Spatial Analysis of Pneumonia in Toddlers on Sumatra Island Using Geographically Weighted Poisson Regression

R N S L Gaol[1], M B Potenza[1], N F Ihsan[1], G A F Pratama[1], and S M Berliana[2,*]

[1] Department of Statistics, Politeknik Statistika STIS, Jakarta, Indonesia
[2] Research Unit of Sustainable Development Goals, Politeknik Statistika STIS, Jakarta, Indonesia

**Corresponding author's email*: sarni@stis.ac.id

Abstract: Pneumonia remains a leading cause of mortality among toddlers (aged 1 to less than 5 years) in Indonesia, with notable spatial disparities across Sumatra Island. This study examines factors influencing pneumonia incidence in toddlers using a Geographically Weighted Poisson Regression (GWPR) model to capture local variations in the effects of community health centers, complete basic immunization coverage, exclusive breastfeeding rates, and low birth weight (LBW) prevalence. Analyzing 2022 cross-sectional data from 154 districts/cities on Sumatra, the global Poisson regression model confirmed all predictors as statistically significant at the 5% level. The GWPR model with a fixed Gaussian kernel outperformed the global model, revealing five regional clusters with distinct combinations of significant variables. The dominant cluster (140 locations) showed significant effects from all predictors, while smaller clusters (14 locations) highlighted localized patterns, such as reliance on immunization and breastfeeding in rural areas like Rejang Lebong. These findings underscore the need for tailored interventions to address regional disparities in toddler pneumonia.
Key Words: Geographically Weighted Poisson Regression, Local Health Interventions, Pneumonia in Children, Spatial Analysis.

1. Introduction

Pneumonia is a respiratory infection that poses a significant challenge for children's health, particularly in the toddler age group. Pneumonia is an acute infection that affects lung tissue, disrupting the oxygen exchange process essential for survival. Pneumonia in toddlers is characterized by symptoms such as coughing, difficulty breathing, and rapid breathing. The criteria for a fast breathing rate vary by age group, namely more than or equal to 60 breaths per minute for infants under 2 months, more than or equal to 50 breaths per minute for ages 2 to less than 12 months, and more than or equal to 40 breaths per minute for ages 1 to less than 5 years [1]. To date, pneumonia control efforts are still prioritized in toddlers because of their high vulnerability.

According to the World Health Organization, Indonesia ranked ninth in the world for infant pneumonia mortality in 2019, with a mortality rate of 32 per 1,000 live births. It means that approximately 2 to 3 children die every hour from pneumonia [2]. The 2022 Ministry of Health report indicates that the number of pneumonia cases in infants increased by 7.38 percent from the previous year, suggesting an increasing disease burden amid efforts to improve the national health system [1]. Specifically, the increase in pneumonia cases in infants presents an obstacle to achieving Sustainable Development Goal 3.2, which aims to ensure a healthy life and improve the well-being of all age groups

[3]. This increase not only affects the health aspect but also highlights the suboptimal basic health services and inadequate public awareness in disease prevention. Given these alarming statistics, it is crucial to explore effective preventive measures. Immunization is one of the most effective strategies for preventing pneumonia. Several vaccines are effective in preventing pneumonia, depending on the cause, including those for measles, Haemophilus influenzae type b (Hib), and pneumococcal conjugate vaccine (PCV). The Indonesian government launched a national PCV immunization program in 2022 to prevent pneumonia in children. This program is expected to significantly reduce morbidity and mortality from pneumonia [4].

Building on the role of immunization, a toddler's immunization status also plays a significant role in the incidence of pneumonia. a relationship between nutritional status, immunization status, and the incidence of pneumonia in infants. Toddlers with poor nutritional status are at higher risk of developing pneumonia due to weakened immune systems [5].

In addition to immunization and nutrition, early feeding practices can further influence a child's susceptibility to the disease. Exclusive breastfeeding is one of the most significant protective factors in preventing pneumonia in toddlers. There is a significant negative association between exclusive breastfeeding and the incidence of pneumonia in toddlers. Exclusive breastfeeding provides optimal protection against various infectious diseases, including pneumonia, because it contains antibodies and immune substances essential for protecting infants from harmful pathogens [6].

Another critical factor that intersects with these vulnerabilities is the condition at birth. Low birth weight (LBW) is defined as a baby born weighing less than 2,500 grams, regardless of age. This condition is a significant factor in increasing infant mortality, morbidity, and disability, and has long-term impacts on their future lives. Low birth weight babies have immature immune systems and underdeveloped organs, making them more susceptible to respiratory infections, including pneumonia. Low birth weight (LBW) is one of the factors examined for its association with the incidence of pneumonia in toddlers. Low birth weight babies have smaller lung capacity and suboptimal respiratory function, making them more susceptible to complications when infected with pneumonia-causing pathogens.[7].

Although pneumonia is a national health problem, the distribution of cases is uneven across Indonesia, suggesting that location-specific factors may influence its incidence and thus requiring localized modeling to analyze its determinants. These variations arise because each location has specific characteristics that influence the number of pneumonia cases. This unevenness is evident in the striking differences between islands, provinces, and even districts/cities. One fascinating region to observe is the island of Sumatra. In this region, the distribution of pneumonia cases in toddlers shows inequality. For example, in 2022, the South Sumatra Provincial Health Office recorded a total of 6,663 cases of pneumonia in toddlers, while the Bengkulu Provincial Health Office recorded only 439 cases [8][9]. This inequality is also evident at the district/city level; for example, the Health Office report shows that in Lahat Regency, Musi Rawas Regency, and Lubuk Linggau City, there were no cases of pneumonia in toddlers. At the same time, Palembang City had a very high number of cases, namely 2,838 cases [8].

Based on this background of national challenges, preventive factors, and regional disparities, this study aims to analyze the factors influencing the number of pneumonia cases in toddlers on Sumatra Island using a spatial approach. Due to the uneven distribution of pneumonia cases across regions, the analytical approach used in this study is Geographically Weighted Poisson Regression (GWPR). This approach is expected to capture local variations in the influence of the independent variables considered, such as the number of health facilities, the percentage of complete basic immunizations, the percentage of exclusive breastfeeding, and the percentage of low birth weight (LBW) infants. The results of this study are expected to contribute to the formulation of more targeted health intervention policies, particularly in addressing childhood pneumonia on Sumatra Island.

2. Research Method

2.1. Poisson Regression Model

The Poisson distribution is a key statistical tool for modeling count data, particularly for events that occur independently and infrequently within a fixed interval of time or space [10]. For a discrete random variable Y with intensity parameter λ, the probability mass function is defined as follows.

$$P(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, y = 0, 1, 2, \dots; \lambda \geq 0 \tag{1}$$

From this foundation, the Poisson regression model is developed to relate the expected count $E(Y_i)$ to a set of independent variables. The natural logarithm of the expected value is expressed as a linear combination of independent variables

$$ln\big(E(Y_i)\big) = \boldsymbol{x}_i^T \boldsymbol{\beta} \tag{2}$$

Here, $\mathbf{x}_i = [1, x_{1i}, x_{i2}, \dots, x_{pi}]^T$ represents the vector of independent variables (including an intercept) for the i^{th} observation, and $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \dots, \beta_p]^T$ denotes the vector of regression coefficients. Consequently, the expected count is modeled as

$$E(Y_i) = \lambda_i = exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) \tag{3}$$

This framework enables the analysis of how various factors influence the occurrence of rare events, such as disease incidence, by capturing their effects through the regression coefficients.

2.2. Multicollinearity Detection

Multicollinearity occurs when two or more independent variables are highly correlated with each other [11], which can affect the stability of regression coefficient estimates. One common method used to detect multicollinearity is the Variance Inflation Factor (VIF). VIF is a measure that indicates how much the variability of regression coefficient estimates increases due to collinearity between independent variables in a regression model [12].

The VIF is calculated for each independent variable in the model and measures the correlation between that variable and the other independent variables in the model. The VIF value is obtained using the equation

$$VIF_k = \frac{1}{1 - R_k^2} \tag{4}$$

where R_k^2 is the coefficient of determination of the regression of the k^{th} independent variable against all other independent variables in the model. Values of $VIF < 5$ are generally considered not to cause collinearity problems [13].

2.3. Dispersion Testing

Dispersion describes the variability in data beyond what is explained by the mean. In Poisson regression, it specifically refers to the relationship between the mean and variance of the count data distribution [14]. The Poisson model assumes equidispersion, where the variance equals the mean.

$$Var(Y) = \mu \tag{5}$$

If the variance significantly exceeds the mean, overdispersion occurs, which can compromise the model's accuracy. Conversely, underdispersion arises when the variance is less than the mean. In this study, we assume the data satisfy the equidispersion assumption. However, this assumption is a limitation, and future research could explore alternative models, such as negative binomial regression or generalized Poisson regression, to address potential overdispersion or underdispersion.

2.4. Spatial Heterogeneity

Spatial heterogeneity occurs when the relationship between independent and dependent variables varies across regions. A key consideration in spatial heterogeneity is determining when differences are statistically significant and when they are insignificant, for example, due to small sample sizes [15]. To test for spatial heterogeneity, the Lagrangian Multiplier (LM) approach can be used. This test is performed by dividing the data into subsets and using Ordinary Least Squares (OLS) regression to test the following null hypothesis.

$H_0 : \sigma_1{}^2 = \sigma_2{}^2 = \ldots = \sigma_n{}^2 = \sigma^2$ (no spatial heterogenity)

$H_1 : At\ least\ one\ of\ \sigma_i{}^2 \neq \sigma^2$ (spatial heterogenity exists)

The LM test statistic is calculated as follows

$$LM = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T;\ f \sim \chi_k^2 \tag{6}$$

$$f = f_i = \left(\frac{e_i^2}{\sigma^2} - 1 \right) \tag{7}$$

where

e_i^2 : Squared residual from Ordinary Least Squares (OLS) regression for the i^{th} observation.

Z : An $(n \times (k+1))$ matrix of normalized covariate vectors for each observation.

σ^2 : The common variance under the null hypothesis.

The null hypothesis H_0 is rejected if the LM test statistic exceeds the critical value of the chi-square distribution with k degrees of freedom or if the p-value is less than the significance level α, indicating the presence of spatial heterogeneity across regions.

2.5. Spatial Weighting Matrix

A spatial weighting matrix, as described by LeSage [16], is an $n \times n$ matrix that quantifies the spatial relationships between observational units, typically based on either contiguity or a distance-based function between locations. This matrix is critical in spatial models like Geographically Weighted Poisson Regression (GWPR) to account for spatial dependencies. Commonly used kernel functions for constructing the spatial weighting matrix include the following [17]

1) Gaussian Kernel Function

   The Gaussian kernel assigns weights to all observations, with weights decreasing exponentially as the distance from the focal point increases, approaching but never reaching zero. The Gaussian kernel is defined as

$$w(d_{ij}) = exp\left(-\frac{1}{2} \left(\frac{d_{ij}}{h} \right)^2 \right) \tag{8}$$

   where

   d_{ij} : The distance between location *i* and location *j*.

   h : The bandwidth, controlling the spatial range of influence.

2) Bi-Square Kernel Function

   The bi-square kernel assigns weights that decrease with distance but drop to zero beyond a specified bandwidth, making it suitable for local analyses with a defined spatial extent. The bi-square kernel is expressed as

$$w(d_{ij}) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2, & \text{if } |d_{ij}| < h \\ 0, & \text{others} \end{cases} \tag{9}$$

Two approaches to kernel selection are fixed and adaptive kernels. In a fixed kernel, the bandwidth h is constant across all locations, resulting in a uniform spatial influence area for each point. Conversely, in an adaptive kernel, the bandwidth adjusts based on the number of neighboring observations, expanding in areas with lower observation density and contracting in denser areas to ensure a consistent number of neighbors [17].

2.6. Geographically Weighted Poisson Regression (GWPR) Model

Geographically Weighted Regression (GWR) is a spatial analysis technique designed to detect spatial nonstationarity, where relationships between variables vary across geographic locations [20]. Unlike global regression models, which assume uniform relationships across space, GWR estimates location-specific parameters, revealing local patterns that may be obscured in traditional models. This approach incorporates spatial information, such as location coordinates and independent variable attributes, to produce regression coefficients that vary by region, enabling spatially nuanced analysis of factors like pneumonia incidence across Sumatra. The GWR model for a given location is mathematically expressed as

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i;\ i = 1,2,\ldots,n \tag{10}$$

where

y_i	:	The dependent variable at location i.
x_{ik}	:	The value of the k^{th} independent variable at location i.
$\beta_0(u_i, v_i)$	:	The intercept at location i with coordinates (u_i, v_i).
$\beta_k(u_i, v_i)$	:	The regression coefficient for the k^{th} independent variable at location i.
(u_i, v_i)	:	The geographic coordinates of location i.
ε_i	:	The residual for location i.

Geographically Weighted Poisson Regression (GWPR) extends GWR to accommodate count data following a Poisson distribution, making it suitable for modeling outcomes like the number of pneumonia cases [21]. GWPR allows regression coefficients to vary spatially, capturing local variations in the relationship between the dependent variable and independent variables, such as health facility availability or immunization rates. The GWPR model is formulated as

$$\lambda_i(u_i, v_i) = \exp\left(\beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik}\right) \tag{11}$$

where $\lambda_i(u_i, v_i)$ is the expected count (mean of the Poisson distribution) at location i. This formulation enables the GWPR model to estimate spatially varying effects, providing insights into how factors influence pneumonia incidence differently across regions like Sumatra.

2.7. Model Evaluation

Evaluating the performance of spatial models like GWPR involves assessing how well the model balances predictive accuracy and complexity, particularly when calibrating spatial parameters such as bandwidth [17]. A primary metric for this purpose is the Corrected Akaike Information Criterion (AICc), which adjusts the standard AIC to account for small sample sizes, preventing overfitting. The AICc is calculated as follows [18]

$$AICc = AIC + \frac{(2p(p+1))}{(n-p-1)} \tag{12}$$

$$\text{AIC} = -2ln(L) + 2p \tag{13}$$

where
p : The number of parameters estimated in the model.
n : The sample size.
L : The maximum likelihood of the model.

The AICc penalizes complex models more heavily than AIC, especially for smaller datasets, ensuring a balance between explanatory power and parsimony. In the context of GWPR for analyzing pneumonia cases across Sumatra, a lower AICc indicates a model that effectively captures spatial variations while avoiding overfitting.

Another goodness-of-fit measure is the Bayesian Information Criterion (BIC), proposed by Schwarz (1978) [19], defined as

$$\text{BIC} = -2ln(L) + p\ ln(n) \tag{14}$$

A lower BIC value signifies a better-fitting model, with BIC imposing a stronger penalty on complexity due to the logarithmic sample size term, favoring simpler models compared to AICc. While both metrics evaluate model performance, BIC prioritizes parsimony, whereas AICc allows more flexibility for including additional variables, which is advantageous in complex spatial analyses.

The best model was selected using the corrected Akaike Information Criterion (AICc). AICc is a modification of AIC that imposes an additional penalty for models with a large number of parameters, especially when the sample size is relatively small. The AICc measure for model evaluation is calculated similarly to Equation (13).

To further assess how well the GWPR model explains variation in the data, the percent deviance explained is used, as formulated below

$$\%\text{Deviance explained} = \left(1 - \frac{\text{Deviance Model}}{\text{Deviance Null}}\right) \times 100 \tag{15}$$

Here, model deviance represents the deviance of the fitted GWPR model, calculated based on the log-likelihood of the model with the estimated parameters. Null deviance corresponds to the deviance of a baseline model, which includes only the intercept (i.e., a model assuming a constant mean without independent variables). A higher percentage of deviance explained indicates that the model captures a greater proportion of the variability in the data, reflecting better fit. In the context of this study, this metric quantifies how effectively the GWPR model accounts for spatial patterns in pneumonia incidence across Sumatra.

2.8. Operatiional Definition of Variables

This study utilizes secondary data sourced from BPS-Statistics Indonesia and Provincial Health Offices across Sumatra, covering all 154 districts/cities on the island. The data are cross-sectional, collected for the year 2022, with the district/city level as the unit of analysis. The study includes one dependent variable and four independent variables, as outlined in Table 1.

Table 1. Operatiional definition of variables.

Variable	Unit	Definition	Source
Number of pneumonia cases in toddlers (Y)	Count	The number of reported pneumonia cases among toddlers (children aged 1 to less than 5 years) in 2022.	Provincial Health Office
Number of health centers (X_1)	Count	The total number of community health centers (Puskesmas) providing primary healthcare services.	BPS-Statistics Indonesia

Percentage of complete basic immunization (X_2)	Percent	The proportion of children receiving complete basic immunizations, protecting against diseases such as tuberculosis, diphtheria, pertussis, tetanus, polio, hepatitis B, and measles.	Provincial Health Office
Percentage of exclusive breastfeeding (X_3)	Percent	The proportion of infants receiving only breast milk, without additional food or drink (except medically necessary supplements, such as medicines or vitamins in liquid form), for the first six months of life.	Provincial Health Office
Percentage of low birth weight (X_4)	Percent	The proportion of infants born weighing less than 2,500 grams, regardless of gestational age.	Provincial Health Office

3. Result and Discussion

3.1. Overview of Pneumonia Cases in Toddlers on Sumatra Island

Table 2 summarizes the descriptive statistics of the variables used to analyze pneumonia cases in toddlers (children aged 1 to less than 5 years) across districts/cities on Sumatra Island in 2022. Among the variables, the percentage of low birth weight (LBW) and the number of community health centers (Puskesmas) exhibited the lowest variability, with averages of 4.04% (standard deviation: 5.32) and 17.53 units (standard deviation: 9.31), respectively. This suggests relatively uniform distributions of these factors across regions. In contrast, the number of pneumonia cases in toddlers showed substantial variability, with an average of 220.20 cases per district/city and a standard deviation of 476.21, indicating significant interregional differences in disease burden.

Tabel 2. Summary of descriptive statistics.

Variable	Maximum	Minimum	Average	Std. Dev.
Number of pneumonia cases in toddlers (Y)	2838	0	220.20	476.21
Number of health centers (X_1)	51	0	17.53	9.31
Percentage of complete basic immunization (X_2)	131.31	0	77.82	26.59
Percentage of exclusive breastfeeding (X_3)	92.9	8.080	55.64	21.01
Percentage of low birth weight (X_4)	34.1	0	4.04	5.32

The spatial distribution of pneumonia cases in toddlers on Sumatra Island is illustrated in Figure 1. Cases are predominantly concentrated in the southern region, with notable clusters in central and eastern Sumatra. Palembang City and Batam City recorded the highest incidences, with 2,838 and 2,765 cases, respectively, in 2022. In contrast, several districts, particularly in North Sumatra Province, reported zero cases. This uneven distribution highlights significant regional disparities in pneumonia prevalence, likely influenced by variations in healthcare access, socioeconomic conditions, and environmental factors.

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

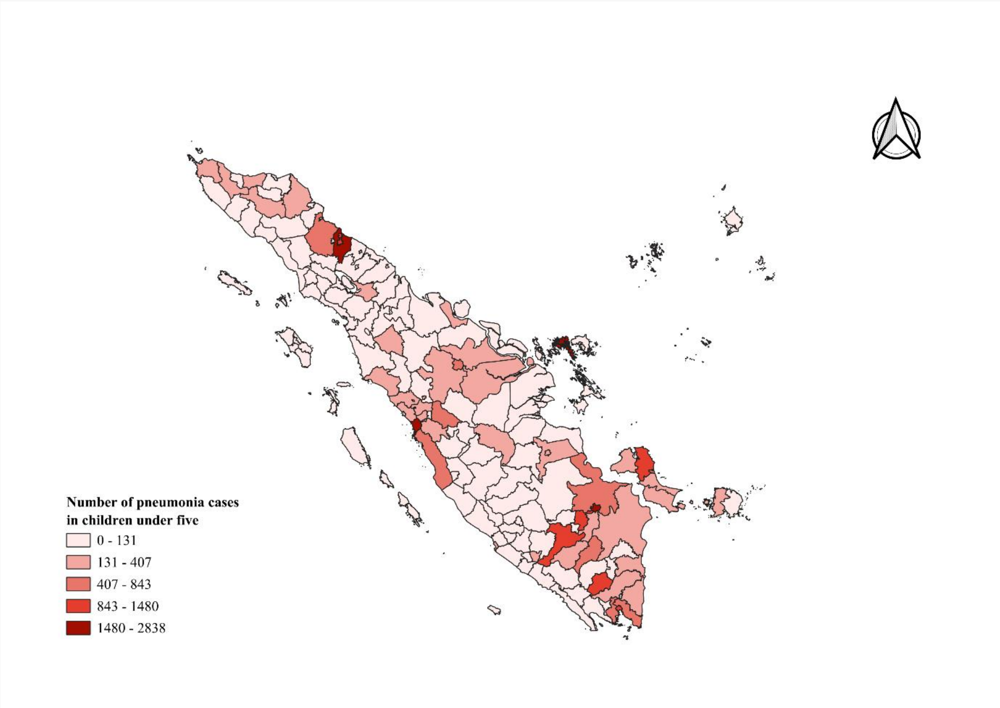

Figure 1. Distribution of pneumonia cases in toddlers, Sumatra Island, 2022.

3.2. Determinants of Pneumonia Cases in Toddlers

Prior to fitting the Poisson regression model, a multicollinearity test was conducted to ensure the independence of the predictor variables. Table 3 presents the Variance Inflation Factor (VIF) values, all of which are below 5, indicating no significant multicollinearity that could bias the model's estimates.

Table 3. VIF for multicollinearity assessment.

Variable	VIF
Number of health centers (X_1)	1.0217
Percentage of complete basic immunization (X_2)	1.0076
Percentage of exclusive breastfeeding (X_3)	1.0203
Percentage of low birth weight (X_4)	1.0477

The parameter estimates from the global Poisson regression model are shown in Table 4, with the model expressed as

$$\hat{\lambda} = exp(4{,}2987 - 0{,}0090\ {X_1}^* - 0{,}0050\ {X_2}^* + 0{,}0306\ {X_3}^* - 0{,}0640{X_4}^*) \quad (16)$$

All independent variables were statistically significant at the 5% significance level ($\alpha = 0.05$). For instance, adding one community health center (X_1) reduces the expected number of pneumonia cases in toddlers by a factor of 0.9910 (odds ratio), holding other variables constant. Similarly, higher percentages of complete basic immunization (X_2) and low birth weight (X_4) are associated with reduced pneumonia cases, while a higher percentage of exclusive breastfeeding (X_3) is associated with an increase, possibly reflecting regional confounding factors or data-specific patterns.

Table 4. Parameter estimates for the global Poisson regression model.

Variable	Est.	SE	z-stat	Odds Ratio
Intercept	4.2987	0.0292	147.1433*	73.6077

Number of health centers (X_1)	-0.0090	0.0007	-13.5180*	0.9910
Percentage of complete basic immunization (X_2)	-0.0050	0.0002	-26.7294*	0.9950
Percentage of exclusive breastfeeding (X_3)	0.0306	0.0003	92.0348*	1.0311
Percentage of low birth weight (X_4)	-0.0640	0.0014	-45.4254*	0.9380

*) Significant at $\alpha = 0.05$

3.3. *Spatial Heterogeneity Assessment*

To determine whether the relationships between the independent variables and pneumonia cases in toddlers vary across Sumatra, a spatial heterogeneity test was conducted. Table 5 presents the results of the Breusch-Pagan (BP) test, which yielded a statistically significant p-value (0.0471) at the 5% level, rejecting the null hypothesis of spatial homogeneity. This confirms the presence of spatial heterogeneity, justifying the use of a spatially varying model like GWPR.

Tabel 5. Spatial heterogeneity test results.

Spatial effect	BP-Test Statistic	p-value
Heteroscedasticity	9.6323	0.0471

3.4. *Comparison of Weighting Functions in GWPR*

In developing the GWPR model, various weighting functions were evaluated to identify the most optimal configuration. Table 6 compares the performance of fixed and adaptive kernel functions based on the Corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC). The fixed Gaussian kernel produced the lowest AICc (33,907.04) and BIC (33,992.69) with an optimal bandwidth of 1.156, indicating superior model fit. Consequently, this kernel was selected for subsequent spatial analyses.

Table 6. Comparison of weighting functions in GWPR models.

Model	AICc	BIC	Bandwidth
Adaptive bisquare	37,200.94	37,275.56	50.8456
Adaptive Gaussian	60,123.29	60,154.88	50.8456
Fixed bisquare	56,661.86	56,704.75	5.7994
Fixed Gaussian	33,907.04	33,992.69	1.156

3.5. *Model Comparison: Global vs. Local*

To assess whether the GWPR model outperforms the global Poisson regression model, a comparison was made using AICc and percent deviance explained, as shown in Table 7. The GWPR model with the fixed Gaussian kernel yielded a significantly lower AICc (33,907.0366 vs. 69,824.4110) and a higher percent deviance explained (58.59% vs. 14.51%) compared to the global model. These results indicate that the increased complexity of the GWPR model is justified by its superior ability to capture spatial variations in pneumonia incidence.

Table 7. Comparison of global and local models.

Model	AICc	*Percent deviance explained*
Global Poisson regression	69,824.411	14.51%
GWPR (Fixed Gaussian kernel)	33,907.0366	58.59%

3.6. *Spatial Analysis of GWPR Parameter Estimates*

The parameter estimates from the GWPR model with the fixed Gaussian kernel are summarized in Table 8. The percentage of exclusive breastfeeding (X_3) exhibits a positive influence on pneumonia cases across all districts/cities, suggesting a complex relationship that may reflect regional socioeconomic or health reporting factors. Conversely, the number of community health centers (X_1), percentage of complete basic immunization (X_2), and percentage of low birth weight (X_4) generally show negative associations with pneumonia cases in most regions, though their effects vary spatially, reflecting local differences in healthcare infrastructure and population characteristics.

Table 8. Summary of parameter estimates for the fixed Gaussian kernel GWPR model.

Variable	Mean	St. Dev.	Minimum	Median	Maximum
Intercept	4.7712	3.3512	-31.1389	5.1131	8.0077
Number of health centers (X_1)	-0.0227	0.0507	-0.1858	-0.0159	0.0814
Percentage of complete basic immunization (X_2)	-0.0060	0.0324	-0.0457	-0.0104	0.3543
Percentage of exclusive breastfeeding (X_3)	0.0288	0.0238	0.0024	0.0227	0.2090
Percentage of low birth weight (X_4)	-0.1322	0.2611	-0.8460	-0.0431	0.2147

Figure 2 illustrates the spatial distribution of districts/cities on Sumatra Island grouped by combinations of significant variables influencing pneumonia cases in toddlers (aged 1 to less than 5 years) in 2022, based on the fixed Gaussian kernel GWPR model. The majority of areas, comprising 140 districts/cities, fall into a dominant group (Group 1) where all independent variables—number of community health centers (X_1), percentage of complete basic immunization (X_2), percentage of exclusive breastfeeding (X_3), and percentage of low birth weight (LBW) (X_4)—are statistically significant at the 5% level. This widespread significance highlights the critical role of healthcare infrastructure, preventive behaviors, and birth-related factors in explaining spatial variations in pneumonia incidence across most of Sumatra.

Four smaller groups, totaling 14 locations, exhibit distinct combinations of significant variables, indicating localized deviations from the dominant pattern. For instance, Rejang Lebong Regency in Bengkulu Province (Group 5) is unique, with only the percentage of complete basic immunization (X_2) and exclusive breastfeeding (X_3) showing significant effects. In this rural area, the lack of significance for community health centers (X_1) and LBW (X_4) may reflect limited healthcare infrastructure or underreporting of LBW cases, emphasizing the prominence of immunization and breastfeeding as key drivers of pneumonia incidence. The complete grouping of districts/cities based on significant variables, as depicted in Figure 2, is detailed in Table 9.

3.7. Local Model Interpretation: Case of Palembang City

The GWPR model yields location-specific parameter estimates, reflecting the spatial heterogeneity in factors influencing pneumonia cases in toddlers across Sumatra. As an example, the model for Palembang City—where the highest number of pneumonia cases (2,838) was recorded in 2022—is formulated as

$$\hat{\lambda} = exp(5{,}8696 - 0{,}0131\, {X_1}^* - 0{,}0078\, {X_2}^* + 0{,}0137\, {X_3}^* - 0{,}0380\, {X_4}^*) \tag{19}$$

All independent variables are statistically significant at the 5% level ($\alpha = 0.05$) in Palembang City, indicating their collective role in explaining local pneumonia incidence. Specifically, an increase of one community health center (X_1) is associated with a reduction in the expected number of pneumonia cases by a factor of $\exp(-0.0131) = 0.9870$, holding other variables constant. This underscores the vital role of accessible primary healthcare facilities in urban settings like Palembang, where high population density can strain resources but also amplify the impact of expanded services in preventing respiratory infections.

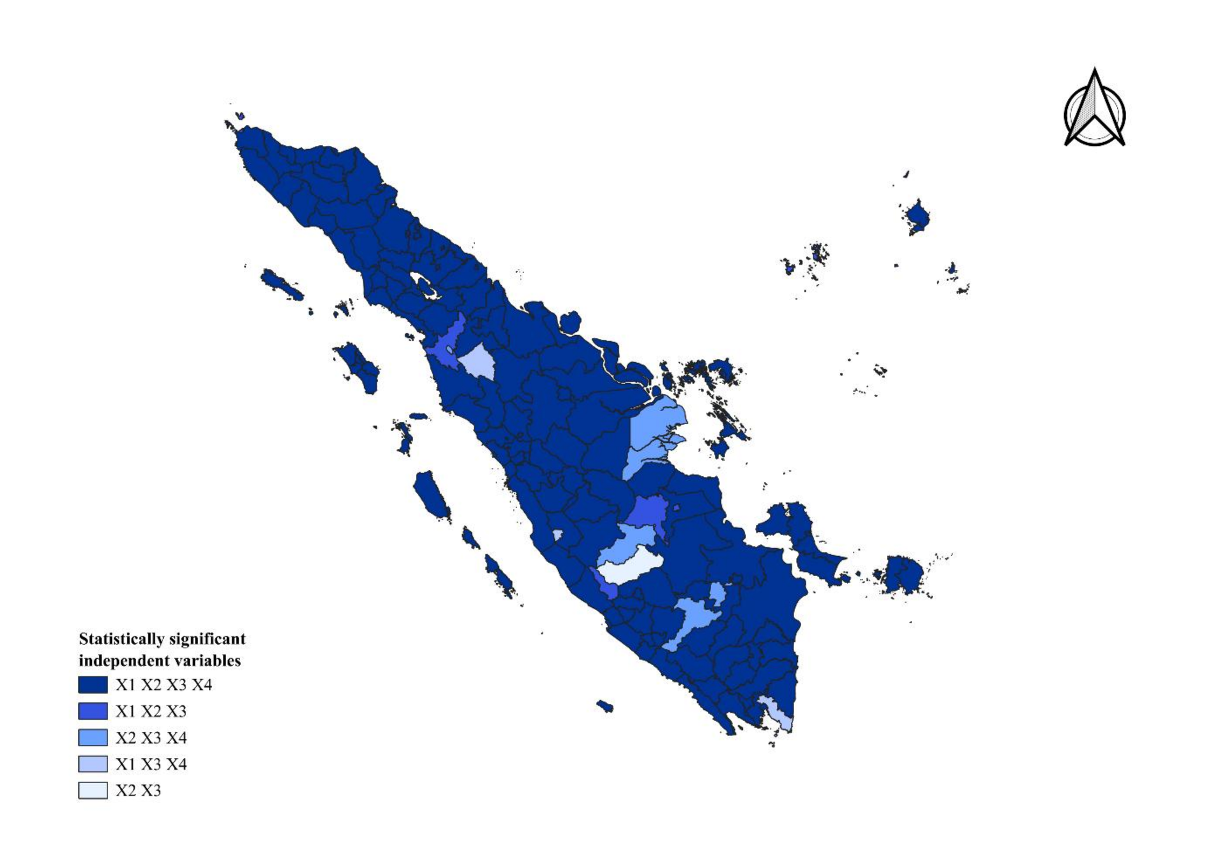

Figure 2. Spatial grouping of districts/cities based on significant variables in the fixed Gaussian kernel GWPR model.

Similarly, a 1% increase in complete basic immunization coverage (X_1) is linked to a decrease in expected pneumonia cases by a factor of $\exp(-0.0078) = 0.9922$, assuming other factors remain unchanged. Idealistiana (2025) found that providing immunization to infants and toddlers has been proven effective in reducing the number of cases of pneumonia [22]. This also aligns with evidence from recent studies in Indonesia, which demonstrate that full immunization schedules, including vaccines against pneumococcal and Haemophilus influenzae type b (Hib) pathogens, significantly lower pneumonia morbidity in toddlers [23].

Unexpectedly, a 1% increase in exclusive breastfeeding rates (X_3) is associated with a slight increase in expected pneumonia cases by a factor of $\exp(0.0137) = 1.0138$, holding other variables constant. While exclusive breastfeeding is generally protective against infectious diseases, including pneumonia, this positive association may arise from confounding factors such as environmental exposures in urban Palembang. For instance, despite high breastfeeding rates, toddlers may face heightened risks from poor air quality due to industrial pollution, traffic emissions, and seasonal haze from nearby peatland fires, which exacerbate respiratory vulnerabilities [24] [25]. Recent analyses in Indonesia highlight that non-exclusive or suboptimal breastfeeding practices, combined with urban environmental stressors, can diminish protective effects and even correlate with increased ARI incidence in polluted areas [26]. This finding emphasizes the need for integrated interventions that pair breastfeeding promotion with environmental improvements, such as stricter air quality regulations in densely populated cities like Palembang.

Conversely, a 1% increase in the low birth weight (LBW) rate (X_4) is associated with a decrease in expected pneumonia cases by a factor of $\exp(-0.038) = 0.9627$, assuming other variables are held constant. Although LBW is typically a risk factor for pneumonia due to immature immune systems and respiratory complications, this counterintuitive negative association may stem from heightened medical attention and early interventions for LBW infants in urban healthcare systems. In Palembang, LBW toddlers often receive intensive monitoring, nutritional support, and prompt treatment, which could mitigate infection risks more effectively than in general populations [27]. Evidence from Indonesian studies supports this, noting that targeted care for LBW infants at primary health centers can reduce acute respiratory infection (ARI) prevalence through enhanced surveillance and vaccination adherence [28]. However, this highlights a potential selection bias, where LBW cases are more likely to be documented and managed, underscoring the importance of preventive measures like prenatal nutrition programs to address LBW root causes.

Table 9. Grouping of districts/cities based on combinations of significant variables in the fixed Gaussian kernel GWPR model.

Group	Significant variables	Total	District/City
1	X_1, X_2, X_3, X_4	140	Simeulue, Aceh Singkil, Aceh Selatan, Aceh Tenggara, Aceh Timur, Aceh Besar, Aceh Barat, Aceh Tengah, Pidie, Bireuen, Aceh Utara, Aceh Barat Daya, Gayo Lues, Aceh Tamiang, Nagan Raya, Aceh Jaya, Bener Meriah, Pidie Jaya, Langsa, Lhokseumawe, Subulussalam, Nias, Mandailing Natal, Tapanuli Selatan, Tapanuli Utara, Tapanuli Tengah, Toba, Asahan, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Humbang Hasundutan, Pakpak Bharat, Samosir, Serdang Bedagai, Batu Bara, Labuhanbatu Selatan, Labuhanbatu Utara, Nias Utara, Nias Barat, Sibolga, Tanjung Balai, Pematang Siantar, Tebing Tinggi, Medan, Binjai, Padang Sidempuan, Kepulauan Mentawai, Pesisir Selatan, Regency Solok, Sijunjung, Padang Pariaman, Agam, Lima Puluh City, Pasaman, Solok Selatan, Dharmasraya, Pasaman Barat, City Padang, City Solok, Sawahlunto, Padang Panjang, Bukittinggi, Payakumbuh, Pariaman, Indragiri Hulu, Indragiri Hilir, Pelalawan, Siak, Kampar, Rokan Hulu, Bengkalis, Rokan Hilir, Kepulauan Meranti, Pekanbaru, Dumai, Kerinci, Merangin, Sarolangun, Batanghari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat, Tebo, Bungo, Jambi, Sungai Penuh, Ogan Komering Ulu, Ogan Komering Ilir, Muara Enim, Lahat, Musi Rawas, Musi Banyuasin, Banyuasin, Ogan Komering Ulu Selatan, Ogan Ilir, Empat Lawang, Penukal Abab Lematang Ilir, Musi Rawas Utara, Palembang, Prabumulih, Pagar Alam, Lubuk Linggau, Bengkulu Selatan, Bengkulu Utara, Kaur, Seluma, Muko Muko, Lebong, Kepahiang, Bengkulu, Lampung Barat, Tenggamus, Lampung Selatan, Lampung Timur, Lampung Tengah, Lampung Utara, Way Kanan, Tulang Bawang, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, Pesisir Barat, Bandar Lampung, Metro, Bangka, Bangka Barat, Bangka Tengah, Bangka Selatan, Belitung Timur, Pangkal Pinang, Karimun, Bintan, Natuna, Lingga, Kepulauan Anambas, Tanjung Pinang
2	X_1, X_2, X_3	7	Banda Aceh, Sabang, Labuhanbatu, Nias Selatan, Padang Lawas Utara, Gunungsitoli, Bengkulu Tengah
3	X_1, X_3, X_4	3	Padang Lawas, Tanah Datar, Belitung
4	X_2, X_3, X_4	3	Kuantan Singingi, Ogan Komering Ulu Timur, Batam
5	X_2, X_3	1	Rejang Lebong

4. Conclusion

In 2022, pneumonia cases in toddlers (aged 1 to less than 5 years) across Sumatra's 154 districts/cities averaged 220.20 cases per district/city, with significant regional disparities, notably high incidences in urban centers like Palembang (2,838 cases) and Batam (2,765 cases). The Geographically Weighted Poisson Regression (GWPR) model with a fixed Gaussian kernel proved optimal for capturing spatial variations, revealing that most areas (140 locations) were significantly influenced by all predictors—number of community health centers, complete basic immunization, exclusive breastfeeding, and low birth weight (LBW)—while 14 locations showed unique combinations, highlighting the need for localized interventions to address diverse regional drivers of pneumonia.

Stakeholders should prioritize tailored interventions, such as expanding health centers in urban high-incidence areas, implementing mobile immunization units in rural regions like Rejang Lebong, and addressing environmental risks (e.g., air pollution from peatland fires) to enhance breastfeeding benefits. Future research should explore negative binomial models to address potential overdispersion, incorporate temporal and microbial data, and include socioeconomic covariates to better understand regional disparities. The full set of 154 location-specific GWPR model results, including parameter estimates, is available upon request from the corresponding author. Requests should include clear reasons for access, such as policy development or academic research.

References

[1] Kementerian Kesehatan RI, *Profil Kesehatan Indonesia 2022*, Jakarta, 2023. Accessed: May 31, 2025. [Online]. Available: https://kemkes.go.id/app_asset/file_content_download/17029583366581150083 45c5.53299420.pdf

[2] L. Sari, "Mom's Knowledge and Practice on Prevention Pneumonia in Toddlers," *Indonesian Journal of Global Health* Research, vol. 4, no. 4, p. 769, 2022, doi: 10.37287/ijghr.v4i4.1254.

[3] United Nations, "Goal 3: Ensure healthy lives and promote well-being for all at all ages", Accessed: May 31, 2025. [Online]. Available: https://www.un.org/sustainabledevelopment/health/

[4] Widyawati, "Indonesia tegaskan komitmen pencegahan pneumonia di forum internasional," Jan. 30, 2020. Accessed: May 31, 2025. [Online]. Available: https://twitter.com/KemenkesRI

[5] V. Alvionita, S. Sulfatimah, A. Astuti, and N. Nurfitri, "Hubungan status gizi dan status imunisasi dengan kejadian pneumonia pada bayi," *Ahmar Metastasis Health Journal*, vol. 1, no. 4, pp. 137–143, Mar. 2022, doi: 10.53770/amhj.v1i4.92.

[6] N. L. P. D. Wardani, M. Rismawan, and P. A. R. Darmayanti, "Hubungan pemberian asi eksklusif dan perilaku merokok keluarga dengan kejadian pneumonia balita," *Jurnal Mutiara Kesehatan Masyarakat*" , vol. 7, no. 1, pp. 13–19, Jul. 2022, doi: 10.51544/jmkm.v7i1.2980.

[7] M. Erliandani, R. I. P. Priono, S. Ruqayyah, and A. F. Benvenuto, "Hubungan riwayat pemberian asi eksklusif, berat badan lahir rendah, dan kondisi fisik rumah dengan angka kejadian pneumonia pada balita," *Jambura Journal of Health Sciences and Research*, vol. 5, no. 2, pp. 746–754, Apr. 2023, doi: 10.35971/jjhsr.v5i2.19141.

[8] Dinas Kesehatan Provinsi Sumatera, "Profil kesehatan provinsi sumatera selatan tahun 2022," Sumatra Selatan, 2023. Accessed: May 31, 2025. [Online]. Available: www.dinkes.sumselprov.go.id.

[9] Dinas Kesehatan Provinsi Bengkulu, "Profil kesehatan provinsi bengkulu tahun 2022," Bengkulu, 2023. Accessed: May 31, 2025. [Online]. Available: http://ppid.bengkuluprov.go.id/storage/dokumen/62KW9RpJMDOTF2TUUNMEyZtIUjaS1WYACXlvrm03.pdf

[10] A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count Data*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139013567.

[11] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics, Sixth Edition* Pearson, 2013. Accessed: May 31, 2025. [Online]. Available: http://ndl.ethernet.edu.et/bitstream/123456789/27657/1/Barbara%20G.%20Tabachnick_2013.pdf

[12] S. Weisberg, *Applied Linear Regression*. Wiley, 2005. doi: 10.1002/0471704091.

[13] N. Shrestha, "Detecting multicollinearity in regression analysis," *Am J Appl Math Stat*, vol. 8, no. 2, pp. 39–42, Jun. 2020, doi: 10.12691/ajams-8-2-1.

[14] L. Fahrmeir, T. Kneib, S. Lang, and B. D. Marx, "Regression models," in *Regression*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2021, pp. 23–84. doi: 10.1007/978-3-662-63882-8_2.

[15] R. Haining, *Spatial Data Analysis*. Cambridge University Press, 2003. doi: 10.1017/CBO9780511754944.

[16] J. P. Lesage, *The Theory and Practice of Spatial Econometrics*. 1999. Accessed: May 31, 2025. [Online]. Available: https://www.researchgate.net/publication/266218273

[17] V. Loonis and M.-P. de Bellefon, *Handbook of Spatial Analysis: Theory and Application with R*. 2018. Accessed: May 31, 2025. [Online]. Available: www.insee.fr

[18] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," In B. N. Petrov, & F. Csaki, (Eds.), *Proceedings of the 2nd International Symposium on Information Theory*, pp. 199–213.1998, doi: 10.1007/978-1-4612-1694-0_15.

[19] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, Mar. 1978, doi: 10.1214/aos/1176344136.

[20] A. Fotheringham and P. Rogerson, *The Sage Handbook of Spatial Analysis*. 1 Oliver's Yard, 55 City Road, London England EC1Y 1SP United Kingdom : SAGE Publications, Ltd, 2009. doi: 10.4135/9780857020130.

[21] D. S. Wicaksono et al., “Modeling factors affecting educated unemployment on java island using geographically weighted poisson regression model,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 1, pp. 0615–0626, Mar. 2024, doi: 10.30598/barekengvol18iss1pp0615-0626.
[22] E. Sunarsi and L. Idealistiana, “Pengaruh pemberian imunisasi pcv terhadap kejadian pneumonia pada bayi balita di wilayah kerja upt puskesmas pulo ampel,” *MAHESA : Malahayati Health Student Journal*, vol. 5, no. 4, pp. 1562–1573, Apr. 2025, doi: 10.33024/mahesa.v5i4.17022.
[23] E. Sidabutar, Ansariadi, Wahiduddin, N. Bustan, Stang, and A. B. Birawida, “Analysis of risk factor for pneumonia in children less than five years in Makassar,” *J. Educ. Health Promot.*, vol. 13, no. 1, Jan. 2024, doi: 10.4103/jehp.jehp_727_23.
[24] M. Gedefaw and R. Berhe, “Determinates of childhood pneumonia and diarrhea with special emphasis to exclusive breastfeeding in north achefer district, northwest ethiopia: a case control study,” *Open J. Epidemiol.*, vol. 05, no. 02, pp. 107–112, 2015, doi: 10.4236/ojepi.2015.52014.
[25] S. Hossain and S. Mihrshahi, “Exclusive breastfeeding and childhood morbidity: a narrative review,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 22. MDPI, Nov. 01, 2022. doi: 10.3390/ijerph192214804.
[26] V. N. Sutriana, M. N. Sitaresmi, and A. Wahab, “Risk factors for childhood pneumonia: a case-control study in a high prevalence area in Indonesia,” *Clin. Exp. Pediatr.*, vol. 64, no. 11, pp. 588–595, 2021, doi: 10.3345/CEP.2020.00339.
[27] E. L. Haksari, “Historical perspectives: low birthweight and preterm infants in indonesia,” *Neoreviews*, vol. 20, no. 10, pp. e548–e560, Oct. 2019, doi: 10.1542/neo.20-10-e548.
[28] V. A. Puspamaniar, R. A. Setyoningrum, and D. Susanti, “Low birth weight as risk factor of pneumonia child in primary health care of surabaya,” 2019. [Online]. Available: https://e-journal.unair.ac.id/juxt