SPATIAL MODELING OF POVERTY IN BENGKULU PROVINCE WITH MIXED GEOGRAPHICALLY WEIGHTED REGRESSION

Sigit Nugroho[1], Dyah Setyo Rini[2*], Tommy Jomecho[3], Cinta Rizki Oktarina[4], Stevy Cahya Pratiwi[5], Elisabeth Evelin Karuna[6]

[1,2,4,5,6]*Department of Mathematics, Faculty of Mathematics and Natural Science, University of Bengkulu*
[3]*Statistics Indonesia-Bengkulu Province, Indonesia*
Wr. Supratman Street, Kandang Limun, Bengkulu City, 38371, Indonesia

Corresponding author's e-mail: *dyah.setyorini@unib.ac.id

ABSTRACT

The percentage of poor people in Bengkulu Province is high from year to year. The poverty rate in Bengkulu Province also tends to fluctuate. If there is a decrease in the poverty rate, the decrease is relatively small. Poverty in the regions of Bengkulu Province also varies from district to district, subdistrict, and village to village, because poverty data is spatial data that varies regionally. The diversity of poverty data in Bengkulu Province is influenced by spatial effects, namely spatial dependency and spatial heterogeneity. Spatial dependency occurs due to spatial error correlation in cross section data, while spatial heterogeneity occurs due to random area effects, which is the difference between one region and another. Therefore, classical methods are not qualified enough to analyze the resulting diversity. This research will model the poverty of each district/city in Bengkulu Province using Mixed Geographically Weighted Regression (MGWR), because this method is quite complex in modeling data that contains spatial heterogeneity and variations in geospatial data. This modeling aims to identify and analyze poverty indicators in Bengkulu Province spatially, namely based on poverty data in each district/city in Bengkulu Province. The results showed that by using the MGWR method, the variables that locally influence the percentage of extreme poor people in each district/city in Bengkulu Province are Female Head of Household (X_1) and not having a water heater (X_2). Meanwhile, the variable that has a global effect on the percentage of the extreme poor in each district/city in Bengkulu Province is not having a flat screen television (X_3).

1. INTRODUCTION

The problem of poverty is one of the fundamental problems that is complex and multidimensional in nature so that it becomes the center of attention of the government in any country. In Indonesia, the government has great concern for the creation of a just and prosperous society, so that the implemented development programs always include poverty alleviation efforts within it. The Central Bureau of Statistics (BPS)states that poverty is seen as an economic inability to meet basic food and non-food needs measured in termsof expenditure. Poor people are people who have an average monthly per capita expenditure below the poverty line. In September 2022, the number of poor people in Indonesia reached 26,36 million people (9,57 percent). This number increased by 0,14 million people against September 2021. The percentage of poor people in urban areas is 7,53 percent. Meanwhile, the percentage of poor people in rural areas is 12.36 percent [1].

The percentage of poor people in Indonesia varies for each region. This difference is caused by the inequality between regions in Indonesia. The largest percentage of poor people is located in the Maluku and Papua Islands, at 20,10 percent, while the lowest percentage of poor people is located in Kalimantan Island, at 5,9 percent. In terms of numbers, most of the poor are still in Java (13,94 million people), while the lowest number of poor people are in Kalimantan (1,00 million people) [1]. The percentage of poor people in Sumatra Island in September 2022 was 9,47 percent (5,67 million people). In other words, Sumatra Island is an island where there are still many poor people.

Bengkulu Province is a province located on the island of Sumatra. Bengkulu Province ranks sixth as thepoorest province in Indonesia in September 2022 based on the percentage of poor people, which amounted to 14,34 percent. This percentage of 14.34 percent increased by 0.09 percent from September 2021 [2]. The percentage of poor people in Bengkulu Province is high from year to year. The poverty rate in Bengkulu Province also tends to fluctuate. If there is a decrease in the poverty rate, the decrease is relatively small. Poverty that occurs in areas in Bengkulu Province also varies for each district/city, subdistrict, and village, because poverty data isspatial data that varies regionally.

The diversity of poverty data in Bengkulu Province is influenced by spatial effects, namely spatial dependence and spatial heterogeneity. Spatial dependency occurs due to the spatial correlation of errors in the cross-section data, while spatial heterogeneity occurs due to the random region effect, which is the difference between one region and another [3]. Therefore, classical methods are not qualified enough to analyze the resulting diversity. One method that can overcome these problems is *Geographically Weighted Regression* (GWR). GWR model is a spatial modeling that produces parameters that are local to each point and location where the data is observed. However, not all regression coefficients in the GWR model vary spatially. Therefore, the GWR model was developed into a *Mixed Geographically Weighted Regression* (MGWR) model. The MGWR model is a combination of global linear regression and GWR [4]. The use of Mixed Geographically Weighted Regression (MGWR) compared to regular GWR can have some strong scientific reasons, especially related to the complexity and variation of geospatial data. The following are several scientific reasons that support the use of MGWR, namely the existence of Spatial Heterogeneity: MGWR allows a more flexible model by taking into account spatial heterogeneity, namely variations in patterns and relationships between variables that can vary in various geographic locations. This is important because physical and social environments are often heterogeneous and MGWR can capture these varying patterns better than GWR which only takes local patterns into account. Non-Stationarity: The concept of non-stationarity refers to changes in the relationships between variables spatially or temporally. MGWR can capture this non-stationarity better than regular GWR, because it can model spatially and temporally varying effects [5]. Adjustment for Global Trends: Although MGWR takes into account local effects that vary by location, it can also account for global trends that may exist in the data. This is important because global trends can provide important context and information in understanding the phenomenon being studied.

This research will model poverty in each district/city in Bengkulu Province using *Mixed GeographicallyWeighted Regression* (MGWR). This modeling will identify and analyze poverty indicators in Bengkulu Province spatially, which is based on poverty data in each district/city in Bengkulu Province. The resulting indicators will be global and local indicators. Global indicators are indicators of poverty in Bengkulu Province in general, while local indicators are indicators of poverty in each district/city in Bengkulu Province. The poverty indicators obtained at each location (district/city) will affect the policies given to each region in poverty alleviation efforts.

Research on poverty in Bengkulu Province has been conducted by [6], conducted research on poverty modeling in Bengkulu Province using Small Area Estimation and Penalized Spline Semiparametric Regression. This study produced a poverty model in Bengkulu Province, namely a linear P-Spline model with 1 knot. [7] also conducted research on poverty in Bengkulu Province. The study produced factors that influence poverty in each sub-district in Bengkulu Province using the *Geographically Weighted Logistic Regression* method. The results showed that the number of education facilities affected poverty in 103 subdistricts in Bengkulu Province, while in the other 10 subdistricts, poverty was influenced by the number of families without electricity, the number of education facilities, the number of health facilities, the number of JAMKESMAS recipients, and the number of SKTM. This is due to the spatial heterogeneity effect of the data. [8] also conducted research on Mixed Geographically Weighted Regression (MGWR) on Poverty Rates in Central Java. The results of the research show that there are five predictor variables that have a significant effect on the level of poverty in districts/cities in Central Java, namely minimum wage (X_1), percentage of the number of families working in the agricultural sector (X_2), percentage of the number of families who utilize the health insurance program. National Jamkesmas (X_4), percentage of families who have defecation facilities (X_6), and inflation (X_8).

This research is a follow-up research from [6] [7]. In this follow-up research, the method used will produce global and local poverty indicators in Bengkulu Province. This research is expected to be the basis of government policy in alleviating poverty in Bengkulu Province in terms of the district/city area.

2. RESEARCH METHODS

2.1 Data and Data Sources

The data used in this study are secondary data obtained from the Bengkulu Province Central Bureau of Statistics (Susenas 2022). The variables used in this study are the percentage of extreme poor people in each district/city in Bengkulu Province in 2022 as the response variable (Y). Meanwhile, the predictor variables used are Female Head of Household (X_1), Not having a water heater (X_2), and not having a flat screen television (X_3) with the unit being the number of households. This data will be used to model, identify, and analyze the indicators that influence poverty in Bengkulu Province using the MGWR method. The choice of this variable is of course due to the diversity of poverty data that occurs in Bengkulu Province which is influenced by spatial effects, namely spatial dependency and spatial heterogeneity.

2.2 Regression Analysis

According to [9], regression analysis deals with the study of the dependence of one variable, namely the dependent variable on one or more independent variables. The form of a regression model with k predictor variables and *n* observations is as follows:

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_j + \varepsilon_i \tag{1}$$

with $i = 1,2,\dots,n$
Y_i : the response variable-i
X_j : the predictor variable-j
ε_i : error-i
β_j : regression model parameters
β_0 : intercept

2.3 Testing for Spatial Heterogeneity

Testing for spatial heterogeneity refers to the presence of diversity in the relationship across regions [10]. In almost every case it is assumed that we will get a different relationship at each observation location. As a result, global parameters estimated from geographic data do not describe well the geographic

phenomenon at a particular location. Spatial heterogeneity can be identified using the *Breusch-Pagan* test [11].The hypothesis used is as follows:

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (no spatial heterogeneity)
H_1: at least one $\sigma_i^2 \neq \sigma^2$ (spatial heterogeneity exists)
Test statistics:

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T \sim \chi^2_{(p)} \tag{2}$$

with vector elements f is $f_i = \frac{e_i^2}{\sigma^2} - 1$ and Z is a matrix of size $n \times (p + 1)$ which contains vectors that have been normalized standard for each observation. H_0 is rejected if the $BP > \chi^2_{(p)}$ or *p-value* $< \alpha$ with p is the number of independent variables, which means there exists spatial heterogeneity [12].

2.4 Spatial Weighting Function

In spatial analysis, parameter estimates at longitude and latitude points (u_i, v_i) will be more influenced by points that are close than points that are further away [13]. Therefore, the selection of spatial weights used in estimating parameters becomes very important. There are several methods that can be used to determine the weighting value, one of which is used in this research is the Adaptive Gaussian Kernel weighting function, because this kernel function has a different bandwidth at each observation location [14]. The Kernel Adaptive Gaussian weighting function can be calculated using the following equation:

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right] \tag{3}$$

with:

$$d_{ij} = \sqrt{(u_i, u_j)^2 + (v_i, v_j)^2} \tag{4}$$

Based on this formula, d_{ij} is the euclidean distance between locations (u_i, v_i) to the location (u_j, v_j) and h is a smoothing parameter or bandwidth [15]. Determining the optimum bandwidth is selected using the CV (Cross Validation) method [16]. The CV calculation in MGWR is the same as GWR which is calculated based on the average of the dependent and independent variables for the whole time. The optimum bandwidth is obtained if the resulting CV value is the minimum. The following is the formula for the Cross Validation method [17]:

$$CV = \sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(h)^2] \tag{5}$$

With y_i is the dependent variable at the observation location-i dan $\hat{y}_{\neq i}(h)$ is the predicted value of the observation location-i without involving the location-i with bandwidth h.

2.5 Geographically Weighted Regression (GWR)

According to [18], Geographically Weighted Regression (GWR) is a development of multiple linear regression models. This model is a linear regression model that produces parameter estimates that are local to each location where the data is obtained. The Geographically Weighted Regression (GWR) model was developed to overcome the influence of parameter heterogeneity resulting from spatial models caused by differences in the conditions of each location. The parameters resulting from the GWR model will be different in each location. The model of Geographically Weighted Regression (GWR) can be written as follows [19]:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{6}$$

2.6 Bandwidth

The choice of bandwidth in the GWR model is important because the accuracy of the model to the data is influenced by the bandwidth. The bandwidth can be analogized as the radius of a circle, so that an observation location point that is within the radius of the circle is still considered influential in shaping the

parameters at the observation location point to-i. Therefore, the method of selecting the optimum bandwidth is very important to use for estimating the right weight function.

One of the methods used to determine the optimum bandwidth is the Cross Validation (CV) method as follows:

$$CV(b) = \sum_{i=1}^{n}[y_i - \hat{y}_{i \neq 1}(b)]^2 \tag{7}$$

with $\hat{y}_{i \neq 1}(b)$ is the estimated value y_i, where the observation at point location-i is omitted from the estimation process. The optimum *bandwidth* value is obtained when the CV is minimum.

2.7 Mixed Geographically Weighted Regression (MGWR)

The level of spatial diversity in some coefficients could not be significant, or it could be ignored. Consequently, GWR model was developed to the mix geographycal weighted regression (MGWR) in which it was the combination of linear regression model and GWR model [20]. The MGWR model is the development of a multivariate spatial model with local parameter estimators (GWR) for each observation location [21]. In the MGWR model some coefficients in the GWR model are assumed to be constant for all observations while others vary according to the location of observation. Based on the GWR model in the equation, if not all predictor variables have a local effect, but some have a global effect, then such a model is called a Mixed Geographically Weighted Regression (MGWR) model. In the MGWR model, some GWR coefficients are assumed to be constant for all locations while others vary according to the location of data observation. The MGWR model can be expressed as follows:

$$y_i = \sum_{k=1}^{q} \beta_k x_{ik} + \sum_{k=q+1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i , i = 1,2,..,n \tag{8}$$

with assumption $\varepsilon_i \sim N(0, \sigma^2)$

y_i : observation value of the response variable-i
x_{ik} : the observation value of the predictor variable at observation -i
u_i, v_i : coordinate point (longitude, latitude) of the location-i
ε_i : error-i
$\beta_1, ..., \beta_q, \beta_{q+1}(u_i, v_i), ...,, \beta_p(u_i, v_i)$: MGWR model parameters

In MGWR modeling, there are several tests that need to be carried out. These tests can be seen in the Table 1 [22]:

Table 1. MGWR Testing

Name of test	Hypothesis testing	Statistic testing	Criteria
MGWR model suitability test	$H_0: \beta_k(u_i, v_i) = \beta_k$ (The MGWR model is not different from the Global Regression Model $H_1: \beta_k(u_i, v_i) \neq \beta_k$ (The MGWR model is different from the Global Regression Model)	$F_1 = \dfrac{y^T[(I - H) - (I - S)^T(I - S)]y/v_1}{y^T(I - S)^T(I - S)y/u_1}$	Reject H_0 if the test statistic > table statistic or $p_{value} < \alpha$
Simultaneous testing of global predictor variable parameters	$H_0: \beta_{q+1} = \beta_{q+2}$ $= \cdots = \beta_p = 0$ H_1: There is at least one $\beta_k \neq 0$	$F_2 = \dfrac{y^T[(I - S_l)^T - (I - S_l) - (I - S)^T(I - S)]y/r_1}{y^T(I - S)^T(I - S)y/u_1}$	Reject H_0 if the test statistic > table statistic or $p_{value} < \alpha$

Simultaneous testing of local predictor variable parameters	$H_0: \beta_1(u_i, v_i)$ $= \beta_2(u_i, v_i) = \cdots$ $= \beta_q(u_i, v_i) = 0$ H_1: There is at least one $\beta_k(u_i, v_i) \neq 0$	$F_3 = \dfrac{y^T \left[(I - S_g)^T - (I - S_g) - (I - S)^T (I - S) \right] y / t_1}{y^T (I - S)^T (I - S) y / u_1}$	Reject H_0 if the test statistic > table statistic or $p_{value} < \alpha$
Partial testing on global variables	$H_0: \beta_k = 0$ $H_1: \beta_k \neq 0$	$t = \dfrac{\hat{\beta}_k}{\hat{\sigma}\sqrt{g_{kk}}}$	Reject H_0 if the test statistic > table statistic or $p_{value} < \alpha$
Partial testing on local variables	$H_0: \beta_k(u_i, v_i) = 0$ $H_1: \beta_k(u_i, v_i) \neq 0$	$t = \dfrac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{m_{kk}}}$	Reject H_0 if the test statistic > table statistic or $p_{value} < \alpha$

2.8 Research Activity Design

The design or stages in conducting research with the MGWR method are as follows:

1. Data collection

   The data used in this research is secondary data in the form of raw data related to poverty variables in Bengkulu Province based on Susenas 2022 data. At this stage of data collection, the researcher collected the coordinate points of each district/city in Bengkulu Province, where the reference point was the central coordinate point of each district/city.

2. Data Exploration

   Data exploration was conducted with the aim of providing a complete and accurate description of the poverty condition in Bengkulu Province. This exploration was carried out by mapping poverty conditions in Bengkulu Province using data on the percentage of subdistrict poverty and coordinate points that had been collected with Arcview GIS software. At this stage, assumptions were also tested as a condition for using spatial methods, including spatial autocorrelation, multicollinearity, and spatial heterogeneity using the R 4.0.2 program.

3. MGWR Modeling

   The first step in MGWR modeling is to perform multiple linear regression modeling along with classical assumption testing. In testing classical assumptions, when the data used has spatial heterogeneity, it can be continued with GWR modeling. GWR modeling begins with the calculation of the optimum bandwidth and weight matrix of each observation location studied. Furthermore, the weight matrix will be used to estimate the GWR model parameters using WLS estimation. The parameter estimation results will then be tested whether they have a location effect or not. If there is a variable that does not have a location effect, then it can be assumed as a global coefficient. Then next, a spatial variability test will be conducted to determine the global and local coefficients that will be used in MGWR modeling. MGWR modeling is performed using an adaptive gaussian kernel weighting function. At this stage, the global and local indicators that determine poverty in Bengkulu Province will be obtained as well as the thematic map of the poverty model obtained (the stage of identifying indicators that affect poverty). In addition, the parameter estimation of the MGWR model will also be generated at this stage.

3. RESULTS AND DISCUSSION

3.1 Classical Assumption Test

3.1.1   Normality Test

The normality test used is the Anderson-Darling test, with the following hypothesis:

H_0: Data obtained from a normally distributed population
H_1: Data obtained from a population that is not normally distributed
with test statistics:

$$A = -n - \frac{1}{n}\sum_{i=1}^{n}[2i-1]\left[\ln\left(F\left(\frac{x_i - \bar{x}}{s}\right)\right) + \ln\left(1 - F(Z_{n+1-i})\right)\right]$$

The test results show that the statistical value of the Anderson-Darling test obtained is $A = 0.215$ with a $p_{value} = 0.7863 > \alpha = 0.17$. So H_0 is accepted, meaning that at a test significance level of 17% it can be said that the data comes from a normally distributed population.

3.1.2   Spatial Heterogeneity Test

Spatial heterogeneity can be identified using the *Breusch-Pagan* test [11].The hypothesis used is as follows:

H_0: There is no inequality of residual variance between one observation to another (homoscedasticity)
H_1: There is inequality of residual variance between one observation to another (spatial heterogeneity)
with test statistics:

$$\text{BP} = \frac{1}{2}f^T Z(Z^T Z)^{-1} Z^T \sim \chi^2_{(p)}$$

The test results show that the statistical value of the Breusch-Pagan test obtained is $BP = 5.0332$ with $p_{value} = 0.1694 < \alpha = 0.17$. Then H_0 rejected, meaning that at the real level of testing 17% It can be said that there is an inequality of residual variance between one observation to another or there is spatial heterogeneity. The choice of alpha of course depends on the data characteristics, analysis objectives, and the desired trade-off between local sensitivity and global representation in GWR and MGWR modeling. The use of an alpha of 17% in the GWR and MGWR modeling in this research is due to the condition of the data which takes into account aspects of Spatial Heterogeneity and also with an alpha of 17% the GWR and MGWR models are believed to give considerable weight to local influences, namely, the influence of neighbors. nearest neighbors, are able to provide clearer information from the overall (global) data, and can help improve model resistance to high local variability, such as when there are outlier observations that may have a significant influence on local estimates.

3.1.3   Multicollinearity Test

The multicollinearity test aims to determine whether or not there is a relationship between the independent variables. Multiple regression analysis is said to be good if there is a relationship between one independent variable and another, if there is no such relationship then there is a multicollinearity problem. Multicollinearity testing can be seen from the VIF value, using package "car" in R-studio software and the following results are obtained:

Table 2. Multicollinearity Test

Variables	VIF Value
Female Head of Household (X_1)	2.83
Does not have a water heater (X_2)	2.69
Does not have a flat screen television (X_3)	1.11

The regression model is said to have no multicollinearity if $VIF < 10$ and the regression is said to have multicollinearity if the value of $VIF \geq 10$. Based on **Table 2** which is an analysis of the data, the value of $VIF < 10$ for all independent variables so that it can be said that there is no multicollinearity problem in the data.

3.1.4   Autocorrelation test

The autocorrelation test used is the Durbin-Watson test, with the following hypothesis:

H_0: There is no autocorrelation between one observation and another.

H_1: There is autocorrelation between one observation to another observation.

with test statistics:

$$d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

The test results show that the statistical value of the Durbin-Watson test obtained is $d = 2.6818$ with $p_{value} = 0.8382 > \alpha = 0.17$. Then H_0 accepted, meaning that at the real level of testing 17% it can be said that there is no autocorrelation between one observation to another.

3.2   Geographically Weighted Regression

In testing spatial heterogeneity, it is known that there is spatial heterogeneity in this research data, so GWR modeling will be carried out next.

3.2.1   Calculation of *Euclidean* Distance

The first thing that must be done before GWR modeling is to form a weight matrix (W_{ij}), where the weight matrix is formed by substituting the optimum *bandwidth* value and *euclidean* distance between locations (d_{ij}) distance between locations (u_i, v_i). Then the *euclidean* distance calculation is done first. The calculation of *euclidean distance* requires the *latitude* and *longitude* values of each district/city in Bengkulu province which are then calculated using the *Euclidean* distance formula.

$$d_{ij} = \begin{pmatrix} 0 & 0.96 & 1.86 & 0.54 & 0.67 & 2.65 & 1.30 & 0.88 & 1.41 & 0.95 \\ 0.96 & 0 & 1.05 & 1.50 & 0.58 & 1.85 & 0.42 & 0.22 & 0.61 & 0.57 \\ 1.86 & 1.05 & 0 & 2.40 & 1.20 & 0.80 & 0.63 & 1.01 & 0.46 & 0.94 \\ 0.54 & 1.50 & 2.40 & 0 & 1.20 & 3.18 & 1.85 & 1.42 & 1.95 & 1.47 \\ 0.67 & 0.58 & 1.20 & 1.20 & 0 & 1.98 & 0.73 & 0.37 & 0.78 & 0.27 \\ 2.65 & 1.85 & 0.80 & 3.18 & 1.98 & 0 & 1.42 & 1.81 & 1.26 & 1.71 \\ 1.30 & 0.42 & 0.63 & 1.85 & 0.73 & 1.42 & 0 & 0.42 & 0.21 & 0.55 \\ 0.88 & 0.22 & 1.01 & 1.42 & 0.37 & 1.81 & 0.42 & 0 & 0.55 & 0.35 \\ 1.41 & 0.61 & 0.46 & 1.95 & 0.78 & 1.26 & 0.21 & 0.55 & 0 & 0.53 \\ 0.95 & 0.57 & 0.94 & 1.47 & 0.27 & 1.71 & 0.55 & 0.35 & 0.53 & 0 \end{pmatrix}$$

Bandwidth is a circle with a radius from the center point of the observed location, which means that locations around the bandwidth value still have a considerable influence on the observed location. The bandwidth value for each location is listed in **Table 3**, where the bandwidth value is the value obtained using the initial unit (radius) and Bandwidth in km^2 is the bandwidth value calculated in km^2 units.

Table 3. **Bandwidth Value of Each Location**

Regency/City	Bandwidth	Bandwidth in km^2
Bengkulu Selatan	1.0723	119.1102
Rejang Lebong	0.7209	80.2498
Bengkulu Utara	1.0266	114.2800
Kaur	1.6070	178.8896
Seluma	0.7501	83.5003
Mukomuko	1.8257	203.2350
Lebong	0.6646	73.9826
Kepahiang	0.6511	72.4798

Regency/City	Bandwidth	Bandwidth in km^2
Bengkulu Tengah	0.6637	73.8824
Bengkulu City	0.6851	76.2646

3.2.2 Geographic Weighting Matrix

The weighting function calculation will use the Euclidean distance **Equation (3)** and the bandwidth value in **Table 2** and then substitute it into the adaptive Gaussian kernel function formula in **Equation (4)**. This weighting matrix will have different values for each observation location, so a 10×10 matrix will be formed.

$$W(u_i, v_i) = \begin{pmatrix} 1 & 0.44 & 0.04 & 0.77 & 0.67 & 0.00 & 0.22 & 0.50 & 0.50 & 0.17 \\ 016 & 1 & 0.11 & 0.01 & 0.51 & 0.00 & 0.70 & 0.90 & 0.48 & 0.53 \\ 0.03 & 0.34 & 1 & 0.00 & 0.25 & 0.54 & 0.68 & 0.37 & 0.81 & 0.43 \\ 0.89 & 0.41 & 0.10 & 1 & 0.57 & 0.01 & 0.26 & 0.45 & 0.22 & 0.42 \\ 0.44 & 0.54 & 0.07 & 0.07 & 1 & 0.00 & 0.38 & 0.77 & 0.33 & 0.87 \\ 0.12 & 0.35 & 0.82 & 0.04 & 0.30 & 1 & 0.54 & 0.37 & 0.61 & 0.41 \\ 0.02 & 0.66 & 0.40 & 0.00 & 0.29 & 0.00 & 1 & 0.66 & 0.90 & 0.50 \\ 0.15 & 0.88 & 0.08 & 0.00 & 0.71 & 0.00 & 0.65 & 1 & 0.48 & 0.74 \\ 0.01 & 0.42 & 0.61 & 0.00 & 0.25 & 0.02 & 0.90 & 0.50 & 1 & 0.51 \\ 0.14 & 0.49 & 0.15 & 0.00 & 0.85 & 0.00 & 0.52 & 0.76 & 0.53 & 1 \end{pmatrix}$$

3.2.3 GWR Parameter Estimation

The weight matrix is used to estimate the GWR model parameters by incorporating the weight values into the calculation. Parameter coefficient values $\hat{\beta}(u_i, v_i)$ in the GWR model is obtained using WLS estimation. Each parameter value is estimated at the observation point, so each observation point has a different parameter value. The following is an example of GWR model parameter estimation for Kaur regency:

Table 4. Parameter Estimation of GWR Model of Kaur Regency

Variable	β	t_{count}	t_{table}	Description
Intercept	-41.8815	2.86		Reject H_0
X_1	-0.3700	5.60	2.06	Reject H_0
X_2	0.1535	4.95		Reject H_0
X_3	-0.0348	3.48		Reject H_0
		$R^2_{Squared} = 98.33\%$		

The parameter estimation results in **Table 4** show that there are 2 variables that have a negative effect on the Percentage of Extreme Poor Population, namely the Female Head of Household, which means that every one-unit increase in the variable will decrease the Percentage of Extreme Poor Population by one-unit X_1 will decrease the Percentage of Extreme Poor Population by 0.3700 and Does not have a flat screen television, which means that every one-unit increase in the variable will decrease the Percentage of Extreme Poor Population by X_3 will reduce the Percentage of Extreme Poor Population by 0.0348 and other variables have a positive effect, namely the variable Does not have a water heater, which means that every one-unit increase in the variable will increase the Percentage of Extreme Poor Population by X_2 will increase the Percentage of Extreme Poor Population by 0.1535. Meanwhile, to get a significant variable in the model, the value of t_{count} will be compared with t_{table}. If the value of $t_{count} > t_{table}$ then the three independent variables have a significant influence on the Percentage of Extreme Poor Population. With real level 17% The GWR model obtained for the Percentage of Extreme Poor Population in Kaur Regency is:

$$\hat{Y}_{Kaur} = -41.8815 - 0.3700X_1 + 0.1535X_2 - 0.0348X_3$$

The GWR model produces a coefficient of determination of 98.33%. This means that the GWR model obtained with a value of R^2 by 98.33% with independent variables, namely, Female Head of Household, does not have a water heater and does not have a flat screen television is able to explain the diversity of the dependent variable of 98.33% while 1.77% the rest is explained by other variables that are not studied.

3.2.4   Model Fit Test GWR

After obtaining the GWR model for neonatal deaths in each location, the next step is to test the suitability of the model to see if the GWR model is more suitable for use when compared to multiple linear regression models.

Table 5. GWR Model Fit Test

F	df_1	df_2	F_{table}	P_{value}	Description
13.608	6	2.327	5.20	0.0512	Reject H_0

The hypothesis used in this test is

H_0: $\beta_k(u_i, v_i) = \beta_k, k = 1,2, \ldots, p$ (there is no significant difference between multiple linear regression models and GWR)

H_1: There is at least one $\beta_k(u_i, v_i) \neq \beta_k$ (there is a significant difference between multiple linear regression models and GWR)

with test statistics:

$$F = \frac{\frac{SSE(H_0) - SSE(H_1)}{V}}{\frac{SSE(H_1)}{\delta_1}} = 13.608$$

The test results show that the statistical value of the F test obtained is $F = 13.608$ with $p_{value} = 0.0512 < \alpha = 0.17$. Then H_0 rejected, which means that there is a significant difference between the multiple linear regression model and GWR with a real level of significance 17% so the GWR model can be used.

3.2.5   GWR Model Variability Test

In testing the significance of the GWR model parameters, it is known that not all variables are significant in each observation location, this has the possibility that there are variables that do not have a location effect. If there is a parameter that does not have a location effect, it is assumed to be a global coefficient or have the same parameter estimate value for all observation locations. Thus, a spatial variability test will be conducted to determine whether there are variable parameters that do not have a location effect, this test is also used to determine the global and local coefficients that will be used in MGWR modeling, with the following hypothesis:

H_0: $\beta_{1k}(u_i, v_i) = \cdots = \beta_{nk}(u_i, v_i) = 0$ for each $k = 1,2, \ldots, p$ (There is no significant difference in the effect of predictor variables X_k between one location and another)

H_1: There is at least one $\beta_k(u_i, v_i) \neq 0$ for each $i = 1,2, \ldots, n$ (There is a significant difference in the effect of predictor variables X_k between one location and another)

The test results can be seen in the **Table 6**:

Table 6. Testing the Variability of the GWR Model

Variable	F_3 statistic	p_{value}	Description	Conclusion
Intercept	4.0079	0.12683		
X_1	9.0752	0.04060	Local	Reject H_0
X_2	7.9483	0.05058	Local	Failed to reject H_0
X_3	1.3504	0.41457	Global	Failed to reject H_0

Based on the results in **Table 6**, it can be seen that, with a real level of 17% obtained that $p_{value} = 0.04 < \alpha = 0.17$. Then H_0 is rejected, which means that there is a significant difference in the effect of the predictor variables X_k between one location and another. So, it can be said that the Female Head of Household variable is a local model.

Based on **Table 6**, the variable that has a significant effect or has a global effect is not having a flat screen television, while the other variables have a local effect, namely the Female Head of Household and not having a water heater. Because not all independent variables have a local effect, the data on the Percentage of Extreme Poor can be modeled using Mixed Geographically Weighted Regression (MGWR).

3.3 Model Mixed Geographically Weighted Regression

Based on the previous GWR analysis, namely the variability test, 2 independent variables were obtained that had a local effect on the Percentage of Extreme Poor Population, namely Female Head of Household (X_1) and not having a water heater (X_2) while 1 independent variable that has a global effect is not having a flat screen television (X_3). Therefore, the MGWR model will be formed.

3.3.1   Testing the Fit of the MGWR Model

Testing the suitability of the MGWR model is carried out to determine whether there is a significant difference between the multiple linear regression model and the MGWR model. This test uses the following hypothesis:

$H_0: \beta_k(u_i, v_i) = \beta_k$ for each $k = 1,2, \dots, p$ and $i = 1,2, \dots, n$ (There is no significant difference between multiple linear regression model and MGWR)

$H_1: \beta_k(u_i, v_i) \neq \beta_k$ for each $k = 1,2, \dots, p$ and $i = 1,2, \dots, n$ (There is a significant difference between multiple linear regression model and MGWR)

The test results can be seen in the **Table 7**:

Table 7. Testing the Suitability of the MGWR Model

F	df_1	df_2	F_{table}	P_{value}	Description
5.1563	7.6354	6.1614	3.0144	0.0291	Reject H_0

Based on the results in **Table 7**, it can be seen that, tith a real level of 17% obtained that $p_{value} = 0.0291 < \alpha = 0.17$. Then H_0 rejected, which means that there is a significant difference between the multiple linear regression model and MGWR with a real level of 17% so the MGWR model can be used.

3.3.2   MGWR Global Parameter Estimation

After knowing that there are local and global variables that have an influence on the Percentage of Extreme Poor in Bengkulu Province, then simultaneous and partial testing is carried out to find out what global variables have a significant effect on the Percentage of Extreme Poor in the districts/cities of Bengkulu province. This test uses the following hypothesis:

$H_0: \beta_{q+1} = \beta_{q+2} = \cdots = \beta_p$, for $k = q + 1, q + 2, \dots, p$ (There is no effect of global variables together on the variable Percentage of Extreme Poor Population)

H_1: There is at least one $\beta_k \neq 0$, for $k = q + 1, q + 2, \dots, p$ (There is a joint influence of global variables on the variable Percentage of Extreme Poor Population)

The results of simultaneous MGWR testing are as follows:

Table 8. MGWR Global Concurrent Test

F_2	df_1	df_2	F_{table}	P_{value}	Description
5.011	2.869	6.1614	3.4633	0.044	Reject H_0

Based on the results in **Table 8**, it can be seen that, with a real level of 17% obtained that $p_{value} = 0.044 < \alpha = 0.17$. Then H_0 rejected, which means that there is an effect of global variables together on the dependent variable with a real level of 17%.

After carrying out simultaneous testing, the next step is to carry out partial testing of MGWR using the variable of not having a flat screen television (X_3), because based on the results of the previous GWR analysis, the independent variable that has a global influence is the variable Not having a flat screen television (X_3). Below is shown the testing hypothesis:

$H_0: \beta_3 = 0$ (The global variable of not having a flat screen television does not have a significant effect on the Percentage of Extreme Poor)

$H_1: \beta_3 \neq 0$ (The global variable of not owning a flat screen television has a significant effect on the Percentage of Extreme Poor)

The results of MGWR Global Partial testing are as follows:

Table 9. MGWR Global Partial Test

Variable	$\widehat{\beta_g}$	t_{hitung}	df	p_{value}	Description
X_3	0.008950912	2.952	6.16	0.012	Reject H_0

Based on the results in **Table 9**, with a real level of 17% obtained that $p_{value} < \alpha$ is the parameter $\hat{\beta}_3$ where p_{value} 0.0123 < 0.17. This means $\hat{\beta}_3$ is significant, which means that there is an influence of the Female Head of Household variable on the Percentage of Extreme Poor Population with a real level of 17% so that the regression model can be used.

3.3.3   MGWR Local Parameter Estimation

After knowing that there are local and global variables that have an influence on the Percentage of Extreme Poor in Bengkulu province, local simultaneous and local partial testing is then carried out to find out which local variables have a significant effect on the Percentage of Extreme Poor in the districts/cities of Bengkulu province, with the following hypothesis:

$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots = \beta_q(u_i, v_i) = 0$, for $k = 1,2, \dots, q$ (There is no effect of local variables on the variable Percentage of Extreme Poor Population together)

H_1: There is at least one $\beta_k(u_i, v_i) \neq 0$, for $k = 1,2, \dots, q$ (There is an influence of local variables on the variable Percentage of Extreme Poor Population together)

The results of the MGWR local simultaneous testing are as follows:

Table 10. MGWR Local Simultaneous Tests

F_3	df_1	df_2	F_{table}	P_{value}	Description
11.0123	8.99	6.16	2.24	0.0038	Reject H_0

Based on the result in **Table 10,** with a real level of 17% obtained that $p_{value} = 0.003 < \alpha = 0.17$. Then H_0 rejected, which means that there is an influence of local variables on the variable Percentage of Extreme Poor Population together with a real level of 5%.

After conducting local simultaneous testing, the next is to conduct local partial testing of MGWR for each variable, with the following results:

Table 11. MGWR Local Partial Test of Kaur Regency

Variable	$\widehat{\beta_g}$	t_{count}	df	t_{table}	Decision	Conclusion
Intercept	−27.37	−2.772			Reject H_0	Significant
X_1	−0.2937	−5.632	6.16	2.06	Reject H_0	Significant
X_2	0.0835	3.601			Reject H_0	Significant

Based on the results in **Table 11**, the MGWR model equation for the Kaur region can be created as follows:

$$\hat{Y}_{Kaur} = -27.37 - 0.2937X_1 + 0.0835X_2 + 0.008950912X_3$$

4. CONCLUSIONS

Based on the results of research using the existing MGWR method, it can be concluded that the variable that has a significant influence globally is X_3. Meanwhile, the variable that has a significant influence locally is X_1 and X_2 where the parameter estimation process for each observation is carried out using the WLS method which will produce different models for each location. This indicates that the MGWR model is very suitable for modeling data on the percentage of extreme poor people, because not all predictor variables have a local influence, but some have a global influence as well.

ACKNOWLEDGMENT

REFERENCES

[1]   BPS, "Profil Kemiskinan di Indonesia Maret 2023," Badan Pusat Statistika, Jakarta, 2023.

[2]   BPS, "Profil Kemiskinan Provinsi Bengkulu," Badan Pusat Statistik, Bengkulu, 2023.

[3]   L. Anselin, "Lagrange Multiplier Test Diagnostics for Spatial," *GeographicalAnalysis,* vol. 20, pp. 1-17, 1988.

[4]   H. Yasin, B. Warsito and A. R. Hakim, "Pemodelan Pertumbuhan Ekonomi di Provinsi Banten menggunakan Mixed Geographically Weighted Regression," *Media Statistika,* vol. 11, no. 1, pp. 53-64, 2018.

[5]   X. Wang, S. Shi, X. Zhao, Z. Hu, M. Hou and L. Xu, "Detecting Spatially Non-Stationary between Vegetation and Related Factors in the Yellow River Basin from 1986 to 2021 Using Multiscale Geographically Weighted Regression Based on Landsat," *MDPI,* vol. 14, no. 24, pp. 1-20, 2022.

[6]   I. Sriliana, E. Sunandi and U. Rafflesia, "The Poverty Modeling Using Small Area Estimation with Semiparametric P-spline (a case study: Poverty in Bengkulu Province)," *Asian Journal of Applied Sciences,* vol. 6, no. 4, pp. 166-174, 2018.

[7]   D. S. Rini, I. Sriliana and H. Fransiska, "The Poverty Modelling in Bengkulu Province using Geographically Weighted Logistic Regression," in *The 2nd International Conference On Science and Technology*, 2019.

[8]   M. Y. Darsyah, R. Warsono and M. F. Agustina, "The Model of Mixed Geographically Weighted Regression (MGWR) for Poverty Rate in Central Java," *International Journal of Applied Mathematics and Statistics,* vol. 536, 2015.

[9]   N. Lutfiani, S. and S. Mariani, "PEMODELAN GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) DENGAN FUNGSI PEMBOBOT KERNEL GAUSSIAN DAN BI-SQUARE," *UNNES Journal of Mathematics,* vol. 8, pp. 82-91, 2019.

[10] S. H. Daulay and E. Simamora, "PEMODELAN FAKTOR-FAKTOR PENYEBAB KEMISKINAN DI PROVINSI SUMATERA UTARA MENGGUNAKAN METODE GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)," *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam,* vol. 2, pp. 47-60, 2023.

[11] P. L. Ari Dewi and I. Zain, "Pemodelan Faktor Penyebab Kecelakaan Lalu Lintas Berdasarkan Metode Geographically Weighted Regression di Jawa Timur," *JURNAL SAINS DAN SENI ITS ,* vol. 5, pp. 58-64, 2016.

[12] S. M. Meutuah, H. Yasin and D. A. I. Maruddani, "PEMODELAN FIXED EFFECT GEOGRAPHICALLY WEIGHTED PANEL REGRESSION UNTUK INDEKS PEMBANGUNAN MANUSIA DI JAWA TENGAH," *JURNAL GAUSSIAN,* vol. 6, pp. 241-250, 2017.

[13] T. W. Utami, A. Rohman and A. Prahutama, "PEMODELAN REGRESI BERGANDA DAN GEOGRAPHICALLY WEIGHTED REGRESSION PADA TINGKAT PENGANGGURAN TERBUKA DI JAWA TENGAH," *MEDIA STATISTIKA,* vol. 9, pp. 133-147, 2016.

[14] A. Maulani, N. Herrhyanto and M. Suherman, "Aplikasi Model Geographically Weighted Regression (GWR) Untuk Menentukan Faktor-Faktor yang Mempengaruhi Kasus Gizi Buruk Anak Balita di Jawa Barat," *Eurekamatika,* vol. 4, pp. 46-63, 2016.

[15] A. . M. Dzikrina and S. W. Purnami, "Pemodelan Angka Prevalensi Kusta dan FaktorFaktor yang Mempengaruhi di Jawa Timur," *JURNAL SAINS DAN SENI POMITS,* vol. 2, pp. 275-281, 2013.

[16] W. Härdle, Smoothing Techniques: With Implementation in S, New York: Cambridge University Press, 1991.

[17] Y. Sutro and S. Martha, "PEMODELAN FIXED EFFECT GEOGRAPHICALLY WEIGHTED PANEL REGRESSION UNTUK INDEKS PEMBANGUNAN MANUSIA DI KALIMANTAN BARAT," *Bimaster,* vol. 9, pp. 413-422, 2020.

[18] A. S. Fotheringham, C. Brunsdon and M. Charlton, Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, England: John Wiley & Sons, 2002.

[19] S. Lee, D. Kang and M. Kim, "Determinants of Crime Incidence in Korea: A Mixed GWR Approach," in *In World Conference of the Spatial Econometrics Association*, 2009.

[20] F. Pongoh, I. M. Sumertajaya and M. N. Aidi, "Geographichal Weighted Regression and Mix Geographichal Weighted Regression," *International Journal of Statistics and Applications,* vol. 5, pp. 1-4, 2015.

[21] D. Hanum and Purhadi, "Faktor-faktor yang Mempengaruhi Morbiditas Penduduk Jawa Timur dengan Multivariate Geographically Weighted Regression (MGWR)," *JURNAL SAINS DAN SENI POMITS,* vol. 2, pp. 189-194, 2013.

[22] A. R. Hakim, H. Yasin and S. , "Pemodelan Persentase Penduduk Miskin di Kabupaten dan Kota di Jawa Tengah dengan Pendekatan Mixed Geographically Weighted Regression," *JURNAL GAUSSIAN,* vol. 3, pp. 575-584, 2014.