Journal of Mathematics and Scientific Computing with Applications
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 7, Number 1, 2026
DOI: 10.53806/jmscowa.v7i1.1562
Page: 127–138

MIXED GEOGRAPHICALLY WEIGHTED THREE-PARAMETER LOG-LOGISTIC REGRESSION FOR MODELING LEPROSY INCIDENCE IN CENTRAL JAVA

Nuranisa[1], Purhadi[2], Santi Wulan Purnami[3]
[1,2,3]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Indonesia

Article Info

Article history:
Received : May 01, 2026
Revised : June 01, 2026
Accepted : July 05, 2026

Keywords:
Berndt-Hall-Hall-Hausman algorithm;
Leprosy Incidence;
LL3R;
MGWLL3R;
Spatial Heterogeneity.

ABSTRACT

Leprosy incidence exhibited non-normal, right-skewed distribution and spatial heterogeneity, requiring a model accommodating both characteristics. This study aimed to develop and apply the Mixed Geographically Weighted Three-Parameter Log-Logistic Regression (MGWLL3R), representing the first integration of the Mixed Geographically Weighted Regression framework with the Three-Parameter Log-Logistic distribution, to identify factors influencing leprosy incidence using cross-sectional data from 35 regencies/cities in Central Java, based on 2024 data obtained from BPS of Central Java. Parameters were estimated using Maximum Likelihood Estimation with the Berndt–Hall–Hall–Hausman algorithm, and the optimal bandwidth was selected using minimum Cross Validation. The best model employed a Fixed Gaussian kernel with bandwidth 0.37 and achieved the lowest AICc of 90.70 compared with the LL3R and GWLL3R models. Proper sanitation access was the only global predictor, whereas unemployment, safe drinking water access, population growth, and mean years of schooling showed spatial variation. These findings support region-specific leprosy control through global and local effects.

Corresponding Author:
Nuranisa,
Department of Statistics,
Institut Teknologi Sepuluh Nopember, Indonesia
Email: nuranisaaaa04@gmail.com

1. INTRODUCTION

Regression is a statistical method used to explain the relationship between response variables and predictor variables. However, classical regression models assume that the response variable is normally distributed and has a linear relationship with the predictor variable [1]. In practice, many positive continuous data are asymmetric and skewed to the right, thus not meeting this assumption. To overcome this problem, a regression model based on the Three-Parameter Log-Logistic Distribution (LL3D) was developed, which has location, scale, and shape parameters, making it more flexible in modeling various data characteristics compared to the two-parameter log-logistic distribution [2].

The development of the LL3D distribution resulted in the Three-Parameter Log-Logistic Regression (LL3R) model, which has been proven to be able to model positive continuous data that is not normally distributed [3]. Previous research by [4] showed that the LL3R model is able to overcome the problem of positive continuous data through the Maximum Likelihood Estimation (MLE) approach. Because the first derivative of the log-likelihood function does not produce a closed form, the parameter estimation process is carried out using numerical optimization methods, such as Berndt–Hall–Hall–Hausman (BHHH) and Conjugate Gradient (CG).

However, the LL3R model is still global in nature because it assumes that the relationship between predictor variables and response variables is the same across all observation areas. In geographically referenced data, this assumption is often not met due to spatial heterogeneity, a condition where the influence of predictor variables on response variables differs across regions [5]. To accommodate this condition, a Geographically Weighted Regression (GWR) approach was developed that allows model parameters to vary based on geographic location. Parameter estimation in GWR is performed using spatial weighting through a kernel function, so that nearby locations have a greater influence than distant locations [6]. Further development was carried out by [7] through the Geographically Weighted Three-Parameter Log-Logistic Regression (GWLL3R) model, which combines LL3R ability to handle non-normal data with GWR ability to accommodate spatial heterogeneity.

Although GWLL3R accounts for spatial heterogeneity by estimating all regression coefficients locally, this approach implicitly assumes that every explanatory variable exhibits spatially varying effects. In practice, however, some predictors may have relatively constant effects across all locations, whereas others vary geographically. Estimating all coefficients as local therefore introduces unnecessary model complexity, increases the number of parameters to be estimated, and may reduce estimation efficiency and model interpretability. Consequently, a modeling framework capable of simultaneously estimating global and local effects is required to provide more parsimonious and informative inference.

To overcome these limitations, a Mixed Geographically Weighted Regression (MGWR) framework was developed that allows global and local parameters to be modeled simultaneously in one model [8]. This framework is based on the concept of spatial heterogeneity, which recognizes that some explanatory variables may exhibit spatially varying effects, while others may have relatively constant effects across regions. Accordingly, predictors with relatively constant effects across regions are modeled as global parameters, whereas predictors whose effects vary across regions are modeled as local parameters. However, the application of the MGWR framework to the LL3D distribution remains unexplored in the current literature. Therefore, this study proposes a Mixed Geographically Weighted Three-Parameter Log-Logistic Regression (MGWLL3R) model as an extension of GWLL3R. The proposed model is expected to be able to produce a more parsimonious and easily interpretable model than GWLL3R.

Leprosy remains a public health problem in Indonesia due to the uneven distribution of cases across regions and the continued discovery of new cases each year [9]. Previous studies [10], [11], [12] have shown that socioeconomic factors, sanitation, education, and access to health services influence the incidence of leprosy. Central Java represents one of the provinces with a high leprosy burden in Indonesia and demonstrates considerable variation in leprosy incidence among its 35 regencies/cities. Based on the national leprosy program report, the incidence rate of leprosy in Central Java increased from approximately 3.9 cases per 100,000 population in 2023 to approximately 4.2 cases per 100,000 population in 2024, indicating that transmission continues despite various control efforts implemented through elimination strategies Differences in social, economic, educational, and environmental characteristics between regions suggest that some factors may exert relatively consistent effects across all regions, whereas others may vary spatially. Consequently, a statistical model capable of simultaneously accommodating both global and local effects is required to accurately describe the spatial heterogeneity of leprosy incidence.

Therefore, this study aims to develop and apply the MGWLL3R model, estimated using the Maximum Likelihood Estimation method and the Berndt-Hall-Hall-Hausman algorithm, to identify the factors influencing leprosy incidence across the 35 regencies and cities of Central Java Province using cross-sectional data from 2024.

2. RESEARCH METHOD

2.1. Data and Research Variables

This study used secondary data obtained from BPS of Central Java for 2024. The dataset consisted of 35 regencies/cities in Central Java Province. The response variable was the the incidence of leprosy cases in each regencies/cities during 2024. The five predictor variables were selected based on previous studies identifying socioeconomic, demographic, educational, and environmental factors as important determinants of leprosy incidence. Therefore, five predictor variables were included in the model because they are theoretically relevant to explaining the spatial variation in leprosy incidence [10], [11], [12], [13]. The research variables are presented in Table 1. All variables were continuous and measured on a ratio scale.

Table 1: Research Variables

Variable	**Description**	**Unit**	**Role**
Y	Leprosy incidence	Cases per 100,000 population	Response
X_1	Unemployment rate	Percent (%)	Predictor
X_2	Households with safe drinking water access	Percent (%)	Predictor
X_3	Population growth rate	Percent (%)	Predictor
X_4	Mean years of schooling	Years	Predictor
X_5	Households with access to proper sanitation	Percent (%)	Predictor

Classical regression models do not account for spatial dependencies among regions. To address this issue, spatial regression models incorporate a spatial weighting matrix (W) constructed from the geographic coordinates (u_i, v_i) of each region. In this study, u_i and v_i denote the longitude and latitude of the i-th regencies/cities, respectively. The geographic coordinates were obtained from the LatLong.net website and were used to calculate the distance between regions matrix for kernel weighting. Consequently, geographically closer regions receive larger weights than more distant regions.

2.2. Three-Parameter Log-Logistic Regression Model (LL3R)

The Three-Parameter Log-Logistic Distribution (LL3D) is a continuous distribution used to model positive and asymmetric data [14]. This distribution has three main parameters, namely ψ as the shape parameter, σ as the scale parameter, and μ as the location parameter. The probability density function is given as follows [15]:

$$f(y \mid \mu, \sigma, \psi) = \begin{cases} \dfrac{\sigma^{-\psi}\psi(y-\mu)^{\psi-1}}{\left(1+\left(\dfrac{y-\mu}{\sigma}\right)^{\psi}\right)^2}, & \sigma > 0,\ \psi \geq 1,\ y \geq \mu, \\ 0, & y < \mu. \end{cases} \tag{1}$$

Three-Parameter Log-Logistic Regression (LL3R) is a regression model used when the response variable Y is assumed to follow a Three-Parameter Log-Logistic distribution [16]. In LL3R, the location parameter (μ) is assumed to be influenced by predictor variables through a link function, such as the log-link function. Consequently, the value of the location parameter may vary across observations according to the characteristics of the predictor variables. Parameter estimation in the LL3R model is generally carried out using the Maximum Likelihood Estimation (MLE) method. The LL3R model can be expressed as follows [4]:

$$\ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \ \Rightarrow\ \mu_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) \tag{2}$$

The LL3R model assumes that the regression parameters are constant across all observation locations. However, in many cases, spatial heterogeneity causes the influence of predictor variables to differ across regions. Therefore, the Geographically Weighted Three-Parameter Log-Logistic Regression (GWLL3R) model is used to accommodate variations in the regression parameters at each location through spatial weighting. The GWLL3R model can be expressed as follows:

$$\mu_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}(u_i, v_i)\right), \qquad i = 1, 2, \ldots, n \tag{3}$$

2.3. Mixed Geographically Weighted Three-Parameter Log-Logistic Regression

Mixed Geographically Weighted Three-Parameter Log-Logistic Regression (MGWLL3R) is an extension of a spatial regression model that combines the concepts of Geographically Weighted Regression (GWR) and Three-Parameter Log-Logistic Regression [8]. This model is applied when the response variable follows a Three-Parameter Log-Logistic distribution and the data exhibit spatial heterogeneity. In the MGWLL3R model, the regression parameters are divided into global and local parameters. According to [5] not all predictor variables necessarily exhibit spatially varying relationships. Predictors with relatively constant relationships across the study area may be modeled using global coefficients, whereas predictors whose effects vary across locations are more appropriately modeled using local coefficients.

The geographical location of each observation is represented by spatial coordinates $(u_i,\ v_i)$, where u_i and v_i denote longitude and latitude coordinates, respectively [5]. The MGWLL3R model can be expressed as follows:

$$\ln(\mu_i) = \mathbf{x}_i^{*T}\boldsymbol{\beta}^*(u_i,v_i) + \mathbf{x}_i^{**T}\boldsymbol{\gamma}, \quad i=1,2,\ldots,n \ \Rightarrow\ \mu_i = \exp\big(\mathbf{x}_i^{*T}\boldsymbol{\beta}^*(u_i,v_i) + \mathbf{x}_i^{**T}\boldsymbol{\gamma}\big) \tag{4}$$

The probability density function is shown as follows:

$$f(y_i \mid \beta^*(u_i,v_i), \sigma^*(u_i,v_i), \psi, \gamma) = \begin{cases} \dfrac{\frac{\psi}{(\sigma^*(u_i,v_i))^{\psi}}\left(y_i - \exp\left(x_i^{*T}\beta^*(u_i,v_i) + x_i^{**T}\gamma\right)\right)^{\psi-1}}{\left(1+\left(\frac{y_i-\exp\left(x_i^{*T}\beta^*(u_i,v_i)+x_i^{**T}\gamma\right)}{\sigma^*(u_i,v_i)}\right)^{\psi}\right)^2}, & y_i > \mu_i, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

where $\sigma^*(u_i,v_i)$ denotes the local scale parameter and $\sigma^*(u_i,v_i) > 0$, ψ denotes the shape parameter with $\psi \geq 1$, and the response variable satisfies $y_i \geq \mu_i$, where $\mu_i = \exp\big(x_i^{*T}\beta^*(u_i,v_i) + x_i^{**T}\gamma\big)$

The estimated parameters in the MGWLL3R model can be written in vector form as $\boldsymbol{\theta}_{\text{MGWLL3R}} = [\beta^{*T}(u_i,v_i)\ \sigma^*(u_i,v_i)\ \psi\ \gamma]^T$ Parameter estimation in the MGWLL3R model is carried out using the MLE approach with the goal of optimizing the LL3R likelihood function. The likelihood and ln- likelihood function is written as follows:

$$L(\boldsymbol{\theta}_{\text{MGWLL3R}}) = \prod_{i=1}^{n} [f(y_i; \boldsymbol{\theta}_{\text{MGWLL3R}})]^{w_{ii}^*} \tag{6}$$

$$\ln l\big(\theta^*_{t,MGWLL3R}\big) = \sum_{i=1}^{n} w_{ii^*} \ln\left(\frac{\dfrac{\psi}{(\sigma^*(u_{i^*},v_{i^*}))^{\psi}}\left(y_i - \exp\left(x_i^{*T}\beta^*(u_{i^*},v_{i^*}) + x_i^{**T}\gamma\right)\right)^{\psi-1}}{\left(1+\left(\dfrac{y_i - \exp\left(x_i^{*T}\beta^*(u_{i^*},v_{i^*}) + x_i^{**T}\gamma\right)}{\sigma^*(u_{i^*},v_{i^*})}\right)^{\psi}\right)^2}\right). \tag{7}$$

The parameter estimates are obtained by taking the gradient of the ln-likelihood function with respect to each parameter and setting the resulting equations equal to zero. However, the resulting gradient equations do not yield closed-form solutions. Therefore, numerical optimization is required to obtain the parameter estimates. In this study, the optimization process is performed using the Berndt-Hall-Hall-Hausman (BHHH) iterative method [17]. BHHH is a numerical optimization method developed by Ernst Berndt, Bronwyn Hall, Robert Hall, and Jerry Hausman to obtain parameter estimates through MLE. The main advantage of BHHH over the Newton Raphson method is that BHHH does not require the explicit second derivative of the ln-likelihood function. The Hessian matrix is approximated using the Outer Product of Gradients (OPG) [18]. Thus, the gradient vector and Hessian matrix of the BHHH algorithm are presented:

$$\mathbf{g}_i\left(\hat{\boldsymbol{\theta}}_{i,\text{MGWLL3R}}\right) = \begin{bmatrix} \dfrac{\partial \ln f(y_i; \boldsymbol{\beta}^*(u_{i^*},v_{i^*}), \sigma^*(u_{i^*},v_{i^*}), \psi, \boldsymbol{\gamma})}{\partial \boldsymbol{\beta}^*(u_{i^*},v_{i^*})} \\ \dfrac{\partial \ln f(y_i; \boldsymbol{\beta}^*(u_{i^*},v_{i^*}), \sigma^*(u_{i^*},v_{i^*}), \psi, \boldsymbol{\gamma})}{\partial \sigma^*(u_{i^*},v_{i^*})} \\ \dfrac{\partial \ln f(y_i; \boldsymbol{\beta}^*(u_{i^*},v_{i^*}), \sigma^*(u_{i^*},v_{i^*}), \psi, \boldsymbol{\gamma})}{\partial \psi} \\ \dfrac{\partial \ln f(y_i; \boldsymbol{\beta}^*(u_{i^*},v_{i^*}), \sigma^*(u_{i^*},v_{i^*}), \psi, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \end{bmatrix} \tag{8}$$

$$\mathbf{H}\left(\hat{\boldsymbol{\theta}}^*_{i,MGWLL3R}\right) = -\sum_{i=1}^{n} w_{ii^*}\, \mathbf{g}\left(\hat{\boldsymbol{\theta}}^*_{i,MGWLL3R}\right)\left[\mathbf{g}\left(\hat{\boldsymbol{\theta}}^*_{i,MGWLL3R}\right)\right]^T. \tag{9}$$

The following steps are used to obtain the parameter estimates $\beta^*(u_i,v_i)$, $\sigma^*(u_i,v_i)$, ψ, and γ using the BHHH algorithm [17]:

1. Determine the initial parameter estimates from the results of the global LL3R model estimation: $\hat{\boldsymbol{\theta}}^{(0)}_{i^*,\text{MGWLL3R}} = \left[\hat{\boldsymbol{\beta}}^{*(0)T}(u_{i^*},v_{i^*})\ \hat{\sigma}^{*(0)}(u_{i^*},v_{i^*})\ \hat{\psi}^{(0)}\ \hat{\boldsymbol{\gamma}}^{(0)T}\right]^T$

2. Substitute the initial parameter vector $\hat{\boldsymbol{\theta}}_{i^*}^{(0)}$ into the gradient vector $\mathbf{g}_i\left(\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}\right)$ and the approximated Hessian matrix $\mathbf{H}\left(\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}\right)$ given in Equations (8) and (9).
3. Perform the iterative estimation starting from $m = 0$. The parameter estimates at the $(m+1)$-th iteration are updated using $\left(\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}^{(m)}\right)^T \mathbf{H}\left(\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}^{(m)}\right) \hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}^{(m)} < 0$
4. Repeat the previous step by setting $m = m + 1$.
5. The iteration is terminated when $\left\|\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}^{(m+1)} - \hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}^{(m)}\right\| < \varepsilon$, where ε is a small positive tolerance value. If this condition is not satisfied, the iteration continues until convergence or the maximum number of iterations is reached.
6. The parameter estimate $\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}} = \hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}^{(M)}$ is obtained.
7. The covariance matrix of parameter estimates $\widehat{\text{cov}}\left(\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}\right) = -\mathbf{H}^{-1}\left(\hat{\boldsymbol{\theta}}_{i^*,\text{MGWLL3R}}\right)$ is obtained
8. The variance of each parameter $\widehat{\text{var}}(\hat{\beta}_{j\,i^*,\text{MGWLL3R}})$ is obtained from the corresponding diagonal element of the covariance matrix.
9. Finally, the standard error of each parameter estimate is calculated as the square root of its corresponding variance.

Hypothesis testing consists of simultaneous and partial tests. Simultaneous parameter testing is conducted to examine whether the predictor variables collectively have a significant effect on the response variable. The hypothesis used for the simultaneous test are:

$H_0 : \beta_j^*(u_i, v_i) = \gamma_m = 0, \quad i = 1, 2, \ldots, n,$

$H_1 :$ at least one $\beta_j^*(u_i, v_i) \neq \gamma_m \neq 0; \quad j = 1, 2, \ldots, k, \; m = 1, 2, \ldots$

The test statistic for this hypothesis is as follows:

$$G^2_{MGWLL3R} = -2 \ln \left(\frac{L(\hat{\omega}_{MGWLL3R})}{L(\hat{\Omega}_{MGWLL3R})} \right) \sim \chi^2_{\alpha,df}. \tag{10}$$

The test statistic $G^2_{MGWLL3R}$ is Chi-square distributed with a significance level of $\alpha = 0.05$. Reject H_0 if $G^2_{MGWLL3R} > \chi^2_{\alpha,df}$. If the simultaneous test rejects H_0, the subsequent step is to perform partial testing to evaluate the individual effect of each predictor variable on the response variable and to determine the significance of these effects. The hypotheses for local parameter testing are formulated as follows:

Table 2: Partial Test of the MGWLL3R Model

Parameter	$\beta_j^*(u_i)$	$\sigma^*(u_i)$	ψ	γ_l
Hypothesis	$H_0 : \beta_j^*(u_i) = 0$ $H_1 : \beta_j^*(u_i) \neq 0$	$H_0 : \sigma^*(u_i) = 0$ $H_1 : \sigma^*(u_i) > 0$	$H_0 : \psi = 1$ $H_1 : \psi > 1$	$H_0 : \gamma_l = 0$ $H_1 : \gamma_l \neq 0$
Test Statistic	$Z_{\beta_j^*(u_i)} = \dfrac{\hat{\beta}_j^*(u_i)}{SE\left(\hat{\beta}_j^*(u_i)\right)}$	$Z_{\sigma^*(u_i)} = \dfrac{\hat{\sigma}^*(u_i)}{SE(\hat{\sigma}^*(u_i))}$	$Z_\psi = \dfrac{\hat{\psi}}{SE(\hat{\psi})}$	$Z_{\gamma_l} = \dfrac{\hat{\gamma}_l}{SE(\hat{\gamma}_l)}$
Critical Region	$\lvert Z_{\beta_j^*(u_i)}\rvert > Z_{\alpha/2}$	$Z_{\sigma^*(u_i)} > Z_\alpha$	$Z_\psi > Z_\alpha$	$\lvert Z_{\gamma_l}\rvert > Z_{\alpha/2}$
Decision	Reject H_0	Reject H_0	Reject H_0	Reject H_0

The test statistic applied in the partial test is the Z-statistic. The null hypothesis is rejected if the p-value is less than 0.05 ($\alpha = 5\%$).

The data were analyzed using R software. Parameter estimation for the LL3R, GWLL3R, and MGWLL3R models was performed using the MLE method with the BHHH algorithm. The modeling procedure begins with descriptive analysis, followed by testing the LL3 distribution of the response variable using the Kolmogorov–Smirnov test. Multicollinearity among predictor variables is then assessed using the Variance Inflation Factor (VIF), after which the LL3R model is estimated. Spatial heterogeneity is subsequently examined using the Breusch–Pagan test. If spatial heterogeneity is detected, the GWLL3R model is estimated using four kernel functions, namely fixed gaussian, fixed bisquare, adaptive gaussian, and adaptive bisquare. For each kernel function, the optimal bandwidth is determined by minimizing the Cross-Validation (CV) criterion, and the kernel function with the best performance is selected. The predictor variables are then classified into global and local variables. Finally, the MGWLL3R model is estimated using the selected optimal kernel function and bandwidth, and the model results are interpreted. All hypothesis tests were conducted at a significance level of $\alpha = 0.05$ (5%).

3. RESULT AND ANALYSIS

3.1. Descriptive Analysis

This study uses data from 35 regencies/cities in Central Java, where the response variable is the incidence of leprosy cases. All variables were checked for missing values and potential outliers prior to analysis. No missing values were identified, and no observations were excluded from the dataset. The results are shown in Table 3 below.

Table 3: Data Description

Variable	N	Min	Max	Mean	Variance
Leprosy incidence (Y)	35	0.12	13.08	3.18	11.68
Unemployment rate (X_1)	35	2.35	8.35	4.52	2.23
Households with safe drinking water access (X_2)	35	87.09	99.93	95.96	14.69
Population growth rate (X_3)	35	0.24	1.28	0.96	0.07
Mean years of schooling (X_4)	35	6.41	11.48	8.32	1.89
Households with proper sanitation access (X_5)	35	52.40	98.54	86.57	117.55

Based on Table 3, the incidence of leprosy ranges from 0.12 to 13.08 cases, with a mean of 3.18 and a variance of 11.68, indicating considerable variability among districts/cities. The predictor variables also exhibit substantial variation. Proper sanitation access ranges from 52.40% to 98.54%, while safe drinking water access ranges from 87.09% to 99.93%. Population growth rate varies between 0.24% and 1.28%, and mean years of schooling ranges from 6.41 to 11.48 years. Furthermore, multicollinearity among predictor variables was examined using the Variance Inflation Factor (VIF). Since all VIF values are below 5, no serious multicollinearity is present, and all predictor variables can be included in the model.

Next, a distributional fit test was conducted to verify whether the response variable follows a LL3D. Figure 1 presents the histogram of the leprosy incidence rate data for the 35 regencies/cities in Central Java along with the fitted LL3D. The x-axis represents the leprosy incidence rate, while the y-axis represents the frequency. The estimated parameters of $\psi = 1.95$, $\sigma = 1.71$, and $\mu = 0.02$. The fitted density curve closely follows the observed data distribution, suggesting an adequate fit.

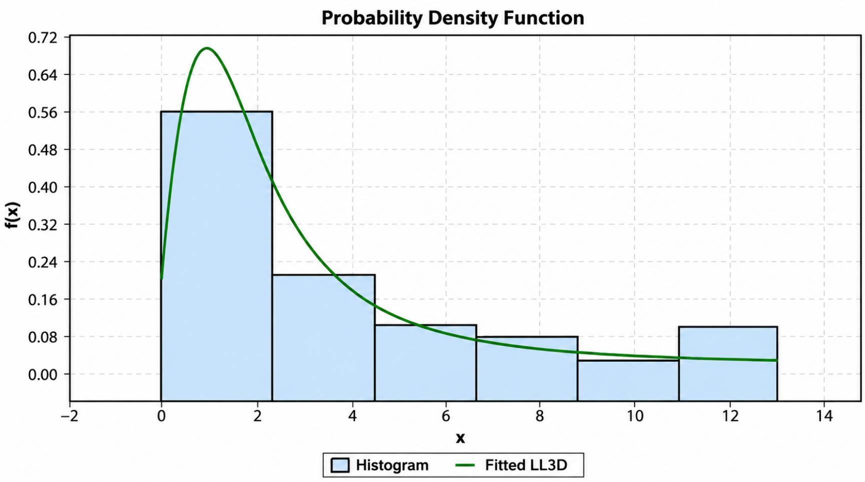

Figure 1. Histogram of the LL3D

This finding is supported by the Kolmogorov Smirnov test, which yielded a test statistic of 0.08 and a p-value of 0.98. Since the p-value exceeds 0.05, the null hypothesis cannot be rejected, indicating that the response variable follows a LL3D distribution. Therefore, the analysis proceeded with the LL3R model.

3.2. Modeling using the LL3R Model

The parameters of the LL3R model were estimated using the MLE method. The estimated parameters, along with their standard errors, Z statistics, and p values, are presented in Table 4.

Table 4: Parameter Estimates of the LL3R Model

	Estimate	**Std. Error**	**Z-value**	**P-value**
β_0	-1.72	0.53	-3.23	0.00*
β_1	0.63	0.28	2.29	0.02*
β_2	0.76	0.32	2.34	0.02*
β_3	0.89	0.45	1.98	0.04*
β_4	-1.73	0.59	-2.93	0.00*
β_5	2.79	0.49	5.64	0.00*
σ	0.93	0.23	3.99	0.00*
ψ	1.51	0.32	4.75	0.00*

Table 4 presents the parameter estimates of the LL3R model. Based on the partial significance tests, all predictor variables X_1, X_2, X_3, X_4, and X_5, have a significant effect on leprosy incidence at the 5% significance level. Subsequently, the Breusch Pagan test was performed to detect spatial heterogeneity. The obtained p-value of 0.004 indicates significant spatial heterogeneity among regions. Therefore, the GWLL3R model is utilized to accommodate spatial heterogeneity in the data.

3.3. Modeling using the GWLL3R Model

The GWLL3R model was estimated using four kernel functions, namely fixed gaussian, fixed bisquare, adaptive gaussian, and adaptive bisquare. For each kernel function, the optimal bandwidth was determined by minimizing the CV. The adaptive gaussian kernel yielded the lowest CV value of 307.40 and was therefore selected as the optimal kernel. The corresponding optimal bandwidth was 0.30, which represents the proportion of nearest neighbors used in the weighting of the adaptive gaussian kernel. Using the selected kernel, the GWLL3R model produced local parameter estimates for each of the 35 regencies/cities. A summary of these estimates is presented in Table 5.

Table 5: Summary of Local Parameter Estimates of the GWLL3R Model

Parameter	**Min**	**Median**	**Max**	**Mean**
β_0	-2.22	-1.33	-1.27	-1.54
β_1	0.61	0.76	1.23	0.86
β_2	0.11	0.42	0.78	0.39
β_3	-0.42	0.36	0.95	0.37
β_4	-1.81	-1.45	-1.08	-1.38
β_5	2.03	2.66	3.49	2.64

Table 5 summarizes the distribution of the local parameter estimates across the 35 regencies/cities. The varying ranges of the regression coefficients indicate that the effects of the predictor variables differ spatially, confirming that the relationships between the predictors and leprosy incidence are not constant across locations.

Subsequently, simultaneous hypothesis testing was conducted using the likelihood ratio test. The test results showed a p-value of 1.61×10^{-8}, indicating that the null hypothesis was rejected at the 5% significance level. Therefore, at least one predictor variable simultaneously has a significant influence on leprosy incidence. Then, partial hypothesis tests were performed locally for each regency/city. The results indicate that the significant predictor variables differ across locations, reflecting spatial variation in the relationships between the predictors and leprosy incidence. Based on the results of the local significance tests, the regencies/cities were classified into several groups according to the combination of significant predictor variables, as presented in Table 6.

Table 6: Regional Groups Based on Significant Predictors of the GWLL3R Model

Regencies/Cities	**Total**	**Significant Predictor**
Jepara	1	X_1, X_2, X_3, X_4, X_5
Tegal Regency	1	X_1, X_3, X_4, X_5
Banyumas, Pemalang, Brebes	3	X_1, X_3, X_5

Regencies/Cities	Total	Significant Predictor
Wonosobo, Magelang Regency, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Demak, Semarang Regency, Batang, Magelang City, Surakarta City, Salatiga City, Semarang City	20	X_1, X_4, X_5
Cilacap, Purbalingga, Banjarnegara, Kebumen, Purworejo, Temanggung, Pekalongan Regency, Pekalongan City, Tegal City	9	X_1, X_5
Kendal	1	X_5

Table 6 shows that the significant predictor variables vary across regencies/cities, indicating spatial heterogeneity in the factors associated with leprosy incidence. Based on the combination of significant predictor variables, the 35 regencies/cities were classified into six regional groups. Among the predictor variables, X_5 was significant in all 35 regencies/cities, whereas X_1 was significant in 34 regencies/cities. In contrast, X_2, X_3, and X_4 were significant only in selected regions. These findings suggest that the effect of X_5 is relatively consistent across the study area, whereas the effects of the remaining predictor variables exhibit greater spatial variation. Accordingly, the GWLL3R results were used to identify candidate global and local variables. The candidate global and local variables were subsequently estimated using the back-fitting algorithm in the MGWLL3R model, where X_5 was specified as the global variable, while X_1, X_2, X_3, and X_4 were specified as local variables.

Table 6 shows that the significant predictor variables vary across regencies/cities, indicating spatial heterogeneity in the factors associated with leprosy incidence. Based on the combination of significant predictor variables, the 35 regencies/cities were classified into six regional groups. Among the predictor variables, X_5 was significant in all 35 regencies/cities, whereas X_1 was significant in 34 regencies/cities. In contrast, X_2, X_3 and X_4 were significant only in selected regions. These findings suggest that the effect of X_5 is relatively consistent across the study area, whereas the effects of the remaining predictors exhibit greater spatial variation. Accordingly, the GWLL3R results were used to identify candidate global and local variables, where X_5 was specified as the global variable and X_1, X_2, X_3 and X_4 were specified as local variables.

3.4. Modeling Using The MGWLL3R Model

The MGWLL3R model was estimated to accommodate both global and local effects. Global and local variables were identified as candidate variables from GWLL3R. A total of 141 parameter estimates were obtained for the 35 regencies/cities, including local regression coefficients and a global coefficient. The MGWLL3R model was estimated using four kernel functions, namely fixed gaussian, fixed bisquare, adaptive gaussian, and adaptive bisquare. For each kernel function, the optimal bandwidth was determined by minimizing the CV. The fixed gaussian kernel yielded the lowest CV value of 288.19 and was therefore selected as the optimal kernel. The optimal bandwidth was 0.37, which represents which represents the spatial distance over which neighboring observations influence the local parameter estimation. Table 7 summarizes the distribution of the estimated parameters across all locations.

Table 7: Summary of Local Parameter Estimates of the MGWLL3R Model

Parameter	Min	Median	Max	Mean
β_0	-2.35	-1.69	-1.47	-1.76
β_1	-0.43	0.79	1.12	0.68
β_2	-0.49	0.55	1.05	0.43
β_3	-0.89	0.71	2.48	0.54
β_4	-2.27	-1.51	-0.34	-1.44

Table 7 summarizes the distribution of the local parameter estimates of the MGWLL3R model across all observation locations. The coefficient of the unemployment rate (β_1) varies from -0.43 to 1.12, suggesting that its effect differs across locations and is generally positive. Similarly, the coefficient of the households with access to safe drinking water (β_2) ranges from -0.49 to 1.05, with an average of 0.43. The population growth rate (β_3) exhibits the widest range of local coefficients, from -0.89 to 2.48. In contrast, the coefficient of mean years of schooling (β_4) is consistently negative, ranging from -2.27 to -0.34 with a mean of -1.44, indicating that higher mean years of schooling is associated with a lower value of the response variable in all locations, although the magnitude of the effect varies spatially.

Based on the simultaneous significance test, the MGWLL3R model produced a p-value of 3.04×10^{-10}, leading to the rejection of the null hypothesis. This result indicates at least one parameter significantly influences leprosy incidence. Therefore, partial significance tests were conducted for each location to determine the significant local predictors. An example of the partial significance test results for one observation location is presented in Table 8.

Table 8: Parameter Estimates of the MGWLL3R Model for Rembang

	Estimate	**Std. Error**	**Z-value**	**P-value**
β_0	-1.53	0.53	-2.88	0.00*
β_1	0.81	0.27	2.92	0.00*
β_2	0.41	0.32	1.27	0.20
β_3	0.26	0.44	0.57	0.56
β_4	-1.54	0.59	-2.62	0.01*
γ	2.83	0.15	-18.28	0.00*

Table 8 presents an example of the local parameter estimates and the partial significance test results of the MGWLL3R model for Rembang Regency. At the 5% significance level, the intercept (β_0), unemployment rate, mean years of schooling, and the global coefficient (Y), representing the households with access to proper sanitation, are statistically significant, whereas the households with access to safe drinking water and population growth rate are not significant.

The positive coefficient of the unemployment rate (X_1) suggests that regions with higher unemployment tend to experience a higher incidence of leprosy, which may reflect poorer socioeconomic conditions that can limit access to healthcare services and delay early diagnosis and treatment. Conversely, the negative coefficient of mean years of schooling (X_4) indicates that regencies/cities with higher average years of schooling tend to have lower leprosy incidence. This finding suggests that higher educational attainment at the regional level is associated with more favorable social and health conditions that may support disease prevention.

It should be noted that the positive coefficient of households with access to proper sanitation (X_5) differs from the expected theoretical relationship. Although the Pearson correlation between X_5 and leprosy incidence is positive, the correlation is very weak ($r = 0.063$), indicating that this relationship should be interpreted with caution. One possible explanation is that regencies/cities with higher sanitation coverage may also have better healthcare infrastructure and more active case detection systems. In addition, this study used aggregated data at the regencies/cities level, so the estimated relationship may also have been influenced by other regional factors that were not included in the model. Therefore, the positive coefficient should not be interpreted as evidence that improved sanitation increases the incidence of leprosy. Instead, it may reflect the characteristics of the study area and the combined influence of multiple factors related to leprosy incidence. Overall, the findings of this study are consistent with previous studies [10], [11], [12], which identified socioeconomic and environmental factors as significant determinants of leprosy incidence. While the identified factors are generally in agreement with the existing literature, this study extends previous research by showing that the effects of these factors vary across regencies/cities through the application of the MGWLL3R model.

Subsequently, the MGWLL3R model was estimated by specifying X_5 as a global variable and X_1, X_2, X_3, and X_4 as local variables. Based on the partial significance tests, the regencies/cities were classified into several groups according to the combination of significant local predictors, as shown in Table 9.

Table 9: Regional Groups Based on Significant Predictors of the MGWLL3R Model

Regencies/Cities	**Total**	**Significant Predictor**
Banyumas, Purbalingga, Pemalang, Tegal Regency, Brebes, Tegal City	6	X_1, X_2, X_3, X_4, X_5
Pati, Kudus, Jepara	3	X_2, X_3, X_4, X_5
Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang	6	X_1, X_4, X_5
Cilacap	1	X_1, X_5
Banjarnegara, Batang, Boyolali, Demak, Kebumen, Kendal, Klaten, Magelang Regency, Pekalongan Regency, Purworejo, Semarang Regency, Sukoharjo, Temanggung, Wonosobo, Magelang City, Pekalongan City, Salatiga City, Semarang City, Surakarta City	19	X_5

Based on Table 9, classifies the regencies/cities in Central Java into 5 groups according to the significant predictors identified by the MGWLL3R model. The results reveal substantial spatial heterogeneity in the determinants of leprosy, indicating that the influence of socioeconomic and demographic factors is not uniform across the study area. The first group, consisting of Banyumas, Purbalingga, Pemalang, Tegal Regency, Brebes, and Tegal City, is characterized by all local predictors with the global predictor being statistically significant, suggesting that leprosy incidence in these regions is influenced by a combination of unemployment, access to safe drinking water, population growth, educational attainment, and proper sanitation. In contrast, Pati, Kudus, and Jepara show no significant effect of the unemployment rate, whereas Wonogiri, Karanganyar, Sragen, Grobogan, Blora, and Rembang are mainly influenced by unemployment, education, and proper sanitation. Cilacap forms a distinct group in which only unemployment and proper sanitation remain significant. These findings suggest that the relative importance of risk factors differs across regencies/cities, likely reflecting variations in local socioeconomic conditions, demographic characteristics, and environmental settings.

Therefore, the MGWLL3R model provides valuable insights into the spatially varying effects of the explanatory variables. Visually, the spatial distribution of the regional groups is presented in Figure 2. The map illustrates the classification of regencies/cities based on the combination of significant local predictors identified by the MGWLL3R model.

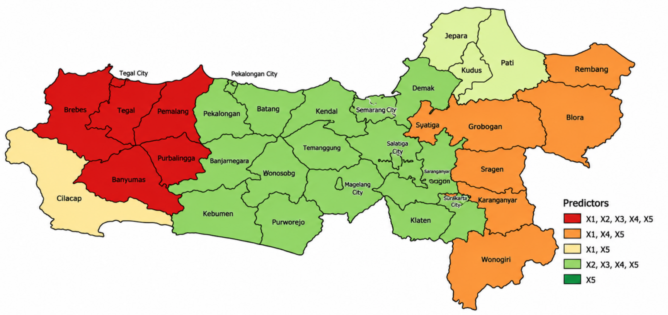

Figure 2. Regional Groups Based on Significant Predictors of Leprosy Incidence in Central Java Province, 2024

Subsequently, the three models were compared based on the corrected Akaike Information Criterion (AICc). The best model was selected based on the smallest AICc value, as a lower AICc indicates a better balance between model fit and complexity. The comparison of AICc values for the LL3R, GWLL3R, and MGWLL3R models is presented in Table 10.

Table 10: Best Model Selection

Model	**AICc**
LL3R	127.09
GWLL3R	104.21
MGWLL3R	90.70

Based on Table 10 the MGWLL3R model with the fixed gaussian kernel weighting function produced the lowest AICc value compared with the GWLL3R and LL3R models. This result indicates that the MGWLL3R model provides a better representation of the spatial variation in leprosy incidence across Central Java Province in 2024. Therefore, the MGWLL3R model was selected as the best model because it achieves a superior balance between goodness-of-fit and model complexity. Furthermore, by allowing both global and local effects, the MGWLL3R model is more capable of capturing spatial heterogeneity than the LL3R and GWLL3R models.

4. CONCLUSION

This study demonstrated that the MGWLL3R model effectively captured both spatial heterogeneity and global effects in modeling leprosy incidence across regencies/cities in Central Java. The regencies/cities were classified into five regional groups based on the significant predictors. Households with access to proper sanitation (X_5) was identified as the only globally significant predictor, whereas the unemployment rate (X_1), households with access to safe drinking water (X_2), population growth rate (X_3), and mean years of schooling (X_4) exhibited spatially varying effects. These findings indicate that sanitation improvement should remain a priority across Central Java, while interventions addressing the other significant factors should be tailored to the characteristics of each region. The results may support the Central Java Provincial Public health Office and the national leprosy elimination program in developing region-specific intervention strategies. This study was limited to data from one province and one year. Future studies are encouraged to use multi-year data and apply Geographically and Temporally Weighted Regression (GTWR) to account for both spatial and temporal heterogeneity, providing a better understanding of how the effects of risk factors on leprosy incidence change over time.

REFERENCES

[1] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th ed. New York: McGraw-Hill Irwin, 2005, https://books.google.com/books?id=q0-PAAAACAAJ.

[2] W. A. M. Shehata, M. M. Abdullah, and M. K. A. Refaie, "A novel four-parameter log-logistic model: Mathematical properties and applications to breaking stress, survival times and leukemia data," *Pakistan Journal of Statistics and Operation Research*, vol. 18, no. 1, pp. 133–149, 2022, https://doi.org/10.18187/pjsor.v18i1.3268.

[3] S. Suyitno *et al.*, "Geographically weighted weibull regression modeling on dissolved oxygen data to analyze river water quality in east kalimantan," *MethodsX*, vol. 16, p. 103745, 2025, https://doi.org/10.1016/j.mex.2025.103745.

[4] T. Sulistyaningsih, Purhadi, and S. Andari, "Analysis of hospitalization-recovery duration for dhf patients at rsud haji, east java utilizing the three-parameter log-logistic regression," in *Proceedings of the 9th International Conference on Business and Industrial Research (ICBIR)*, 2024, pp. 1305–1310, https://doi.org/10.1109/ICBIR61386.2024.10875875.

[5] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, England: John Wiley & Sons, 2002, https://onlinelibrary.wiley.com/doi/book/10.1002/9780470996268.

[6] E. M. Mujiarti, Y. Yundari, and N. M. Huda, "Pemodelan geographically weighted regression pada angka partisipasi sekolah di kalimantan barat tahun 2022," *Jurnal Gaussian*, vol. 13, no. 1, pp. 36–47, 2024, https://doi.org/10.14710/j.gauss.13.1.36-47.

[7] N. Mawadah, Purhadi, and S. Andari, "Geographically weighted log-logistic 3-parameter regression model for stunting prevalence: A case study on sulawesi island, indonesia," in *Proceedings of the 8th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2025, pp. 552–557, https://doi.org/10.1109/ISRITI68345.2025.11393132.

[8] N. Chamidah *et al.*, "Spatial modeling of food security index in central java using mixed geographically weighted regression," *ZERO: Jurnal Sains, Matematika dan Terapan*, vol. 9, no. 1, pp. 247–256, 2025, https://doi.org/10.30829/zero.v9i1.25044.

[9] A. N. A. M. Pertiwi and F. Syahrul, "Risk factors for leprosy: A systematic review," *Indonesian Journal of Public Health*, vol. 19, no. 3, pp. 575–589, 2024, https://doi.org/10.20473/IJPH.V19I3.2024.575-589.

[10] C. Andreas, H. Horidah, R. Sulistiana, D. Venosia, and N. Chamidah, "Pemodelan prevalensi penyakit kusta di jawa timur dengan pendekatan geographically weighted regression," *Jurnal Aplikasi Statistika dan Komputasi Statistik*, vol. 14, no. 1, pp. 33–48, 2022, https://doi.org/10.34123/jurnalasks.v14i2.351.

[11] M. Makful, Y. Handayani, and F. Nugraha, "Spatial model of geographic distribution of leprosy cases in east java province, indonesia," *Unnes Journal of Public Health*, vol. 13, no. 2, pp. 93–104, 2024, https://doi.org/10.15294/ujph.v13i2.4188.

[12] T. Saifudin *et al.*, "Leprosy case modeling in east java using spatial regression with queen contiguity weighting," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 3, pp. 2141–2154, 2025, https://doi.org/10.30598/BAREKENGVOL19ISS3PP2141-2154.

[13] Kementerian Kesehatan Republik Indonesia, "Program pengendalian kusta: Ran eliminasi kusta 2023–2027," Kementerian Kesehatan Republik Indonesia, Jakarta, Indonesia, Tech. Rep., 2023, https://nlrindonesia.or.id/wp-content/uploads/2025/01/RAN-Eliminasi-Kusta-2023-2027-1.pdf.

[14] V. P. Singh, "Three-parameter log-logistic distribution," in *Entropy-Based Parameter Estimation in Hydrology*. Springer, 1998, pp. 297–311, https://doi.org/10.1007/978-94-017-1431-0_18.

[15] V. P. Singh, H. Guo, and F. X. Yu, "Parameter estimation for 3-parameter log-logistic distribution (lld3) by pome," *Stochastic Hydrology and Hydraulics*, vol. 7, no. 3, pp. 163–177, 1993, https://doi.org/10.1007/BF01585596.

[16] A. H. Muse, S. M. Mwalili, and O. Ngesa, "On the log-logistic distribution and its generalizations: A survey," *International Journal of Statistics and Probability*, vol. 10, no. 3, pp. 93–110, 2021, https://doi.org/10.5539/ijsp.v10n3p93.

[17] E. N. Safitri, A. R. Hakim, and Suparti, "Pemodelan regresi gamma menggunakan metode optimasi broyden-fletcher-goldfarb-shanno (bfgs): Studi kasus pencemaran sungai di kota semarang," *Jurnal Gaussian*, vol. 14, no. 2, pp. 247–256, 2025, https://doi.org/10.14710/J.GAUSS.14.2.247-256.

[18] P. R. Hargandi, Purhadi, and A. Choiruddin, "Parameter estimation and hypothesis testing of gtw compound correlated bivariate poisson regression model: A theoretical development," *Jambura Journal of Mathematics*, vol. 7, no. 2, pp. 138–143, 2025, https://doi.org/10.37905/JJOM.V7I2.32712.