , J., Nasrin, D., Roose, A., Blackwelder, W., & Levine, M. (2017). Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. *Vaccine*, *35*(49), 6783-6789.

Lawson, A. (2010). Hotspot detection and clustering: ways and means. *Environmental and Ecological Statistics*, *17*, 231–245.

Lawson, A., & Rotejanaprasert, C. (2014). Childhood Brain Cancer in Florida: A Bayesian Clustering Approach. *Statistics and Public Policy*, *1*(1), 99-107.

Lee, J., Han, G., Fulp, W., & Giuliano, A. (2012). Analysis of overdispersed count data: application to the Human Papillomavirus Infection in Men (HIM) Study. *Epidemiology & Infection*, *140*(6), 1087–1094.

Leroux, B., Lei, X., & Breslow, N. (1999). *Estimation of disease rates in small areas: a new mixed model for spatial dependence*. In M. Halloran, & D. Berry, Statistical Models in Epidemiology, the Environment and Clinical Trials (pp. 135–178). New York: Springer-Verlag.

Li, R., Lai, Y., Feng, C., Dev, R., Wang, Y., & Hao, Y. (2020). Diarrhea in Under Five Year-old Children in Nepal: A Spatiotemporal Analysis Based on Demographic and Health Survey Data. *International Journal of Environmental Research and Public Health*, 17(2140), 1-17.

Ministry of Health RI, (2019). *Profil Kesehatan Indonesia Tahun 2018*. Jakarta: Ministry of Health RI

Mohebbi, M., Wolfe, R., & Forbes, A. (2014). Disease Mapping and Regression with Count Data in the Presence of Overdispersion and Spatial Autocorrelation: A Bayesian Model Averaging Approach. *International Journal of Environmental Research and Public Health, 11*, 883-902.

Ngesa, O., Mwambi, H., & Achia, T. (2014). Bayesian Spatial Semi-Parametric Modeling of HIV Variation in Kenya. *PLoS ONE*, 9(7), 1-11.

Ntirampeba, D., Neema, I., & Kazembe, L. (2018). Modelling spatio-temporal patterns of disease for spatially misaligned data: An application on measles incidence data in Namibia from 2005-2014. *PLoS ONE*, *13*(9), e0201700.

Otsuka, Y., Agestika, L., Widyarani, Sintawardani, N., & Yamauchi, T. (2019). Risk Factors for Undernutrition and Diarrhea Prevalence in an Urban Slum in Indonesia: Focus on Water, Sanitation, and Hygiene. *American Journal of Tropical Medicine and Hygiene, 100*(3), 727–732.

Paul, P. (2020). Socio-demographic and environmental factors associated with diarrhoeal disease among children under five in India. *BMC Public Health*, *20*(1886), 1-11.

Pena, D., & SaNchez , I. (2007). Measuring The Advantages Of Multivariate Vs. Univariate Forecasts. *Journal of Time Series Analysis*, 28(6), 886-906.

Rahayu, A., Darmawan, G., & Jaya, I. (2023). Calculation of the Risk Index for Diarrhea, ISPA, and Pneumonia in Toddlers in the City of Bandung Using Geographically Weighted Principal Component Analysis. *Indonesian Journal of Advanced Research* (IJAR), *2*(4), 285 - 300.

Raza, O., Mansournia, M., Foroushani, A., & Holakouie-Naieni, K. (2020). Exploring spatial dependencies in the prevalence of childhood diarrhea in Mozambique using global and local measures of spatial autocorrelation. *Medical Journal of The Islamic Republic of Iran*, *34*(59), 1-7.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B*, *71*(2), 319–392.

Simpson, D., Rue, H., Riebler, A., Martins, T., & Sørbye, S. (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, *32*(1), 1–28.

Spiegelhalter, D., Best, N., Carlin, B., & Linde , A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Socieity Statistical Methodlogy B*, *64*(4), 583-639.

Sudasman, F., Bachtiar, A., Laelasari, E., & Ciptaningtyas, R. (2019). *Factors Associated With The Risk Of Diarrhea In Children Under Five In Bandung, West Java*. International Conference on Public Health (pp. 1-5). Surabaya: Masters Program in Public Health, Graduate School, Universitas Sebelas Maret.

Ugarte, M. D., Adin, A., Goicoa, T., & Militino, A. F. (2014). On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research, 23*(6), 507–530. doi:10.1177/0962280214527528

Ugarte, M., Goicoa, T., Etxeberria, J., & Militino, A. (2012). Projections of cancer mortality risks using spatio-temporal P-spline models. Stat Methods Med Res, 21(5), 545-560.

Walker, C., Rudan, I., Liu, L., Nair, H., Theodoratou, E., Bhutta, Z., . . . Black, R. (2013). Global burden of childhood pneumonia and diarrhoea. *Thelancet*, *381*, 1405-1416.

Wang, X., Ryan, Y., & Faraway, J. (2018). *Bayesian Regression Modeling with INLA*. Boca Raton: CRC Press.

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, *11*, 3571-3594.

WHO. (2017). Diarrhoeal disease. Retrieved from WHO: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease#:~:text=five%20years%20old.-,Diarrhoeal%20disease%20is%20the%20second%20leading%20cause%20of%20death%20in,that%20are%20necessary%20for%20survival. Accessed 1 July 2023