BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 1765–1778
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp1765-1778

SPATIAL REGRESSION APPROACH TO MODELLING POVERTY IN JAVA ISLAND 2022

Maria A. Hasiholan Siallagan[1], Novi Hidayat Pusponegoro[2*]

[1,2] Politeknik Statistika STIS
Jl. Otto Iskandardinata No. 64C, Jakarta, Indonesia

*Corresponding author's e-mail: * **novie@stis.ac.id***

ABSTRACT

Article History:

Received: 29th January 2024
Revised: 22th March 2024
Accepted: 1th July 2024
Published: 1st September 2024

Keywords:

Generalized Additive Model;
Smoothing Function;
Spatial.

Geographically Weighted Regression (GWR) model is a powerful tool for analyzing spatial patterns in data. However, the standard form of a spatial model that uses a single bandwidth calibration may be unrealistic because the response-predictor relationship may be either linear or nonlinear. To address this issue, the Multiscale GWR (MSGWR) model offers improved model performance by employing Generalized Additive Model (GAM) with varying bandwidth or smoothing function for each covariate in the model. This research aims to analyze the Percentage of Poor Population (PPP) on Java Island in 2022 using the geospatial models and related socioeconomic and demographic attributes, such as Open Unemployment Rate, Human Development Index, Labor Force Participation Rate, and GRDP Per capita to identify the best model in explaining the spatial pattern and to find out the determinant of PPP on Java Island in 2022. This study uses secondary data from Statistics Indonesia. The findings reveal that the MSGWR model provides the highest R^2 and smallest AICc value compared to single bandwidth models, specifically the GWR and MXGWR models. Furthermore, the MSGWR model indicates that HDI has a significant negative effect on PPP, whereas LFPR has a significant positive effect on PPP across all districts in Java Island in 2022.

How to cite this article:

M. A. H. Siallagan and N. H. Pusponegoro., "SPATIAL REGRESSION APPROACH TO MODELLING POVERTY IN JAVA ISLAND 2022," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 1765-1778, September, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Poverty is a multidimensional problem that can portray the well-being of a region. Statistics Indonesia (Badan Pusat Statistik/BPS) data show that the majority of Indonesia's impoverished population, amounting to 52.8% as of September 2022, resides on Java Island. The Percentage of Poor Population (PPP) as one of the poverty indicators is influenced by various factors and exhibits spatial patterns [1], [2], [3], [4]. PPP analysis with spatial patterns provides a more comprehensive interpretation since the policies implemented in one region may influence neighboring regions as well [5].

Major things in spatial analysis are assessing spatial patterns, such as spatial autocorrelation and spatial heterogeneity. Spatial autocorrelation occurs due to spill-over effects, which initially arise in certain areas and then spread to other areas meanwhile, spatial heterogeneity arises due to differences that are inherent in each spatial unit [5]. Some previous studies have proven that the Geographically Weighted Regression (GWR) provides a more powerful model compared to global regression models since the assumptions are not always fulfilled as spatial variation is not stationary [6], [7], [8], [9].

GWR is a (local) modeling technique to estimate regression models with spatially varying relationships [10]. However, different specifications of models in GWR yield different estimates for parameter results. Global variables, kernel selection, and spatial scale have an impact on the bandwidth size, which ultimately affects the GWR parameter estimation [11], [12]. A smaller bandwidth size yields more locally oriented parameters, while a larger bandwidth produces parameters that approximate a global model [13]. Misleading the parameter estimation can be caused by these inappropriate model specifications. It derives misinterpretation in the analysis of the relationship between response and predictor variables.

The GWR model assumes that all predictor variables exhibit local influences on the response variable. Yet, some predictor variables may operate at a local scale, while others may operate at a broader regional scale [14]. The MXGWR model combines Ordinary Least Squares (OLS) and GWR models, accommodating situations where certain predictor variables have global influences while others have local influences [15]. Thus, compared to OLS and GWR models, the MXGWR model provides a more comprehensive interpretation [16]. However, GWR and MXGWR are the standard form of spatial models that use a single bandwidth calibrate. This may be unrealistic because it implicitly assumes that each response-to-predictor relationship operates at the same spatial scale. To address this issue, the Multiscale GWR (MSGWR) model offers improved model performance compared to OLS, GWR, and MXGWR by using varying bandwidth for each covariate in the model [14], [17]. This model utilized the generalized additive model (GAM) to explore spatially non-stationary relationships. Therefore, it is crucial to understand the influence of different model specifications on model performance and identify a model that delivers the best performance.

Previous studies proved that the poverty-causing factors operate simultaneously at local and global scales [3], [18]. However, the major limitation of those studies is that they did not fully assess the relationship between poverty and socioeconomic-demographic attributes. Thus, we need to explore their effects on poverty using a spatial approach not only with local and mixed parameters with a single bandwidth but also using multiscale bandwidth. Therefore, this research aims to analyze PPP on Java Island in 2022 using the geospatial models and related socioeconomic and demographic attributes using GWR, MXGWR, and MSGWR models, to identify the best model for explaining the spatial pattern, and to find out the determinant of PPP on Java Island in 2022.

2. RESEARCH METHODS

To obtain the goals of the study, we apply the geospatial models to Statistics Indonesia’s data. The dataset consists of socioeconomic and demographic variables from 118 districts in Java Island 2022. The dependent variable of this paper is the percentage of the poor population (PPP) as a proxy of poverty. Then, the independent variables are the Open Unemployment Rate (OUR), Human Development Index (HDI), Labour Force Participation Rate (LFPR), and Gross Regional Domestic Product per Capita (GRDP per capita). All data are secondary data obtained from Statistics Indonesia. The variables’ operational definitions of the dataset are shown in **Table 1** below.

Table 1. Operational Variable Definitions

Variable	Definition	Unit
PPP	Percentage of poor population compared to the total population.	Percentage
OUR	Percentage of the unemployed population compared to the labor force.	Percentage
HDI	An index that measures three dimensions of human development, such as health, education, and expenditure [19].	Index
LFPR	Percentage of the labor force compared to the population aged > 10 years.	Percentage
GRDP per capita	Total gross products generated in a region divided by the regional population using current prices based on the 2010 base year.	Billion Rupiah

This paper employs descriptive analysis to identify spatial patterns in PPP, while inferential analysis is conducted to compare the spatial models between local and mixed parameters with single bandwidth and multiscale bandwidth. The research was conducted in several steps, which are as follows: 1) Analysing the spatial autocorrelation and heterogeneity in PPP on Java Island in 2022 using thematic maps; 2) Validating the presence of spatial autocorrelation through the global Moran's I statistical test and assessing spatial heterogeneity using the Breusch Pagan test; 3) Estimating OLS, GWR, MXGWR, and MSGWR models with Adaptive Bi-square and Fixed Gaussian kernels to examine the impact of the kernel on AICc and R^2 values; 4) Comparing the GWR model with the MXGWR model to assess the influence of globally influential variables on parameter estimation; 5) Comparing the GWR model with the MSGWR model to investigate the impact of scale on parameter estimation; 6) Interpreting the best-fitting model that provides insights into the relationship between predictor variables and PPP.

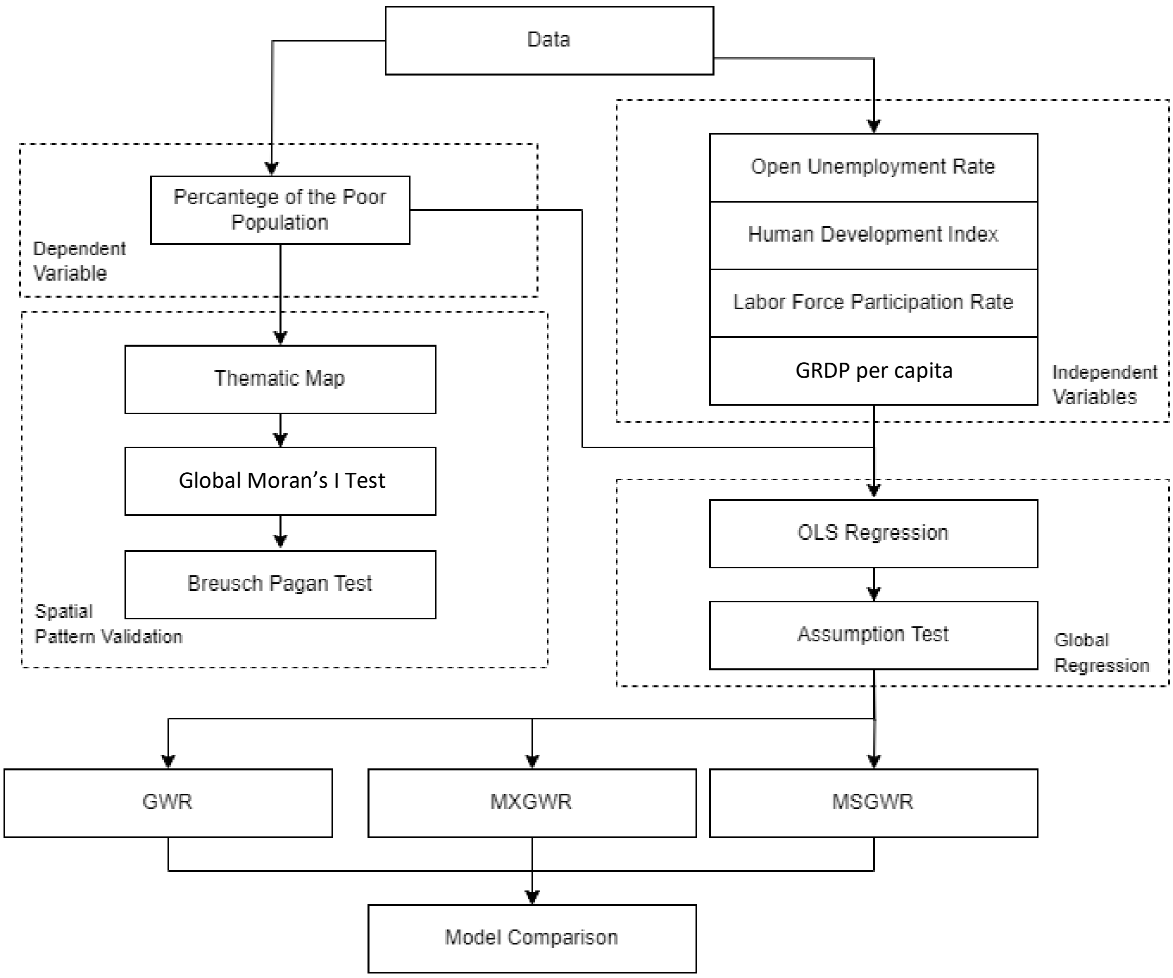


Figure 1. Research Flowchart

2.1 Spatial Pattern Identification

Moran's I provides a global measure of spatial autocorrelation, which means it assesses the overall spatial pattern in the entire dataset rather than focusing on specific locations [5]. Due to its ease of interpretation, Moran's Index is employed to analyze the autocorrelation pattern in PPP with the following formula.

$$\boldsymbol{I} = \frac{\frac{\boldsymbol{n}}{\boldsymbol{s_0}}\sum_i^n \sum_j^n \boldsymbol{w_{ij}}(\boldsymbol{x_i}-\overline{\boldsymbol{x}})(\boldsymbol{x_j}-\overline{\boldsymbol{x}})}{\sum_j^n(\boldsymbol{x_i}-\overline{\boldsymbol{x}})^2} \tag{1}$$

where $S_0 = \sum_i^n \sum_j^n w_{ij}$, w_{ij} is element of spatial weighting matrix, n is number of observations, x_i is value at i-th location, $\bar{x}$ is the mean value of x_i from n locations. If $I > 0$ there is positive spatial autocorrelation. If $I = 0$ there is no autocorrelation and if $I < 0$ there is negative spatial autocorrelation. The greater absolute value of I means a stronger spatial correlation.

Spatial heterogeneity can be tested using the Breusch-Pagan test with the following statistical tests.

$$BP = \left(\frac{1}{2}\right)\boldsymbol{f}^T\boldsymbol{Z}\,(\boldsymbol{Z}^T\boldsymbol{Z}\,)^{-1}\,\boldsymbol{Z}^T\boldsymbol{f} \sim \chi_p^2 \tag{2}$$

where $f = \left(\frac{e_i^2}{\sigma^2}-1\right)$, e_i are least square residuals for i-th observation, $Z_{n\,x\,(p+1)}$ is a matrix of predictor variables containing standardized normalized vectors (z) for each observation. If $BP > \chi_p^2$ then the null hypothesis is rejected identifying that there is spatial heterogeneity so that the GWR model and its expansion can be used.

2.2 Geographically Weighted Regression (GWR) Model and Its Extension

The global regression model assumes a constant relationship between response variables and predictors between locations, with the following formula.

$$\boldsymbol{Y_i} = \sum_j^n \boldsymbol{\beta_j X_{ij}} + \boldsymbol{\varepsilon_i} \tag{3}$$

where i is the index of the observation and j is the index of the predictor variable, β_j is regression coefficient, X_{ij} is the predictor variable, Y_i is the response variable and ε_i is the error component. In the global regression model, all parameters are assumed to have a global or stationary effect on location. To capture the complexity of the PPP pattern, non-stationary spatial aspects must be included in the model so a spatial model is needed. Geographically Weighted Regression (GWR) is a spatial model that can capture local spatial variations [20] with the following formulation.

$$\boldsymbol{Y_i} = \sum_j^p \boldsymbol{\beta_{ij}(u_i, v_i) X_{ij}} + \boldsymbol{\varepsilon_i} \tag{4}$$

where (u_i, v_i) is the geographical location of i-th observation and parameter $\beta_k(u_i, v_i)$ is a function of (u_i, v_i) for i-th observation.

The GWR model assumes that all predictor variables locally influence the response variable. The PPP pattern is caused by complex factors that often have global and local influences simultaneously. The MXGWR model is an extension of the GWR which can simultaneously explain global and local spatial relationships [20] with the following equation.

$$\boldsymbol{Y_i} = \sum_k^q \boldsymbol{\gamma_k X_{ik}} + \sum_j^p \boldsymbol{\beta_{ij}(u_i, v_i) X_{ij}} + \boldsymbol{\varepsilon_i} \tag{5}$$

with k is the globally valid index of the predictor variable and j is the index of the predictor variable that applies locally to Y_i.

The GWR and MXGWR models assume that local relationships vary over the same spatial scale. MSGWR is an extension of GWR which facilitates conditional relationships between response variables and predictor variables varying at different spatial scales [21] with the following formula.

$$\boldsymbol{Y_i = \sum_{j}^{p} \beta_{bwj}(u_i, v_i) X_{ij} + \varepsilon_i} \tag{6}$$

with bwj as the bandwidth used for calibrating the j conditional relationship. MSGWR is a form of the generalized additive model (GAM) which calibrates using a back-fitting algorithm [22]. If MSGWR is formulated into GAM, it is obtained as follows.

$$Y_i = \sum_{j}^{p} f_{ij} + \varepsilon_i \tag{7}$$

where f_{ij} have the same role as $\beta_{bwj}(u_i, v_i)X_{ij}$ in **Equation (6)**. f_{ij} is j-th additive component which acts as a smoothing function on the j-th predictor variable the i-th observation**.** The model calibration process will produce a bandwidth for each j-th predictor variable. This difference in bandwidth indicates the difference in spatial scale. It captures the effect of the scale used in the spatial model-building process. Thus, the MSGWR model can explain spatial heterogeneity patterns more accurately [22].

2.3 Weighting Function Selection

In the GWR model, the kernel has an important role in determining the weighting function to be used in modeling. The next weighting function will produce different parameter estimates according to location. Kernels in GWR are divided into two types, namely fixed and adaptive kernels. The Adaptive kernel uses a different bandwidth value for each spatial observation while the Fixed kernel uses the same bandwidth value. The kernels that will be compared in this study are the Adaptive Bi-square and Fixed Gaussian kernels with the following equations.

Adaptive Bi-square

$$\boldsymbol{w_{ij} = \begin{cases} \left[1 - \left(\dfrac{d_{ij}}{b_{i(q)}}\right)^2\right]^2, & d_{ij} < b \\ 0, & others \end{cases}} \tag{8}$$

Fixed Gaussian

$$\boldsymbol{w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right]^2} \tag{9}$$

where d_{ij} is the distance between the i-th and j-th observation. This study uses the Euclidean distance as follows $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$. Then, b is the bandwidth value that will limit the number of nearest neighbors affecting the kernel. The greater the bandwidth value of a predictor variable, the more global the influence of that variable on the response variable. This research will use Golden Search to find the best bandwidth value, with the smallest AICc criteria.

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{10}$$

with $AIC = 2k - 2\log(L)$ [23], where k is the number of predictor variables and L is the likelihood function of the model. The Golden Search is an iterative process to find the smallest AICc in the minimum and maximum range of observation locations [10].

Additionally, to compare the models, it is crucial to assess them based on various criteria as follows: 1) R^2 and AICc as a measure of goodness of fit, 2) the Simultaneous Test has a p-value< 0.05, meaning that there are predictor variables that significantly affect the response variable. 3) Variance Inflation Factor (VIF) < 10 to indicate there is no multicollinearity between predictor variables [24]. 4) The Jarque-Bera test is significant at α = 5% to show that the residuals are normally distributed [25]. 5) Morans' I and Breush Pagan tests are significant at α =5% [5].

3. RESULTS AND DISCUSSION

3.1. Distribution of the Percentage of Poor Population (PPP)

To initially understand the spatial distribution of PPP, a thematic map illustrating PPP across districts/cities on Java Island is presented in Figure 2. The PPP is visualized in a gradient of colors, where the higher the PPP of a regency/city, the map color becomes increasingly red, while the lower the PPP, the map color turns greener.

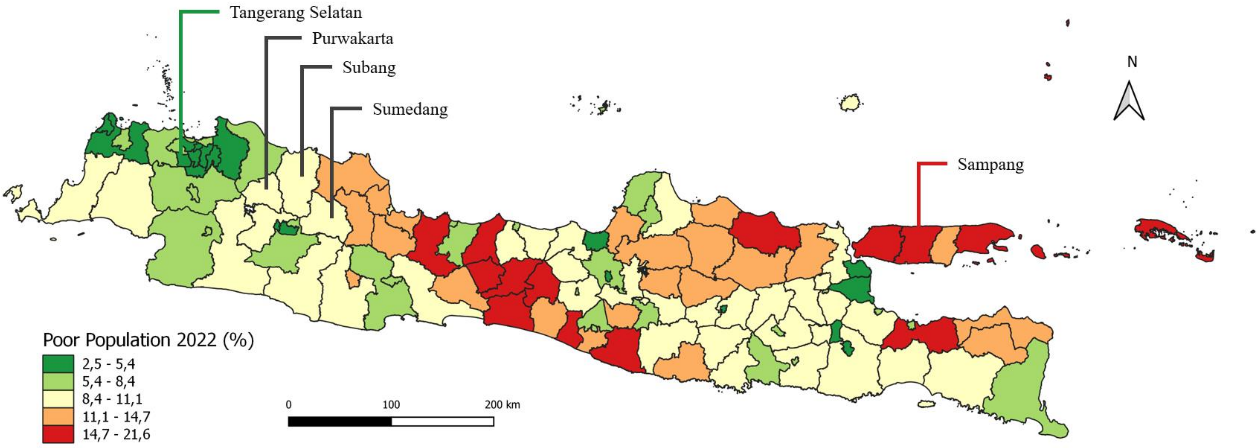


Figure 2. Thematic map on the percentage of poor population in Java Island in 2022

From Figure 2, it can be seen that PPP in Java varies according to districts/cities. The district/city with the lowest PPP was South Tangerang (2.5%) and the highest was Sampang (21.16%). It shows that districts/cities with lower PPP tend to be surrounded by districts/cities that also have a low PPP. This condition occurs in the western and eastern parts of the island of Java. Meanwhile, in the central part of the island of Java, the districts/cities that have a high PPP are also surrounded by regencies/cities with a high PPP. This condition indicates that there is a positive autocorrelation in PPP.

In addition, several regions that are close to each other have different levels of poverty even though they belong to the same category. As in West Java province, many districts/cities have PPP categories ranging from 8.4 to 11.1, but with varying levels of PPP. For instance, in Subang (9.75%), the PPP is lower than in Sumedang (10.14%), but slightly higher than in Purwakarta (8.65%), yet all fall under the same category, indicated by the color yellow on the thematic map. This diversity of PPP values indicates the occurrence of spatial heterogeneity.

3.2. Spatial Autocorrelation and Heterogeneity

According to the descriptive analysis above, we emphasize the existence of spatial autocorrelation and spatial heterogeneity in PPP data. Testing for spatial autocorrelation and spatial heterogeneity can be carried out using Moran's I and Breusch-Pagan tests, respectively.

Table 2. Assumption Tests

Assumption Testing	Test	Test Statistics	p-value
Spatial Autocorrelation	Moran's I	4.780	0.000
Spatial Heteroscedasticity	*Breusch-Pagan*	9.722	0.045
Normality	*Jarque Bera*	4.512	0.104

In Table 2 the p-value of the Moran's I test is less than the 5% significance level, it represents that there is a spatial autocorrelation in PPP. Similar results were obtained in the Pagan Breusch test, so it was proven that there is spatial heterogeneity in PPP. The existence of spatial autocorrelation violates the assumptions of OLS estimation which causes the estimation results to be biased, thus we use the geospatial model to explain the PPP by considering the spatial heterogeneity effects. This paper conducts the GWR, MXGWR, and MSGWR models that assume the predictor variable must be normally distributed.

Table 3. VIF and Pearson Correlation of Predictor Variables

Variable	OUR	HDI	LFPR	GDP per capita
VIF	1.724	1.294	1.845	1.268
Pearson Correlation	-0.403**	-0.657**	0.471**	-0.319**

Thus, we need to conduct some tests to evaluate the normality and non-multicollinearity assumption. Table 2 shows that the p-value from the Jarque-Bera test is greater than 5%, indicating that the PPP variable is normally distributed. In Table 3, it can be observed that all predictor variables meet the assumption of non-multicollinearity (VIF < 10) and are significantly correlated with PPP. Based on the categorization of correlation coefficients in [26], OUR has a moderate negative relationship with PPP, HDI has a strong negative relationship with PPP, LFPR has a moderate positive relationship with PPP, and GRDP per capita has a weak negative relationship with PPP. Variables showing a negative correlation (OUR, HDI, and GRDP per capita) suggest that as their values rise, PPP values decline. Conversely, a variable with a positive correlation (LFPR) suggests that as it increases, PPP also increases. Since each predictor variable meets the assumption of non-multicollinearity and is significantly correlated with PPP, these variables are deemed suitable for use as predictor variables.

3.3. Spatial Models of The Percentage of Poor Population (PPP)

This research aims to identify the best model for explaining the spatial pattern of PPP and related socioeconomic and demographic attributes on Java Island in 2022 using GWR, MXGWR, and MSGWR models. Furthermore, several model scenarios were conducted to explore the impact of kernel usage on model performance, thus we employ the Adaptive Bi-square and Fixed Gaussian kernels.

Table 4. R^2 and AICc Values for Modeling Scenarios

Scenario	Model	Kernel	R^2	AICc
1	OLS	-	0.543	574.689
2	GWR	Adaptive Bi-square	0.667	567.492
3	MXGWR	Adaptive Bi-square	0.655	566.115
4	MSGWR	Adaptive Bi-square	0.675	551.646
5	GWR	Fixed Gaussian	0.722	558.502
6	MXGWR	Fixed Gaussian	0.718	557.846
7	MSGWR	Fixed Gaussian	0.762	533.989

It can be observed in Table 4, that the R^2 value increases and the AICc decreases for all spatial models. This indicates that the GWR model and its extensions are better suited compared to the OLS model. However, the GWR model does not consider modeling with both global and local predictor variables simultaneously, which is why the MXGWR model will be used. It can be seen that the MXGWR model has a lower AICc value than GWR, indicating that MXGWR is more parsimonious than the GWR model. However, the MXGWR model does not take into account the spatial scale differences for each predictor variable, so the MSGWR model will be used. MSGWR successfully increases the R^2 and decreases the AICc, demonstrating that MSGWR is a better fit for the data. Among the various modeling scenarios, it is evident that the MSGWR model with the Fixed Gaussian kernel has the lowest AICc value (533.989) and the highest R^2 (0.762). Therefore, the MSGWR model with the Fixed Gaussian kernel is selected as the best model for modeling the influence of predictor variables on PPP.

Furthermore, different kernels affected the performance of the spatial models generated as well. In Table 4, it can be seen that the models using the Fixed Gaussian kernel overall provide higher performance compared to the models using the Adaptive Bi-square kernel. Hence, for further comparative analysis, the Fixed Bi-square kernel will be used.

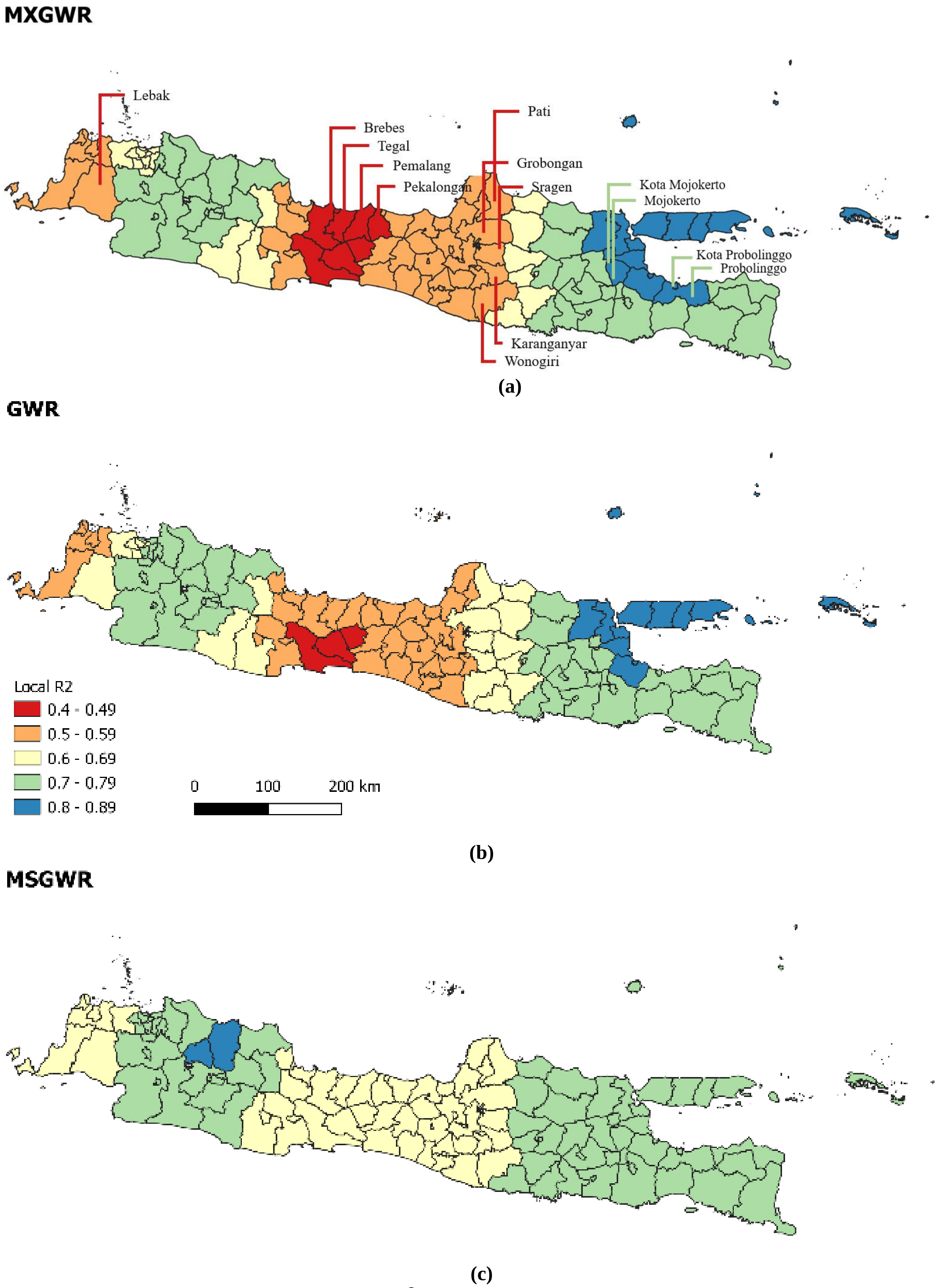


Figure 3. **Thematic Maps of R^2 for MXGWR (a), GWR (b) and MSGWR (c) models**

In **Figure 3,** the spatial distribution of R^2 values in the GWR, MXGWR, and MSGWR models are shown. It can be observed that the GWR and MXGWR models exhibit a similar pattern of R^2 distribution. However, it can be seen in **Figure 3** (b) that when estimated with the MXGWR model, there are 10 districts with lower R^2 compared to the GWR model, namely Lebak, Brebes, Tegal, Pemalang, Pekalongan, Pati, Grobongan, Sragen, Karanganyar, and Wonogiri. Additionally, four districts have a higher R^2 compared to the GWR model, namely Mojokerto City, Mojokerto, Probolinggo City, and Probolinggo. Therefore, despite providing a similar pattern, the implementation of global variables in the MXGWR model still affects the model's performance.

On the other hand, the R^2 distribution pattern given by GWR and MXGWR is significantly different from the MSGWR model (**Figure 3**). The MSGWR model increases the R^2 value in many districts, resulting in no R^2 values falling within the range of 0.4-0.49 and 0.5-0.59. However, some districts are more accurately estimated with GWR/MXGWR compared to MSGWR. Thus, it can be said that the implementation of different spatial scales in the MSGWR model can either improve or decrease the model's performance partially.

Table 5. Parameter Estimation Summary

Model	Variable	Mean	STD	Min	Median	Max	Bandwidth
GWR	Intercept	33.793	10.007	9.064	41.548	52.111	0.794
	OUR	-0.147	0.216	-0.826	0.004	0.240	
	HDI	-0.426	0.103	-0.708	-0.347	-0.273	
	LFPR	0.114	0.109	-0.036	0.152	0.556	
	GRDP per capita	-0.003	0.007	-0.023	-0.0004	0.009	
MXGWR	Intercept	34.522	11.897	5.145	43.583	53.963	0.746
	OUR	-0.153	0.214	-0.913	-0.012	0.279	
	HDI	-0.440	0.107	-0.717	-0.366	-0.248	
	LFPR	0.117	0.121	-0.055	0.160	0.615	
	GRDP per capita			0.001			
MSGWR	Intercept	26.939	1.706	22.96	26.894	32.109	0.330
	OUR	0.097	0.000	0.097	0.097	0.097	17.250
	HDI	-0.369	0.000	-0.37	-0.369	-0.369	17.250
	LFPR	0.132	0.000	0.132	0.132	0.132	17.250
	GRDP per capita	0.001	0.002	-0.002	0.001	0.004	2.970

Table 5 provides a summary of parameter estimates from GWR, MXGWR, and MSGWR models. The GWR and MSGWR models assume local influence for each predictor variable, while MXGWR facilitates the use of both local and global variables in a single model. From the MXGWR model, it can be observed that the variable GRDP per capita has a global interaction with PPP, while the other predictors affect PPP locally. The estimated coefficient value for GRDP per capita in the MXGWR model is 0.001, and this influence applies to all districts. On the other hand, OUR has varying influences with a minimum value of 0.193 dan maximum value of 0.279, HDI ranges from -0.717 to -0.248, and LFPR ranges from -0.055 to 0.16.

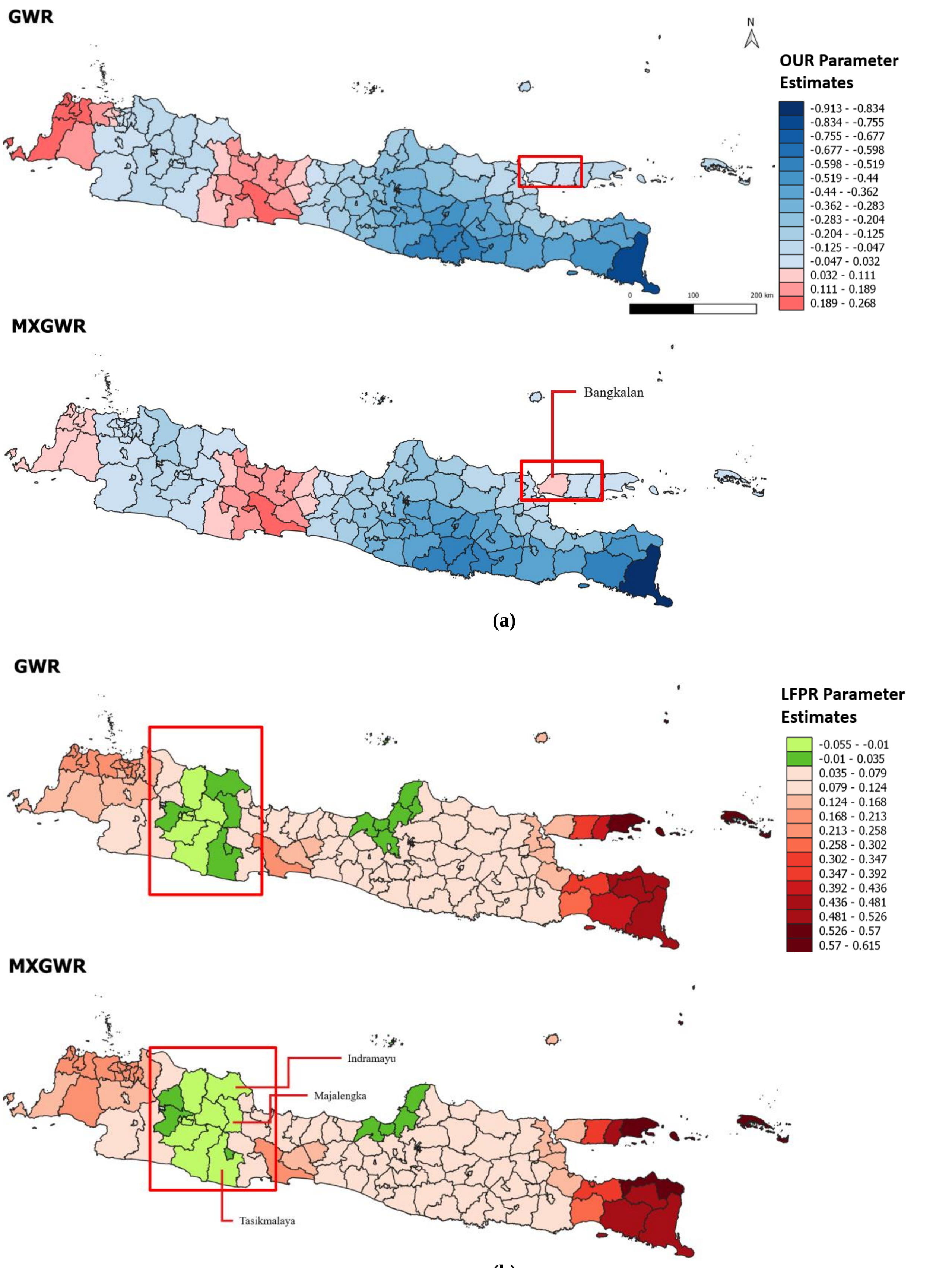


Figure 4. Thematic Maps of Estimated Parameter Outcomes for OUR (a) and LFPR(b) in GWR and MXGWR models

Figure 4 displays the parameter estimation mapping of GWR and MXGWR for two predictor variables, namely OUR and LFPR. It can be observed that the pattern of OUR's influence and LFPR's influence on PPP tends to be similar between the GWR and MXGWR models. However, in **Figure 4** (a), there is one district that is estimated to have a different direction of relationship, which is Bangkalan. In the

GWR model for Bangkalan, OUR has a negative influence (-0.125 - -0.047), whereas in the MXGWR model, OUR has a positive influence (0.032 – 0.111). Additionally, several districts/cities are estimated to have the same direction of relationship but with different magnitudes. For example, in **Figure 4** (b) for Tasikmalaya, Majalengka, and Indramayu. These areas have LFPR parameter estimates in light green color (-0.055 - -0.01) in the MXGWR model, whereas in the GWR model, they are in dark green (-0.01 – 0.035). This indicates that although the estimation patterns produced by the GWR and MXGWR models are similar, there are still differences in the estimation results in terms of both the magnitude and direction of the relationships.

Next, this study will compare the GWR and MSGWR models to examine the influence of spatial scale on parameter estimation. In **Table 5**, it is observed that the GWR and MXGWR models use the same bandwidth for each variable. This issue can be addressed through the utilization of MSGWR, as it enables the application of distinct bandwidths for each variable. The relationship between predictor variables and PPP in the MSGWR model shows local spatial effects, with the model formation process varying at the scale used. The variables OUR, HDI, and LFPR have an impact on PPP with an optimal bandwidth of 17 nearest neighbors (**Table 5**). For example, Magetan has 17 nearest neighbors, namely Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobongan, Blora, Surakarta, Pacitan, Ponorogo, Trenggalek, Tulungagung, Nganjuk, Madiun, Ngawi, Bojonegoro, Kediri, and Madiun City. On the other hand, the variable GRDP per capita affects PPP at a relatively closer distance, with an optimal bandwidth of 3 nearest neighbors. For instance, Magetan has 3 nearest neighbors, namely Madiun, Ngawi, and Madiun City. This indicates that OUR, HDI, and LFPR in a district/city have broader or more global influence on other districts/cities compared to the influence exerted by GRDP per capita.

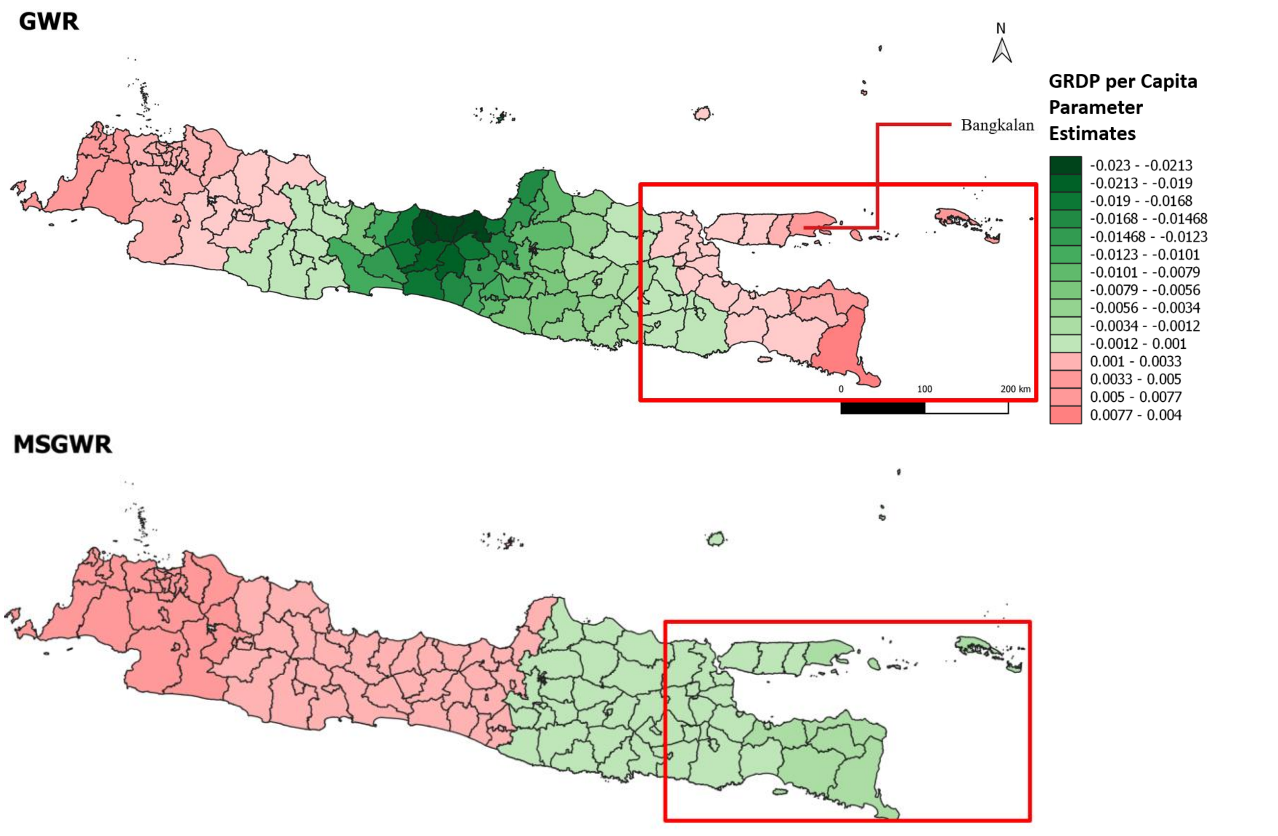
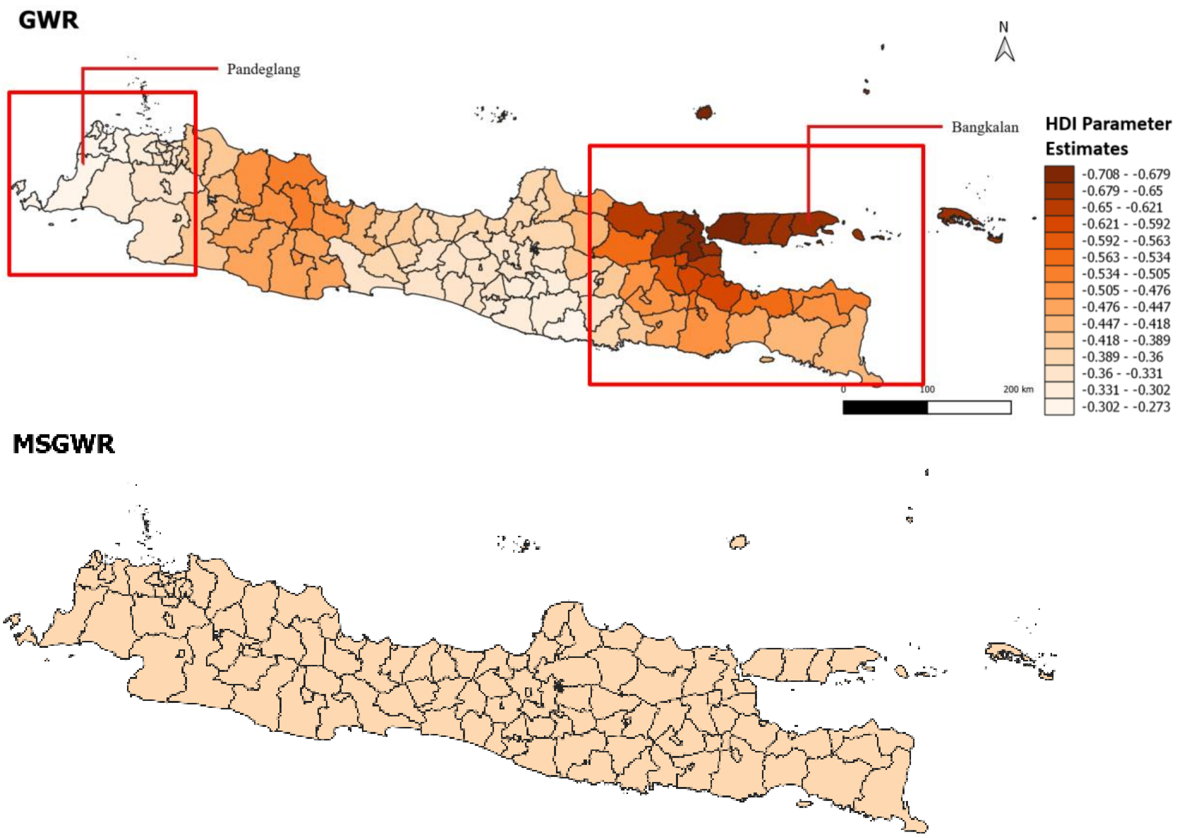


Figure 5. **Thematic Maps of Estimated Parameter Outcomes for GRDP per capita from GWR and MSGWR Models**



(b)

Figure 6. Thematic Maps of Estimated Parameter Outcomes for HDI from GWR and MSGWR Models

To gain a clearer understanding of the spatial scale variation, thematic maps of parameter estimation from the GWR and MGWR models are shown in **Figure 5** and **Figure 6**. It can be observed that there are differences in the parameter estimation results for GRDP per capita and HDI between the GWR and MSGWR models because of the difference in bandwidth usage. GWR employs a single bandwidth, which is the average of the bandwidth obtained from calibration results. On the other hand, the MSGWR model finds and uses optimal bandwidth for each relationship in the model.

In **Figure 5**, the parameter estimation results for GRDP per capita differ significantly between the GWR and MSGWR models. The GWR model has a smaller bandwidth (0.794) compared to the bandwidth for GRDP per capita in MSGWR (2.970), as seen in **Table 5**. Consequently, GWR provides more locally-focused parameters compared to MSGWR. Moreover, these differences in results indicate that there are some districts/cities estimated with different relationships in both models. For instance, in **Figure 6**, in the eastern part of Java Island, the GWR model predicts a positive relationship between GRDP per capita and PPP for Bangkalan and its surrounding areas, marked in red color. However, the estimation using the MSGWR model shows a negative relationship, marked in green color. This demonstrates that the use of different bandwidths leads to different parameter values and directions of relationships.

In **Figure 6**, the parameter estimation results for HDI also differ significantly between the GWR and MSGWR models. The local parameter estimation for HDI from the MSGWR model appears uniform within the range of -0.38 to -0.36. On the other hand, the GWR model identifies spatial variation in the distribution pattern of HDI parameter estimates. For example, in **Figure 6**, in the western part of Java Island, such as in Pandeglang and its surrounding areas, there is a lower influence of HDI on PPP, indicated by lighter colors on the map. On the other hand, in the eastern part of Java Island, like Bangkalan and its neighboring areas, there is a stronger influence of HDI on PPP, indicated by darker colors on the map. This significant disparity in results between the GWR and MSGWR models is caused by the difference in bandwidth values used. In this case, the GWR bandwidth (0.794) is lower than the MSGWR bandwidth (17.250). MSGWR with bandwidth equal to 17.250 states that the HDI of one area will affect the PPP of its 17 neighboring areas. Therefore, MSGWR provides more globally oriented parameter estimates compared to GWR. This indicates that the larger the difference in bandwidth used between models, the greater the disparity in estimation results.

Hence, it can be concluded that spatial models such as GWR, MXGWR, and MSGWR are highly sensitive to the properties of the model used. The use of local variables, different kernels, and scales leads to different parameter estimation results and performance. Therefore, selecting the spatial model with the best model properties is crucial to obtaining accurate and targeted policies.

3.4. Fitted Models of PPP

Table 4 shows that MSGWR with the Fixed Gaussian kernel is the best model with the highest R^2 and lowest AICc. The parameter estimation results of this model indicate that the HDI and LFPR variables have a significant influence on PPP in all districts/cities. An increase in HDI will decrease PPP, while an increase in LFPR will increase PPP. The following is one MSGWR regression model with an R^2 of 0.78 for the city of Surabaya.

$$PPP = 27.627 + 0.096\ OUR - 0.369\ HDI^{**} + 0.132\ LFPR^{*} - 0.0007\ GRDPpercapita$$

*p<0.1. **p<0.05

A R^2 value of 0.78 means that 78% of the variation in PPP in Surabaya can be explained by OUR, HDI, LFPR, and GRDP per capita, while the remaining 12% is explained by other variables that are not included in the model. The regression coefficient for HDI is -0.369^{**}, which means that HDI has a negative and significant effect on PPP. For every unit increase in HDI, PPP decreases by 0.369% ceteris paribus. It emphasizes that human development increases economic productivity at both national and regional levels, thereby increasing people's income [27].

On the other hand, the regression coefficient for LFPR is 0.132^{*}, indicating that LFPR has a positive and significant effect on PPP. For every unit increase in LFPR, PPP increases by 0.132% ceteris paribus. This is caused by the majority of workers in Indonesia working in informal sector [28], indicates that many workers have not received fair wages or are not working full-time, resulting in underutilization of labor and low-income [29]. Furthermore, a significant increase in LFPR among highly educated individuals increases unemployment in Indonesia [30]. Similar studies also indicate that the increasing TPAK of males reduces labor absorption in the informal sector [29]. The subsequent increase in unemployment leads to a slowdown in economic growth [31], causing PPP to rise.

4. CONCLUSIONS

The research shows that the differences in the model properties used in spatial regression models (such as the implementation of global variables, different kernels, and scales) influence the bandwidth value. This bandwidth value subsequently impacts the model's performance and parameter estimations. The MXGWR model has a smaller AICc than OLS and GWR because it can explore predictor variables that have global or local influences on the response variable simultaneously. The presence of global variables makes the model more parsimonious. Furthermore, the calibration in the MSGWR model facilitates the use of optimal bandwidths that vary for each relationship in the model. Therefore, MSGWR improves model performance in terms of goodness-of-fit and prediction accuracy compared to GWR and MXGWR models. The MSGWR model, as the best model explaining the relationship between predictor variables and PPP, states that HDI and LFPR have significant negative and positive effects, respectively, on PPP in all districts in Java Island 2022. Since poverty is a complex problem over time, a comprehensive study is needed, not only involving relationships between variables, considering the spatial effects but also over time. So, further research is necessary to observe those effects simultaneously.

REFERENCES

[1] I. Y. Prasada, T. Fatma, M. Yulhar, and T. A. Rosa, “Determinants of poverty rate in Java Island : Poverty alleviation policy,” vol. 18, no. December, pp. 95–104, 2020.

[2] S. A. Gultom and A. P. Utomo, “Analisis Kemiskinan di Pulau Jawa Saat Pandemi Covid-19 Tahun 2020 dengan Metode Geographically Weighted Regression (GWR),” pp. 1349–1358, 2020.

[3] R. N. Harahap, “Implementasi Geographically Weighted Regression (GWR) dan Mixed Geographically Weighted Regression (MGWR) Dalam Perhitungan Penduduk Miskin,” 2022.

[4] S. Zakaria, N. A. Rahman, and N. Zaini, “Spatial analysis on distribution patterns of poverty incidences in Peninsular Malaysia,” *AIP Conf Proc*, vol. 2184, no. December, 2019, doi: 10.1063/1.5136415.

[5] L. Anselin, “Local Indicators of Spatial Association—LISA,” *Geogr Anal*, vol. 27, no. 2, pp. 93–115, 1995, doi: 10.1111/j.1538-4632.1995.tb00338.x.

[6] F. Gao, C. languille, K. karzazi, M. Guhl, B. Boukebous, and S. Deguen, “Efficiency of fine scale and spatial regression in modelling associations between healthcare service spatial accessibility and their utilization,” *Int J Health Geogr*, vol. 20, no. 1, pp. 1–19, 2021, doi: 10.1186/s12942-021-00276-y.

[7] T. Herbiansyah, I. Yahya, Baharuddin, Agusrawati, Ruslan, and L. Laome, “PEMODELAN PRODUK DOMESTIK REGIONAL BRUTO DI INDONESIA DENGAN GEOGRAPHICALLY WEIGHTED REGRESSION,” *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, vol. 07, no. 1, pp. 23–33, 2018.

[8] R. B. Prasetyo, “Analisis Spasial pada Aglomerasi Industri Manufaktur di Pulau Jawa,” *Jurnal Sipil*, vol. 1, no. August, pp. 1–35, 2015.

[9] W. C. D. Weku, “Eksplorasi efektifitas model spasial untuk menjelaskan hubungan antara penduduk dan infrastruktur terhadap kesejahteraan masyarakat Kota Manado,” *Aiti*, vol. 17, no. 2, pp. 130–142, 2021, doi: 10.24246/aiti.v17i2.130-142.

[10] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression : Analysis of Spatially Varying Relationship*. 2002.

[11] F. Chen and C. L. Mei, “Scale-adaptive estimation of mixed geographically weighted regression models,” *Econ Model*, vol. 94, no. February 2020, pp. 737–747, 2021, doi: 10.1016/j.econmod.2020.02.015.

[12] J. C. Martori, R. Lagonigro, and R. Iglesias-Pascual, “Social status and air quality in Barcelona: A socio-ecological approach,” *Sustain Cities Soc*, vol. 87, no. June, p. 104210, 2022, doi: 10.1016/j.scs.2022.104210.

[13] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, “Geographically weighted regression: a method for exploring spatial nonstationarity,” *Geogr Anal*, vol. 28, no. 4, pp. 281–298, 1996, doi: 10.1111/j.1538-4632.1996.tb00936.x.

[14] A. S. Fotheringham, W. Yang, and W. Kang, “Multiscale Geographically Weighted Regression (MGWR),” *Ann Am Assoc Geogr*, vol. 107, no. 6, pp. 1247–1265, 2017, doi: 10.1080/24694452.2017.1352480.

[15] C. L. Mei, S. Y. He, and K. T. Fang, “A note on the mixed geographically weighted regression model,” *J Reg Sci*, vol. 44, no. 1, pp. 143–157, 2004, doi: 10.1111/j.1085-9489.2004.00331.x.

[16] C. Jin, J. Xu, and Z. Huang, “Spatiotemporal analysis of regional tourism development: A semiparametric Geographically Weighted Regression model approach,” *Habitat Int*, vol. 87, no. March, pp. 1–10, 2019, doi: 10.1016/j.habitatint.2019.03.011.

[17] H. Yu, A. S. Fotheringham, Z. Li, T. Oshan, W. Kang, and L. J. Wolf, “Inference in Multiscale Geographically Weighted Regression,” *Geogr Anal*, vol. 52, no. 1, pp. 87–106, 2020, doi: 10.1111/gean.12189.

[18] M. Y. Darsyah and R. Wasono, “Pemodelan Mixed Geographically Weighted Regression pada Tingkat Kemiskinan di Provinsi Jawa Tengah The Model of Mixed Geographically Weighted Regression (MGWR) for Poverty Level in ... Pemodelan Mixed Geographically Weighted Regression pada Tingkat Kemi,” no. June 2016, 2014.

[19] BPS, “Indeks Pembangunan Manusia 2022,” 2022.

[20] B. Lu, M. Charlton, P. Harris, and A. S. Fotheringham, “Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data,” *International Journal of Geographical Information Science*, vol. 28, no. 4, pp. 660–681, 2014, doi: 10.1080/13658816.2013.865739.

[21] W. Yang, “An extension of geographically weighted regression with flexible bandwidths,” p. 167, 2014.

[22] A. S. Fotheringham and T. M. Oshan, “Geographically weighted regression and multicollinearity: dispelling the myth,” *J Geogr Syst*, vol. 18, no. 4, pp. 303–329, 2016, doi: 10.1007/s10109-016-0239-5.

[23] Akaike, “A New Look at the Statistical Model Identification,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90-A, no. 12, pp. 2762–2769, 1974, doi: 10.1093/ietfec/e90-a.12.2762.

[24] R. M. O’Brien, “A caution regarding rules of thumb for variance inflation factors,” *Qual Quant*, vol. 41, no. 5, pp. 673–690, Oct. 2007, doi: 10.1007/s11135-006-9018-6.

[25] T. 1967- Thadewald and H. 1939- Büning, *Jarque-Bera test and its competitors for testing normality a power comparison*.

[26] Sugiyono, “Metode Penelitian Kuantitatif Kualitatif dan R&D,” 2010.

[27] L. L. N. El Hasanah, J. L. Panjawa, P. Prastowo, and J. A. Prakoso, “How Human Development Effect Poverty Alleviation in Origin and Expansion Regions?,” *MIMBAR : Jurnal Sosial dan Pembangunan*, no. 10, pp. 90–98, 2022, doi: 10.29313/mimbar.v0i0.8514.

[28] BPS, “Badan Pusat Statistik.” Accessed: Oct. 06, 2023. [Online]. Available: https://ppukab.bps.go.id/indicator/6/252/1/-sakernas-jumlah-dan-persentase-penduduk-berumur-15-tahun-ke-atas-yang-bekerja-menurut-formal-informal.html

[29] C. Birgitta, “ANALISIS PENGARUH TPAK, PENGANGGURAN, UPAH BERSIH, DAN PENDIDIKAN TERHADAP PENYERAPAN TENAGA KERJA SEKTOR INFORMAL DI INDONESIA TAHUN 2015-2020,” 2021.

[30] R. Kartika and I. Muslim, “ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PENGANGGURAN DI INDONESIA,” pp. 64–66, 2020.

[31] M. Asfia, *Ekonomi Makro*. Buku Ajar Ekonometrika, Yogyakarta: Deepublish., 2006.