Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:131
https://doi.org/10.28919/cmbn/8783
ISSN: 2052-2541

REGIONAL CLASSIFICATION BASED ON MATERNAL MORTALITY RATE USING A ROBUST SEMIPARAMETRIC GEOGRAPHICALLY WEIGHTED POISSON REGRESSION MODEL

FITRIAYU, ANNA ISLAMIYATI*, ERNA TRI HERDIANI

Department of Statistics, Faculty of Mathematical and Natural Sciences, Hasanuddin University, Makassar, 90245, Indonesia

Abstract: The Semiparametric Geographically Weighted Poisson Regression (SGWPR) model advances the GWPR model, combining local and global parameters relative to location. Outliers are sometimes encountered when analyzing data using the GWPR model. These outliers can be identified as they differ significantly from other sample points. Outliers can impact the estimation results, leading to biased parameter estimates. One approach to addressing outliers is the robust M method. This study aims to classify regions based on the parameter estimates of the robust SGWPR model applied to maternal mortality rate data in East Java Province using Tukey's Bisquare weighting. The outcome of this research is the classification of regions based on significant factors influencing maternal mortality rates in East Java Province in 2021.

Keywords: SGWPR; robust M; Tukey bisquare; maternal mortality rate.

2020 AMS Subject Classification: 62J12.

---

*Corresponding author
E-mail address: annaislamiyati701@gmail.com
Received July 26, 2024

1. INTRODUCTION

Spatial data is information linked to specific locations or places on Earth [1]. This type of data encompasses details about coordinates, geographic attributes, the geometry of objects (such as points, lines, or polygons), and additional attributes or information related to these objects [2]. Spatial decision-making is based on the analysis of field data [3]. Over time, spatial data has become an essential tool for development planning and resource management, considering each region's conditions. Due to its dependency on geographic location, spatial data exhibits spatial heterogeneity, leading to variations in regression coefficients across different locations. Consequently, the same predictor variable may yield different responses at various observation sites [4]. Therefore, to transform spatial data into valuable information, regression models that incorporate the effects of spatial heterogeneity are necessary.

According to Hailegebreal's research, Geographically Weighted Regression (GWR) is a statistical model developed to address spatial heterogeneity and can be used to assess predictor variables [5]. Fotheringham describes GWR as a linear regression model that provides parameter estimates for each location or observation point, thus allowing for different interpretations at each location [6]. Additionally, the GWR model is based on the framework of a simple regression model [7]. However, for discrete data in the response variable, Nakaya recommends using Geographically Weighted Poisson Regression (GWPR) [8].

Some regression coefficients in the GWPR model do not vary spatially, while others differ across regions [9]. Therefore, the GWPR model has been advanced into the Semiparametric Geographically Weighted Poisson Regression (SGWPR) model. This model incorporates the spatial aspect and can address issues of spatial correlation by combining local Poisson regression with location-specific parameters. According to Nakaya, the SGWPR model is obtained by integrating location-varying parameters as local variables with constant parameters for each location as global variables [9]. In practice, outliers are sometimes found in the SGWPR model [10]. These outliers can lead to inconsistent parameter estimates, indicated by large standard errors when using the least squares method [11]. To address outliers, robust estimators have been developed, such as M-estimation (Maximum Likelihood Type) [12], LTS estimation (Least Trimmed Squares) [13], S estimation [14], LD method (Likelihood Displacement) [15], and MM estimation [16]. However, these studies have not considered the potential spatial factors in the data, often assuming that most detected outliers are in the predictor variables [17]. Therefore, this study will estimate data containing outliers using the Robust M method.

Maternal mortality rate is a critical indicator of societal well-being in a country [18]. Addressing maternal mortality is a significant challenge in improving human development [19]. The issue is not only caused by medical factors but also by non-medical factors such as socioeconomic conditions, transportation, education, and the environment. These factors vary between regions, so addressing maternal mortality requires consideration of inter-regional relationships or spatial effects [20]. The approach to solving maternal mortality issues cannot be generalized across regions because influencing factors may differ. Several studies on maternal mortality have used the GWPR model to identify and categorize the factors affecting maternal mortality rates. One such study by Martafiyah utilized the GWPR model to identify the factors influencing maternal mortality in each region, but it found some regression coefficients did not vary spatially and identified outliers in the data [20].

Therefore, this study aims to understand the interrelationship of factors affecting maternal mortality in East Java Province by classifying regions based on these factors using the SGWPR model with Robust M estimation. In this research, parameter estimation of the Robust SGWPR model is conducted at each observation location [21], using Tukey's Bisquare weighting, which considers geographic elements or location as weights in estimating model parameters.

2. PRELIMINARIES

2.1 Robust Semiparametric Geographically Weighted Poisson Regression (SGWPR) Model

The SGWPR model is an extension of the GWPR model that combines local and constant (or global) parameters concerning location. In the SGWPR model, the response variable (y) is estimated using the predictor variable (x), with each regression coefficient $\left(\beta(u_i, v_i)\right)$

depending on the geographic location and a constant β. Geographical locations are denoted by (u_i, v_i), representing the coordinates of the i-th location (coordinates on a map). Thus, the SGWPR model can be expressed as follows [9]:

$$y_i \sim \text{ Poisson } (\mu_i) \text{ with}$$

$$\mu_i = \exp\left(\beta_0(u_i, v_i) + \sum_{j=1}^{k}(\beta_j x_{ij}) + \sum_{j=k+1}^{l}(\beta_j(u_i, v_i)x_{ij}) + \varepsilon_i\right)$$

The SGWPR model can be written as:

$$y_i = \exp\left(\beta_0(u_i, v_i) + \sum_{j=1}^{k}(\beta_j x_{ij}) + \sum_{j=k+1}^{l}(\beta_j(u_i, v_i)x_{ij}) + \varepsilon_i\right) \tag{1}$$

where y_i represents the observed value of the i-th response variable, x_{ij} denotes the observed value of the j-th predictor variable at the location (u_i, v_i), $\beta_0(u_i, v_i)$ is the intercept of the regression model, $\beta_j(u_i, v_i)$ is the regression coefficient for the j-th predictor variable at each location (u_i, v_i), (u_i, v_i) indicates the latitude and longitude coordinates of the j-th point at a geographic location, β_j is the global regression coefficient, and x_{ij} is the observed value of the global predictor variable. Equation (1) can be linearized using the natural logarithm, resulting in the following equation:

$$\ln(y_i) = \beta_0(u_i, v_i) + \sum_{j=1}^{k}(\beta_j x_{ij}) + \sum_{j=k+1}^{l}(\beta_j(u_i, v_i)x_{ij}) + \varepsilon_i \tag{2}$$

Using a matrix approach, Equation (2) can be expressed in the following form:

$$\begin{aligned}\ln(y_i) = \beta_0(u_i, v_i) + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \beta_{k+1}(u_i, v_i)x_{i,k+1} \\ +\beta_{k+2}(u_i, v_i)x_{i,k+2} + \cdots + \beta_l(u_i, v_i)x_{i,l} + \varepsilon_i, \quad i = 1,2,\dots,n\end{aligned} \tag{3}$$

Equation (3) can be changed to:

$$\ln(y_i) = \boldsymbol{X}_g\beta_g + \boldsymbol{X}_m\beta_m(u_i, v_i) + \varepsilon_i \tag{4}$$

Where $\boldsymbol{X}_g$ represents the observed values of the global predictor variable, β_g denotes the global regression coefficient, $\boldsymbol{X}_m$ indicates the observed values of the predictor variable at the location (u_i, v_i) and $\beta_m(u_i, v_i)$ is the regression coefficient for the predictor variable at each location (u_i, v_i).

The SGWPR model in Equation (4) comprises global and local parameters. To facilitate the estimation process, these parameters are estimated separately. The global parameter of the SGWPR model is as follows:

$$\ln(y_i) = \boldsymbol{X}_g\beta_g + \varepsilon_{gi}$$
$$\varepsilon_{gi} = \ln(y_i) - \boldsymbol{X}_g\beta_g \tag{5}$$

While the local parameters of the SGWPR model are as follows:

$$\ln(y_i) = \boldsymbol{X}_m\beta_m(u_i, v_i) + \varepsilon_{li}$$
$$\varepsilon_{li} = \ln(y_i) - \boldsymbol{X}_m\beta_m(u_i, v_i) \tag{6}$$

In this study, the SGWPR model is assumed to contain outliers. For the i-th data point and n observations that include outliers, the global parameter of the SGWPR model containing outliers is assumed to be as follows:

$$\rho(\ln(y_i)) = \rho\boldsymbol{X}_g\beta_g + \rho(\varepsilon_{gi})$$

$$\rho(\varepsilon_{gi}) = \rho(\ln(y_i)) - \rho\boldsymbol{X}_g\beta_g$$

The SGWPR model containing outliers is estimated using the Robust M method to obtain the global parameter estimates. Thus, the forecast of the worldwide regression model with outliers by minimizing the objective function (minimizing the residual ρ), is given by the following equation:

$$SSE = \ln {y_i}^T \rho \ln y_i - 2{\boldsymbol{X}_g}^T{\beta_g}^T\rho \ln y_i + {\boldsymbol{X}_g}^T{\beta_g}^T\rho\boldsymbol{X}_g\beta_g \tag{7}$$

To minimize Equation (7), partial differentiation with respect to ${\boldsymbol{\beta}_g}^T$ is performed, and the resulting equation is set to zero. This yields the estimator β_g as follows:

$$\hat{\beta}_{g(OLS)} = \left({\boldsymbol{X}_g}^T\rho\boldsymbol{X}_g\right)^{-1}{\boldsymbol{X}_g}^T\rho \ln y_i \tag{8}$$

In Equation (8), the parameter β is obtained through the Ordinary Least Squares (OLS) process. Therefore, the residuals in Equation (6) can be determined as follows:

$$\varepsilon_{gi} = \ln(y_i) - \boldsymbol{X}_g\hat{\beta}_{g(OLS)}$$

In Equation (8), ρ is a parameter that includes outliers. This parameter can be determined by assuming $\rho = \psi$ as an influence function, thereby transforming Equation (8) into:

$$\hat{\beta}_g = \left(\boldsymbol{X}_g{}^T \psi \boldsymbol{X}_g\right)^{-1} \boldsymbol{X}_g{}^T \psi \ln y_i \tag{9}$$

According to Draper, the influence function of the weighting function is expressed as follows [22]:

$$W_i = W\left(\varepsilon_{gi}{}^*\right) = \frac{\psi\left(\varepsilon_{gi}{}^*\right)}{\varepsilon_{gi}{}^*} \tag{10}$$

With $\varepsilon_{gi}{}^*$ representing the residual standardized against the biased standard deviation estimate $(\hat{\sigma})$ of ε_{gi}, the following is obtained:

$$\varepsilon_{gi}{}^* = \frac{\varepsilon_{gi}}{\hat{\sigma}} \tag{11}$$

To obtain the value of $\varepsilon_{gi}{}^*$, the standard deviation of the residuals $\hat{\sigma}$ must first be calculated. This leads to Equation (12) as follows:

$$\varepsilon_{gi}{}^* = \frac{\ln(y_i) - \boldsymbol{X}_g \hat{\beta}_{g(OLS)}}{\frac{MAD(x)}{0{,}6745}} \tag{12}$$

From the weighting process in Equation (10), unbiased estimates are expected to be obtained since the influence function has been standardized. Thus, Equation (9) can be transformed into:

$$\hat{\beta}_g = \left(\boldsymbol{X}_g{}^T \boldsymbol{W_i} \boldsymbol{X}_g\right)^{-1} \boldsymbol{X}_i{}^T \boldsymbol{W_i} \ln y_i$$

Where $\boldsymbol{W_i}$ is an $n \times n$ weighting matrix with diagonal elements containing weights $\boldsymbol{W_1}, \boldsymbol{W}_2, \dots, \boldsymbol{W}_n$. In this study, the weighting function used is the Tukey Bisquare weighting function. This weighting function is expressed as follows:

$$W_i = \begin{cases} \left[1 - \left(\frac{\varepsilon_i{}^*}{c}\right)^2\right]^2, & |u_i| < c \\ 0, & |u_i| \geq c \end{cases}$$

with $c = 4.685$ [23].

In each iteration, $\boldsymbol{W_i}$ will change, resulting in estimates $\hat{\beta}_g{}^0, \hat{\beta}_g{}^1, \dots, \hat{\beta}_g{}^m$. For a parameter with m being the number of parameters to be estimated, the estimator $\hat{\beta}_g{}^0$ is given by:

$$\hat{\beta}_g{}^0 = \left(\boldsymbol{X}_g{}^T \boldsymbol{W}_i^0 \boldsymbol{X}_g\right)^{-1} \boldsymbol{X}_i{}^T \boldsymbol{W}_i^0 \ln y_i$$

where $\boldsymbol{W}_i^0$ is the initial $n \times n$ weighting matrix containing weights $\boldsymbol{W}_1^0, \boldsymbol{W}_2^0, \dots, \boldsymbol{W}_n^0$. Therefore, the subsequent estimators can be written as follows:

$$\hat{\beta}_g{}^1 = \left(\boldsymbol{X}_g{}^T \boldsymbol{W}_i^0 \boldsymbol{X}_g\right)^{-1} \boldsymbol{X}_i{}^T \boldsymbol{W}_i^0 \ln y_i$$

Then, recalculating the weights of $\boldsymbol{W}_i{}^1$ using $\hat{\beta}_g{}^1$ results in:

$$\hat{\beta}_g{}^2 = \left(\boldsymbol{X}_g{}^T \boldsymbol{W}_i^1 \boldsymbol{X}_g\right)^{-1} \boldsymbol{X}_i{}^T \boldsymbol{W}_i^1 \ln y_i$$

For parameters up to m (the number of parameters to be estimated), the weights $\boldsymbol{W}_i$ can subsequently be expressed as:

$$\boldsymbol{W}_i{}^{m-1} = \frac{\psi\left(\dfrac{\ln y_i - \boldsymbol{X}_g \hat{\beta}_{g(OLS)}{}^{m-1}}{\hat{\sigma}}\right)}{\dfrac{\ln y_i - \boldsymbol{X}_g \hat{\beta}_{g(OLS)}{}^{m-1}}{\hat{\sigma}}} \tag{13}$$

From Equation (13) we get $\hat{\boldsymbol{\beta}}_g{}^m$ as follows

$$\hat{\beta}_g{}^m = \left(\boldsymbol{X}_g{}^T \boldsymbol{W}_i{}^{m-1} \boldsymbol{X}_g\right)^{-1} \boldsymbol{X}_i{}^T \boldsymbol{W}_i{}^{m-1} \ln y_i$$

For $\boldsymbol{W}_i{}^m$, the given weights, the estimator can be obtained as follows:

$$\hat{\beta}_g{}^{m+1} = \left(\boldsymbol{X}_g{}^T \boldsymbol{W}_i{}^m \boldsymbol{X}_g\right)^{-1} \boldsymbol{X}_i{}^T \boldsymbol{W}_i{}^m \ln y_i \tag{14}$$

This process will continue iterating until a convergent estimator is obtained, that is when the difference between $\hat{\beta}_g{}^m$ and $\hat{\beta}_g{}^{m+1}$ approaches zero, where m represents the number of iterations. A higher value of m indicates that the estimator is approaching convergence.

Once the global regression model parameters are estimated, the next step is to determine the local parameters of the SGWPR model that contain outliers. The local parameters that include outliers are as follows:

$$\rho(\ln y_i) = \rho \boldsymbol{X}_m \beta_m(u_i, v_i) + \rho\left(\varepsilon_{gi}\right)$$

$$\rho(\varepsilon_i) = \rho(\ln y_i) - \rho \boldsymbol{X}_m \beta_m(u_i, v_i)$$

To estimate the local parameters of the SGWPR model containing outliers, these parameters are calculated using the Robust M method, following a similar procedure for estimating the global parameters. This results in the following estimator for the local parameters:

$$\hat{\beta}_m(u_i, v_i)^{m+1} = \left(\boldsymbol{X}_i{}^T \boldsymbol{W}_i{}^{*m} \boldsymbol{X}_l\right)^{m-1} \boldsymbol{X}_i{}^T \boldsymbol{W}_i{}^{*m} \ln y_i \tag{15}$$

From the two equations, Equation (14) and Equation (15), the SGWPR model containing outliers is obtained as follows:

$$(\ln y_i) = \boldsymbol{X}_g \hat{\beta}_g^{\,m+1} + \boldsymbol{X}_i \hat{\beta}_g (u_i, v_i)^{m+1}$$

$$y_i = \exp\left(\boldsymbol{X}_g \hat{\beta}_g^{\,m+1} + \boldsymbol{X}_i \hat{\beta}_g (u_i, v_i)^{m+1}\right)$$

2.2 Hypothesis Testing of SGWPR Model Parameters

There are two types of tests for the significance of parameters in the SGWPR model: simultaneous testing and partial testing [7]. Simultaneous testing assesses whether the predictor variables collectively influence the response variable [12]. This overall test comprises two components: testing the global parameters and then testing the local variable parameters. The hypotheses for the global and local variable parameters in the SGWPR model are as follows:

$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_k(u_i, v_i) = 0$

H_1: there is at least one $\beta_j(u_i, v_i) \neq 0, \quad j = 0,1,2, \dots, k$

The test statistic used is the likelihood ratio test, which is expressed as follows:

$$G = -2\, ln \left[\frac{\ell(\hat{\Omega}_{sgwpr})}{\ell(\hat{\omega}_{sgwpr})}\right] = -2\left(\ell(\hat{\Omega}_{sgwpr}) - \ell(\hat{\omega}_{sgwpr})\right)$$

where $\hat{\Omega}_{sgwpr} = \{\hat{\beta}_0(u_i, v_i), \hat{\beta}_1(u_i, v_i), \hat{\beta}_2(u_i, v_i), \dots, \hat{\beta}_k(u_i, v_i)\}, \quad i = 1,2,3, \dots, k$ represents the log-likelihood for the model that includes all predictor variables and $\hat{\omega}_{sgwpr} = \{\hat{\beta}_0(u_i, v_i) \quad i = 1,2,3, \dots, k\}$ denotes the log-likelihood for the model without predictor variables. For the simultaneous (overall) significance test, the criteria are as follows: at a 5% confidence level ($\alpha = 5\%$), reject H_0 if $G > \chi^2_{df;\alpha}$ [24].

Next, the partial significance testing of the SGWPR model parameters assesses whether each predictor variable has an individual effect on the response variable. This partial testing consists of two components: the global and local variable parameters. The hypotheses for the partial testing of the global variable parameters in the SGWPR model are as follows:

$H_0 \; : \; \beta_j = 0$

H_1 : there is at least one $\beta_j \neq 0$

For each $j = q + 1, q + 2, \dots, k$

At a confidence level of α, H_0 should be rejected if the absolute value of $|t_{g\_stat}|$ exceeds the critical value $t_{\alpha/2\,;n-(p+1)}$ [10].

The subsequent hypothesis test seeks to determine which local variables significantly affect the response variable within the SGWPR model. The hypotheses considered are as follows:

$H_0: \beta_j(u_i, v_i) = 0$

$H_1: \beta_j(u_i, v_i) \neq 0$

for each $j = q + 1, q + 2, \dots, k$

To assess the significance of local variable parameters in the SGWPR model, the criterion is to reject H_0 at a confidence level of α if $|t_{stat}|$ exceeds $Z_{\alpha/2}\,; n - (p + 1)$. This implies that the influence of the j-th predictor variable on the response variable (y) at location (u_i, v_i) is considered significant within the model [10].

2.3 Research Method

This research uses secondary data, specifically the Maternal Mortality Rate (MMR) for East Java from 2021, which was obtained from the official website of the East Java Provincial Health Office. The data was derived from the 2022 Health Office profile, including information from 2021. The study includes 38 observation units across East Java, comprising 29 districts and 9 cities [25]. The analysis features one response variable, the maternal mortality rate, and five predictor variables related to factors affecting this rate [26]. Details of the research variables are outlined in Table 1 below.

Table 1. Research Variable

Research Variable	Code	Variable Name
Response Variable	(y)	Maternal mortality rate
Predictor Variables	(x_1)	Health workers handle complicated deliveries
	(x_2)	Health services for pregnant women (K1) and (K4)
	(x_3)	Pregnant women receive Fe1 and Fe3 tablets
	(x_4)	Married women under 17 years old
	(x_5)	Postpartum mothers obtain vitamin A

The data analysis process in this study includes several steps. Descriptive analysis is performed to provide a preliminary understanding of the maternal mortality rate. Subsequently, the data is analyzed using the Robust SGWPR model. Following this, parameter estimation for the Robust SGWPR model is carried out. After obtaining the estimates, significance tests are conducted collectively and individually for global and local variables. The results of these significance tests are then used to conclude regional clustering based on maternal mortality rates using the robust semiparametric geographically weighted Poisson regression model.

3. MAIN RESULTS

3.1 Deskriptif Data

Descriptive statistics summarize and illustrate the characteristics of the data, providing initial insights that prepare the data for more in-depth analysis. The table below displays the results of the descriptive statistical analysis.

Table 2. Descriptive Statistical Analysis of Data

Variable	Minimum	Maximum	Mean	Standard Deviation	Variance
y	1.00	49.00	23.05	13.08	171.24
x_1	8.17	53.26	24.86	11.76	138.36
x_2	69.78	100.00	87.79	7.14	51.01
x_3	65.60	134.10	93.25	17.02	289.91
x_4	0.00	6.68	0.86	1.61	2.59
x_5	67.60	98.20	83.80	7.10	50.44

Table 2 shows that the average Maternal Mortality Rate (y) in East Java for 2021 is 16.89 per 100,000 live births, with the highest recorded value at 49.00 and the lowest at 1.00. The average percentage of deliveries with complications managed by healthcare professionals (x_1) is 24.86%, ranging from a maximum of 53.26% to a minimum of 8.17%. The average percentage of pregnant women receiving antenatal care services (K1) and (K4) (x_2) stands at 87.79%, with values ranging from a maximum of 100% to a minimum of 69.78%. For the percentage of pregnant women receiving Fe1 and Fe3 tablets (x_3), the average is 93.25%, with a range between a maximum of 134.10% and a minimum of 65.60%. The average percentage of married women under the age of 17 (x_4) is 0.87%, with a maximum of 6.68% and a minimum of 0%. Additionally, the average percentage of postpartum mothers receiving vitamin A (x_5) is 83.80%, ranging from 98.20% to a minimum of 67.60%.

2.4 Estimating the Parameters of the Robust SGWPR Model on Maternal Mortality Data

To estimate the parameters of the Robust SGWPR model, the initial step involves identifying outliers using the DfFITS (Difference in Fits) or Standardized DfFITS methods. The following presents the results of the outlier identification:

Table 3. DfFITS Value

Data	$DfFITS$	$\|DfFITS\|$	Inf.	Data	$DfFITS$	$\|DfFITS\|$	Inf.
1	-1.546	1.546	*Outlier*	20	0.121	0.121	No
2	-0.726	0.726	*Outlier*	21	-0.086	0.086	No
3	-1.150	1.150	*Outlier*	22	-0.246	0.246	No
4	0.122	0.122	No	23	0.612	0.612	No
5	-0.005	0.005	No	24	0.183	0.183	No
6	0.356	0.356	No	25	0.798	0.798	*Outlier*
7	3.665	3.665	*Outlier*	26	-0.112	0.112	No
8	0.271	0.271	No	27	-0.131	0.131	No
9	0.359	0.359	No	28	-0.255	0.255	No
10	-0.461	0.461	No	29	0.181	0.181	No
11	1.375	1.375	*Outlier*	30	-0.258	0.258	No
12	-0.623	0.623	No	31	-0.279	0.279	No
13	-1.174	1.174	*Outlier*	32	-0.171	0.171	No
14	-2.971	2.971	*Outlier*	33	0.583	0.583	No
15	-0.356	0.356	No	34	-1.137	1.137	*Outlier*
16	0.827	0.827	*Outlier*	35	-1.127	1.127	*Outlier*
17	-0.221	0.221	No	36	-0.823	0.823	*Outlier*
18	2.311	2.311	*Outlier*	37	-0.283	0.283	No
19	0.152	0.152	No	38	0.106	0.106	No

According to Table 3, data points with DfFITS values exceeding 0.72 include points 1, 2, 3, 7, 11, 13, 14, 16, 18, 25, 34, 35, and 36, considered outliers with significant influence on the regression model. Following the outlier identification, the next step involves assessing which predictor variables have global and local impacts by applying the GWPR model variability test. In GWPR modeling, the initial procedure is to determine the optimal bandwidth (h) for each observation location through the cross-validation (CV) method.

Table 4. Provides the Optimal Bandwidth Values for Each District and City in East Java

Districts/Cities	Bandwidth	Districts/Cities	Bandwidth
Bangkalan	2.543	Lumajang	0.071
Banyuwangi	0.754	Madiun	1.745
Blitar	0.637	Magetan	0.263
Bojonegoro	0.278	Malang	0.081
Bondowoso	0.235	Mojokerto	0.976
Gresik	0.406	Nganjuk	0.550
Jember	1.522	Ngawi	0.566
Jombang	1.101	Pacitan	1.011
Kediri	0.948	Pamekasan	1.738
Batu City	1.781	Pasuruan	0.864
Blitar City	0.537	Ponorogo	0.185
Kediri City	0.915	Probolinggo	0.144
Madiun City	0.291	Sampang	1.224
Malang City	0.091	Sidoarjo	0.413
Mojokerto City	0.812	Situbondo	0.087
Pasuruan City	1.960	Sumenep	0.614
Probolinggo City	1.501	Trenggalek	1.746
Surabaya City	0.623	Tuban	1.356
Lamongan	2.205	Tulungagung	0.709

Once the optimal bandwidth values for each district and city in East Java have been determined, the subsequent step involves calculating the Euclidean distances (d_{ij}) and the weight matrix (w_{ij}) for each observation location (u_i, v_i). The Euclidean distances and weights are computed for each location using an adaptive Gaussian kernel weighting function. The weight matrix is then utilized to estimate the GWPR model parameters by incorporating these weights into the calculations. The parameter estimates $\hat{\beta}_j(u_1, v_1)$ for the GWPR model are derived using Weighted Least Squares (WLS), providing different parameter values for each observation site.

After establishing the GWPR model, the next phase involves testing the model's variability to identify which predictor variables influence the model globally and locally. The results of the GWPR model variability test, based on the processed data, are as follows:

Table 5. Results of the GWPR Model Variability Test

Variable	$F_{statistic}$	F_{table}
x_1	5.101	2.210
x_2	4.382	2.299
x_3	1.539	2.372
x_4	3.483	2.954
x_5	0.143	2.427

According to Table 5, variables x_1, x_2, and x_4 have $F_{statistic}$ greater than the critical F_{table}, which implies that these variables complications are managed by healthcare professionals (x_1), antenatal care services (K1) and (K4) (x_2), and married women under 17 years old (x_4) exhibit local effects. In contrast, the$F_{statistic}$ for variables x_3 and x_5 are less than the critical F-value, indicating that the global effects are associated with pregnant women receiving Fe1 and Fe3 tablets (x_3) and postpartum mothers receiving vitamin A (x_5). Given the global and local influences of these variables the presence of outliers in the maternal mortality data, and its influencing factors for East Java in 2021, these variables will be grouped using the SGWPR model with the Robust M approach.

The existence of outliers in the maternal mortality data leads to reduced accuracy of the SGWPR model in predicting actual values. Consequently, it is necessary to enhance the SGWPR model to make it more robust against outliers. This will be achieved by estimating a robust SGWPR model using the Robust-M method, which involves an iterative process to establish $\hat{\beta}_g$ and $\hat{\beta}_m(u_i, v_i)$ as initial values from the WLS and OLS models, the iteration aims to obtain converge parameter estimates with a convergence criterion of 0.001. For the Robust M method, the Tukey Bisquare weighting function with a constant $c = 4.686$ is used in the parameter estimation. The global parameter estimates for the SGWPR model using the Robust M method achieve convergence by the 12th iteration, as detailed in Table 6 below:

Table 6. Global Parameter Estimates for the SGWPR Model

Parameter	$\hat{\beta}_g$
x_3	-0.328
x_5	0.372

Following this, Table 7 presents an example of the estimation of local variable parameters for the SGWPR model using maternal mortality data from East Java in 2021, specifically for the Bangkalan Regency. The local parameter estimates for the SGWPR model at the Bangkalan Regency location, obtained using the Robust M method, are detailed as follows:

Table 7. Local Parameter Estimates for the SGWPR Model

Location	Parameter	$\hat{\beta}_m(u_1, v_1)$
Bangkalan Regency	Constant	2.837
	x_1	-0.082
	x_2	-0.268
	x_4	0.124

According to Tables 6 and 7, the SGWPR model with the Robust M method produces different results for each observation site in the maternal mortality data. For instance, at the first location, Bangkalan Regency, the SGWPR model using the Robust-M method is as follows:

$$y_1 = 2.837 - 0.082x_1 - 0.268x_2 - 0{,}328x_3 + 0.124x_{41} + 0.372x_5$$

According to the SGWPR model for Bangkalan Regency, if healthcare professionals manage the factors of complications (x_1), antenatal care services (K1) and (K4) (x_2), pregnant women receiving Fe1 and Fe3 tablets (x_3), married women under 17 years old (x_4), and postpartum mothers receiving vitamin A (x_5) are held constant or excluded, the maternal mortality rate in Bangkalan Regency for 2021 is 2.84%.

An increase of one percent in the factor of complications managed by healthcare professionals (x_1), With other factors held constant, the percentage of complications managed by healthcare professionals in Bangkalan will be reduced by 0.08%. Similarly, a one percent increase in the factor of antenatal care services (K1) and (K4) (x_2) will decrease the percentage of these services in Bangkalan by 0.03%. A one percent increase in the factor of pregnant women receiving Fe1 and Fe3 tablets (x_3) will result in a 0.33% decrease in the percentage of women receiving these tablets in Bangkalan. Conversely, a one percent increase in the factor of married women under 17 years old (x_4) will lead to a 0.12% increase in this percentage in Bangkalan. Finally, a one percent increase in the factor of postpartum mothers receiving vitamin A (x_5) will increase the percentage of women receiving vitamin A in Bangkalan by 0.04%.

The subsequent step is to conduct a simultaneous or overall test of the SGWPR model parameters for global and local variables. The first overall test for the global variables produced an $F_{statistic}$ of 28.452, compared to a critical F_{table} of 18.307, indicating that $F_{statistic} > F_{table}$. This suggests that the global variables significantly affect the response variable simultaneously. The second overall test for the local variables resulted in an $F_{statistic}$ of 27.851, with the same critical F_{tabel} of 18.307, showing that $F_{hitung} > F_{tabel}$. This indicates that the local variables also have a significant simultaneous impact on the response variable.

The next stage involves partial or individual testing of the SGWPR model parameters to determine which factors influence maternal mortality. The results of the partial test for the global variables are as follows:

Table 8. Partial Testing of Global Parameters for the SGWPR Model

Variable	$t_{g\_statistic}$	t_{table}
x_3	-4.877	2.032
x_5	5.021	

According to Table 8, the $t_{g\_statistic}$ for all global variables are greater than the critical t-value, indicating that $|t_{g\_statistic}| > t_{table}$. Therefore, it can be concluded that the variables related to pregnant women receiving Fe1 and Fe3 tablets (x_3) and postpartum mothers receiving vitamin A (x_5) individually have a significant effect on maternal mortality in Bangkalan Regency. The next phase involves partial testing of the SGWPR model for local variables, with the results presented in Table 9 below:

Table 9. Partial Testing of Local Parameters for the SGWPR Model

Variable	$t_{m\_statistic}$	t_{table}
Constant	62.462	2.032
x_1	-1.242	
x_2	-2.914	
x_4	1.200	

Based on the results presented in Table 9, the t-values for all local variables were assessed. It can be concluded that two local parameters significantly influence maternal mortality rates in Bangkalan Regency: complications during childbirth managed by healthcare professionals (x_1) and women married under the age of 17 (x_4).

3.2 Regional Classification Based on Significant Variables

The SGWPR model produced for each observation location will vary depending on the coefficients of global and significant local variables affecting the response variable. The classification of districts/cities based on significant global and local variables in the SGWPR model is presented in the following table:

Table 10. Classification of Districts/Cities Based on Significant Variables

Group	Districts/Cities	Significant Variable
1	Bojonegoro, Jombang, Madiun, Magetan, Mojokerto, Nganjuk, Ngawi, Probolinggo, Tuban, Tulungagung, Kediri City, Madiun City, Mojokerto City, Pasuruan City	x_3, x_4, x_5
2	Banyuwangi, Bondowoso, Jember, Lamongan, Pemekasan, Sampang, Situbondo, Sumenep, Batu City	x_1, x_2, x_3, x_4, x_5
3	Blitar, Pacitan, Ponorogo, Trenggalek	x_1, x_3, x_4, x_5
4	Lumajang, Pausuran, Sidoarjo, Probolinggo City	$x_2. x_3, x_5$
5	Gresik, Kediri, Malang City	x_3, x_5
6	Bangkalan, Surabaya City	x_2, x_3, x_4, x_5
7	Malang	x_1, x_2, x_3, x_5
8	Blitar City	x_1, x_3, x_5

Based on Table 10, the classification of districts/cities according to significant variables results in eight distinct groups. It is observed that in East Java, there are eight optimal models for district/city classification. The global variables influencing maternal mortality rates in East Java for 2021 include pregnant women receiving Fe1 and Fe3 tablets (x_3) and women receiving vitamin A (x_5), which are consistent across all districts/cities. In contrast, local variables, such as complications during childbirth, managed by healthcare professionals (x_1), maternal healthcare services (K1) and (K4) (x_2), women married under the age of 17 (x_4), and women receiving vitamin A (x_5) exhibit varying values across different districts/cities in East Java. Additionally, the mapping of the SGWPR model for significant variables using the Robust M method for maternal mortality data in East Java for the year 2021 is illustrated in Figure 1. It shows that the predictor variables significantly influencing maternal mortality rates in East Java for 2021 vary across observation locations and tend to cluster in specific areas. The first group includes 14 districts/cities: Bojonegoro, Jombang, Madiun, Magetan, Mojokerto, Nganjuk, Ngawi, Probolinggo, Tuban, Tulungagung, Kediri City, Madiun City, Mojokerto City, and Pasuruan City. In this group, the significant factors are the percentage of pregnant women receiving Fe1 and Fe3 tablets, women married under 17, and postpartum women receiving vitamin A. The second group comprises 9 districts/cities, including Banyuwangi, Bondowoso, Jember, Lamongan, Pamekasan, Sampang, Situbondo, Sumenep, and Batu City. For these areas, significant variables include the percentage of childbirth complications managed by healthcare professionals, maternal healthcare services (K1) and (K4), pregnant women receiving Fe1 and Fe3 tablets, women married under 17, and postpartum women receiving vitamin A.

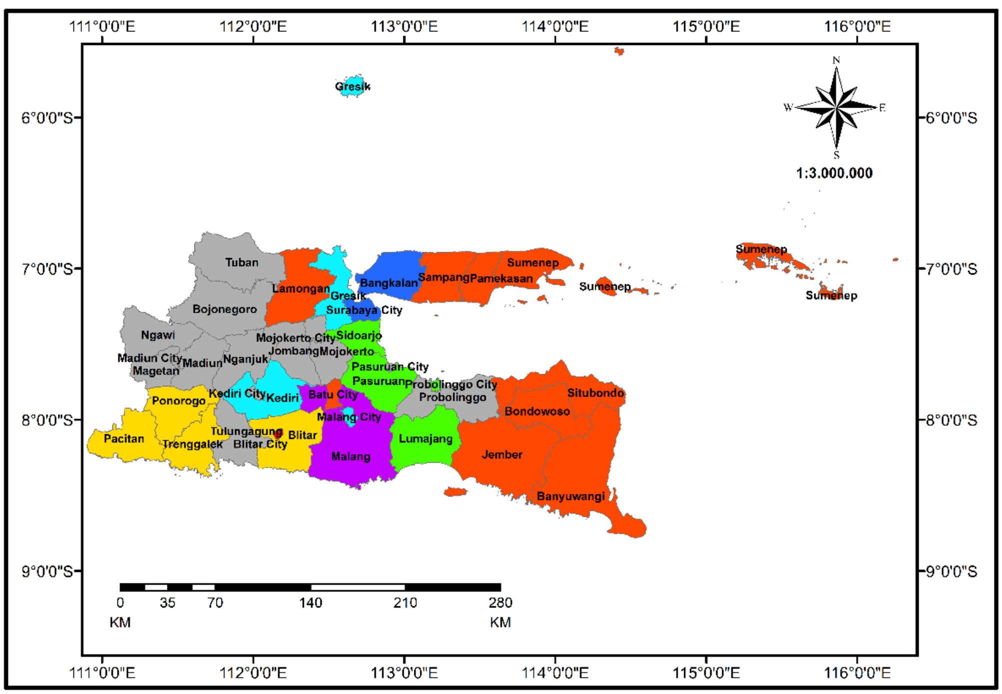

Figure 1. Classification of Districts/Cities Based on Significant Variables Affecting Maternal Mortality Rates in East Java for the Year 2021

The third group comprises four districts: Blitar, Pacitan, Ponorogo, and Trenggalek. In this group, the significant factors include the percentage of childbirth complications managed by healthcare professionals, pregnant women receiving Fe1 and Fe3 tablets, women married under 17, and postpartum women receiving vitamin A. The fourth group includes four districts/cities: Lumajang, Pasuruan, Sidoarjo, and Probolinggo City. The significant factors here are the percentage of maternal healthcare services (K1) and (K4), pregnant women receiving Fe1 and Fe3 tablets, and postpartum women receiving vitamin A. The fifth group comprises three districts/cities: Gresik, Kediri, and Malang City. In these areas, significant factors are the percentage of pregnant women receiving Fe1 and Fe3 tablets and postpartum women receiving vitamin A. The sixth group contains two districts/cities: Bangkalan and Surabaya. The significant factors for this group are the percentage of maternal healthcare services (K1) and (K4), pregnant women receiving Fe1 and Fe3 tablets, women married under 17, and postpartum women receiving vitamin A. The seventh group consists of one district: Malang. The significant factors here are the percentage of childbirth complications managed by healthcare professionals, maternal healthcare services (K1) and (K4), pregnant women receiving Fe1 and Fe3 tablets, and postpartum women receiving vitamin A. The final group, the eighth, includes one city: Blitar. The significant factors are the percentage of childbirth complications managed by healthcare professionals, pregnant women receiving Fe1 and Fe3 tablets, and postpartum women receiving vitamin A. Based on this description, it can be concluded that neighboring regions tend to share similar factors influencing maternal mortality rates in East Java for the year 2021.

4. CONCLUSIONS

Regional classification based on significant variables affecting maternal mortality rates using a robust semiparametric geographically weighted Poisson regression model results in eight distinct groups. These are: Group One: Districts/Cities with significant variables x_3, x_4 and x_5 include Bojonegoro, Jombang, Madiun, Magetan, Mojokerto, Nganjuk, Ngawi, Probolinggo, Tuban, Tulungagung, Kediri City, Madiun City, Mojokerto City, and Pasuruan City. Group Two: Districts/Cities with significant variables x_1, x_2, x_3, x_4 and x_5 are Banyuwangi, Bondowoso, Jember, Lamongan, Pamekasan, Sampang, Situbondo, Sumenep, and Batu City. Group Three: Districts with significant variables x_1, x_3, x_4, and x_5 include Blitar, Pacitan, Ponorogo, and Trenggalek. Group Four: Districts/Cities with significant variables x_2, x_3, and x_5 are Lumajang, Pasuruan, Sidoarjo, and Probolinggo City. Group Five: Districts/Cities with significant variables x_3 and x_5 include Gresik, Kediri, and Malang City. Group Six: Districts/Cities with significant variables x_2, x_3, x_4, and x_5 are Bangkalan and Surabaya City. Group Seven: The district with significant variables x_1, x_2, x_3, and x_5 is Malang. Group Eight: The city with significant variables x_1, x_3, and x_5 is Blitar City. This classification indicates that different regions exhibit unique combinations of significant variables affecting maternal mortality rates in East Java for 2021.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest.

REFERENCES

[1] A. Agresti, An introduction to categorical data analysis, Wiley, 2007.

[2] J. Aguero-Valverde, P.P. Jovanis, Spatial analysis of fatal and injury crashes in Pennsylvania, Accid. Anal. Prev. 38 (2006), 618–625. https://doi.org/10.1016/j.aap.2005.12.006.

[3] Y. Leung, C.L. Mei, W.X. Zhang, Statistical tests for spatial nonstationarity based on the geographically weighted regression model, Environ. Plan. A Econ. Space 32 (2000), 9–32. https://doi.org/10.1068/a3162.

[4] C. Mei, Gegrapically weighted regression technique for spasial data analysis, Master Dissertation, Xi'an Jiaotong University, 2005.

[5] S. Hailegebreal, F. Haile, Y. Haile, et al. Using geographically weighted regression analysis to assess predictors of home birth hot spots in Ethiopia, PLOS ONE 18 (2023), e0286704. https://doi.org/10.1371/journal.pone.0286704.

[6] A.S. Fotheringham, C. Brundson, M. Charlton, Geographically weighted regression, Wiley, 2002.

[7] F. Zhao, L.L.M. Chow, X. Liu, A transit ridership model based on geographically weighted regression and service quality variabels, Report D097591, Florida International University, 2005.

[8] Z. Li, W. Wang, P. Liu, et al. Using geographically weighted poisson regression for county-level crash modeling in California, Saf. Sci. 58 (2013), 89–97. https://doi.org/10.1016/j.ssci.2013.04.005.

[9] T. Nakaya, A.S. Fotheringham, C. Brunsdon, et al. Geographically weighted Poisson regression for disease association mapping, Stat. Med. 24 (2005), 2695–2717. https://doi.org/10.1002/sim.2129.

[10] C. Chen, Robust regression and outlier detection with the robustreg procedure, Paper 265-267, SAS Institute, Cary, (2002).

[11] S.W. Lin. A comparative study of robust estimater in regression, in: Proceeding of the 11 th Annual Conference of Pasific Decision Sciences Instute, 569-572, 2006.

[12] C. Chen, Robust Regression and outlier detection with the ROBUSTREG procedure, SUGI Paper No.265-27, SAS Institute, Cary, 2002.

[13] G.E.P. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. Ser. B: Stat. Methodol. 26 (1964), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

[14] V. Barnett, T. Lewis. Outliers in statistical data, Wiley, 1994.

[15] F. Santos, Modern methods for old data: An overview of some robust methods for outliers detection with applications in osteology, J. Archaeol. Sci. Rep. 32 (2020), 102423. https://doi.org/10.1016/j.jasrep.2020.102423.

[16] C. Leys, O. Klein, Y. Dominicy, C. Ley, Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance, J. Exp. Soc. Psychol. 74 (2018), 150–156. https://doi.org/10.1016/j.jesp.2017.09.011.

[17] A.C. Obikee, G.U. Ebuh, H.O. Obiora-Ilouno, Comparison of outlier techniques based on simulated data, Open J. Stat. 04 (2014), 536–561. https://doi.org/10.4236/ojs.2014.47051.

[18] M.F. MacDorman, E. Declercq, H. Cabral, et al. Recent increases in the u.s. maternal mortality rate: disentangling trends from measurement issues, Obstet. Gynecol. 128 (2016), 447–455. https://doi.org/10.1097/AOG.0000000000001556.

[19] L.E. Suryani, P. Purhadi, Analysis of factors affecting the number of infant and maternal mortality in East Java using geographically weighted bivariate generalized Poisson regression, Inferensi 1 (2018), 63. https://doi.org/10.12962/j27213862.v1i2.6726.

[20] S. Martafiyah, C. Supriadi, T. Saifudin, Modeling the maternal mortality rate in Indonesia using geographically weighted Poisson regression approach, AIP Conf. Proc. 2975 (2023), 080011. https://doi.org/10.1063/5.0181066.

[21] S. Ji, Y. Wang, Y. Wang, Geographically weighted poisson regression under linear model of coregionalization assistance: Application to a bicycle crash study, Accid. Anal. Prev. 159 (2021), 106230. https://doi.org/10.1016/j.aap.2021.106230.

[22] N.R. Draper, H. Smith, Applied regression analysis, Second Edition Translation, PT. Gramedia Pustaka Utama, Jakarta, 1992.

[23] J. Fox. Robust regression, Wiley, 2002.

[24] R. Erdkhadifa, Mixed geographically weighted poisson regression model in the number of maternal mortality, J. Mantik 6 (2022), 2685-4236.

[25] Kementerian Kesehatan Republik Indonesia, Profil kesehatan Jawa Timur tahun 2022, Jawa Timur, 2022.

[26] Kementerian Kesehatan Republik Indonesia, Profil kesehatan Indonesia tahun 2022, Jakarta, 2022.