Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2020, 2020:77

https://doi.org/10.28919/cmbn/4991

ISSN: 2052-2541

MODELING POSITIVE COVID-19 CASES IN BANDUNG CITY BY MEANS GEOGRAPHICALLY WEIGHTED REGRESSION

ELIN MARHAMAH*, I. GEDE NYOMAN MINDRA JAYA

Department of Statistics, Universitas Padjadjaran, Bandung, Indonesia

Abstract: Coronavirus disease 2019 (COVID-19) has rocked the world since the beginning of 2020, Indonesia is no exception. Bandung, as one of the metropolitan cities and one of the largest cities in West Java, has been exposed to the virus since early March 2020, among the impacts is the occurrence of panic buying, mass unemployment due to layoffs, and increased crime rates. The statistical analysis used for modeling the number of COVID-19 cases is the Geographically Weighted Regression (GWR) model. The GWR model is the development of a linear regression model that produces local model parameter estimators for each observation location. The aim of the study was to model the number of COVID-19 cases in the period of March to June 4, 2020, in Bandung. The results showed that variations in the variable population size, distance to the capital city, number of ODP, number of PDP, and number of health facilities in the GWR model were able to explain variations in the number of positive cases of COVID-19 in Bandung City.

Keywords: COVID-19; bandwidth; CV; exponential kernel function; GWR.

2010 AMS Subject Classification: 91D20.

---

*Corresponding author

E-mail address: elin19001@mail.unpad.ac.id

Received September 2, 2020

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus is spread between people during close contact, that is, by tiny droplets produced through coughing, sneezing, or talking [1]. The origin of the virus came from Wuhan China and spread to other countries, causing a pandemic. Besides being able to cause death, COVID-19 also affects social and economic stability and has an impact on the political turmoil of a country. In the case of this pandemic, the social impact is already visible because of the economic downturn it has caused. For example panic buying activities, mass unemployment due to layoffs, increasing crime rates, and many other cases [2].

Indonesia is one of the countries exposed to the COVID-19 virus since early March 2020. As of June 4, 2020, there were 28,818 cases recorded with 1,721 deaths [3]. Likewise in the city of Bandung as one of the largest cities in West Java, there were 342 cases with 38 deaths [4]. One of the economic impacts that occurred, which has been recorded by the Bandung City Manpower Office, as of May 1, 2020, as many as 3,396 people had been laid off and 5,804 people were temporarily discharged [2]

Based on this background, it is necessary to analyze the COVID-19 cases in Bandung City as an effort to prevent and reduce the number of positive cases of COVID-19. One of the statistical analyzes used for modeling is the Geographically Weighted Regression (GWR) model [5]. The GWR model is the development of a linear regression model. In the linear regression model, only parameter estimators that apply globally are produced, while in the GWR model, local model parameter estimators are produced for each observation location [5]. The purpose of this study is to model the number of COVID-19 cases in the period of March 2020 to June 4, 2020, in Bandung.

2. MATERIAL AND METHOD

Material

This study uses secondary data, namely positive numbers, ODP and PDP of COVID-19 cases from the Bandung City COVID-19 Information Center. While data on population, distance to the capital and number of health facilities in each sub-district, were obtained from the Central Statistics Agency of Bandung City. The research variables used are presented in Table 1.

Table 1. Research variables

Variable	Notation	Units
Number of positive cases of COVID-19	Y	Number
Population	X_1	Number
Distance of the district to the capital	X_2	Km
Number of ODP	X_3	Number
Number of PDP	X_4	Number
Number of Health Facilities	X_5	Number

Method

Geographically Weighted Regression (GWR) model is a technique that assumes that the regression parameter is varying spatially. By using GWR, spatial variety in parameter estimation value will be discovered, so that different and valuable interpretation can be obtained for each research location.

According to Anselin (1988) [6], spatial heterogeneity in classical regression analysis is something that deserves special attention. The spatial heterogeneity may be caused by spatial units' conditions in one research area which is basically not homogeneous. For example, in territorial income rate may be different.

The GWR model is a developed global linear regression model where the basic idea is based on nonparametric regression. This model is a local linear regression model that produces local model parameter approximation for every location depending on where the data is collected so that every geographical location has its own regression parameter value. GWR model can be written as following [6]:

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

where:

y_i : the observed value response variable for the i-th location

i : $1,2,\ldots,n$

(u_i, v_i) : geographical positioning coordinates (longitude, latitude) for the i-th observation location

$\beta_k(u_i, v_i)$: k-th predictor variable regression coefficient for the i-th observation location

x_{ik} : k-th predictor variable observed value for the i-th observation location

ε_i : the i-th observation error that is assumed as identical, independent, and normally distributed with zero mean and constant variance σ^2.

One of the aspects of GWR is that the estimated parameters depend in part on the weighting function or the kernel used so that the optimal bandwidth selection is very important [5, 7]. One of the criteria that can be used in choosing bandwidth is Cross Validation with the following formula:

$$CV = \sum_{i=1}^{n}\left(y_i - \hat{y}_{\neq i}(b)\right)^2 \qquad (2)$$

where:

$\hat{y}_{\neq}(b)$: approximated value

yi : observation at point i excluded from the calibration process [4].

In GWR, the weighting function follows Tobler's law, namely the closer the data is to location i, the stronger the effect will be in predicting parameters at the location compared to the data at other locations farther away [8]. The weighting function in this study is the Fixed Exponential Kernel with the following formula:

$$w_j(u_i, v_i) = exp\left(-\frac{d_{ij}}{h}\right) \qquad (3)$$

where :

$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$

d_{ij} : Euclidean distance between location (u , v_i) and location (u_j , v_j)

h : optimal bandwidth.

3. RESULT AND DISCUSSION

Descriptive Analysis

The factors that are thought to affect the number of positive cases of COVID-19 in Bandung City in this study consist of the population (X_1), Distance of the district to the capital (X_2), number of ODP (X_3), number of PDP (X_4) and number of health facilities (X_5). The analysis of the descriptions of the five predictor factors is as follows.

Table 2 Variable Descriptive Analysis

Variable	Mean	Standard Deviation	Minimum	Maximum
Number of positive cases *(Y)*	10.30	7.56	1	38
Population (X_1)	83.60	32.36	28	153.4
Distance of the district to the capital (X_2)	7.64	4.47	1.4	18.7
Numbers of ODP (X_3)	127.30	73.58	25	302
Numbers of PDP (X_4)	7.33	4.36	1	18
Numbers of Health Facilities (X_5)	6.20	2.07	1	10

Table 1 shows a description of 31 sub-districts in Bandung City, where the number of positive cases of COVID-19 (Y) has the lowest number of 1 case and the highest is 38 cases with an average of 10.3 cases and a standard deviation of 7.56 cases. Total population (X_1) has an average number of 83.6 million people with a standard deviation of 32.36 million people. The distance to the capital city (X_2) averages 7.64 km with a standard deviation of 4.47 km. The average number of ODP (X_3) is 127.3 people with a standard deviation of 73.58 people. The average number of PDP (X_4) is 7.3 people with a standard deviation of 4.36 people. Meanwhile, health facilities (X_5) have an average number of 6.2 units with a standard deviation of 2.07 units.

The distribution of positive cases of COVID-19 in Bandung City is shown as follows.

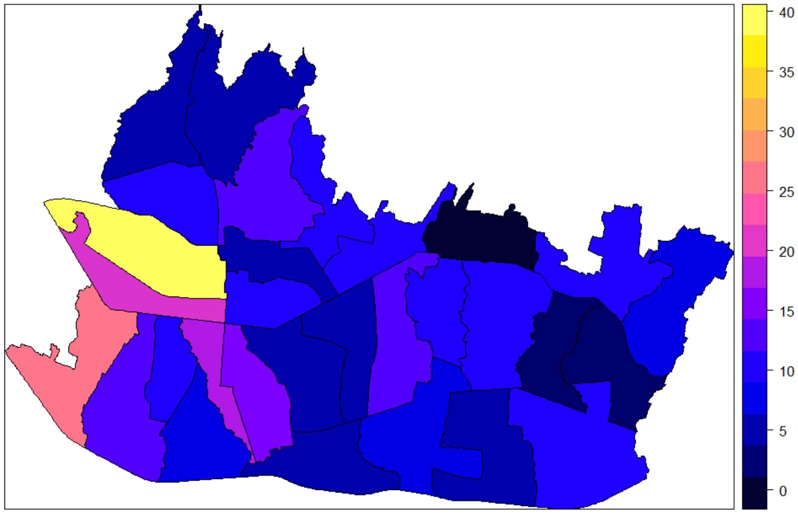

Figure 1. Distribution of Positive cases of COVID-19 in Bandung City (March 2020 to June 4, 2020)

Based on the distribution of positive cases of COVID-19, the number of positive cases varies in 30 sub-districts in Bandung City. The highest number was in the west-center of Bandung city (Cicendo sub-district) and the lowest case was in the north of Bandung (Mandalajati sub-district). The number of positive cases is generally between 5-15 cases, spread across throughout the city of Bandung.

Regression Analysis with OLS

Regression analysis using OLS obtained the following results:

Table 3 Parameter Estimate with OLS

Variable	$\hat{\beta}$	t	p-Value
Intercept	10.6	5.806	0.0803
Population	0.196	0.048	0.0004
Distance of the district to the capital	-0.379	0.295	0.2113
Numbers ODP	-0.029	0.019	0.1478
Numbers PDP	-0.874	0.302	0.0080
Numbers Health Facilities	-0.874	0.676	0.3831

The result of the simultaneous parameter test in Table 3 shows the value of F = 5.446 with p-value = 0.0017, which means that the parameter is significant simultaneously. The regression model using OLS has R^2 = 0.4339 which indicates that the variation of the predictor variables is able to explain 43.39% of the variation in the number of positive cases of COVID-19 in West Java.

Testing Dependence and Spatial Heterogeneity

Testing the spatial dependence using the Moran's index obtained a p-value of 1.769814e-13 which is smaller than the significance level (α = 0.05), meaning that there is dependency between locations. Simultaneous heterogeneity test using the Breusch Pagan statistical test obtained a p-value of 0.08905 which is smaller than the significance level (α = 0.1), which means that the variance in each location is different (heterogeneous). From the test results, it can be seen that the dependency and spatial heterogeneity assumptions are met, so that the analysis can be continued with Geographically Weighted Regression (GWR).

Geographically Weighted Regression (GWR) Model

The optimum bandwidth calculation results obtained through the Cross Validation (CV) technique were 0.09776 with CV = 24.07398. The bandwidth value of 0.09776 indicates that the area or location that is still around 0.09776 degrees from the point of observation still has an influence on the number of positive cases of COVID-19 at the observation location. The results of the GWR model parameter estimation are shown as follows.

Table 4 Result Parameter Estimate GWR Model

Variable	$\hat{\boldsymbol{\beta}}$ **Value**		
	Minimum	Median	Maximum
Intercept	-0.142	-0.048	0.078
Population	0.514	0.826	1.141
Distance of the district to the capital	-0.428	-0.220	-0.130
Numbers of ODP	-0.558	-0.229	0.040
Numbers of PDP	-0.724	-0.505	-0.349
Numbers of Health Facilities	-0.235	-0.147	0.020

From Table 4, it is known that the predictor variables that have a negative relationship with the response variable in the GWR model are the distance to the capital and the amount of PDP. While the population has a positive relationship with the response variable and the number of ODP and health facilities has a variable relationship between negative and positive with response variables in several observation locations. The sum of square residual value of the GWR model is 6.343, which this value is smaller than the sum of square residual of the OLS regression model. The coefficient of determination (R^2) of the GWR model is 0.4761, higher than the R^2 of the OLS regression model. This shows that the variation in the predictor variables in the GWR model is able to explain 47.61% of the variation in the number of positive cases of COVID-19 in Bandung. Furthermore, the significance of the variables based on location was carried out using the *t* test, the results were shown as follows:

Table 5 Significance of Variables based on Location

Sub-District	Numbers of Sub-district	Significance Variable
Andir, Babakan Ciparay, Bandung Kulon, Cicendo, Sukajadi	5	X_1, X_2, X_3 X_4
Astanaanyar, Bandung Kidul, Bandung Wetan, Bojongloa Kaler, Bojongloa Kidul, Lengkong, Regol, Sumur Bandung	8	X_1,X_3 X_4
Sukasari	1	X_1,X_2 X_4
Antapani, Arcamanik, Batununggal, Buah Batu, Cibeunying kaler, Cibeunying kidul, Cibiru, Cidadap, Cinambo, Coblong, Gedebage, Kiaracondong, Mandalajati, Panyileukan, Rancasari, Ujungberung	16	X_1, X_4

Based on Table 5, the variables X_1, X_2, X_3 and X_4 affect the number of positive cases of COVID-19 in 5 sub-districts, variables X_1, X_3 and X_4 affect the number of positive cases of COVID-19 in 8 districts, variables X_1, X_2 and X_4 affect the number of positive cases of COVID-19 in 1 district and variables X_1 and X_4 affect the number of positive cases of COVID-19 in 16 districts.

4. CONCLUSION

Modeling the number of positive cases of COVID-19 in the City of Bandung in the period March 2020 to June 4 2020 with predictor variables including population, distance to the capital, number of ODP, number of PDP and number of health facilities used the Geographically Weighted Regression (GWR) model.

Analysis of the GWR model with a fixed exponential kernel weighting matrix used the optimum bandwidth of 0.09776 with CV = 24.07398. The coefficient of determination (R^2) from the GWR model is 0.4761 which shows that the variation in the predictor variables in the GWR model is able to explain 47.61% of the variation in the number of positive cases of COVID-19 in Bandung, where the number of positive cases of COVID-19 in the Bandung City is strongly influenced by the population and the number of PDP.

For the next research is important to consider the overdispersion problem in GWR modeling and sparsity. The overdispersion can be solved by mean geographically weighted negative binomial (GWNBR) [9] and the sparsity problem can be solve using Bayesian approaches [10]. The models might be improved by considering the climate variables as the potential risk factors.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] WHO. Naming the coronavirus disease (COVID-19) and the virus that causes it. Geneva: WHO, (2020). Retrieve at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(COVID-2019)-and-the-virus-that-causes-it

[2] Simbolon H. Liputan6. 2020. Dampak Pandemi Corona Covid-19 (2020). Retrieve at: https://www.liputan6.com/regional/read/4248156/dampak-pandemi-corona-covid-19-ada-38-juta-warga-jabar-perlu-disubsidi

[3] Worldometer. COVID-19 Coronavirus Pandemic. (2020). Retrieve at: https://www.worldometers.info/ coronavirus/

[4] Bandung. Center Information of COVID-19 (2020). Reterive at: https://covid19.bandung.go.id/data.

[5] A.S. Fotheringham, C. Brunsdon, M. Charlton, Geographically Weighted Regression. John Wiley, Chichester. (2002).

[6] L. Anselin, Spatial econometrics: methods and models. Kluwer Academic Publishers, Dorddrecht. (1988).

[7] P. Purhadi, H. Yasin, Mixed Geographically Weighted Regression Model (Case Study: The Percentage of Poor Households in Mojokerto 2008). Eur. J. Sci. Res. 69 (2012), 188–196.

[8] C. Ramdhani, I.G.N.M. Jaya, Y. Suparman. Malaria Morbidity Modelling in Papua Province, 2016 Using Geographically Weighted Regression (GWR) Method. World Appl. Sci. J. 36 (5) (2018), 703-709.

[9] R. Fitriani, I.G.N.M. Jaya, Spatial modeling of confirmed COVID-19 pandemic in East Java province by geographically weighted negative binomial regression, Commun. Math. Biol. Neurosci. 2020 (2020), Article ID 58

[10] I.G.N.M. Jaya, B.N. Ruchjana, A.S. Abdullah, T. Toharudin, Spatial distribution of tuberculosis disease among men and women in Bandung city, Indonesia, Commun. Math. Biol. Neurosci. 2020 (2020), Article ID 53.