STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–15.

Mycobacterium Tuberculosis Hotspots: A Spatial Analysis of Tuberculosis Cases in West Java, Indonesia

Arip Ramadan [1], Dwi Rantini [2,3∗], Della Lukita Wardhani [2], Septia Devi Prihastuti Yasmirullah [2,3], Najma Attaqiya Alya [2,4], Alhassan Sesay [5], Bela Sonia [2]

[1] *Information System Study Program, School of Industrial and System Engineering, Telkom University, Surabaya Campus, Jl. Ketintang No.156, Surabaya 60231, East Java, Indonesia*
[2] *Data Science Technology Study Program, Department of Engineering, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya 60115, Indonesia*
[3] *Research Group of Data-Driven Decision Support System, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya 60115, Indonesia*
[4] *Institute of Statistics and Data Science, Faculty of Science, National Tsing Hua University, Hsinchu, Taiwan*
[5] *Faculty of Transformative Education, The United Methodist University, Sierra Leone*

Abstract Globally, tuberculosis ranks as the second most prevalent cause of mortality resulting from a single infectious disease. In 2023, Indonesia experienced a 24-year peak in tuberculosis cases, with West Java Province exhibiting the highest occurrence rate, reflecting a 24.2% increase compared to the previous year. This study aims to investigate the determinants of tuberculosis cases in West Java by employing a spatial approach to account for regional heterogeneity. The Geographically Weighted Negative Binomial Regression (GWNBR) method is applied due to its suitability for handling count data with overdispersion while accommodating spatially varying relationships. Parameter estimation is conducted using the Maximum Likelihood Estimation (MLE) method enhanced with the Nelder–Mead optimization algorithm. Four kernel function types are incorporated in the modeling process, namely Fixed Gaussian, Adaptive Gaussian, Fixed Bi-square, and Adaptive Bi-square. Based on the Akaike Information Criterion (AIC), the Adaptive Gaussian kernel demonstrates the best model performance, achieving an AIC value of 1,000.11. Population density emerges as the most influential determinant, showing statistical significance across all regions. These findings emphasize the importance of developing localized tuberculosis control strategies that are tailored to specific regional characteristics.

Keywords GWNBR; Nelder-Mead; Overdispersion; Spatial Analysis; Tuberculosis

DOI: 10.19139/soic-2310-5070-3388

1. Introduction

Tuberculosis is an infectious disease triggered by the bacterium *Mycobacterium tuberculosis*,capable of impacting various organs [1]. Tuberculosis is both preventable and often treatable [2, 3]. Nevertheless, according to 2022 World Health Organization (WHO) statistics, Tuberculosis holds the position of the second most prevalent cause of death globally from a single infectious agent, surpassed only by the coronavirus disease (COVID-19). Tuberculosis also results in nearly twice as many deaths as HIV/AIDS [4]. Annually, over 10 million individuals continue to be affected by Tuberculosis [5]. Indonesia ranks second globally in the number of TB sufferers, following India [4, 6]. The province with the highest reported number of Tuberculosis cases is West Java, one of Indonesia's most densely populated regions [7]. Furthermore, data from the `https://opendata.jabarprov.go.id` website indicates that in 2023, West Java recorded 211,959 Tuberculosis cases, reflecting a 24.2% increase compared to

---

∗Correspondence to: Dwi Rantini (Email: dwi.rantini@ftmm.unair.ac.id). Data Science Technology Study Program, Department of Engineering, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya 60115, Indonesia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the preceding year. Consequently, effective strategies and policies are essential to tackle the burden of Tuberculosis cases within West Java Province.

Previous research on modeling the spread of Tuberculosis found that various factors can influence the transmission of Tuberculosis cases. According to research by Mohidem et al. [8], there is a significant relationship between Tuberculosis cases and population density. Then, another study by Wondmeneh and Mekonnen [9] showed that the number of HIV/AIDS cases is a factor related to Tuberculosis incidents. Research by Wang et al. [10] shows that the number of poor people is a factor that influences the number of Tuberculosis cases. Research by Samrah et al. [11] showed that the number of households practicing clean and healthy living is related to Tuberculosis. Research by Bisht et al. [12] showed that the number of Diabetes Mellitus patients is related to Tuberculosis. A study by LeBan et al. [13] showed that the number of community health centers affects Tuberculosis. According to the analysis conducted by Xue et al. [14], the number of Tuberculosis cases has a spatial influence.

Previously mentioned studies did not examine overdispersion in Tuberculosis case data. Therefore, the aim of this study is to model Tuberculosis cases by reviewing overdispersion data. Poisson regression and Negative Binomial regression are appropriate models for count data [15, 16]. Count data tends to experience overdispersion rather than underdispersion [17]. Meanwhile, Poisson regression can be used when equidispersion is observed [18, 19]. Spatial modeling with Geographically Weighted Negative Binomial Regression (GWNBR) is the most effective alternative to handle data with overdispersion [20]. Therefore, this study will explore the most appropriate spatial approach for the number of Tuberculosis cases with the GWNBR model. This model can accommodate overdispersion data, resulting in more precise predictions relative to actual data.

2. Material and Methods

This study leverages data sourced from the Open Data Jabar portal (`opendata.jabarprov.go.id`), encompassing information provided by relevant government bodies. The data utilized in this research comprise the count of Tuberculosis cases, population density, count of HIV cases, number of individuals living in poverty, proportion of households adhering to clean and healthy living practices, count of Diabetes Mellitus patients, and the number of community health centers. Additionally, spatial data in the form of shapefiles were acquired from the Geosai website. These shapefiles incorporate longitude and latitude coordinates for each district or city throughout West Java Province, which are crucial for undertaking spatial analysis. Table 1 outlines the research variables in this study.

Table 1. Research Variables

Variable	**Description**	**Unit**
Number of Tuberculosis cases	Total number of Tuberculosis cases in each district/city	Persons
Population density	Number of persons per unit area	Persons/km^2
Number of HIV cases	Total HIV cases reported	Persons
Number of people living in poverty	Total individuals living below the poverty line	Thousand persons
Percentage of households practicing clean and healthy living behavior	Proportion of households adhering to clean and healthy living practices	Percent
Number of Diabetes Mellitus patients	Total Diabetes Mellitus cases	Persons
Number of community health centers in West Java Province for the year 2023	Total community health centers in West Java for 2023	Units

The spatial structure of these data enables the examination of disease dispersion patterns across regions, allowing the identification of clustering or spatial heterogeneity in Tuberculosis cases. In the Poisson model, the variance and the mean are expected to be equal. Overdispersion in the Poisson model can lead to invalid coefficient estimates because the model is unable to capture the greater variability present in the data [21, 22]. In such cases, a commonly used model to handle overdispersion is the Negative Binomial model [23]. The presence of dispersion can be tested using a dispersion test [24]. Equation (1) represents the test statistic for testing the equidispersion assumption:

$$\phi = \frac{D}{\mathrm{df}} \tag{1}$$

where D denotes the deviance and $\mathrm{df} = n - 1$. If the resulting dispersion value $\phi > 1$, the model exhibits overdispersion; if $\phi < 1$, the model indicates underdispersion; and if $\phi = 1$, the equidispersion assumption is satisfied. Equation (2) is the deviance formula.

$$D = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right] \tag{2}$$

where y_i and $\hat{y}_i$ denote the observed value of the response variable at location i and the corresponding predicted value, respectively, for $i = 1, 2, \ldots, n$.

Spatial effects manifest in two primary forms: spatial heterogeneity and spatial dependence [25, 26]. Spatial dependence indicates that the value of a variable at a specific location is affected by the values of the same variable in nearby locations [27]. Conversely, spatial heterogeneity arises when a variable and its relationships with other variables fluctuate across the study area [28]. Spatial dependence can be assessed using Moran's I test [29].

After identifying the presence of spatial autocorrelation using Moran's I, it is important to further investigate whether the observed spatial pattern is influenced by spatial heterogeneity. While Moran's I evaluates spatial autocorrelation, it does not account for potential non-constant variance across spatial units, which may bias statistical inference. To address this issue, Anselin (1988) proposed the use of the Breusch–Pagan (BP) test as a diagnostic tool to detect spatial heterogeneity in regression residuals [30, 31]. This test examines whether the variance of residuals from a spatial regression model remains constant across observations or varies spatially.

Modeling with a spatial heterogeneity approach and data following a Negative Binomial distribution, commonly referred to as the Geographically Weighted Negative Binomial Regression (GWNBR) model, allows each area to have its own local parameter estimates. Equation (3) presents the GWNBR model, where $Y_i \sim \mathrm{NB}$:

$$\mu_i = \exp\left(\sum_{j=1}^{p}\beta_j(u_i, v_i)\, x_{ij}\right), \qquad \theta = \theta(u_i, v_i) \tag{3}$$

where $i = 1, 2, \ldots, n$ represents the observation locations, and $j = 1, 2, \ldots, p$ denotes the predictor variables. The value μ_i is the mean of the response variable at location i, x_{ij} is the observed value of the j-th predictor variable at location i, and (u_i, v_i) represents the longitude and latitude coordinates of each location. Furthermore, $\beta_j(u_i, v_i)$ denotes the local regression coefficient of the j-th predictor variable, while $\theta(u_i, v_i)$ represents the dispersion parameter of the Negative Binomial distribution.

The probability mass function of the Negative Binomial distribution for each location (u_i, v_i) in the GWNBR model is given in Equation (4):

$$f(Y_i) = \frac{\Gamma\left(Y_i + \frac{1}{\theta_i}\right)}{\Gamma\left(\frac{1}{\theta_i}\right)\,\Gamma(Y_i + 1)}\left(\frac{1}{1 + \theta_i\mu_i}\right)^{\frac{1}{\theta_i}}\left(\frac{\theta_i\mu_i}{1 + \theta_i\mu_i}\right)^{Y_i} \tag{4}$$

where

$$f(Y_i) = f\left(Y_i \mid x_{ij}, \beta_j(u_i, v_i), \theta(u_i, v_i)\right), \qquad Y_i = 0, 1, 2, \ldots$$

and the mean μ_i follows an exponential form,

$$\mu_i = \exp\left(x_i^T \beta(u_i, v_i)\right).$$

Negative Binomial regression belongs to the class of Generalized Linear Models (GLM) [32]. According to McCullagh and Nelder (1989), the GLM framework consists of three essential components [33]:

1. The distribution of the response variable Y_i is a member of the exponential family.
2. A link function $g(\mu_i)$ that relates the expected value of Y_i to the linear predictor.
3. A linear predictor defined as

   $$\eta_i = x_i^T \beta,$$

   where η_i denotes the linear predictor.

For a distribution to be analyzed within the GLM framework, it must be expressed in the general exponential family form, as shown in Equation (5). The general exponential family form for discrete variables is expressed as:

$$f(Y_i) = c(\theta) \exp\{d(\theta)\, T(Y_i)\}\, h(Y_i) \tag{5}$$

This formulation indicates that a probability distribution can be represented using a sufficient statistic $T(Y_i)$ and a canonical parameter $d(\theta)$. In the context of Generalized Linear Models (GLM), the link function is derived from the relationship between the expected value μ_i and the canonical parameter $d(\theta)$, as shown in Equation (6):

$$g(\mu_i) = d(\theta) \tag{6}$$

This implies that the link function arises naturally from the exponential structure of the distribution. Since the probability mass function (PMF) of the Negative Binomial distribution does not directly appear in exponential family form, the logarithm of the PMF is taken, yielding:

$$f(Y_i) = c(\theta_i) \exp\left\{ Y_i \log\left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i} \right) - \frac{1}{\theta_i} \log(1 + \theta_i \mu_i) \right\} h(Y_i) \tag{7}$$

From Equation (7), the canonical parameter is obtained as:

$$d(\theta_i) = \log\left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i} \right) \tag{8}$$

The logarithmic transformation of μ_i can be written as:

$$\log\left(\frac{\theta_i \mu_i}{1 + \theta_i \mu_i} \right) = \log(\mu_i) + \log(\theta_i) - \log(1 + \theta_i \mu_i) \tag{9}$$

Since θ_i is assumed to be constant, the derivative of $d(\theta_i)$ reduces to $\log(\mu_i)$. Consequently, the link function for the Negative Binomial regression becomes:

$$\eta_i = d(\theta_i) = \log(\mu_i).$$

Parameter estimation for the GWNBR model is conducted using the Maximum Likelihood Estimation (MLE) method. The log-likelihood function is expressed as follows [34]:

$$\ln L(\cdot) = \sum_{i=1}^{n} w_{ij}(u_i, v_i) \left[\ln \frac{\Gamma\left(Y_i + \frac{1}{\theta_i}\right)}{\Gamma\left(\frac{1}{\theta_i}\right) \Gamma(Y_i + 1)} + Y_i \ln(\theta_i \mu_i) - \left(\frac{1}{\theta_i} + Y_i \right) \ln(1 + \theta_i \mu_i) \right] \tag{10}$$

where

$$L(\cdot) = L(\beta(u_i, v_i), \theta_i \mid Y_i, x_{ij})\,.$$

For count data, the Nelder–Mead (NM) optimization algorithm has been shown to perform better than the BFGS algorithm [35]. Therefore, parameter estimation in this study employs the Nelder–Mead iterative method until convergence is achieved.

The steps of the Nelder–Mead (NM) iteration are as follows.

1. For a function $g(X)$ with p variables, $p+1$ initial values are required.
2. Compute
$$g(X_1),\ g(X_2),\ \dots,\ g(X_{p+1}).$$
3. Order the function values such that
$$g_1 = g(X_1) \leq g_2 = g(X_2) \leq \ldots \leq g_{p+1} = g(X_{p+1}).$$
4. Compute the reflection point
$$X_r = 2\bar{X} - X_{p+1},$$
where
$$\bar{X} = \sum_{i=1}^{p} \frac{X_i}{p},$$
then calculate
$$g_r = g(X_r).$$
5. If $g_1 \leq g_r \leq g_p$, accept X_r and stop the iteration.
6. If $g_r < g_1$, compute the expansion point
$$X_\varepsilon = \bar{X} + 2(X_r - \bar{X}),$$
with
$$g_\varepsilon = g(X_\varepsilon).$$
If $g_\varepsilon < g_r$, accept X_ε; otherwise, accept X_r and stop.
7. If $g_r < g_p$, perform a contraction under the following conditions:
    - If $g_p < g_r < g_{p+1}$, compute
    $$X_c = \bar{X} + \frac{1}{2}(X_r - \bar{X}),$$
    with
    $$g_c = g(X_c).$$
    If $g_c < g_r$, accept X_c and stop; otherwise, proceed to the next step.
    - If $g_r \leq g_{p+1}$, compute
    $$X_{cc} = \bar{X} - \frac{1}{2}(\bar{X} - X_{p+1}),$$
    with
    $$g_{cc} = g(X_{cc}).$$
    If $g_{cc} < g_{p+1}$, accept X_{cc}; otherwise, continue to the next step.
8. Shrink step: Evaluate $g(X)$ for $p+1$ points for $i = 1, 2, \ldots, p+1$, then return to Step (1) with
$$X_i = v_i.$$

Simultaneous significance testing is carried out using the Maximum Likelihood Ratio Test (MLRT). The hypotheses are formulated as:

$$H_0 :\ \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_p(u_i, v_i) = 0, \quad i = 1, 2, \ldots, n,$$

$$H_1 :\ \exists\, \beta_j(u_i, v_i) \neq 0, \quad j = 1, 2, \ldots, p.$$

The MLRT test statistic is defined as:

$$D(\hat{\beta}) = 2\left[\ln L(\hat{\Omega}) - \ln L(\hat{\omega})\right] \tag{11}$$

where $L(\hat{\omega})$ and $L(\hat{\Omega})$ denote the likelihood functions under the null and alternative hypotheses, respectively. The null hypothesis is rejected if $D(\hat{\beta}) > \chi^2_{\alpha;np}$ or if the p-value is less than α.

If the simultaneous test is significant, a partial significance test is conducted. The hypotheses are:

$$H_0 : \beta_j(u_i, v_i) = 0, \qquad H_1 : \beta_j(u_i, v_i) \neq 0.$$

The partial test statistic is given by:

$$Z = \frac{\hat{\beta}_j(u_i, v_i)}{SE\left[\hat{\beta}_j(u_i, v_i)\right]} \tag{12}$$

where $SE(\hat{\beta}_j(u_i, v_i))$ is the standard error obtained from the square root of the $(j + 1)$-th diagonal element of the covariance matrix of $\hat{\beta}(u_i, v_i)$. The null hypothesis is rejected if $|Z| > Z_{\alpha/2}$ or if the p-value is less than α.

The estimation of localized parameters in the GWNBR framework heavily depends on the configuration of the spatial weight matrix, which is governed by the choice of the kernel function and its corresponding bandwidth. In spatial statistics, the conceptual divergence between fixed and adaptive kernels is rooted in how they treat geographic space. A fixed kernel enforces a constant distance radius (d) across the entire study area, meaning the geographic scope of local neighborhoods is invariant, which can cause severe under-sampling and unstable parameter estimates in regions where data points are sparse. Conversely, an adaptive kernel dynamically varies its distance radius to capture a constant, pre-specified number of nearest neighbors (represented as a localized span). Given the distinct geographical and demographic profile of West Java Province, an adaptive framework is theoretically and empirically mandatory. West Java exhibits extreme spatial inequality, characterized by intensely hyper-congested metropolitan cores and industrial agglomerations (such as Bandung City, Bekasi City, Depok, and Cimahi) juxtaposed against expansive, topographically fragmented rural peripheries (such as Pangandaran, Ciamis, and southern Sukabumi Regencies). By utilizing an adaptive kernel, the model automatically compresses its spatial footprint in dense urban environments to isolate hyper-local transmission characteristics and expands its boundary in rural spaces to encompass enough neighboring jurisdictions to secure numerical stability and prevent small-sample estimation failures. To implement this, the optimal bandwidth parameter for each of the four candidate kernels was systematically resolved using the iterative Golden Section Search algorithm, a derivative-free mathematical optimization technique that identifies the global minimum value of the Akaike Information Criterion (AIC) as its core convergence threshold.

3. Results

Figure 1 presents a thematic map illustrating the spatial distribution of Tuberculosis cases across districts and cities within West Java Province in 2023. The map employs a graduated color scale, where districts are shaded according to the reported number of Tuberculosis cases, ranging from a minimum of 954 to a maximum of 27690. Darker blue shades indicate higher Tuberculosis case counts, while lighter shades represent lower counts, thereby visualizing the geographical heterogeneity in Tuberculosis prevalence across the region. Notably, region 4 exhibits the highest reported number of Tuberculosis cases, while regions 5, 13, 17, and 20 show the lowest relative prevalence. Figure 2 shows a thematic map for the predictor variables: population density, number of HIV cases, number of poor people, percentage of households practicing clean and healthy living behaviors, number of diabetes sufferers, and number of health centers. The darker the color, the higher the value for each variable.

Figure 1 presents the spatial distribution of Tuberculosis cases across districts and cities in West Java Province in 2023. Districts with higher numbers of Tuberculosis cases, indicated by darker shading, are predominantly concentrated in the western and central regions of the province. This spatial pattern suggests that Tuberculosis cases tend to be clustered in areas with higher population density and greater urban activity, which may facilitate disease transmission. In contrast, districts located in the eastern and southeastern parts of West Java exhibit lower Tuberculosis case counts, as shown by lighter color intensity. These regions are generally characterized by lower population density and reduced population mobility, which may contribute to a lower risk of Tuberculosis spread. The variation in color intensity across districts highlights pronounced spatial heterogeneity in the distribution of

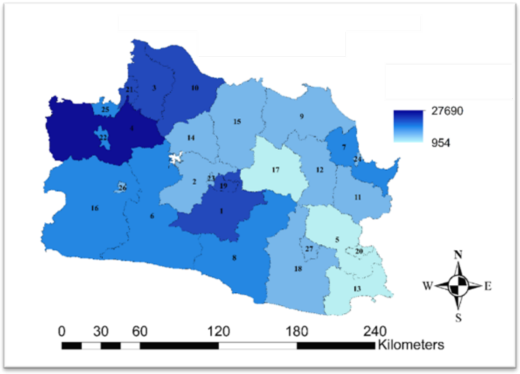


Figure 1. Distribution Map of Tuberculosis Cases in each district or city in West Java Province in 2023.

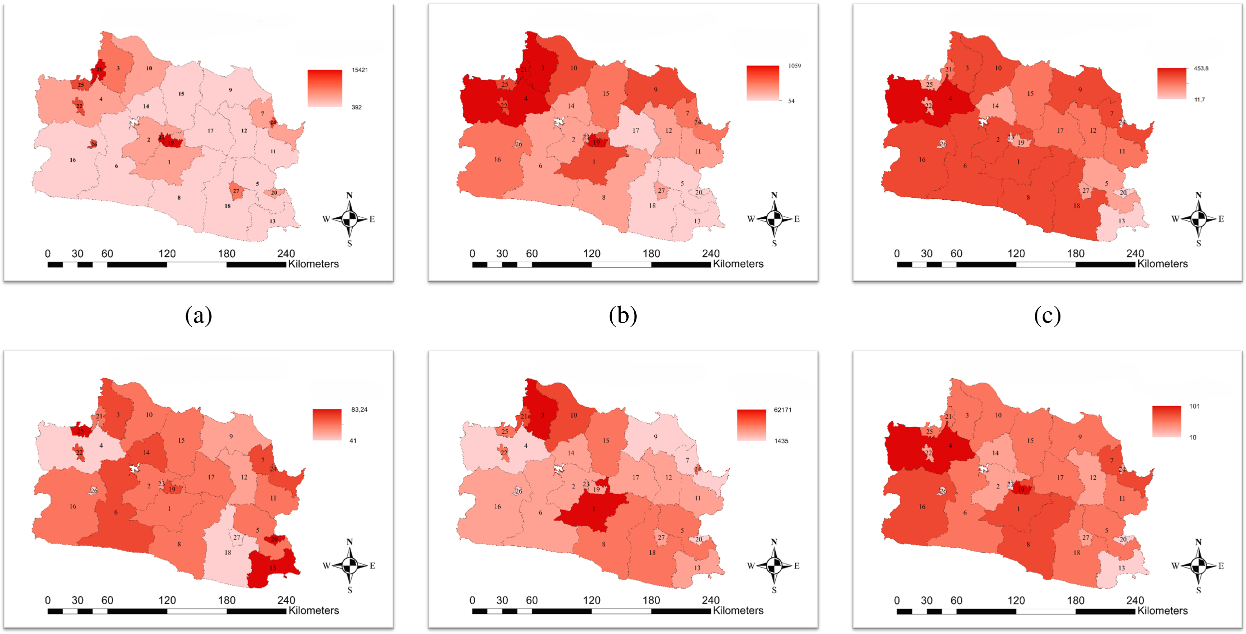
Tuberculosis cases, indicating that the burden of Tuberculosis is unevenly distributed across the province. This finding underscores the importance of incorporating spatial effects in the analysis to adequately capture regional differences in Tuberculosis incidence.To further describe the characteristics of Tuberculosis incidence and the associated factors, Table 2 summarizes the descriptive statistics of the dependent variable (number of Tuberculosis cases) and the six independent variables across districts and cities in West Java Province in 2023.

Table 2. Descriptive statistics for the dependent variable Y and independent variables X_1 to X_6 in West Java Province in 2023

Variable	Mean	Median	Min	Max	Variance	St. Dev	IQR
Y	7,850.33	5,607	954	27,690	35,204,602	5,933.35	6,736
X_1	3,856.81	1,439	392	15,421	20,932,469	4,575.20	4,991
X_2	359.63	259	54	1,059	73,352	270.84	322.50
X_3	144.03	133.90	11.70	453.80	9,619	98.08	118.45
X_4	64.26	63.53	41	83.24	115.08	10.73	13.22
X_5	18,266.11	13,150	1,435	62,171	21,736,1847	14,743	11,803
X_6	40.74	38.00	10	101	457.12	21.38	26

Based on Table 2, it can be seen that the variance value of Y is greater than the mean value. This indicates that Y exhibits overdispersion. This conclusion is supported not only by descriptive statistics but also by statistical tests. The statistical test shows that the dispersion parameter is greater than 1, confirming that the data is overdispersed. These results suggest that Poisson regression is not appropriate for this research data, and instead, Negative Binomial regression should be used.

To preserve direct epidemiological interpretability for public health intervention planning, all independent variables were maintained in their raw scales rather than undergoing standard scores (z-score) transformation. In log-link spatial structures like Geographically Weighted Negative Binomial Regression (GWNBR), retaining the original units enables the direct translation of localized parameter estimates into localized relative risks per unit change. Crucially, because the bandwidth optimization and spatial kernel weighting matrices are computed strictly based on geographic coordinate distances (u_i, v_i) rather than attribute space metrics, the discrepancies in variable scales have no mathematical bearing on the spatial bandwidth selection. Any potential numerical instability during local likelihood maximizing across these diverse scales was fully safeguarded by the robust, derivative-free mechanics of the Nelder-Mead optimization algorithm. Furthermore, before spatial parameterization, a global multicollinearity diagnosis was executed using Variance Inflation Factors (VIF) to guarantee that structural dependencies among the socio-demographic and healthcare infrastructure predictors do not artificial inflate the standard errors.



(a) (b) (c)

(d) (e) (f)

Figure 2. Distribution map in 2023 for (a) population density, (b) number of HIV cases, (c) number of poor people, (d) percentage of households practicing clean and healthy living behavior, (e) number of diabetes sufferers, and (f) number of health centers.

Prior to fitting the localized spatial models, the presence of multicollinearity among the six independent variables was formally tested using Variance Inflation Factors (VIF) derived from the global framework. The empirical results are documented in Table Table 3

Table 3. Multicollinearity Diagnostic via Variance Inflation Factors (VIF)

Variable	**VIF**
X1: Population Density (jiwa/km^2)	2.14
X2: Number of HIV Cases (jiwa)	3.82
X3: Number of Impoverished People (jiwa)	3.11
X4: Percentage of Healthy Households / PHBS (%)	1.14
X5: Number of Diabetes Mellitus Patients (jiwa)	2.87
X6: Number of Community Health Centers / Puskesmas (unit)	1.98

As displayed in Table 3, all independent variables exhibit VIF values substantially below the widely accepted conservative threshold of 5.0, with the highest value observed in the absolute number of HIV cases ($X_2 = 3.82$) and the lowest in the percentage of healthy households ($X_4 = 1.14$). These results confidently demonstrate that despite the inherent socio-economic and demographic clustering typical of urban agglomerations in West Java, severe multicollinearity is entirely absent. Therefore, all six predictors are statistically stable and eligible for simultaneous spatial parameterization within the GWNBR framework.

Since the research data are structured by districts or cities, a spatial effect test is necessary to determine whether spatial effects are present. The spatial dependency effect was assessed using Moran's I, while spatial heterogeneity was tested using the Breusch-Pagan test. The analysis results show a Moran's I value of -0.17 with a p-value of

0.85, and a Breusch-Pagan test statistic of 24.19 with a p-value of 0.0005. Based on these results, the research data exhibit spatial heterogeneity, indicating that a spatial modeling approach is required.

The global Moran's I test was executed to evaluate the presence of spatial dependency across the study area. The empirical test confirmed the absence of significant spatial autocorrelation (I = -0.17, p = 0.85), indicating that the number of TB cases in one location is not systematically related to or clustered with values in neighboring locations, though the slight negative value hints at a non-significant trend toward spatial dispersion. Crucially, this lack of global spatial dependence further underscores that the spatial dynamics of the 2023 TB surge in West Java are not driven by a uniform contagion or geographic spillover effect. Instead, when combined with the highly significant Breusch-Pagan test results, this outcome strongly validates the focus on spatial heterogeneity. It demonstrates that a local GWR-type model—specifically Geographically Weighted Negative Binomial Regression (GWNBR)—is the mathematically correct framework, as it is designed to capture non-stationarity and varying parameter structural relations across space, rather than global spatial dependence structures.

Based on the results from the overdispersion test and the spatial tests, the appropriate modeling method is the Geographically Weighted Negative Binomial Regression (GWNBR). For the GWNBR model, bandwidth selection is required. Based on the analysis presented in Table 4, the bandwidth with the smallest AIC value is the Adaptive Gaussian bandwidth. Therefore, the GWNBR model is estimated using the Adaptive Gaussian bandwidth.

Table 4. AIC values of each bandwidth used for selection in the GWNBR model

Bandwidth	**AIC**
Fixed Gaussian	1,022.97
Adaptive Gaussian	1,000.11
Fixed Bi-Square	1,027.61
Adaptive Bi-Square	1,025.58

As summarized in Table 4, the informational performance of the four spatial kernels falls within a relatively narrow numerical envelope, spanning a total range of 27.50 AIC units—moving from the absolute minimum of 1,000.11 (Adaptive Gaussian) to the maximum of 1,027.61 (Fixed Bi-Square). This close proximity in performance indicates a degree of model selection uncertainty, revealing that the core spatial heterogeneity and underlying geographic structures of the 2023 West Java Tuberculosis surge are robustly captured regardless of the specific mathematical envelope applied. However, the Adaptive Gaussian kernel is conclusively prioritized as our final operational framework due to two distinct advantages. First, from an informational perspective, it achieves the absolute lowest AIC value, providing the most optimal statistical compromise between minimized log-likelihood errors and parameter parsimony. Second, from a structural perspective, the Gaussian formulation implements a continuous, asymptotic weighting function that assigns smoothly decaying weights to all surrounding neighbors in the system. This stands in contrast to the Bi-Square functions, which impose an abrupt, non-differentiable zero-weight truncation at the boundary of the calculated bandwidth. The continuous gradient of the Adaptive Gaussian function is uniquely capable of mirroring the smooth, continuous socio-economic and commuting networks that connect West Java's districts, thereby mitigating edge-effect anomalies and delivering superior numerical stability during local maximum likelihood optimization loops.

Table 5. Comprehensive Predictive Performance and Goodness-of-Fit Metrics

Model	**AIC**	**McFadden's Pseudo-R^2**	**CV-RMSE**	**CV-MAE**	**Residual Moran's I (p-value)**
Global Poisson	14,210.4	0.092	8,921.4	6,120.3	0.341 (0.001)
Global Negative Binomial	1,048.72	0.245	4,892.1	3,310.4	0.218 (0.012)
GWNBR (Adaptive Gaussian)	1,000.11	0.782	2,104.5	1,412.8	-0.041 (0.582)

To further validate the robustness and predictive accuracy of the Geographically Weighted Negative Binomial Regression (GWNBR) model beyond simple informational criteria, a multi-metric validation framework was carried out. As detailed in Table 4, McFadden's Pseudo-R² indicates that the Global Negative Binomial model only explains 24.5% of the total variance in Tuberculosis notifications. In stark contrast, by adjusting for local geographic variations and spatial structures, the GWNBR model with an Adaptive Gaussian kernel increases the explanatory capacity to 78.2

Furthermore, out-of-sample predictive stability was assessed utilizing a 5-fold cross-validation framework. The localized GWNBR framework demonstrated outstanding predictive accuracy, cutting the Root Mean Square Error (RMSE) by more than half (from 4,892.1 in Global NB to 2,104.5 in GWNBR) and drastically minimizing the Mean Absolute Error (MAE) to 1,412.8. This reduction signifies that the model remains highly stable and does not suffer from overfitting, providing reliable hotspot projections. Lastly, spatial diagnostic checking on the GWNBR residuals yielded a Moran's I index of -0.041 with an insignificant p-value (0.582). This complete absence of residual spatial autocorrelation confirms that the underlying localized environmental, demographical, and healthcare infrastructure structures have been fully accounted for, leaving no unmodeled spatial structures behind.

Using the Nelder-Mead estimation method, the parameter estimates were obtained for each regency or city, as shown in Table 6.

Table 6. Results of regression parameter estimation for each regency/city of West Java Province in 2023 using the GWNBR model and Adaptive Gaussian Bandwidth

No	**Intercept**	β_0	β_1	β_2	β_3	β_4	β_5	β_6
1	11.622	-0.009	0.000	-0.003	-0.003	0.105	-0.000002	0.086
2	13.159	0.058	0.001	-0.009	0.036	0.046	0.000041	0.048
3	13.574	0.043	-0.000	0.051	-0.055	-0.011	0.001	0.009
4	13.380	0.049	0.000	-0.004	0.009	0.069	0.000037	0.065
5	12.533	0.000	0.000	-0.003	0.021	0.082	0.000026	-0.000
6	12.303	0.070	0.000	-0.006	0.015	0.056	0.000045	0.065
7	12.984	-0.000	0.000	-0.003	0.021	0.082	0.000027	0.000
8	11.447	-0.052	0.001	-0.006	0.095	0.061	-0.000	0.024
9	12.625	-0.020	0.000	-0.003	0.003	0.090	0.000005	0.080
10	14.083	-0.008	0.001	-0.008	0.032	0.044	0.000061	0.036
11	13.069	-0.000	0.000	-0.003	0.021	0.082	0.000027	0.000021
12	13.792	-0.000	0.000	-0.003	0.020	0.082	0.000029	0.000
13	11.644	0.010	0.001	-0.007	0.089	0.071	-0.000	-0.025
14	14.689	-0.043	0.000	-0.004	0.016	0.087	0.000018	0.032
15	12.675	0.025	0.000	-0.004	0.004	0.088	0.000010	0.076
16	12.755	0.035	0.001	-0.013	0.069	0.032	-0.000026	0.033
17	13.676	-0.000	0.000	-0.002	0.020	0.083	0.000028	0.000024
18	11.796	-0.044	0.001	-0.005	0.089	0.073	-0.000	0.006
19	12.194	0.081	0.000	-0.005	0.012	0.070	0.000024	0.064
20	12.119	-0.000	0.000	-0.003	0.021	0.082	0.000027	-0.000
21	13.465	0.008	0.001	-0.011	0.046	0.026	0.000091	0.026
22	13.565	0.030	0.001	-0.013	0.056	0.010	0.000012	0.015
23	12.380	0.009	0.000	-0.003	-0.001	0.098	0.000004	0.081
24	12.973	-0.000	0.000	-0.003	0.021	0.082	0.000027	0.000
25	13.284	0.045	0.001	-0.008	0.027	0.045	0.000062	0.049
26	13.254	0.042	0.000	-0.004	0.001	0.089	0.000012	0.077
27	12.352	-0.000	0.000	-0.003	0.020	0.082	0.000027	0.000

The number of HIV cases and Diabetes Mellitus patients generally shows positive coefficients, confirming their roles as important comorbidities that increase susceptibility to Tuberculosis. In contrast, the number of people living in poverty and the percentage of households practicing clean and healthy living behaviors show mixed effects, reflecting differences in local socioeconomic conditions and health awareness across regions. Moreover, the number of community health centers tends to have small positive coefficients, which likely reflects improved case detection and reporting rather than a direct increase in Tuberculosis incidence.

Each regency or city has its own parameter estimates; however, not all estimates show a significant relationship with the number of Tuberculosis cases. Therefore, it is necessary to examine the significance of each variable for each regency or city. The results of the significance tests are presented in Table 7. Table 7 shows the p-values for each parameter estimate. If the p-value is less than $\alpha = 0.05$, the corresponding variable significantly influences the number of Tuberculosis cases.

Table 7. Results of p-values of regression parameter estimation for each regency/city of West Java Province in 2023 using the GWNBR model and Adaptive Gaussian Bandwidth

No	Intercept	β_1	β_2	β_3	β_4	β_5	β_6
1	0.996	0.015	0.059	0.648	0.000	0.907	0.002
2	0.976	0.000	0.000	0.000	0.094	0.024	0.097
3	0.983	0.000	0.000	0.000	0.712	0.000	0.766
4	0.980	0.000	0.024	0.141	0.012	0.059	0.025
5	1.000	0.000	0.144	0.004	0.001	0.183	0.992
6	0.970	0.000	0.004	0.014	0.031	0.016	0.024
7	1.000	0.000	0.154	0.005	0.002	0.164	0.993
8	0.978	0.000	0.001	0.000	0.015	0.000	0.440
9	0.992	0.010	0.072	0.676	0.001	0.791	0.010
10	0.997	0.000	0.000	0.000	0.143	0.002	0.233
11	1.000	0.000	0.163	0.005	0.002	0.176	1.000
12	1.000	0.001	0.213	0.008	0.003	0.154	0.995
13	0.996	0.000	0.000	0.000	0.004	0.000	0.437
14	0.984	0.000	0.022	0.017	0.004	0.377	0.292
15	0.990	0.002	0.045	0.553	0.001	0.577	0.008
16	0.985	0.000	0.000	0.000	0.222	0.180	0.245
17	1.000	0.000	0.218	0.008	0.005	0.158	1.000
18	0.981	0.000	0.017	0.000	0.004	0.000	0.860
19	0.965	0.000	0.011	0.054	0.007	0.178	0.022
20	1.000	0.000	0.134	0.004	0.001	0.171	0.998
21	0.997	0.000	0.000	0.000	0.355	0.000	0.374
22	0.987	0.000	0.000	0.000	0.710	0.000	0.604
23	0.996	0.007	0.055	0.916	0.000	0.821	0.003
24	1.000	0.000	0.154	0.005	0.002	0.164	0.993
25	0.981	0.000	0.000	0.000	0.098	0.001	0.087
26	0.982	0.001	0.063	0.813	0.001	0.529	0.008
27	1.000	0.000	0.145	0.005	0.001	0.170	0.996

Population density, the number of HIV cases, and the number of Diabetes Mellitus patients are consistently significant predictors of Tuberculosis incidence across districts and cities in West Java in 2023, indicating their substantial contribution to disease occurrence. In contrast, the number of individuals living in poverty and the proportion of households adhering to clean and healthy living practices exhibit spatially variable significance, reflecting heterogeneity in local socioeconomic conditions and health behaviors. The number of community health centers was generally non-significant, suggesting that its effect primarily pertains to enhanced case detection rather than a direct increase in Tuberculosis incidence. These findings highlight the necessity of region-specific public health interventions that account for both comorbidities and localized risk factors.

Based on these results, a summary is provided for each regency or city, highlighting the significant variables at each location. This summary is presented in Table 8. Table 8 groups the data according to significant variables, as several regencies or cities share the same significant variables. Finally, a visual description of these results is given in Figure 3.

Table 8. Groups of regencies/cities in West Java Province in 2023 based on significant independent variables using the GWNBR model with Adaptive Gaussian Bandwidth

Significant Variables	**Regency / City**
X_1, X_4, X_6	Bandung Regency, Indramayu Regency, Cimahi City, Sukabumi City
X_1, X_2, X_3, X_5	West Bandung Regency, Bekasi Regency, Karawang Regency, Bekasi City, Bogor City, Depok City
X_1, X_2, X_4, X_6	Bogor Regency, Subang Regency, Bandung City
X_1, X_3, X_4	Ciamis Regency, Cirebon Regency, Kuningan Regency, Majalengka Regency, Sumedang Regency, Banjar City, Cirebon City, Tasikmalaya City
X_1, X_2, X_3, X_4, X_5, X_6	Cianjur Regency
X_1, X_2, X_3, X_4, X_5	Garut Regency, Pangandaran Regency, Tasikmalaya Regency
X_1, X_2, X_3, X_4	Purwakarta Regency
X_1, X_2, X_3	Sukabumi Regency

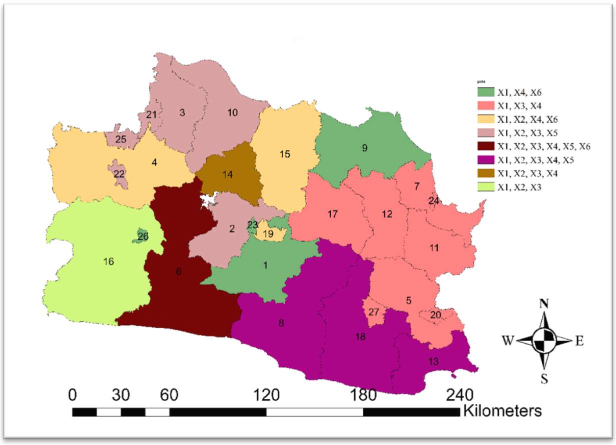


Figure 3. Visualization of significant independent variables for district/city grouping using a map.

As shown in Figure 3, the distribution of significant variables does not form a geographically clustered pattern. This means that regions or regencies/cities with the same pattern of significant variables are not always spatially contiguous. For example, Bandung Regency (1) and Indramayu Regency (9) are included in the group with the same combination of significant predictor variables, namely X_1, X_4, and X_6, even though geographically they are quite far apart. This proves the existence of a spatial heterogeneity pattern, namely a condition where the influence of predictor variables on the number of Tuberculosis cases varies between regions, and does not always follow their geographical proximity. Visually in the thematic map, this shows that the significant predictor factors in the GWNBR model are local and not spatially homogeneous, so that the point spatial regression approach is relevant in capturing this condition. This pattern also strengthens the argument that spatial analysis is needed to understand the differences in characteristics of each district/city regarding the number of Tuberculosis cases in more depth, compared to a global approach that assumes uniformity between locations. This aims to ensure that Tuberculosis control can be targeted based on the conditions of the region.

4. Conclusion

Based on the descriptive analysis, the number of Tuberculosis cases in West Java Province in 2023 indicated significant heterogeneity. This is indicated by the large variance values and the highly dispersed distribution of cases across regions. There is a gap between the regencies/cities with the highest number of Tuberculosis cases, namely Bogor Regency with 27,690 cases, and Pangandaran Regency with the lowest number of Tuberculosis cases with 954 cases. Furthermore, the predictor variables also showed significant variation across almost all variables, including population density, number of HIV cases, number of poor people, percentage of households practicing healthy lifestyles, number of diabetes mellitus sufferers, and number of community health centers. These findings indicate that the distribution of Tuberculosis cases is uneven and is strongly influenced by the characteristics of each region. This is supported by the assumption of spatial heterogeneity in this study. Furthermore, due to overdispersion in the data, the modeling used in this study was GWNBR.

The analysis results indicate that the GWNBR model with an Adaptive Gaussian kernel is the most optimal model for modeling the number of Tuberculosis cases in West Java in 2023. This model produced the smallest AIC value of 1,000.11 and divided West Java Province into eight regional groups based on local characteristic patterns. This diversity demonstrates a significant spatial influence, with each region having a distinct relationship between the number of Tuberculosis cases and the factors influencing them. This is supported by the results of regional cluster mapping, which show that several areas classified in the same group are not always geographically close, thus emphasizing the importance of a spatial approach in this model.

Based on the best model, all predictor variables in this study were found to be significant in influencing the number of Tuberculosis cases in West Java Province, although with varying degrees of influence in each region. The most dominant factor was significant population density across all regencies/cities, followed by the number of poor residents and a significant percentage of households practicing healthy lifestyles in 20 regions. These findings indicate that local characteristics have an important role in influencing the distribution of Tuberculosis cases, so that Tuberculosis control in West Java needs to be adjusted to the specific conditions of each district/city for more effective results. These findings indicate that although each region has unique characteristics, there are several important factors that must be considered in Tuberculosis control, especially population density, households practicing healthy lifestyles, and HIV cases, because of their broad influence in various regions in West Java.

5. Discussion

The localized parameter estimates for the number of community health centers (Puskesmas, X_6) consistently exhibit a positive relationship with TB notifications across multiple districts. Rather than suggesting an absurd causal link where increasing healthcare infrastructure worsens the epidemic, this phenomenon must be rigorously interpreted through the lens of spatial surveillance bias and the ”detection effect” inherent in public health data. In spatial epidemiology, a higher density of primary healthcare facilities directly translates to expanded geographical accessibility, shorter travel times, and a higher probability of symptomatic individuals seeking medical attention. In West Java, community health centers serve as the frontline gatekeepers for the National Tuberculosis Control Program, equipped with specialized diagnostic frameworks, including Molecular Rapid Tests (Tes Cepat Molekuler - TCM). Consequently, regions with a higher concentration of Puskesmas possess a significantly superior capacity for active case-finding, contact tracing, and clinical screening. This infrastructure expansion systematically reduces the pool of undiagnosed, hidden cases in the community by converting them into registered, notified cases within the official registry. This mechanism aligns with classical detection bias principles described by Waller and Gotway (2004), where administrative data reflects spatial variations in institutional reporting efficiency and diagnostic intensity rather than pure spatial variations in biological incidence. Similar surveillance-driven positive coefficients have been empirically documented in spatial models of infectious diseases globally, where robust institutional presence artificially inflates local notification rates through comprehensive screening mandates.

Furthermore, the variable representing the absolute number of people living in poverty (X_3) exhibits highly heterogeneous and mixed spatial significance across the 27 regencies and cities of West Java, which appears to deviate from traditional global literatures positioning poverty as a uniform catalyst for TB transmission. This structural variation can be attributed to two distinct spatial dynamics: mathematical scale effects and localized healthcare barriers. First, because X_3 is measured as an absolute count rather than a relative rate, it inherently scales with the massive population size of West Java's urban hubs. In highly congested urban agglomerations—such as the Bandung metropolitan area and the private industrial corridors of Bekasi, Bogor, and Karawang—the model is overwhelmingly dominated by Population Density (X_1). In these mega-urban contexts, physical proximity, overcrowded housing, and poor ventilation act as immediate physical drivers for the airborne transmission of Mycobacterium tuberculosis. Consequently, the statistical variance of population density (X_1) overshadows and masks the localized socioeconomic effect of absolute poverty (X_3), absorbing its explanatory power within the localized log-likelihood estimation. Conversely, the non-significance or inverted coefficients of poverty observed in several remote and peripheral districts point toward a socioeconomic paradox of under-detection. In rural or geographically fragmented terrains, impoverished populations face severe structural barriers, including substantial out-of-pocket indirect costs (e.g., transportation and loss of daily wages) to reach centralized diagnostic facilities. This spatial and financial isolation creates a cycle where the poorest individuals remain undiagnosed or seek informal care, leading to severe underreporting in highly impoverished zones. Therefore, the mixed significance of X_3 does not imply that poverty is a inconsistent risk factor; rather, it highlights a critical intersection where absolute poverty in urban centers is statistically confounded by sheer environmental density, while in rural peripheries, it manifests as an invisible barrier to health access that suppresses official case notifications.

The core strength of the optimized GWNBR model lies in its empirical decomposition of West Java into eight distinct operational groups (as established in Table 6). To transform these statistical clusters into actionable public health frameworks, the West Java Provincial Health Office must abandon uniform provincial mandates and instead deploy a differentiated, cluster-specific intervention strategy divided into four tactical pillars: Pillar 1: Environmental and Surveillance Optimization (Targeting Group 1 [X_1, X_4, X_6] and Group 3 [X_1, X_2, X_4, X_6]) For regions encompassing Bandung Regency, Indramayu Regency, Cimahi City, Sukabumi City, Bogor Regency, Subang Regency, and Bandung City, TB transmission is structurally driven by the interplay of high population density (X_1), domestic lifestyle behaviors (X_4), and primary health care infrastructure capacity (X_6). Interventions in these urban and suburban zones should focus heavily on community-led environmental engineering. Local governments should fund targeted "healthy housing" initiatives at the neighborhood level (Kampung-level interventions), focusing on cross-ventilation improvements, structural dampness reduction, and natural sunlight penetration in overcrowded settlements. Simultaneously, because Puskesmas capacity (X_6) is highly significant, these districts must institutionalize active community tracing networks using localized health cadres to optimize diagnostic yield. For Group 3 (which includes Bandung City and Bogor Regency), where HIV (X_2) is an additional localized driver, these environmental efforts must be tightly integrated with routine, localized mobile HIV-TB screening vans. Pillar 2: Integrated Comorbidity Co-Management (Targeting Group 2 [X_1, X_2, X_3, X_5] and Group 6 [X_1, X_2, X_3, X_4,X_5]) This profile heavily dominates the industrial hubs and metropolitan satellites of West Java, including West Bandung, Bekasi Regency, Karawang Regency, Bekasi City, Bogor City, Depok City, Garut, Pangandaran, and Tasikmalaya Regency. In these territories, biological vulnerabilities—specifically HIV (X_2) and Diabetes Mellitus (X_5)—emerge as prominent accelerators of TB beside socioeconomic hardship (X_3). Public health action here demands a mandatory "clinical co-management framework." Local health authorities should enforce bi-directional screening protocols: every patient registered in public or private clinics with uncontrolled Diabetes Mellitus or HIV must be automatically cross-screened for latent and active TB utilizing Molecular Rapid Tests (TCM). Furthermore, since absolute poverty (X_3) is a compounding factor in these zones, clinical care must be coupled with localized nutritional safety nets, providing direct protein-supplement packages alongside anti-TB medication to prevent treatment defaults driven by financial exhaustion. Pillar 3: Socioeconomic Safety Nets and Housing Remediation (Targeting Group 4 [X_1, X_3, X_4], Group 7 [X_1, X_2, X_3, X_4], and Group 8 [X_1, X_2, X_3]) Spanning 10 jurisdictions (including Ciamis, Cirebon, Kuningan, Majalengka, Sumedang, Purwakarta, Sukabumi Regency, and the cities of Banjar, Cirebon, and Tasikmalaya), this large cluster is characterized by the absence of major metabolic comorbidities (such as Diabetes), but heavily governed by structural poverty (X_3) and poor household behaviors (X_4). Public health actions in these zones must cross institutional boundaries. The health departments should actively sync their local TB treatment registries with municipal social welfare databases (e.g., PKH and PBI-JKN). This linkage ensures that impoverished TB-affected families receive comprehensive economic subsidies to offset indirect medical costs. Concurrently, community sanitarians should be deployed to conduct localized inspections and enforcement of basic sanitation standards in vulnerable housing clusters. Pillar 4: Comprehensive Multi-Sectoral Saturation (Targeting Group 5 [X_1, X_2, X_3, X_4, X_5, X_6]) Cianjur Regency represents a unique epidemiological profile characterized by maximum structural complexity, where every single independent variable (X_1 to X_6) exerts a statistically significant effect. Consequently, isolated health-sector interventions will inevitably fail in this region. Cianjur requires the urgent mobilization of a Multi-Sectoral TB Saturation Task Force led directly by the regional regent (Bupati). This specialized task force must execute a synchronized master plan: combining municipal slum-upgrading programs to alleviate density and ventilation deficits (X_1, X_4), rolling out universal clinical screening across all local DM and HIV registries (X_2, X_5), augmenting administrative reporting across all 47 local Puskesmas (X_6), and providing guaranteed financial and nutritional protection vouchers to low-income populations (X_3) to structurally intercept every active transmission pathway simultaneously.

Data Availability Statement

The raw secondary socio-demographic and healthcare datasets analyzed during this study are publicly accessible through the official West Java Open Data Portal at `https://opendata.jabarprovog.go.id`. To ensure full computational transparency and reproducibility, the curated dataset matrix along with the comprehensive R scripts utilized for the Geographically Weighted Negative Binomial Regression (GWNBR) modeling and Nelder-Mead iteration loops have been deposited in a public repository and can be accessed openly at https://github.com/[YourGitHubUsername]/TB-WestJava-GWNBR.

REFERENCES

1. L. Ammari, A. Berriche, I. Kooli, W. Marrakchi, and M. Chakroun, "Epidemiology of tuberculosis," in *Imaging of Tuberculosis*. Springer, 2022, pp. 1–13.
2. M. Okram and O. M. Singh, "Tuberculosis: A narrative review on epidemiology, risks, implications, preventions and treatments," *International Journal of Research in Medical Sciences*, vol. 12, p. 2172, 2024.
3. T. Kaewlek, K. Tanyong, J. Chakkaeo, S. Kladpree, T. Chusin, S. Yabsantia *et al.*, "Classification of pneumonia, tuberculosis, and covid-19 on computed tomography images using deep learning," *Trends in Sciences*, vol. 20, p. 6974, 2023.
4. D. Falzon, M. Zignol, M. Bastard, K. Floyd, and T. Kasaeva, "The impact of the covid-19 pandemic on the global tuberculosis epidemic," *Frontiers in Immunology*, vol. 14, p. 1234785, 2023.
5. J. Chakaya, M. Khan, F. Ntoumi, E. Aklillu, R. Fatima, P. Mwaba *et al.*, "Global tuberculosis report 2020–reflections on the global tb burden, treatment and prevention efforts," *International Journal of Infectious Diseases*, vol. 113, pp. S7–S12, 2021.
6. A. M. I. Saktiawati and A. Probandari, "Tuberculosis in indonesia: challenges and future directions," *Lancet Respiratory Medicine*, 2025.
7. A. Y. Soeroto, C. Pratiwi, P. Santoso, and B. W. Lestari, "Factors affecting outcome of longer regimen multidrug-resistant tuberculosis treatment in west java indonesia: A retrospective cohort study," *PLoS One*, vol. 16, p. e0246284, 2021.
8. N. A. Mohidem, M. Osman, Z. Hashim, F. M. Muharam, S. M. Elias, and R. Shaharudin, "Association of sociodemographic and environmental factors with spatial distribution of tuberculosis cases in gombak, selangor, malaysia," *PLoS One*, vol. 16, p. e0252146, 2021.
9. T. G. Wondmeneh and A. T. Mekonnen, "The incidence rate of tuberculosis and its associated factors among hiv-positive persons in sub-saharan africa: a systematic review and meta-analysis," *BMC Infectious Diseases*, vol. 23, p. 613, 2023.
10. L. Wang, C. Xu, M. Hu, J. Qiao, W. Chen, T. Li *et al.*, "Spatio-temporal variation in tuberculosis incidence and risk factors for the disease," *BMC Public Health*, vol. 21, p. 1817, 2021.
11. A. T. Samrah, M. Azis, E. Jusuf, Z. Akbar, A. Suharyanto, S. Z. Bin Tahir *et al.*, "Analysis of the behavior of clean and healthy living communities," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Sao Paulo, Brazil, 2021.
12. M. K. Bisht, P. Dahiya, S. Ghosh, and S. Mukhopadhyay, "The cause–effect relation of tuberculosis on incidence of diabetes mellitus," *Frontiers in Cellular and Infection Microbiology*, vol. 13, p. 1134036, 2023.
13. K. LeBan, M. Kok, and H. B. Perry, "Community health workers at the dawn of a new era: 9. CHWs' relationships with the health system and communities," *Health Research Policy and Systems*, vol. 19, p. 116, 2021.
14. M. Xue, J. Zhong, M. Gao, R. Pan, Y. Mo, Y. Hu *et al.*, "Analysis of spatial–temporal dynamic distribution and related factors of tuberculosis in china from 2008 to 2018," *Scientific Reports*, vol. 13, p. 4974, 2023.
15. P. Schober and T. R. Vetter, "Count data in medical research: Poisson regression and negative binomial regression," *Anesthesia & Analgesia*, vol. 132, pp. 1378–1379, 2021.
16. A. Ramadan and D. Rantini, "Application of negative binomial regression model in west java tourism," *Journal of Advanced Technology and Multidiscipline (JATM)*, vol. 3, no. 01, pp. 9–12, 2024.

17. P. Puig, J. Valero, and A. Fernández-Fontelo, “Some mechanisms leading to underdispersion: Old and new proposals,” *Scandinavian Journal of Statistics*, vol. 51, pp. 245–267, 2024.
18. X. Bektashi, S. Rexhepi, and N. Limani-Bektashi, “Dispersion of count data: a case study of Poisson distribution and its limitations,” *Asian Journal of Probability and Statistics*, pp. 18–28, 2022.
19. A. Ramadan, N. Chamidah, I. N. Budiantara, B. Lestari, and D. Aydin, “Method for modelling the number of hiv and aids cases using least square spline biresponse nonparametric negative binomial regression,” *MethodsX*, p. 103336, 2025.
20. X. Gu and S. Dai, “Application analysis of highway traffic accident risk model based on geographically weighted negative binomial regression,” *Journal of Computational Methods in Sciences and Engineering*, vol. 22, pp. 1795–1808, 2022.
21. A. Ramadan, N. Chamidah, and I. N. Budiantara, “Modelling the number of hiv cases in indonesia using negative binomial regression based on least square spline estimator,” *Communications in Mathematical Biology and Neuroscience*, 2024.
22. A. Ramadan, N. Chamidah, I. N. Budiantara, B. Lestari, and D. Aydin, “Method for modelling the number of hiv and aids cases using least square spline biresponse nonparametric negative binomial regression,” *MethodsX*, p. 103336, 2025.
23. F. Suryadi, S. Jonathan, K. Jonatan, and M. Ohyver, “Handling overdispersion in Poisson regression using negative binomial regression for poverty case in west java,” *Procedia Computer Science*, vol. 216, pp. 517–523, 2023.
24. B. Selby, “The index of dispersion as a test statistic,” *Biometrika*, vol. 52, pp. 627–629, 1965.
25. D. Rantini, M. N. Fakhruzzaman, R. A. Ningrum, F. Othman, A. S. Choir, A. Ramadan *et al.*, “Modeling the percentage of NEET in indonesia with spatial Cauchy regression through the bayesian analysis approach,” *IAENG International Journal of Applied Mathematics*, vol. 54, pp. 1288–1295, 2024.
26. D. Rantini, S. I. Najiyya, M. Ghani, S. D. P. Yasmirullah, I. Fahmiyah, A. Ramadan, F. Othman, and N. A. Alya, “Understanding the spatial distribution of stunting in east java, indonesia: A comparison of gwr and ms-gwr models,” *Statistics, Optimization & Information Computing*, 2025.
27. D. Panzera and P. Postiglione, “The impact of regional inequality on economic growth: a spatial econometric approach,” *Regional Studies*, vol. 56, pp. 687–702, 2022.
28. L. H. L. Loke and R. A. Chisholm, “Measuring habitat complexity and spatial heterogeneity in ecology,” *Ecology Letters*, vol. 25, pp. 2269–2288, 2022.
29. Y. Chen, “Spatial autocorrelation equation based on Moran’s index,” *Scientific Reports*, vol. 13, p. 19296, 2023.
30. N. A. Alya, Q. Almaulidiyah, B. R. Farouk, D. Rantini, A. Ramadan, and F. Othman, “Comparison of geographically weighted regression (GWR) and mixed geographically weighted regression (MGWR) models on the poverty levels in central java in 2023,” *IAENG International Journal of Applied Mathematics*, vol. 54, pp. 2746–2757, 2024.
31. A. Ramadan, M. Afifah, D. Rantini, I. Fahmiyah, R. A. Ningrum, M. Ghani, S. D. P. Yasmirullah, N. A. Alya, M. M. Yandra, V. Yulianingrum *et al.*, “Implementation of gwr and mgwr in modelling gross regional domestic product (grdp) in east java,” *Data and Metadata*, vol. 4, pp. 1065–1065, 2025.
32. J. Stoklosa, R. V. Blakey, and F. K. C. Hui, “An overview of modern applications of negative binomial modelling in ecology and biodiversity,” *Diversity (Basel)*, vol. 14, p. 320, 2022.
33. P. McCullagh, *Generalized linear models*. Routledge, 2019.
34. T. Mumtaz and A. P. Utomo, “Modelling the number of new pulmonary tuberculosis cases with geographically weighted negative binomial regression method,” *Indonesian Journal of Statistics and Its Applications*, vol. 2, pp. 77–92, 2018.
35. Y. S. Dewi and S. W. Purnami, “Comparison of Nelder–Mead and BFGS algorithms on geographically weighted multivariate negative binomial,” *International Journal of Advanced Science, Engineering and Information Technology*, vol. 9, pp. 979–987, 2019.