

Comparison of Adaptive Bisquare And Adaptive Gaussian Weighting Functions In Poverty Case Modeling Using Geographically Weighted Regression In Papua Province

Aidawayati Rangkuti¹, Nurtiti Sunusi^{2*}, Arya Winanda Tahir³

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences
Hasanuddin University, Indonesia

^{2,3}Department of Statistics, Faculty of Mathematics and Natural Sciences
Hasanuddin University, Indonesia

*Email: nurtitisunusi@unhas.ac.id

Abstract: Papua has the largest proportion of poor people in Indonesia. Predicting the poverty cases in Papua Province in 2021 using Geographically Weighted Regression (GWR) analysis, this study compared the adaptive bi-square and adaptive gaussian weighting functions. The SUSENAS data from the BPS's 2021 National Socio-Economic Survey were used. The analysis step that was conducted was using the OLS method. Using the test results, one significant variable was discovered. Additionally, the GWR approach was used for testing, and the values R^2 and AIC of the GWR models with adaptive bi-square and adaptive gaussian weighting functions were compared. The value R^2 and AIC of the GWR model with the bi-square adaptive weighting function were 94.5% and 147.0325, based on the findings. Using a gaussian adaptive weighting function, the GWR model's value R^2 and AIC were 66.6% and 184.26. The GWR model with bi-square adaptive weighting function has the highest value R^2 and the lowest AIC.

Keywords: Poverty, Geographically Weighted Regression, Adaptive Gaussian, Adaptive Bi-square

1. INTRODUCTION

The problem of poverty is still a big problem throughout the history of Indonesia as a country. Poverty is a condition in which there is a shortage of the usual things to have such as food, clothing, shelter and drinking water. These things are closely related to quality of life (Arfiani, 2019). Poverty has made millions of children unable to get quality education, difficulty paying for health, lack of savings and investment, lack of access to public services and lack of jobs and other problems. Poverty causes people to be willing to sacrifice anything for the sake of their safety and necessities of life, even risking their physical strength.

Poverty is a very complex and chronic problem, because it requires a response with the right analysis, involves all components of the problem, and requires an appropriate, sustainable and not temporary coping strategy. Poverty alleviation efforts are carried out by providing basic needs such as food, health and education services, expanding employment opportunities, agricultural development, providing revolving funds through a credit system, infrastructure development and assistance, sanitation counseling and so on. However, everything is still material-oriented, so that in the long run its sustainability depends on budget availability and government commitment.

In the literature, poverty has become an interesting discussion for researchers, but basically it can be divided into three definitions of poverty, namely; absolute poverty, relative

poverty and cultural poverty. Someone belongs to the absolute poor if their income results are below the poverty line, not sufficient to meet the minimum living needs: food, clothing, health, shelter, education. A person who is classified as relatively poor actually lives above the poverty line but is still below the means of the surrounding community. While cultural poverty is closely related to the attitude of a person or group of people who do not want to try to improve their level of life even though there are efforts from other parties to help them.

Based on data obtained from BPS Papua Province (2022), Papua occupies the first position with the largest percentage of poor people among all provinces in Indonesia. The number of poor people in Papua in September 2021 reached 944.49 thousand people. Compared to March 2021, the number of poor people increased by 24.05 thousand people. Meanwhile, when compared to September 2020, the number of poor people increased by 32.36 thousand people. The percentage of poor people in September 2021 was recorded at 27.38 percent, up 0.52 percentage points from March 2021 and up 0.58 percentage points from September 2020.

The analysis to find out the factors that influence the percentage of poor people in Papua is regression analysis. Factors that affect the percentage of poor people in Papua may vary in each district/city or in this case depend on the condition of the region, so that spatial heterogeneity is considered. Analysis of data with spatial heterogeneity cannot be carried out with ordinary regression analysis, so it is necessary to use methods that are able to produce local models according to the

factors that influence each observation location. One method that is capable of producing local models at each observation location is Geographically weighted regression (GWR).

GWR is a classical regression model that is capable of producing different parameters at each point in the observation location. GWR is a development of the global regression model by adding geographic weights at the observation location where the data is taken. Based on research conducted by Haryanto & Andriani (2021), it was concluded that GWR modeling is more effective in describing the number of poor people. In line with that, in research conducted by Firdang et al, (2021), also concluded that the GWR model is better than the global multiple regression model in determining the factors that affect the poverty rate of each province in Indonesia.

In GWR, an observation is given a weight according to the location of the observation so that the weight of an observation varies. The role of weighting is very important because it represents the location of the observation data with one another (Caraka & Yasin, 2017). In a study conducted by Pratiwi et al (2019), it was shown that the GWR model with a fixed gaussian weighting function was better than bi-square adaptive weighting. Almost the same conclusion was obtained in the study of Lutfiani et al, (2019), which showed that fixed gaussian weights were better than fixed bi-square weights. From some literature, it turns out that Gaussian weights are more widely used than fixed bi-square and adaptive square weights. In this study, the authors were interested in seeing a comparison of the GWR model with adaptive gaussian and bi-square adaptive weights.

2. METHOD

This research is a descriptive study using the Geographically Weighted Regression (GWR) method. The data used is the 2021 National Socio-Economic Survey (SUSENAS) data published by BPS. The response variable (y) used is the percentage of poor people in Papua in 2021. The predictor variables that are thought to influence the percentage of poverty in Papua are shown in the following table.

Table 1 Research Variable

Variable	Information
y	Percentage of Poor Population
x ₁	Percentage of Households Having Access to Proper Sanitation
x ₂	Open Unemployment Rate (Percent)
x ₃	Labor Force Participation Rate (Percent)
x ₄	PDRB at Constant Price Based on Expenditures (Rupiah)
x ₅	Average Years of Schooling of Population Over 15 Years (Years)
x ₆	Expenditure per Capita Adjusted (Thousand Rupiah/Person/Year)
x ₇	Human Development Index
x ₈	Life Expectancy (Years)

x ₉	Average Years of Schooling of Population Over 25 Years (Years)
x ₁₀	Percentage of households that have access to safe drinking water

Geographically Weighted Regression (GWR) is a non-stationary technique that can model relationships that vary spatially. GWR is a development of linear regression with different parameters depending on the observation area (Lu et al in Maryani et al, 2022). The linear regression equation is as follows.

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i \quad \dots \dots \dots (1)$$

In GWR modeling, response variables and predictor variables depend on the location of the observations. The GWR equation is as follows.

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i ; i = 1, 2, \dots, n \quad (2)$$

where :

y_i : The observed value of the response variable at the i-th observation location.

x_{ik} : The value of the k-th predictor variable observation at the i-th observation location

β₀(u_i, v_i) : Constant value (intercept) of GWR model.

β_k(u_i, v_i) : The k-th location predictor variable regression coefficient at the i-th observation location

(u_i, v_i) : The geographic coordinates of the i-th observation location

ε_i : The i-observation error.

We can see that equation (2) implicitly assumes that observed data near location i is more influential in estimating β_k(u_i, v_i) than data far from location i. In GWR an observation is weighted with location I so that the weighting is no longer constant but has variations. The role of weights in the GWR model is very important because the weight values represent the location of the observation data with each other (Caraka & Yasin, 2017). Data from observations that are close to I have more weight than data from observations that are located further away. It has been stated that the GWR is a development of linear regression, so that the GWR also has model parameter estimates which can be obtained using the Weighted Least Square as in the following equation.

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (3)$$

W(u_i, v_i) : Diagonal matrix elements of geographic weighting on each observed data for regression point i.

The W weighting can be calculated with the kernel function based on the closeness between the i and n data regression points around it. Gaussian and bi-square are examples of kernel functions that are often used as weights.

1) Gaussian function

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{b} \right)^2 \right] \quad (4)$$

2) Bisquare function

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b} \right)^2 \right]^2, & d_{ij} < b \\ 0, & \text{others} \end{cases} \quad (5)$$

where:

d_{ij} : Euclidean distance between observation point j and regression points i , and

b : Kernel bandwidth

The procedure that was carried out before the GWR analysis was carried out was a multiple linear regression analysis with the conditions that the assumptions applied. Then the spatial heterogeneity test was carried out using the Breush-Pagan test and the dependency test using Moran's I test. The GWR analysis steps are as follows

1. Calculates the Euclidian distance between points in the observation area
2. Determines bandwidth based on minimum AIC criteria
3. Calculating the weight matrix for each point of the observation area with the kernel function
4. Estimating GWR parameters using the optimum bandwidth
5. Comparing R^2 GWR models

3. RESULT AND DISCUSSION

3.1 DESCRIPTION OF POVERTY CASES IN PAPUA, 2021

Cases of poverty in every province in Indonesia are very diverse. This can be influenced by several factors. Based on survey results obtained from the official BPS Papua statistical news (2022), Papua Province is the province with the highest cases of poverty. The percentage of poverty cases by province as of September 2021 is presented in the following figure.

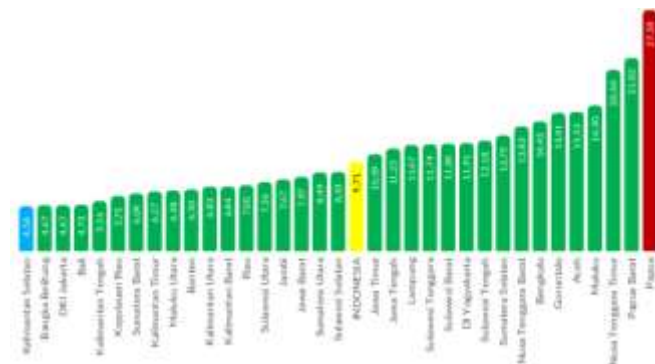


Figure 1. Percentage of Poor Population by Province, September 2021 (BPS Papua)

Based on figure 1, Papua Province is the region with the highest percentage of poor people. The number of poor people in Papua Province in September 2021 reached 944.49 thousand people. Compared to March 2021, the number of poor people increased by 24.05 thousand people.

The number of poor people in Papua Province varies greatly in each district/city. This is caused by geographical conditions or regions and the factors that influence them are also different. In the following, descriptive statistics are presented to see an overview of the characteristics of the variables used.

Table 2. Descriptive Statistics of Research Variables

Variable	Mean	Min	Max
y	28.38	10.16	41.66
x_1	40.287	0	85.726
x_2	3.2238	0	11.6706
x_3	78.4	56.39	97.93
x_4	5540775	767101	69619313
x_5	6.249	1.42	11.57
x_6	6820	3976	14937
x_7	57.7	32.84	80.11
x_8	65.26	55.43	72.36
x_9	6.775	1.923	11.39
x_{10}	63.87	0	98.67

Table 2 shows that the highest percentage of poor people in Papua Province in 2021 is 41.66, to be precise in Intan Jaya Regency, while the lowest percentage of poor people is 10.16, to be precise in Merauke Regency. The mean value shown in the table is 28.38, this means that the average percentage of poor people in Papua Province in 2021 is approximately 28.38.



Figure 2. Mapping of the Percentage of Poor People in Papua Province in 2021

Based on Figure 2, it is shown that the percentage of poor people tends to be higher in the central districts of Papua Province.

3.2 ASSUMPTION OF MULTIPLE LINEAR REGRESSION ANALYSIS

In linear regression analysis, one of the important conditions that must be met is the assumption of normality. To carry out the normality test, you can use the Shapiro Wilk test. Based on the analysis results obtained $p\text{-value} = 0.2161$. By using the initial hypothesis H_0 , namely data errors are normally distributed and H_1 data errors are not normally distributed, then with a significant level $\alpha = 0.05$, the decision to fail to reject H_0 is obtained. This shows that the data used is normally distributed.

The next assumption is multicollinearity. Multicollinearity is a situation where there is a correlation or relationship between two or more predictor variables. If the VIF value is not more than 10, then multicollinearity does not occur. The VIF value of each predictor variable is shown in the following table.

Variable	VIF
x_1	15.2883
x_2	2.88885
x_3	4.56536
x_4	1.95656
x_5	71.5245
x_6	9.84446
x_7	52.9589
x_8	3.12369
x_9	16.1095
x_{10}	1.58197

Based on table 3, it can be seen that the predictor variables x_1 , x_5 , x_7 and x_9 have VIF values greater than 10, which means that these variables contain multicollinearity, so they must be excluded. Following are the VIF values of other predictor variables after x_1 , x_5 , x_7 and x_9 are excluded.

Variable	VIF
x_2	2.44543
x_3	1.97986
x_4	1.76786
x_6	3.45583
x_8	1.53503
x_{10}	1.49144

Based on table 4, it can be seen that the VIF values for all predictor variables are less than 10, which means that there is no multicollinearity between variables. So that the predictor variables that will be used for the purposes of further analysis are x_2 , x_3 , x_4 , x_6 , x_8 and x_{10} .

Analysis of the relationship between variables was carried out to determine the linear relationship between the response variables and the predictor variables. The following presents the results of parameter estimation for each variable used.

Variable	Estimation	$Pr(> t)$	$Pr(> F)$
Intercept	8.85e+00	0.75894	
x_2	2.59e-01	0.66525	
x_3	1.76e-01	0.21265	
x_4	-1.20e-08	0.91957	0.0001149
x_6	-2.83e-03	0.00305	
x_8	4.09e-01	0.30771	
x_{10}	-3.85e-02	0.43397	

The results of the analysis in table 5 show that the open unemployment rate, labor force participation rate and life expectancy are positively correlated with the percentage of poor people, which means that every one-unit increase in these three variables will increase the percentage of poor people. Meanwhile, GRDP based on constant prices according to expenditure, adjusted per capita expenditure and the percentage of households that have access to proper drinking

water shows a negative correlation. This means that every one unit increase in these three variables will reduce the percentage of poor people.

Table 6. Test dependency effect and spatial heterogeneity effect

Test	<i>p-value</i>
<i>Breusch-Pagan</i>	0.7746
<i>Morans'I</i>	4.10e-05

Table 6 shows the results of the spatial effect test that was carried out before the GWR analysis, namely the heterogeneity test using the Breusch-Pagan test and the dependency test using the Morans'I test. The results of the Breusch-Pagan test gave a p-value of 0.77460 indicating that there was no spatial heterogeneity. Morans'I test results provide a p-value of 0.0001183 indicating that there is a spatial dependency effect that occurs in the data. So it can be concluded that the case of poverty in Papua has a spatial influence, so that the analysis using the GWR method can be continued.

3.3 PARAMETER SIGNIFICANCE TEST

To obtain the best global regression model, parameter significance tests were carried out simultaneously or partially

1. Simultaneous Test

Based on the results of the analysis in table 5, a p-value of 0.0001149 is obtained. by using the initial hypothesis H₀, namely that there is no predictor variable that has a significant effect on the model, and H₁, namely that there is at least one predictor variable that has a significant effect on the model, then with a significant level of $\alpha = 0.05$, the decision to reject H₀ is obtained, which means that H₁ is accepted. This shows that there is at least one predictor variable that has a significant effect on the model.

2. Partial Test

Based on the results of the analysis in table 5 and using the initial hypothesis H₀, namely the variable k does not significantly affect the model and H₁, namely the variable k significantly affects the model, then with a significant level of $\alpha = 0.05$ it is obtained that Adjusted Per Capita Expenditure has a significant effect on model. While other variables do not significantly influence the model.

3.4 MODELING OF GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

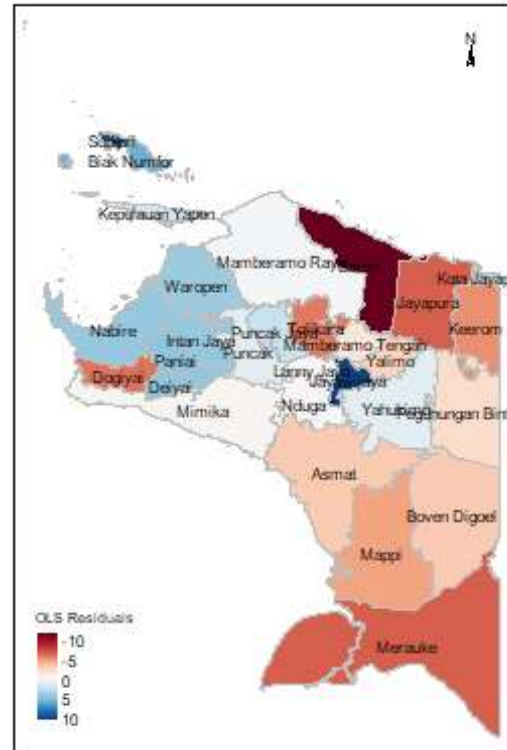


Figure 3. Mapping of Global Model Residuals in Papua Province

Based on Figure 3, it can be seen that some districts/cities in Papua Province have quite high residuals. As a result, errors in drawing conclusions may occur. As in Jayawijaya Regency and Sarmi Regency, both of which have very high residuals, so it could be that per capita expenditure is not the main cause of the increase or decrease in the percentage of poor people in these areas. This can be influenced by geographic conditions in the data, so a suitable method is needed for this case, namely the GWR method.

In making the GWR model, it is necessary to select the optimum weights in order to obtain the best model. Bandwidth selection on weighting is seen based on the smallest AIC value. There are two kernel functions used in this study, namely adaptive bisquare and adaptive gaussian, then these two functions will be compared and the most optimal weighting will be selected.

3.4. GWR MODEL OF ADAPTIF BISQUARE WEIGTING

In the following, the GWR model for each district/city in Papua Province is presented using adaptive weighting bisquare.

Table 7. GWR Model with *Adaptif Bisquare Weighting*

Regency	Intercep	x_6
Asmat	44.437	-0.0025
Biak Numfor	54.3847	-0.0033
Boven Digoel	57.7641	-0.0049
Deiyai	58.6103	-0.0039
Dogiyai	51.9385	-0.0032
Intan Jaya	59.303	-0.0036
Jayapura	26.15	-0.0011
Jayawijaya	35.1721	0.0002
Keerom	37.1385	-0.0020
Kepulauan Yapen	50.0548	-0.0029
Kota Jayapura	26.0466	-0.0011
Lanny Jaya	37.2863	-0.0001
Mamberamo Raya	55.696	-0.0050
Mamberamo Tengah	76.4955	-0.0095
Mappi	49.9679	-0.0039
Merauke	51.7952	-0.0041
Mimika	59.3507	-0.0039
Nabire	47.6295	-0.0027
Nduga *	38.4877	-0.0002
Paniai	61.2065	-0.0040
Pegunungan Bintang	56.3256	-0.0045
Puncak	42.3694	-0.0010
Puncak Jaya	57.5657	-0.0037
Sarmi	79.5112	-0.0099
Supiori	55.0555	-0.0033
Tolikara	34.5569	0.0002
Waropen	69.8536	-0.0058
Yahukimo	33.0493	0.0005
Yalimo	30.3562	0.0009

Based on table 7, the Quasi-global R^2 value is 0.9450383 and the AIC value is 147.0325. The following is a GWR model that can be formed from the table as follows.

$$y = 26.0466 - 0.0011x_6$$

$$y = 35.1721 + 0.00021x_6$$

Then, to see more clearly visually the strength of the GWR model using bi-square adaptive weights for each district/city is presented in the following figure.

Based on Figure 4, it can be seen that the local mapping of R^2 is not evenly distributed in every region in Papua Province. This means that the performance of the GWR model, namely spending per capita in influencing the percentage of poor people, is different for each district/city.

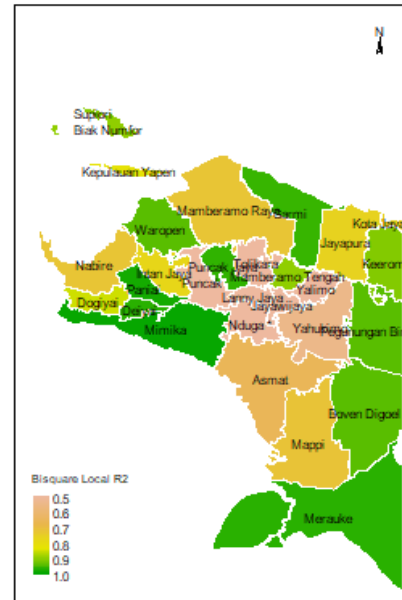


Figure 4. Local R^2 Mapping for Bisquare Adaptive Weighting

3.5 GWR MODEL WITH ADAPTIF GAUSSIAN WEIGHTING

In the following, the GWR model for each district/city in Papua Province is presented using Adaptive Gaussian Weighting.

Table 7. GWR Model with Adaptive Gaussiam Weighting

Regency	Intercep	x_6
Asmat	49.48457	-0.0031032
Biak Numfor	49.06939	-0.0029297
Boven Digoel	48.95708	-0.0030924
Deiyai	49.80093	-0.0030268
Dogiyai	49.78741	-0.0030269
Intan Jaya	49.64494	-0.0029835
Jayapura	47.93177	-0.0029041
Jayawijaya	48.62664	-0.0029585
Keerom	48.02437	-0.0029295
Kepulauan Yapen	49.1459	-0.0029290
Kota Jayapura	47.93367	-0.0029042
Lanny Jaya	48.99115	-0.0029750
Mamberamo Raya	48.6346	-0.0028918
Mamberamo Tengah	48.29455	-0.0028674
Mappi	49.46127	-0.0031507
Merauke	49.45465	-0.0031548
Mimika	49.81234	-0.0030422
Nabire	49.71367	-0.0029974
Nduga *	49.19794	-0.0030116
Paniai	49.73777	-0.0030067
Pegunungan Bintang	48.40328	-0.0029986
Puncak	49.50046	-0.0029738
Puncak Jaya	49.71496	-0.0030070
Sarmi	47.99392	-0.0028653
Supiori	49.08254	-0.0029355
Tolikara	48.66088	-0.0029256

Waropen	49.32124	-0.0029399
Yahukimo	48.51281	-0.0029911
Yalimo	48.22503	-0.0029345

Based on table 8, the Quasi-global R² value is 0.666158 and the AIC value is 184.26. The following is a GWR model that can be formed from the table as follows.

$$y = 49.71496 - 0.003007x_6$$

$$y = 48.66088 - 0.0029256x_6$$

Then, to visually see more clearly the strength of the GWR model using Gaussian adaptive weighting for each district/city region is presented in the following figure.



Figure 5. Local Mapping R² with Adaptive Gaussian Weighting

Based on Figure 5, it can be seen that local mapping of R² is evenly distributed in every region in Papua Province with an average R² of around 60%. This means that the performance of the GWR model, namely expenditure per capita in influencing the percentage of poor people, is almost the same in each district/city.

3.6 COMPARISON OF GWR MODEL

After modeling poverty cases using the GWR method with adaptive bi-square and adaptive gaussian weights, a comparison of the GWR model between the two weights will be carried out. This is done to determine the most optimal weighting in order to obtain the best model for poverty cases in Papua Province in 2021. The criteria for selecting the optimal weighting can be determined by looking at the smallest AIC value and the largest R² value. The following

table presents a comparison of bi-square adaptive and Gaussian adaptive weights as follows.

Table 9. Comparison Between Adaptif Bisquare and Adaptif Gaussian Weighting

Weighting	Bandwidth	AICc	AIC	R ²
Adaptif Bisquare	0.2092122	249.9248	147.0325	0.9450383
Adaptif Gaussian	0.6896361	190.6177	184.26	0.666158

Based on the test results in table 9, it was found that the AIC on the bi-square adaptive weights was smaller than the AIC value on the gaussian adaptive weights. This is also in line with the coefficient of determination (R²) obtained, which shows that the coefficient of determination of the bi-square weight is greater than the Gaussian weight. In the GWR model with adaptive bi-square weighting, R² is 0.9450383 which means that the model formed can explain 94.5% of cases of poverty and 5.5% is explained by other variables not included in the model. Whereas in the GWR model with gaussian adaptive weighting, R² is 0.666158 which means that the model formed is only able to explain cases of poverty of 66.61%.

4. CONCLUSION

Based on the results and discussion, it can be concluded that the GWR model for poverty cases in Papua Province in 2021 with bi-square adaptive weights is more optimal than the gaussian adaptive weights. The GWR model also produces a poverty case model in Papua in 2021 with bi-square adaptive weights which have a coefficient of determination that is greater than the Gaussian adaptive weights. When compared with multiple linear regression, the GWR analysis produces a better model. This is in line with research conducted by Maryani, et al (2022), which concluded that the GWR analysis produced a higher coefficient of determination compared to modeling with linear regression.

Based on the results of the analysis in this study, it is suggested for further research to validate the data first and add significant predictor variables in order to be able to obtain results and models that are in accordance with the actual conditions.

5. REFERENCES

- [1] Arfiani, D. 2019. Berantas Kemiskinan. Semarang: ALPRIN
- [2] Berita Resmi Statistik. 2022. Profil Kemiskinan Provinsi Papua September Tahun 2021. Papua: Badan Pusat Statistik Provinsi Papua.
- [3] Caraka, R. E. & Yasin, H. 2017. Geographically Weighted Regression: Sebuah Pendekatan Regresi Geografis. Yogyakarta: MOBIOUS.
- [4] Firdang, A. P. A., Baharuddin, B. & Laome, L. 2021. Geographically Weighted Regression Dalam Analisis

- Tingkat Kemiskinan di Indonesia. E-Journal BIAStatistics, Vol. 2020, No. 2.
- [5] Haryanto, S. & Andriani, G. A. 2021. Pemodelan Jumlah Penduduk Miskin di Jawa Tengah Menggunakan Geographically Weighted Regression (GWR). Jurnal Litbang Sukowoti, Vol. 4, No. 2.
- [6] Lutfiani, N., Sugirman & Mariani, S. 2017. Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-square. UNNES Journal of Mathematics, Vol. 8, No. 1.
- [7] Maryani, H., Kristiana, L. & Izza, N. 2022. Pemodelan Spasial Kasus Kematian Akibat Covid-19 di Provinsi Jawa Timur Tahun 2020. Jurnal Ilmu Kesehatan, Vol. 6, No. 1.
- [8] Pratiwi, Y. D., Mariani, S. & Hendikawati, P. 2019. Pemodelan Regresi Spasial Menggunakan Geographically Weighted Regression Dengan Pembobot Fixed Kernel Gaussian dan Adaptive Kernel Bisquare. UNNES Journal of Mathematics, Vol. 8, No. 1.
- [9] Yasin, H., Warsito, B. & Hakim, A. R. 2020. Regresi Spasial: (Aplikasi dengan R). Pekalongan: Wade Group.