CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 1219-1232
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Generalized Poisson SAR for Under-Five Pneumonia in Tuban Regency

Joshua Capri Gunawan Sihombing*, Sutikno, and Achmad Choiruddin

Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Abstract

The number of pneumonia cases in children under five in Tuban Regency presents two significant data challenges, namely, overdispersion and spatial dependency. This study aims to develop and apply the Generalized Poisson Spatial Autoregressive (GPSAR) model to address both issues simultaneously. The model was estimated using a MLE-BHHH procedure and validated using 10-fold cross-validation (CV). The results confirm the model's validity and superiority. The GPSAR model outperformed the non-spatial GPR model in terms of goodness-of-fit (AIC: 1301.09 vs. 1312.67) and predictive accuracy (Out-of-Sample CV-RMSE: 8.451 vs. 8.716). Statistically, the structural parameters for spatial lag ($\hat{\rho} = 0.453$) and overdispersion ($\hat{\varphi} = 0.312$) were highly significant. Two predictor variables, exclusive breastfeeding (X_1) and complete basic immunization (X_2), were also found to be statistically significant factors. This research provides a robust regression framework for spatial count data exhibiting overdispersion and offers new insights into pneumonia case determinants in the region.

Keywords: Count data; overdispersion; pneumonia; spatial autoregression; spatial lag

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

1 Introduction

Pneumonia, an acute respiratory infection of the lungs, is one of the leading causes of mortality in children under five globally, where it is recorded to have resulted in approximately 502,000 child deaths annually in 2021 [1]. This is highly prevalent because the immune systems of children under five are not fully developed, particularly in developing countries where access to healthcare services is limited [2], [3], [4]. In Indonesia, pneumonia remains a significant public health problem with a high disease burden.

In modeling disease case data such as pneumonia, which constitutes count data, Poisson Regression is the most fundamental model [5], [6], [7], [8]. The Poisson model is designed based on the fundamental assumption of equidispersion, where the mean value is equal to the variance value [6], [7]. However, this assumption is frequently violated [9], [10]. This is evident in this study's case study, as the pneumonia cases in 328 villages/urban villages in Tuban Regency exhibit characteristics of extreme data overdispersion. Descriptive statistics show the mean number of cases is 4.527 with a variance of 60.286, yielding a Variance-to-Mean Ratio of 13.32 (substantially greater than 1). This condition indicates that the standard Poisson model is inadequate.

*Corresponding author. E-mail: joe.cgsihombing12@gmail.com

Submitted: October 3, 2025 Reviewed: November 10, 2025 Accepted: November 19, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.37288

To address the issue of overdispersion, Generalized Poisson Regression (GPR) emerges as a more flexible alternative to the equidispersion assumption, as it is capable of handling data characteristics exhibiting either overdispersion or underdispersion [11], [12]. When modeling count data exhibiting the severe overdispersion found in this study (Variance-to-Mean Ratio = 13.32), Negative Binomial (NB) regression is often considered a primary alternative to Poisson [8]. Nevertheless, this study selects the Generalized Poisson Regression (GPR) foundation due to its theoretical flexibility. Unlike NBR, which is specifically designed to handle overdispersion, the GPR model can inherently accommodate data with characteristics of both overdispersion ($\varphi > 0$) and underdispersion ($\varphi < 0$).

Furthermore, the empirical justification for not employing more complex *zero-inflated* models (e.g., ZIP or ZINB) is supported by diagnostic tests conducted in this research. The test indicated that the observed number of zeros (85) is not statistically significantly different (p-value = 0.42) from the expected number of zeros (78) under the baseline model. Therefore, GPR was chosen as a parsimonious yet flexible framework.

However, the GPR model, much like the Poisson model, assumes that the observation units in this case, the villages/urban villages in Tuban Regency are independent. This assumption is also often unrealistic for geographical data [13], [14], [15]. This phenomenon, known as spatial dependency, aligns with Tobler's First Law of Geography: "everything is related to everything else, but near things are more related than distant things" [14], [16], [17], [18], [19]. The empirical justification for this is also reinforced by the results of the Moran's I test on the pneumonia case data, which yielded a p<0.001, statistically indicating a highly significant positive spatial dependency pattern.

To address spatial dependency in data experiencing overdispersion, the literature provides several established methods, such as spatial Negative Binomial or spatial Poisson-lognormal models. These approaches typically address dependency by incorporating spatially structured random effects (often using Conditional Autoregressive (CAR) priors) [20], [21].

In addition to these methods, the spatial autoregressive (SAR) framework is frequently used [22], [23]. The SAR model, which incorporates a spatial lag component, has proven effective in capturing spillover effects and spatial interdependence in various fields [24], [25], [26]. Recent spatial modeling in Tuban Regency by Sutikno et al. (2025) [27] has successfully addressed spatial dependency using a Multivariate Spatial Autoregressive (MSAR) framework. However, their approach assumed a normal distribution, which may not fully capture the discrete nature and extreme overdispersion of pneumonia cases. To our knowledge, the integration of the Generalized Poisson structure with spatial lag dependency (the GPSAR model) is still relatively limited in the public health literature.

Driven by this methodological gap and the strong empirical findings from the data, this research aims to develop and apply the Generalized Poisson Spatial Autoregressive (GPSAR) model to determine the factors that significantly influence the incidence of pneumonia in children under five in Tuban Regency, East Java, Indonesia. The predictor variables used include five public health indicators: (1) the percentage of infants receiving exclusive breastfeeding, (2) the percentage of children under five who received complete basic immunization, (3) the percentage of children under five who received vitamin A, (4) the percentage of pregnant women attending prenatal classes, and (5) the percentage of households with access to clean water. Specifically, this research will: (1) derive the parameter estimation procedure using MLE-BHHH, (2) develop the hypothesis testing procedure using MLRT, and (3) apply the model to the empirical data on childhood pneumonia.

2 Methods

In this section, we outline the methodological framework used to develop and validate the GPSAR model. The components are structured sequentially, beginning with the data description, followed

by the model formulation, estimation procedure, and hypothesis testing. Each subsection builds upon the previous one to provide a coherent methodological foundation.

2.1 Data

This study utilizes secondary data from 2023, obtained from the Tuban Regency Government and the Center for Regional Potential Studies and Community Empowerment (PDPM) ITS. The observation units comprise 311 villages and 17 urban villages, totaling 328 observation units across Tuban Regency.

The spatial weight matrix ($\mathbf{W}$) was constructed using a digital map (SHP) of administrative boundaries. Specifically, a first-order queen contiguity matrix of size $n \times n$ (where $n = 328$) was defined, where two units are considered neighbors if they share any common boundary point (vertex or edge). This definition offers greater flexibility than rook contiguity, which relies solely on shared boundaries. Only immediate neighbors were assigned non-zero weights ($w_{ij} = 1$ for neighbors, $w_{ii} = 0$) [27].

Subsequently, the matrix was row-standardized to ensure row sums of unity, assigning each neighbor a weight of $1/m_i$ (where m_i is the neighbor count for unit i). This standardization facilitates the interpretation of the spatial parameter (ρ) as the average spatial influence. Crucially, no isolated units ("islands") were detected in the study area; consequently, every row in $\mathbf{W}$ contains at least one non-zero entry, thereby preventing singularity in the model estimation. The dimensions of the resulting matrix (328×328) are strictly consistent with the unit of analysis [28].

2.2 Generalized Poisson Regression Model

The GPSAR model is an extension of the Generalized Poisson Regression (GPR) model, which was developed by Famoye [12], and can be expressed as follows:

$$P(Y = y_i|x_i) = \left[\frac{\mu_i}{1+\varphi\mu_i}\right]^{y_i} \frac{(1+\varphi y_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i(1+\varphi y_i)}{1+\varphi\mu_i}\right), \quad y_i = 0, 1, 2, ... \tag{1}$$

where y_i is the response variable and φ is the dispersion parameter of y. When $\varphi = 0$, the GPR model is reduced to the standard Poisson regression model. Meanwhile, if $\varphi < 0$ or $\varphi > 0$, the GPR model exhibits underdispersion or overdispersion. To relate the expected value μ_i to the linear predictor $\eta_i = x_i^T\beta$, a link function for the GPR model is established, expressed as follows.

$$\mu_i = \mu(x_i) = q_i \exp(x_i^T\beta) \tag{2}$$

Where q_i is the exposure variable, x_i is the vector of predictor variables and β is the parameter vector of the GPR model for the response variable.

$$\boldsymbol{x}_{i_{(p+1)\times 1}} = \begin{bmatrix} 1 & x_{1i} & \cdots & x_{pi} \end{bmatrix}^T \quad \boldsymbol{\beta}_{(p+1)\times 1} = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_p \end{bmatrix}^T \tag{3}$$

$i = 1, 2, ..., n$.

2.3 Generalized Poisson Spatial Autoregressive (GPSAR) Model

The SAR model for continuous data differs from the model for count data. The SAR model for count data was previously developed by Lambert [29]. An essential component in modeling SAR for count data is the spatial autoregressive link function.

To ensure notational clarity, we first define the linear predictor as an $n \times 1$ vector, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$. The spatially lagged predictor is then $\mathbf{A}^{-1}\boldsymbol{\eta}$, where $\mathbf{A} = (\mathbf{I} - \rho\mathbf{W})$. The link function for the i-th observation, including the exposure component, is defined as:

$$\mu_i = q_i \exp\left((\mathbf{A}^{-1}\boldsymbol{\eta})_i\right) \tag{4}$$

which can be written in full as $\mu_i = q_i \exp\left(((\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta})_i\right)$. Here, q_i is the exposure variable, $\mathbf{A}^{-1}$ is the inverse of matrix $\mathbf{A}$, and $(\mathbf{A}^{-1}\boldsymbol{\eta})_i$ denotes the i-th element of the resulting $n \times 1$ vector. ρ is the spatial lag coefficient, $\mathbf{W}$ is the row-standardized Queen Contiguity weight matrix (as defined in Section 2.1), $\mathbf{X}$ is the $n \times (p+1)$ matrix of predictor variables, and $\boldsymbol{\beta}$ is the $(p+1) \times 1$ vector of regression coefficients. This notation clarifies that the i-th element of the spatially transformed linear predictor is used.

Subsequently, the GPSAR model, which is a combination of the GPR model in equation (1) and the SAR link function in equation (4), yields a model that can be expressed in the following form.

$$P(Y = y_i|x_i) = \left[\frac{\mu_i^{GPSAR}}{1+\varphi\mu_i^{GPSAR}}\right]^{y_i} \frac{(1+\varphi y_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i^{GPSAR}(1+\varphi y_i)}{1+\varphi\mu_i^{GPSAR}}\right), \quad y_i = 0, 1, 2, ... \tag{5}$$

with $\mu_i^{GPSAR} = q_i \exp\left((\mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta})_i\right)$ as in equation (4), where y_i is the response variable and φ is the dispersion parameter of y.

3 Results and Discussion

This section presents the empirical findings of the study. We begin by reporting the parameter estimation results of the GPSAR model, followed by hypothesis testing, model interpretation, and performance evaluation. Each subsection provides complementary insights to support the overall discussion.

3.1 Parameter Estimation of the GPSAR Model

The method used to estimate the parameters of the GPSAR model is the Maximum Likelihood Estimation (MLE) method by maximizing the likelihood function. The parameters to be estimated are $\boldsymbol{\beta}$, ρ, and φ. The estimation steps are as follows. It is important to note that the spatial autoregressive parameter ρ is constrained to a theoretical domain that ensures the matrix $\mathbf{A} = (I - \rho\mathbf{W})$ remains invertible. This domain is defined as $|\rho| < 1/\lambda_{\max}\mathbf{W}$, where $\lambda_{\max}\mathbf{W}$ is the spectral radius of the matrix $\mathbf{W}$. For a row-standardized $\mathbf{W}$ matrix (as used in this application, this bound simplifies to $|\rho| < 1$ [30]. During the optimization procedure (i.e., MLE-BHHH), this parameter domain was strictly enforced, either by constraining the search space or by performing an invertibility check on A at each numerical iteration. This step is necessary to ensure the resulting parameter estimate $\hat{\rho}$ is valid, stable, and mathematically sound.

First, take n random samples, (y_i, x_i) with $i = 1, 2, ..., n$. Then, the likelihood function for the GPSAR model is formed, which can be expressed as follows.

$$L(\bullet) = \prod_{i=1}^{n} P(Y = y_i|x_i) = \prod_{i=1}^{n} \left[\frac{\mu_i^{GPSAR}}{1+\varphi\mu_i^{GPSAR}}\right]^{y_i} \frac{(1+\varphi y_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i^{GPSAR}(1+\varphi y_i)}{1+\varphi\mu_i^{GPSAR}}\right) \tag{6}$$

We will maximize the likelihood function by differentiating it with respect to the parameter $\theta = \begin{bmatrix}\beta & \rho & \varphi\end{bmatrix}$. However, to simplify the process of maximizing the likelihood function, a logarithmic transformation is applied to the function, because the parameter value that maximizes the log-likelihood function is also the parameter value that maximizes the likelihood function itself. The form of the log-likelihood function can be expressed as follows.

$$\ln L(\bullet) = \sum_{i=1}^{n} \left[y_i \ln \frac{\mu_i^{GPSAR}}{1+\varphi\mu_i^{GPSAR}} + (y_i - 1)\ln(1+\varphi y_i) - \ln y_i! - \frac{\mu_i^{GPSAR}(1+\varphi y_i)}{1+\varphi\mu_i^{GPSAR}}\right] \tag{7}$$

Subsequently, equation (7) is partially differentiated with respect to the parameter vector $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta}^T & \rho & \varphi\end{bmatrix}^T$. The resulting gradient vectors are as follows.

$$\frac{\partial \ln L(\bullet)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left\{ \left[\frac{y_i - \mu_i^{GPSAR}}{(1+\varphi\mu_i^{GPSAR})^2} \right] \cdot \left((\mathbf{A}^{-1}\mathbf{X})_i \right)^T \right\} \tag{8}$$

$$\frac{\partial \ln L(\bullet)}{\partial \rho} = \sum_{i=1}^{n} \left\{ \left[\frac{y_i - \mu_i^{GPSAR}}{(1+\varphi\mu_i^{GPSAR})^2} \right] \cdot \left(\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta} \right)_i \right\} \tag{9}$$

$$\frac{\partial \ln L(\bullet)}{\partial \varphi} = \sum_{i=1}^{n} \left[-\frac{y_i \mu_i^{GPSAR}}{1+\varphi\mu_i^{GPSAR}} + \frac{(y_i-1)y_i}{1+\varphi y_i} - \frac{\mu_i^{GPSAR}(y_i - \mu_i^{GPSAR})}{(1+\varphi\mu_i^{GPSAR})^2} \right] \tag{10}$$

To ensure notational clarity in the equations above, the subscript i when applied outside parentheses denotes the i-th element of a vector or the i-th row of a matrix. Specifically, the first term, $(\mathbf{A}^{-1}\mathbf{X})_i$ (in equation 8), refers to the i-th row of the $n \times (p+1)$ matrix resulting from the multiplication $\mathbf{A}^{-1}\mathbf{X}$. Similarly, the second term, $(\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta})_i$ (in equation 9), refers to the i-th scalar element of the $n \times 1$ vector that results from the full matrix-vector multiplication.

Since the form above is not in closed form, the Berndt-Hall-Hall-Hausman (BHHH) iteration is subsequently used. This iteration is the preferred choice because its process does not require the second derivative when obtaining its Hessian matrix. The steps to perform the BHHH iteration are as follows.

1. Determine the initial values for all GPSAR model parameters, $\boldsymbol{\theta} = [\boldsymbol{\beta} \quad \rho \quad \varphi]^T$.
2. Form the gradient vector: $\mathbf{g}(\boldsymbol{\theta}^{(m)}) = [(\frac{\partial ln\ L(\bullet)}{\partial \beta})^T (\frac{\partial ln\ L(\bullet)}{\partial \rho})(\frac{\partial ln\ L(\bullet)}{\partial \varphi})]^T$
3. Form the Hessian matrix: $\mathbf{H}^*(\theta^{(m)}) = -\sum_{i=1}^{n} \mathbf{g}_i(\boldsymbol{\theta}^{(m)})\mathbf{g}_i(\boldsymbol{\theta}^{(m)})^T$ where $\mathbf{g}_i(\boldsymbol{\theta}^{(m)})$ is the gradient vector of the i-th observation.
4. Start the iteration at $m = 0$ using the following equation: $\hat{\boldsymbol{\theta}}^{(m+1)} = \hat{\boldsymbol{\theta}}^{(m)} - H^{*-1}(\hat{\boldsymbol{\theta}}^{(m)})g(\hat{\boldsymbol{\theta}}^{(m)})$ The iteration stops if $||\hat{\boldsymbol{\theta}}^{(m+1)} - \hat{\boldsymbol{\theta}}^{(m)}|| \leq \epsilon$ where ϵ is a very small positive number approaching 0.
5. Repeat the iteration from step 2 onwards with $m = m + 1$.

3.2 Hypothesis Testing of the GPSAR Model

Hypothesis testing for the GPSAR model parameters is conducted simultaneously and partially. The simultaneous hypothesis test uses the MLRT method, and the partial test uses the Wald Test method. To determine the joint influence of the predictor variables on the response variable, a simultaneous test is performed. The hypotheses for the simultaneous test in the GPSAR model are as follows. $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ and H_1: at least one $\beta_k \neq 0$, $k = 1, 2, ..., p$

Consider $\boldsymbol{\Omega}$ as the parameter set under the full model with $\boldsymbol{\Omega} = \{\boldsymbol{\beta}, \rho, \varphi\}$ and $\boldsymbol{\omega}$ as the parameter set under the null hypothesis ($H_0 : \beta_1 = \cdots = \beta_p = 0$), where $\boldsymbol{\omega} = \{\beta_{0\omega}, \rho_\omega, \varphi_\omega\}$.

The log-likelihood for the full model, $\ln L(\hat{\boldsymbol{\Omega}})$, has the same form as equation (7). The log-likelihood under the null hypothesis, $\ln L(\boldsymbol{\omega})$, is specified as follows:

$$\ln L(\boldsymbol{\omega}) = \sum_{i=1}^{n} \left[y_i \ln \frac{\mu_{i\omega}^{GPSAR}}{1+\varphi_\omega\mu_{i\omega}^{GPSAR}} + (y_i - 1)\ln(1+\varphi_\omega y_i) - \ln y_i! - \frac{\mu_{i\omega}^{GPSAR}(1+\varphi_\omega y_i)}{1+\varphi_\omega\mu_{i\omega}^{GPSAR}} \right] \tag{11}$$

To define the mean $\mu_{i\omega}^{GPSAR}$ under the null model, we first define the autoregressive matrix under the null as $\mathbf{A}_\omega = (\mathbf{I} - \rho_\omega\mathbf{W})$. The mean $\mu_{i\omega}^{GPSAR}$ is now correctly and consistently specified as:

$$\mu_{i\omega}^{GPSAR} = q_i \exp\left(\left(\mathbf{A}_\omega^{-1}\mathbf{1}\beta_{0\omega} \right)_i \right)$$

Here, $\mathbf{1}$ is an $n \times 1$ vector of ones. This definition ensures dimensional consistency and specification accuracy.

Parameter estimation under the null hypothesis is performed by differentiating $\ln L(\boldsymbol{\omega})$ with respect to $\beta_{0\omega}$, ρ_ω, and φ_ω. This yields the internally consistent gradient vectors:

$$\frac{\partial \ln L(\omega)}{\partial \beta_{0\omega}} = \sum_{i=1}^{n} \left\{ \left[\frac{y_i - \mu_{i\omega}^{GPSAR}}{(1+\varphi_\omega \mu_{i\omega}^{GPSAR})^2} \right] \cdot \left(\mathbf{A}_\omega^{-1} \mathbf{1} \right)_i \right\} \tag{12}$$

$$\frac{\partial \ln L(\omega)}{\partial \rho_\omega} = \sum_{i=1}^{n} \left\{ \left[\frac{y_i - \mu_{i\omega}^{GPSAR}}{(1+\varphi_\omega \mu_{i\omega}^{GPSAR})^2} \right] \cdot \left(\mathbf{A}_\omega^{-1} \mathbf{W} \mathbf{A}_\omega^{-1} \mathbf{1} \beta_{0\omega} \right)_i \right\} \tag{13}$$

$$\frac{\partial \ln L(\omega)}{\partial \varphi_\omega} = \sum_{i=1}^{n} \left[-\frac{y_i \mu_{i\omega}^{GPSAR}}{1+\varphi_\omega \mu_{i\omega}^{GPSAR}} + \frac{(y_i - 1) y_i}{1+\varphi_\omega y_i} - \frac{\mu_{i\omega}^{GPSAR}(y_i - \mu_{i\omega}^{GPSAR})}{(1+\varphi_\omega \mu_{i\omega}^{GPSAR})^2} \right] \tag{14}$$

Since the form above is not in closed form, the BHHH iteration is subsequently performed. Following this, $\hat{\Omega} = \{\hat{\beta}, \hat{\rho}, \hat{\varphi}\}$ and $\hat{\omega} = \{\hat{\beta}_{\partial\omega}, \hat{\rho}_\omega, \hat{\varphi}_\omega\}$ are obtained, which are then substituted into the following test statistic:

$$G_{GPSAR}^2 = 2[ln\ L(\hat{\Omega}) - ln\ L(\hat{\omega})] \tag{15}$$

In $L(\hat{\Omega})$ is obtained by substituting the parameter estimation results under the population into equation (7), and In $L(\hat{\omega})$ is obtained by substituting the parameter estimation results under the null hypothesis into equation (11). The critical region for the hypothesis test is as follows. $\alpha = P(G_{GPSAR}^2 > \chi_{(\alpha,df)}^2)$ G_{GPSAR}^2 will follow a chi-square distribution as $n \to \infty$ such that the rejection region for H_0 is $G_{GPSAR}^2 > \chi_{(\alpha,df)}^2$ with $df = n(\Omega) - n(\omega) = [(P+1)+2] - [1+2] = p$

After the simultaneous parameter hypothesis testing is performed and a decision to reject H_0 is obtained, partial parameter hypothesis testing is then conducted. Partial hypothesis testing for the parameter ρ is performed first, with the following hypotheses. $H_0 : \rho = 0$ $H_1 : \rho \neq 0$ The test statistic used for the above hypothesis test, using the Wald test, is

$$W_\rho = (\frac{\hat{\rho}}{se(\hat{\rho})})^2 \sim \chi_{(\alpha,1)}^2 \tag{16}$$

The decision criterion is to reject H_0 if $W_\rho > \chi_{(\alpha,1)}^2$. The hypotheses used for the partial parameter test β_k are as follows. $H_0 : \beta_k = 0$ $H_1 : \beta_k \neq 0$ Subsequently, the partial significance test for the parameter β_k is conducted using the following test statistic.

$$W_{\beta_k} = (\frac{\hat{\beta}_k}{se(\hat{\beta}_k)})^2 \sim \chi_{(\alpha,1)}^2 \tag{17}$$

The decision criterion is to reject H_0 if $W_{\beta_k} > \chi_{(\alpha,1)}^2$. Subsequently, the partial significance test for the parameter φ has the following hypotheses $H_0 : \varphi = 0$ $H_1 : \varphi \neq 0$ Subsequently, the partial significance test for the parameter φ is conducted using the following test statistic.

$$W_\varphi = (\frac{\hat{\varphi}}{se(\hat{\varphi})})^2 \sim \chi_{(\alpha,1)}^2 \tag{18}$$

The decision criterion is to reject H_0 if $W_\varphi > \chi_{(\alpha,1)}^2$. Subsequently, the standard error for each parameter can be obtained from the square root of the main diagonal elements of the variance-covariance matrix, which is generally defined as $Cov(\hat{\boldsymbol{\theta}}) = -(\mathbf{H}^{*-1}(\hat{\boldsymbol{\theta}}))$.

3.3 Application

3.3.1 Descriptive Statistics

The regression model proposed in this study involves five predictor variables relevant to public health indicators: (X_1) Percentage of infants receiving exclusive breastfeeding, (X_2) Percentage of children under five who received complete basic immunization, (X_3) Percentage of children

under five who received vitamin A, (X_4) Percentage of pregnant women attending prenatal classes, and (X_5) Percentage of households with access to clean water.

In accordance with the spatial analysis framework used, the unit of analysis in this study is the village/urban village level, encompassing all 328 valid observation units in Tuban Regency. Demographic characteristics (such as the population of children under five, which is used as an exposure variable q_i) vary substantially across each observation unit.

Table 1: Descriptive statistics of the response variable (Number of Pneumonia cases)

Variable	Min	Max	Mean	Std. Deviation	Variance
Pneumonia Cases	0	115	4.527	7.776	60.286

Descriptive statistical analysis was conducted on the dependent variable (Y = Number of Pneumonia cases) based on the 328 observation units (Table 1). A preliminary review revealed extreme data variability, with a range from 0 (case-free areas) to a maximum of 115 cases. The data exhibits a mean of 4.527 but a variance of 60.286. The resulting Variance-to-Mean Ratio (VMR) of 13.32 indicates strong *overdispersion*, confirming that a standard Poisson model is inappropriate.

As a further diagnostic step to validate the model selection, a formal test for *zero-inflation* was conducted . This test determines whether a more complex zero-inflated (ZI) model is necessary. The diagnostic test (conducted using the `DHARMa` package in R) compares the observed number of zeros in the data ($n_{obs} = 85$) against the expected number of zeros ($n_{pred} = 78$) simulated from the baseline model. The result of this test yielded a p-value = 0.42. As this p-value is substantially greater than the $\alpha = 0.05$ significance level, the null hypothesis (that the data is not zero-inflated) cannot be rejected. This finding statistically justifies the decision to use the Generalized Poisson (GP) distribution for this analysis, as a more complex zero-inflated structure is not required.

Subsequently, the distribution pattern of pneumonia cases in children under five can be visualized as shown in Figure 1 below.

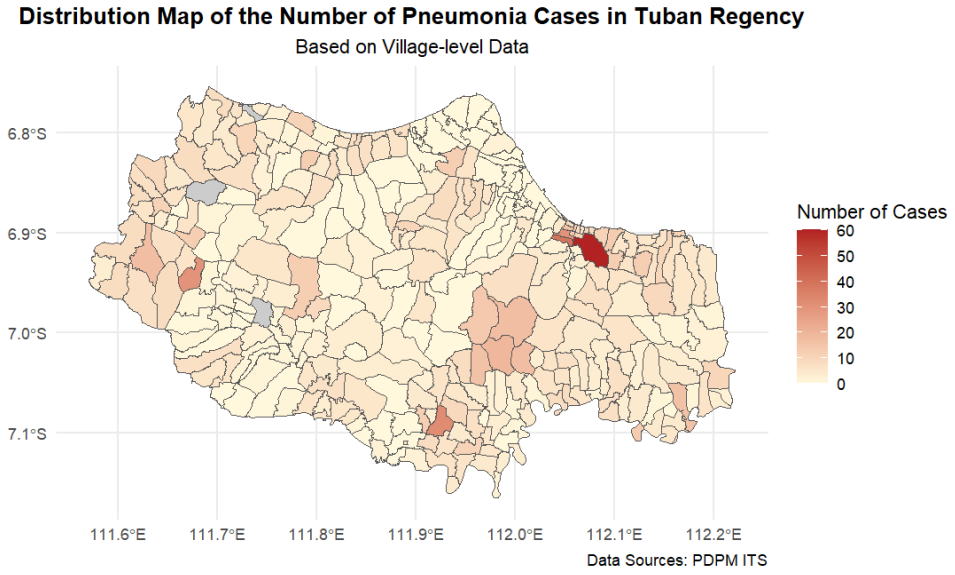

Figure 1: Distribution Map of the Number of Pneumonia Cases
Data Sources: PDPM ITS

This choropleth map visualizes the spatial distribution of the number of Pneumonia cases across 328 observation units (villages/urban villages) in Tuban Regency, which clearly illustrates two crucial data characteristics. First, extreme heterogeneity is observed, wherein the majority

of areas exhibit a very low number of cases (0-10 cases, marked in light yellow), yet several specific villages stand out as 'hotspots' with an extremely high number of cases (dark red). This visual finding has been statistically confirmed as extreme overdispersion, where the data show a mean value (4.527) that is substantially smaller than its variance (60.286), resulting in a Variance-to-Mean Ratio of 13.32. Second, the map also indicates a pattern of spatial clustering, whereby villages with high cases tend to be geographically proximate, rather than randomly dispersed. This visual pattern has also been statistically validated via the Moran's I test, which yielded a $p < 0.001$, indicating the presence of very strong and significant positive spatial dependency. Consequently, the visual findings from this map fully support the methodological justification for utilizing a model that can simultaneously address overdispersion (GPR) and spatial dependency (SAR).

3.3.2 Application of the Generalized Poisson Spatial Autoregressive (GPSAR) Model to Pneumonia Cases in Children Under Five in Tuban Regency

Following the data exploration, the GPSAR model was subsequently applied to the data. The parameter estimation results for all parameters are summarized in Table 2. The table includes parameter estimates, standard errors, p-values, and the 95% Confidence Intervals (CIs) for all parameters.

Table 2: Parameter Estimation Results of the GPSAR Model

Parameter	Estimate	Std. Error	p-value	95% LCL	95% UCL
β_0 (Intercept)	0.01703	0.39475	0.9656	-0.75666	0.79072
$\beta_1(X_1)$	-0.00145	0.00071	0.0422 *	-0.00285	-0.00005
$\beta_2(X_2)$	0.00929	0.00454	0.0409 *	0.00038	0.01820
$\beta_3(X_3)$	0.00040	0.00035	0.2577	-0.00029	0.00109
$\beta_4(X_4)$	0.00159	0.00082	0.0531 .	-0.00002	0.00320
$\beta_5(X_5)$	-0.00257	0.00400	0.5213	-0.01041	0.00528
ρ (Spatial Lag)	0.45229	0.04206	10^{-16} ***	0.36985	0.53473
φ (Dispersion)	0.31225	0.03216	10^{-16} ***	0.24921	0.37529

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1*

The estimation results (Table 2) indicate that both structural parameters of the model ($\hat{\varphi}$ and $\hat{\rho}$) are highly significant ($p < 0.001$), which statistically confirms the validity of adopting the GPSAR model structure for this dataset.

In addition to the structural parameters, the estimation results for the predictor parameters β_k were obtained. The final fitted model can be expressed as follows:

$$\hat{\eta}_i = 0.01703 - 0.00145X_{1i} + 0.00929X_{2i} + 0.00040X_{3i} + 0.00159X_{4i} - 0.00257X_{5i} \tag{19}$$

$$\hat{\mu}_i^{GPSAR} = q_i \exp((I - 0.45229W)^{-1}\hat{\eta}_i) \tag{20}$$

where q_i (the exposure variable) is the number of children under five and $\mathbf{W}$ is the spatial weight matrix.

An in-depth analysis of the statistically significant predictor coefficients (using $\alpha = 0.05$) identified two key factors associated with the pneumonia case rate. Variable X_1 (Percentage of exclusive breastfeeding) demonstrated a significant protective effect ($\hat{\beta}_1 = -0.00145$, p=0.042). The Rate Ratio (RR) is calculated as $\exp(-0.00145) \approx 0.9985$, implying that a one percentage point increase in exclusive breastfeeding coverage is associated with a 0.15% decrease in the pneumonia case rate, ceteris paribus. While Sutikno et al. (2025) [27] also identified exclusive breastfeeding as a significant determinant using an MSAR model, their study reported a counter-intuitive positive association. The GPSAR model employed in this study successfully captures the theoretically expected protective effect (negative association), demonstrating the

advantage of the Generalized Poisson structure in handling overdispersed count data. Despite this difference in direction, both frameworks consistently highlight the critical role of breastfeeding and immunization (X_2) in the region. This finding is supported by its 95% CI for the RR [exp(−0.00285), exp(−0.00005)], or [0.9972, 0.9999], which is entirely below 1.0.

Conversely, variable X_2 (Percentage of complete basic immunization) was also found to be significant ($\hat{\beta}_2 = 0.00929$, p=0.041), but with a positive association. This coefficient yields an RR of exp(0.00929) ≈ 1.0093, indicating that each one percentage point increase in immunization coverage is associated with a 0.93% increase in the pneumonia case rate. The 95% CI for the RR [1.0004, 1.0184] confirms this finding, as the interval is entirely above 1.0. Furthermore, variable X_4 (Percentage of pregnant women attending prenatal classes) was found to be approaching significance (p = 0.053). Its 95% CI of [-0.00002, 0.00320] demonstrates that the interval only barely crosses zero, suggesting a potential positive association that warrants further investigation.

The dispersion parameter $\hat{\varphi}$ was estimated to be 0.31225 (95% CI [0.24921, 0.37529]) with a highly significant p-value ($p < 10^{-16}$)). In the context of the Generalized Poisson model, this positive and statistically significant $\hat{\varphi}$ value provides quantitative evidence of the phenomenon of *overdispersion*. This indicates that the variability in the number of Pneumonia cases across regions substantially exceeds its mean value. This finding confirms that the use of a standard Poisson model (which assumes variance equals the mean) would be inadequate, and the selection of the Generalized Poisson model was appropriate for addressing this excess variability.

The spatial lag parameter $\hat{\rho}$ was estimated to be 0.45229 (95% CI [0.36985, 0.53473]), and was also highly significant ($p < 10^{-16}$). This positive and exceptionally strong ρ value confirms the existence of dominant positive spatial autocorrelation. In practical terms, this means that the number of Pneumonia cases in one village/area is not independent, but is instead significantly influenced by the number of cases in its neighboring villages/areas. There is a strong indication of a clustering effect, whereby regions with high cases tend to be in proximity to other high-case regions. This finding validates the use of the Spatial Autoregressive (SAR) component in the model. Collectively, the significance of both structural parameters demonstrates that Pneumonia cases in Tuban Regency exhibit characteristics of both overdispersion and spatial dependency, thus making the use of the GPSAR model an appropriate choice.

3.3.3 Model Goodness of Fit and Validation

To evaluate the model's performance, a comparison was made against the non-spatial GPR model. This evaluation was based on both in-sample goodness-of-fit and, more importantly, on out-of-sample predictive accuracy. A 10-fold Cross-Validation (CV) analysis was implemented to provide a robust estimate of out-of-sample performance.

We compared both the in-sample goodness-of-fit (AIC, BIC) and the out-of-sample predictive performance of the proposed GPSAR model against the non-spatial GPR model. The results, presented in Table 3, show the performance metrics for both in-sample (trained on $n = 328$) and out-of-sample (the average CV $K = 10$) evaluation.

Table 3: Model Goodness of Fit Comparison (In-Sample vs. Out-of-Sample)

	In-Sample ($n = 328$)				Out-of-Sample ($K = 10$ CV)	
Model	AIC	BIC	MSE	RMSE	MSE (CV)	RMSE (CV)
GPR (Non-Spasial)	1312.67	1339.22	69.45	8.334	75.97	8.716
GPSAR (Spasial)	1301.09	1331.43	61.19	7.823	71.42	8.451

The results (Table 3) confirm the superiority of the spatial model across all metrics. In terms of in-sample goodness-of-fit, the GPSAR model shows a substantially better fit, with lower AIC (1301.09 vs. 1312.67) and BIC (1331.43 vs. 1339.22). This indicates that the inclusion of the

spatial lag parameter (ρ) is highly justified, even after penalizing for the additional parameter ($k = 8$ for GPSAR vs. $k = 7$ for GPR).

This finding is reinforced by the predictive accuracy metrics. The GPSAR model also outperformed the non-spatial GPR model in predictive accuracy on unseen data (Out-of-Sample CV-RMSE 8.451 vs. 8.716). This robustly supports our claim that incorporating the spatial lag component provides a significant and necessary improvement for modeling pneumonia cases in Tuban.

3.3.4 Diagnostic Checks and Sensitivity Analysis

As a final diagnostic check, a spatial autocorrelation test (Moran's I) was performed on the Pearson residuals of the fitted GPSAR model (from Table 2). The test result (Moran's I = 0.214, p-value < 0.001) indicates that some significant spatial autocorrelation remains in the residuals. This suggests that while the GPSAR model is a significant improvement over the GPR model, the spatial dependency structure in the data may be more complex than captured by the first-order SAR component. This is noted as a limitation and an avenue for future research.

Furthermore, to test the robustness of the model findings, a sensitivity analysis was conducted on the spatial weights matrix $\mathbf{W}$. The primary model (Table 2) was estimated using *Queen contiguity*. This analysis re-estimated the full model using a *Rook contiguity* matrix. The results were highly consistent: the Rook model produced similarly dominant structural parameters ($\hat{\rho} = 0.494$; $\hat{\varphi} = 0.352$, both $p < 0.001$) and identical inferential conclusions (i.e., no β predictors were significant at $\alpha = 0.05$). This consistency demonstrates that the study's main findings are robust and not sensitive to the specific choice of contiguity definition.

4 Conclusion

This study developed and applied a Generalized Poisson Spatial Autoregressive (GPSAR) model to analyze pneumonia cases in Tuban Regency, based on 328 village/urban village observations. The analysis confirmed the statistical validity of the model, finding that both overdispersion ($\hat{\varphi} = 0.312$) and spatial autocorrelation ($\hat{\rho} = 0.453$) are highly significant components of the data.

Furthermore, the GPSAR model proved to be superior in both model fit and predictive accuracy compared to its non-spatial (GPR) counterpart. In terms of in-sample goodness-of-fit, the GPSAR model yielded lower AIC (1301.09 vs. 1312.67) and BIC (1331.43 vs. 1339.22), confirming its superiority even after penalizing for the additional spatial parameter. This finding was reinforced by predictive accuracy metrics, where the model also showed a lower in-sample RMSE (7.823 vs. 8.334) and, more importantly, a lower 10-fold cross-validation (Out-of-Sample CV-RMSE: 8.451 vs. 8.716). Substantively, the GPSAR model identified two predictor variables as statistically significant ($\alpha = 0.05$): Percentage of exclusive breastfeeding (X_1) as a protective factor, and Percentage of complete basic immunization (X_2) as a risk factor.

It is important to discuss the *counter-intuitive* finding regarding X_2 (complete basic immunization), which showed a positive association with pneumonia cases. This finding should not be interpreted causally, i.e., that immunization causes pneumonia. A more plausible explanation is the phenomenon of *ecological fallacy* or *omitted variable bias*, common in aggregate-level village data. Villages/urban areas with a high percentage of immunization (X_2) are likely those with better and more active health infrastructure (such as Posyandu or Puskesmas). This better infrastructure also leads to a more sensitive case detection and reporting system. Consequently, X_2 may be acting as a *proxy* for "health reporting quality," where areas with better reporting statistically record a higher number of cases.

This research makes a methodological contribution by applying the GPSAR framework, whose application in public health, to the researchers' knowledge, is still developing. It also provides strong policy implications by highlighting the dominant role of spatial clustering, suggesting that

interventions should be targeted at regional hotspots, not just at individual villages.

Several limitations are acknowledged. First, the diagnostic Moran's I test on the model residuals ($p < 0.001$) indicated that significant spatial autocorrelation remains, suggesting that future research could explore more complex spatial error structures. Second, the study is cross-sectional, relying on data from a single year. Future work, as suggested in the original manuscript, could explore spatio-temporal dynamics or point-process models to capture disease spread more accurately.

Although this approach has provided an overview of the risk distribution in Tuban Regency, future studies are recommended to use point-based data (geocoded home addresses). As demonstrated in the COVID-19 analysis by Choiruddin et al. (2023) [31], the use of point data permits analysis at a finer spatial resolution and avoids the assumption of homogeneous risk within a single administrative area. Thus, modeling using point process methods can reveal dispersion patterns and risk factors at a micro-scale with greater accuracy.

In addition to improvements in data resolution, the methodological development of this cross-sectional and univariate GPSAR model is also wide open. First, the model can be extended to the spatiotemporal domain to analyze the dynamics of disease spread from year to year. Rather than merely applying traditional spatial panel data models to areal (aggregate) data, future research could integrate the point process approach (previously suggested for geocoded data) with deep learning methods. As demonstrated by Choiruddin et al. (2025) [32], the application of probabilistic deep learning to spatio-temporal and highly multivariate Log-Gaussian Cox Process models can offer a substantially more flexible framework. This approach can potentially handle complex data patterns with greater computational efficiency compared to traditional computational statistics methods. Furthermore, for a more holistic understanding, a multivariate approach can be considered to model Pneumonia simultaneously with other relevant childhood illnesses, such as Stunting or Diarrhea, in order to identify interrelated health determinants. For a deeper causal understanding, this model can be integrated into a Spatial Structural Equation Modeling (Spatial SEM) framework, which allows for the modeling of complex causal pathways and latent variables.

CRediT Authorship Contribution Statement

Joshua Capri Gunawan Sihombing: Conceptualization, Methodology, Software, Formal Analysis, Visualization, Writing-Original Draft Preparation. **Sutikno:** Conceptualization, Data Curation, Resources, Methodology, Validation, Supervision. **Achmad Choiruddin:** Conceptualization, Methodology, Supervision, Validation.

Declaration of Generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized generative AI and AI-assisted technologies. Specifically, Google Gemini was employed for writing assistance, including improving grammar, refining paragraph structure, and ensuring linguistic clarity.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding. The authors would like to express their sincere gratitude to their families for their support and encouragement throughout this research.

Data and Code Availability

The data and code supporting the findings of this study are not publicly available due to institutional regulations. However, the data and code can be accessed from the corresponding author upon request and with the approval of the relevant institution.

References

[1] R. G. Bender et al., "Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990–2021: A systematic analysis from the Global Burden of Disease Study 2021," *The Lancet Infectious Diseases*, vol. 24, no. 9, pp. 974–1002, 2024. DOI: 10.1016/S1473-3099(24)00176-2.

[2] D. Manik, W. P. J. Kaunang, and E. M. Mantjoro, "Distribusi kasus dan kematian akibat pneumonia pada balita di Indonesia tahun 2019-2023 [distribution of cases and deaths due to pneumonia in children under five in Indonesia 2019-2023]," *Profetif: Jurnal Kesehatan Masyarakat*, vol. 9, no. 2, pp. 2972–2986, 2025. DOI: 10.31004/prepotif.v9i2.44589.

[3] R. Aprilia, F. Faisal, I. Irwandi, S. Suharni, and E. Efriza, "Tinjauan literatur: Faktor risiko dan epidemiologi pneumonia pada balita [literature review: Risk factors and epidemiology of pneumonia in toddlers]," *Scientific Journal*, vol. 3, no. 3, pp. 166–173, 2024. DOI: 10.56260/sciena.v3i3.144.

[4] V. Alvionita, S. Sulfatimah, A. Astuti, and N. Nurfitri, "Hubungan status gizi dan status imunisasi dengan kejadian pneumonia pada bayi [relationship between nutritional status and immunization status with the incidence of pneumonia in infants]," *Ahmar Metastasis Health Journal*, vol. 1, no. 4, pp. 137–143, 2022. DOI: 10.53770/amhj.v1i4.92.

[5] J. Liu, "Sufficient dimension reduction for Poisson regression," *Econometrics and Statistics*, vol. 34, pp. 109–119, 2025. DOI: 10.1016/j.ecosta.2022.09.001.

[6] S. Subedi and U. J. Dang, "Multivariate Poisson lognormal distribution for modeling counts from modern biological data: An overview," *Computational and Structural Biotechnology Journal*, vol. 24, pp. 1–13, 2025. DOI: 10.1016/j.csbj.2025.03.017.

[7] S. Mardalena, P. Purhadi, J. D. T. Purnomo, and D. D. Prastyo, "Parameter estimation and hypothesis testing of multivariate Poisson inverse Gaussian regression," *Symmetry*, vol. 12, no. 10, p. 1738, 2020. DOI: 10.3390/sym12101738.

[8] S. Q. Aina, M. Fauziyah, and W. P. Nurmayanti, "Comparison of generalized Poisson regression and negative binomial regression models based on Akaike information criterion values," *Statistika*, vol. 25, no. 1, pp. 86–101, 2025. DOI: 10.29313/statistika.v25i1.5402.

[9] P. Puig, J. Valero, and A. Fernández-Fontelo, "Some mechanisms leading to underdispersion: Old and new proposals," *Scandinavian Journal of Statistics*, vol. 51, no. 1, pp. 245–267, 2024. DOI: 10.1111/sjos.12677.

[10] L. P. Fávero, R. d. F. Souza, P. Belfiore, H. L. Corrêa, and M. F. Haddad, "Count data regression analysis: Concepts, overdispersion detection, zero-inflation identification, and applications with R," *Practical Assessment, Research & Evaluation*, vol. 26, p. 13, 2021. DOI: 10.7275/44nn-cj68.

[11] S. M. Berliana, P. Purhadi, S. Sutikno, and S. P. Rahayu, "Parameter estimation and hypothesis testing of geographically weighted multivariate generalized Poisson regression," *Mathematics*, vol. 8, no. 9, p. 1523, 2020. DOI: 10.3390/math8091523.

[12] S. W. Tyas, L. A. Puspitasari, et al., "Geographically weighted generalized Poisson regression model with the best kernel function in the case of the number of postpartum maternal mortality in East Java," *MethodsX*, vol. 10, p. 102 002, 2023. DOI: 10.1016/j.mex.2023.102002.

[13] S. Manda, N. Haushona, and R. Bergquist, "A scoping review of spatial analysis approaches using health survey data in sub-Saharan Africa," *International Journal of Environmental Research and Public Health*, vol. 17, no. 9, p. 3070, 2020. DOI: 10.3390/ijerph17093070.

[14] P. Phang, S. Aslam, J. Labadin, and V. J. Jayaraj, "Spatial autocorrelation analysis of infectious disease incidence rates at state and district level using supra-adjacency weights matrix," *Health*, vol. 13, no. 2, pp. 456–470, 2025. DOI: 10.13189/ujph.2025.130217.

[15] T. Saffary, O. A. Adegboye, E. Gayawan, F. Elfaki, M. A. Kuddus, and R. Saffary, "Analysis of COVID-19 cases' spatial dependence in US counties reveals health inequalities," *Frontiers in Public Health*, vol. 8, p. 579 190, 2020. DOI: 10.3389/fpubh.2020.579190.

[16] K. Vilinová and L. Petrikovičová, "Spatial autocorrelation of COVID-19 in Slovakia," *Tropical Medicine and Infectious Disease*, vol. 8, no. 6, p. 298, 2023. DOI: 10.3390/tropicalmed8060298.

[17] A. Getis, "Reflections on spatial autocorrelation," *Regional Science and Urban Economics*, vol. 37, no. 4, pp. 491–496, 2007. DOI: 10.1016/j.regsciurbeco.2007.04.005.

[18] F. Famoye, J. T. Wulu, K. P. Singh, et al., "On the generalized Poisson regression model with an application to accident data," *Journal of Data Science*, vol. 2, no. 3, pp. 287–295, 2004. DOI: 10.6339/JDS.2004.02(3).167.

[19] J. Cheng, Y. Cui, X. Wang, Y. Wang, and R. Feng, "Spatial characteristics of health outcomes and geographical detection of its influencing factors in Beijing," *Frontiers in Public Health*, vol. 12, p. 1 424 801, 2024. DOI: 10.3389/fpubh.2024.1424801.

[20] Y. Wang and K. M. Kockelman, "A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods," *Accident Analysis & Prevention*, vol. 60, pp. 71–84, 2013. DOI: 10.1016/j.aap.2013.07.030.

[21] A. Rosilala and K. M. Kockelman, "Comparison of negative binomial SAR, SEM, and SARMA methods in modeling the number of malnutrition cases among toddlers in Central Java," *Jurnal Gaussian*, vol. 14, no. 1, pp. 42–53, 2025. DOI: 10.14710/j.gauss.14.1.42-53.

[22] N. A. Cruz, D. A. Romero, and O. O. Melo, "SAR models with specific spatial coefficients and heteroskedastic innovations," *arXiv preprint arXiv:2502.15580*, 2025. DOI: 10.48550/arXiv.2502.15580.

[23] J. Zhao and Y. Pu, "Characterization and estimation of heterogeneous spatial autocorrelation in spatial autoregressive models," *PLoS ONE*, vol. 20, no. 7, e0327316, 2025. DOI: 10.1371/journal.pone.0327316.

[24] P. Pfeiffer, J. Varga, et al., "Quantifying spillovers of coordinated investment stimulus in the EU," *Macroeconomic Dynamics*, vol. 27, no. 7, pp. 1843–1865, 2023. DOI: 10.1017/S1365100522000487.

[25] M. Jankiewicz, "Analysis of spatial dependencies and spatial effects in the relationship between economic growth and unemployment in Europe," *Hungarian Geographical Bulletin*, vol. 74, no. 2, pp. 163–176, 2025. DOI: 10.15201/hungeobull.74.2.3.

[26] K. Wang et al., "Quantifying the spatial spillover effects of non-pharmaceutical interventions on pandemic risk," *International Journal of Health Geographics*, vol. 22, no. 1, p. 13, 2023. DOI: 10.1186/s12942-023-00335-6.

[27] S. Sutikno, P. Purhadi, F. Fachrunisah, and F. D. Cahyoko, "Estimation of parameters and hypothesis testing of multivariate spatial autoregressive model," *MethodsX*, vol. 14, p. 103 294, 2025. DOI: 10.1016/j.mex.2025.103294.

[28] L. P. S. Pratiwi, N. P. N. Hendayanti, and I. K. P. Suniantara, "Perbandingan pembobotan seemingly unrelated regression – spatial durbin model untuk faktor kemiskinan dan pengangguran [comparison of seemingly unrelated regression – spatial durbin model weighting for poverty and unemployment factors]," *Jurnal Varian*, vol. 3, no. 2, pp. 51–64, 2020. DOI: 10.30812/varian.v3i2.596.

[29] D. M. Lambert, J. P. Brown, and R. J. G. M. Florax, "A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application," *Regional Science and Urban Economics*, vol. 40, no. 4, pp. 241–252, 2010. DOI: 10.1016/j.regsciurbeco.2010.04.001.

[30] StataCorp, *Stata spatial autoregressive models reference manual*, College Station, TX: Stata Press, 2017.

[31] A. Choiruddin, F. F. Hannanu, J. Mateu, and V. Fitriyanah, "COVID-19 transmission risk in Surabaya and Sidoarjo: An inhomogeneous marked Poisson point process approach," *Stochastic Environmental Research and Risk Assessment*, vol. 37, no. 6, pp. 2271–2282, 2023. DOI: 10.1007/s00477-023-02393-5.

[32] A. Choiruddin, E. R. F. Sakti, T. D. A. Widhianingsih, J. Mateu, and K. Fithriasari, "Probabilistic deep learning for highly multivariate spatio-temporal log-Gaussian Cox processes," *IEEE Access*, vol. 13, pp. 94 761–94 776, 2025. DOI: 10.1109/ACCESS.2025.3570476.