

Under-five Mortality Rate Modeling Using Geographically and Temporal Weighted Regression (GTWR)

Conita and Tuti Purwaningsih

Department Statistics, Universitas Islam Indonesia, Yogyakarta, Indonesia

E-mail : 13611091@students.uui.ac.id, tutipurwaningsih@gmail.com

Abstract. Indonesia has a relatively high Under Five Mortality Rate. The causes of under-five mortality are low birth weight, pneumonia, education, socio-economic and cultural aspects. Central Java province is included in the province with high mortality rate in 2015 amounted to 11.64 per 1,000 live births (BPS, 2015). One of method to determine the factors who affected the under-five mortality using global regression analysis. Regression analysis of spatial data can be used as it has the nature of dependence between the measurement data by location. The local form of Regression analysis spatial is Geographically Weighted Regression (GWR). Development of methods GWR is Geographically and Temporal Weighted Regression method (GTWR). GTWR was accommodating spatial heterogeneity (location) and temporally (time). GTWR modeling using Gaussian Kernel functional Under Five mortality rates in central java province in 2013 - 2015 shows GTWR model is better than global regression method. Global regression has R^2 value 0.5080 and Mean Square Error (MSE) 27.137, while GTWR method has R^2 value of 0.5210 and MSE 1.084.

1. Introduction

Under-five Mortality Rate in Indonesia is still high compared to developing countries in ASEAN such as Malaysia and Singapore which has been under 10 deaths per 1,000 births. Over the past few years, the results of the 2015 Intercensal Population Survey (SUPAS) show an Under-five Mortality Rate of 26.29 per 1,000 live births, also meeting the 2015 Millennium Development Goals (MDGs) 2015 target of 1,000 live births meaning that it has reached the Millennium Development Goals (MDGs), which is a program of the Indonesian government since 2011. The under-five mortality rate is the mortality of infants aged 0-4 years of life per 1000 births [1]. The direct cause of premature death by World Health Organization (WHO) is Heavy Baby Low Birth, Pneumonia, lack of oxygen (asphyxia), and Malnutrition. The indirect cause of under-five mortality is due to societal conditions such as education, socio-economic and cultural.

The 2015 Intercensal Population Survey (SUPAS) result shows an infant mortality rate of 22.23 per 1,000 live births, meaning that it has reached the 2015 MDG target of 23 per 1,000 live births. Similarly, the 2015 SUPAS Mortality Rate of 26.29 per 1,000 live births has also met the 2015 MDG target of 32 per 1,000 live births. Central Java province is included in the province with high under-five mortality rate with the number of Under-five Mortality Rate at 38 per 1000 births of the highest regency / municipality with the highest Under-five Mortality Rate in Central Java is Grobogan with 37.85 per 1,000 live births, followed by Rembang 36.17 per 1,000 live births, and Kendal 35.04 per 1,000 live birth. The lowest districts / cities with MMR are Semarang 17.03 per 1,000 live births.

2. Literature Review and Methodology

Research of "Modeling the Proportion of Poor People of Districts and Cities in Central Java Province Using Geographically and Temporally Weighted Regression" describes the model of proportion of the poor by taking into account spatial and temporal variations in Central Java Province using the GTWR method. The data used in this study is secondary data about factors that allegedly

affect the level of poverty in Central Java Province obtained from the Central Statistics Agency of Central Java Province. The variables used are regional minimum wage, labor force participation rate, land ownership, population density, poor families, human development index. The result of this research is that the modeling of poverty rate in Central Java using GTWR is better than using global regression. This is because the GTWR model is able to maximize the R2 value of 0.997 with MSE value smaller than the global regression of 4.577. [2]

Research entitled "Modeling of Particle Concentration Dust (PM10) on Air Pollution in Surabaya City with Geographically-Temporally Weighted Regression Method" describes how to model dust particles by accommodating the influence of spatial and temporal heterogeneity on dust particle concentration (PM10). The data used are secondary data on ambient air quality in Surabaya City in 2010 obtained from Environment Agency (BLH) of Surabaya City. There are 5 points of observation which is location of placement of ambient air quality monitoring station (SUF) namely: Achievement Park at Jalan Ketabang Kali (1), Perak Timur at Jalan Selanggor (2), Sukomanunggal Street (3), Gayungan at Jalan Raya Pagesangan 4), and Gebang Putih on Jalan Arif Rachman Hakim (5). However, this research uses only three observation points namely SUF 1 Park Achievement, SUF 4 Gayungan, and SUF 5 Gebang Puti. The result of the research is air pollution condition in Surabaya city in 2010 is still good and GTWR method gives more accurate result than non-spatial regression because it can accommodate the influence of spatial and temporal heterogeneity on dust particle concentration (PM10). [2]

2.1 Material

The data collected are secondary data obtained from Health Service website of Central Java Province (dinkesjateng.goprov.id) and Central Statistics Agency of Central Java website (jateng.bps.go.id). The data obtained will be analyzed using the program R 3.2.5, Tablue Public, and SPSS 16.0. The data used are divided into several categories of variables, namely Health, Socio-Economic, and Environmental Factors. The analysis method used is Geographically and Temporally Weighted Regression (GTWR) and Linear Multiple Regression.

2.2 Methodology

2.2.1 Heterogenitas Spatial Test

Spatial heterogeneity testing is to test the effect of heterogeneity [3]. The hypothesis of Breusch Pagan Test according is as follows

$$BP = \frac{1}{2} f^T z(z^T z)^{-1} z^T f \sim x_{(k+1)}^2 \quad (1)$$

with :

e_i : Error value for i-th observation.

z : Matrix size $n \times (k+1)$ which contains a vector of x that has been standardized for each observation.

f : Vector size $(n \times 1)$.

n : Number of observation areas.

k : The number of explanatory variables.

σ^2 : Variety of remaining e_i .

Criteria for decision-making can also be done by comparing p -value with α . p -value can be obtained from the calculation as follows:

p -value = $P(x_{(k+1)}^2 \geq BP)$ if p -value < α H_0 is rejected, so it can be concluded that there is spatial heterogeneity.

2.2.2 Geographically and Temporal Wiegthed Regression

Geographically and Temporally Weighted Regression (GTWR) Geographically is a development method of Geographically Weighted Regression that accommodates spatial

heterogeneity (location) and temporally (time) [5]. Unlike the standard GWR model, GTWR combines temporal and spatial information in a weighted matrix in identifying the presence of spatial and temporal heterogeneity. GTWR model for free variable p with bound variable y_i on location

$\{(u_i, v_i, t_i)\}$ for each observation is written as follows:

$$Y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^p \beta_k(u_i, v_i, t_i)x_{ik} + \varepsilon_i \quad (2)$$

The following parameter estimates on the GTWR model for each k variable with the observation point $\{(u_i, v_i, t_i)\}$.

$$\tilde{\beta}(u_i, v_i, t_i) = (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) Y \quad (3)$$

with $W(u_i, v_i, t_i) = \text{diag}(\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in})$ and n is the number of observations. Diagonal element α_{ij} ($1 \leq j \leq n$) is a function of spatial-temporal distance at the point of view (u_i, v_i, t_i) . It is assumed that the proximity of the data observation point to the point i in the spatial-temporal coordinate system has a greater influence on the estimation of the parameters $\tilde{\beta}(u_i, v_i, t_i)$ than the data located further from point i . Defining and measuring spatial-temporal proximity in coordinate systems is a major problem in the preparation of the GTWR model because the proximity of the observation point has two elements, namely spatial closeness and temporal closeness.

The spatial-temporal spatial function consists of a combination of spatial distance function and temporal distance function. With the dS spatial distance function and the temporal distance function dT we obtain the following equation:

$$(d^{ST})^2 = \lambda(d^S)^2 + \mu(d^T)^2 \quad (4)$$

where λ and μ determine different scale balancing factors for measuring spatial and temporal distances. So the euclidean distance becomes (Huang, et al, 2010):

$$(d_{ij}^{ST})^2 = \lambda \left\{ (u_i - u_j)^2 + (v_i - v_j)^2 \right\} + \mu (t_i - t_j)^2$$

Based on the equation obtained

$$\begin{aligned} \alpha_{ij} &= \exp \left\{ - \left(\frac{\lambda [(u_i - u_j)^2 + (v_i - v_j)^2] + \mu (t_i - t_j)^2}{h_{ST}^2} \right) \right\} \\ &= \exp \left\{ - \left(\frac{[(u_i - u_j)^2 + (v_i - v_j)^2]}{h_S^2} \right) + \frac{\mu (t_i - t_j)^2}{h_T^2} \right\} \text{ with } h_S^2 = \frac{h_{ST}^2}{\lambda} \text{ and } h_T^2 = \frac{h_{ST}^2}{\mu} \\ &= \exp \left\{ - \left(\frac{(d_{ij}^S)^2}{h_S^2} + \frac{(d_{ij}^T)^2}{h_T^2} \right) \right\} \\ &= \exp \left\{ - \frac{(d_{ij}^S)^2}{h_S^2} \right\} \times \exp \left\{ - \frac{(d_{ij}^T)^2}{h_T^2} \right\} \\ &= \alpha_{ij}^S \times \alpha_{ij}^T \quad \text{d } \alpha_{ij}^S = \exp \left\{ - \frac{(d_{ij}^S)^2}{h_S^2} \right\} \text{ and } \alpha_{ij}^T = \exp \left\{ - \frac{(d_{ij}^T)^2}{h_T^2} \right\} \end{aligned} \quad (5)$$

with :

h_{ST}^2 : parameters of spatial-temporal bandwidth

h_S^2 : spatial bandwidth parameters

h_T^2 : temporal bandwidth parameters

Parameter τ is obtained through the determination coefficient optimization (R^2) method iteratively. So the parameter estimate τ can produce the maximum (R^2). The parameter τ is used to enlarge or minimize the effect of temporal distances on the effects of spatial distances. Then the parameter estimation μ and λ is obtained through the iterative method based on the parameter estimate τ which yields the maximum R^2 value. Let τ be the ratio parameter of $\frac{\mu}{\lambda}$ with $\lambda \neq 0$ then the equation is

obtained:

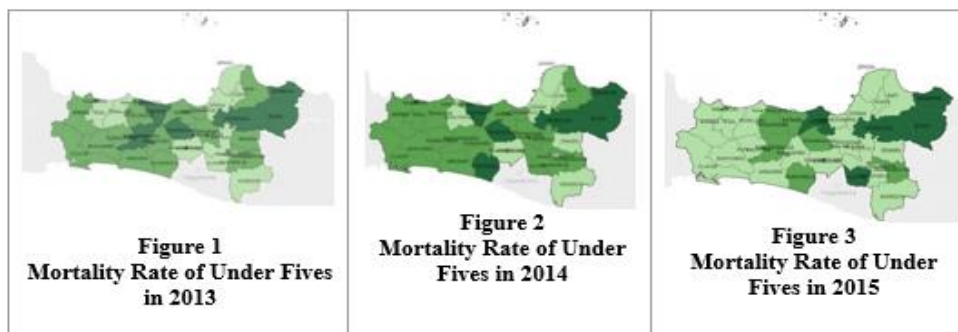
$$\frac{(d_{ij}^{ST})^2}{\lambda} = (u_i - u_j)^2 + (v_i - v_j)^2 + \tau (t_i - t_j)^2 \quad (6)$$

Parameter τ serves to enlarge or minimize the effect of temporal distance on spatial distance. This parameter is derived from the minimum CV criterion through initial initial value τ . Furthermore, parameter estimation μ and λ can be obtained by iterative method based on the estimation result τ that produce minimum CV. [6].

3. Result

3.1 Descriptive

Descriptive data is used to provide an overview of the data so that the information presented becomes easy to understand. This study uses secondary data sourced from Health Office and Central Bureau of Statistics of Central Java Province in 2013 - 2015 [7,8,9]. The observation unit in this research are 29 regencies and 6 cities in Central Java province. In the Figure 1, Figure 2, and Figure 3, the mortality rate of children under five in Central Java year 2013 – 2015 are presented.



In Figure 1, Mortality Rate of Under Fives in 2013, there are 10 areas with high mortality rate, namely Rembang, Blora, Grogoban, Salatiga, Magelang, Temanggung, Batang, Pekalongan, Banjarnegara and Tegal with percentage 31.43% - 39.83%. The mortality rate is 15 regions, namely Sragen, Pati, Karanganyar, Sukoharjo, Semarang, Semarang, Kendal, Purworejo, Kebumen, Purbalingga, Banyumas, Cilacap, Brebes, Pekalongan and Wonosobo with percentage 19.97% - 28.19%. Low toddler mortality rate there are 10 areas of Surakarta, Klaten, Wonogiri, Boyolali, Magelang, Demak, Kudus, Tegal, Pemalang and Jepara with percentage of 19.41%.

In Figure 2 Mortality Rate of Under Fives in 2014 there are 8 areas with high mortality rate, namely Tegal, Batang, Purworejo, Temanggung, Salatiga, Grogoban, Blora, Rembang with percentage 30.31% - 39.49%. Infant mortality rate is 19 regions, namely Brebes, Tegal, Cilacap, Banyumas, Pemalang, Purbalingga, Kebumen, Banjarnegara, Wonosobo, Pekalongan City, Kendal, Semarang City, Magelang City, Semarang, Boyolali, Klaten, Sukoharjo, Karanganyar and Pati listen percentage 18.65% - 27.80%. Low toddler mortality rate is 8 regions, namely Pekalongan, Magelang, Demak, Surakarta, Kudus, Jepara, Sragen and Wonogiri 18.01%.

In Figure 3 Under Fives Mortality Rate of 2015, there are 6 regions: Rembang, Magelang, Blora, Grogoban, Klaten, Kendal with percentage 31.88% - 37.85%. The under-five mortality rate is 8 Pekalongan, Pekalongan, Banjarnegara, Temanggung, Karanganyar, Purworejo, Batang, Tegal with percentage 24.23% - 30.26%. Low toddler mortality rate is 21 regions, namely Cilacap, Brebes, Tegal, Pemalang, Banyumas, Purbalingga, Kebumen, Wonosobo, Magelang, Semarang, Semarang, Boyolali, Sragen, Sukoharjo, Wonogiri, Demak, Jepara, Kudus, Pati, Surakarta, Kota Salatiga with a percentage of 29.06%.

3.2 Linear Regression Analysis Model

From the result of linear regression analysis using software R. 3.2.5 obtained linear regression model as shown in Table 1.

Table 1 Coefficients Output R. 3.2.5

Variabel	Coefficient	P-value	t _{hitung}
Constanta (β_0)	14.8245	0.051	1.974
Toddlers with Pneumonia (x_1)	0.035342	0.12218	1.561
Heavy Baby Low Birth (x_2)	1.592254	1.9e-05	4.525
Toddler Malnutrition (x_3)	1.882620	0.35447	0.931
Pregnant Women who do not complete Coverage K4 (x_4)	-0.015259	0.93648	-0.080
Non-Medical Birth (x_5)	-0.097500	0.55160	-0.598
Baby Exclusive ASI 0-6 Months (x_6)	0.059931	0.09830	1.671
Toddlers are not breastfed (x_7)	-0.440673	0.06117	-1.897
Toddler Weight Red Lines (x_8)	0.760493	0.37702	0.888
Toddlers who do not have Service coverage (x_9)	-0.009179	0.87495	-0.158
Unhealthy House (x_{10})	0.008013	0.31863	1.003
Decent Drinking Water (x_{11})	0.045665	0.50323	0.672
Sanitation Not Eligible (x_{12})	-0.103176	0.02491	-2.282
Posyandu (x_{13})	-0.056840	0.11930	-1.573
Women graduates of Elementary School (SD) (x_{14})	0.139785	0.04972	1.990
Woman of Younger Marriage Status (x_{15})	1.953588	0.00327	3.025
Puskesmas facilities (x_{16})	-0.181876	0.02070	-2.356
poverty (x_{17})	-0.049512	0.76111	-0.305

From Table 1 it can be seen that from 17 independent variables used, five variables that influence Underfive Mortality Rate in Central Java in 2013 - 2015 are Low Birth Weight Babies (x_2), Sanitation Not Eligible (x_{12}), Women graduates of Elementary School (x_{14}), Woman of Younger Marriage Status (x_{15}), and Puskesmas Facility (x_{16}) Therefore the model estimators formed from linear regression analysis are:

$$y = 14.8245 + 1.592254x_2 - 0.103176x_{12} + 0.139785x_{14} + 1.953588x_{15} - 0.181876x_{16}$$

From the data that has been analyzed in get value R^2 equal to 0.5082. The purpose of $R^2 = 0.5082 = 50.82\%$ is to illustrate that 50.82% of the variance occurring in the dependent variable (y) can be explained by the independent (x) variable possessed. The rest of 0.4918 (49.18%) is explained by other variables not included in the model. It is known that if the value of R^2 means greater the better. In this case R^2 is large, good value.

3.3 Overall Test

After the linear regression model is obtained, the next step is to perform simultaneous test, this test aims to test the role of independent variables in the linear regression model together. To perform a simultaneous test can see the output of software R. 3.2.5. After the linear regression analysis using the program R 3.2.5 obtained that p-value = 8.018e-08 therefore p-value < α so that the decision is rejected H_0 . Based on the results of the analysis conducted by using 95% confidence level, it is obtained that the decision taken is reject H_0 which means there is an influence of independent variables on the Infant Mortality Rate in Central Java, as shown in Table 2

Table 2 Value of Overall Test

R _{square}	F _{hitung}	p-value	MSE
0.5082	5.289	8.018e-08	27.137

3.4 Partial Test

Partial test is a test used to know that the parameters of each variable significant (means) or to test the parameters of β_i partially. P-value for Low Birth Weight, Inappropriate Sanitation, Women of Elementary School, Women with Young Age and Puskesmas Facilities, as shown in Table 3.

Table 3 Output R. 3.2.5 for partial test

Variabel	p-value	t _{hitung}
Constanta (β_0)	0.051	1.974
Low Birth Weight Babies (x_2)	1.9e-05	4.525
Sanitation Not Eligible (x_{12})	0.02491	-2.282
Young Married Women's Status (x_{15})	0.00327	3.025
Puskesmas facilities (x_{16})	0.02070	-2.356

Therefore, it can be decided that 5 significant variables reject H_0 . So it can be concluded that the independent variables that influence the variable of Underfive Mortality Rate significantly ie variable Weight Low Birth Wealth, Inappropriate Sanitation, Women with Young Age and Puskesmas Facilities.

3.5 Analysis of Spatial and Temporal Heterogeneity

Spatial heterogeneity occurs due to differences in characteristics of one region with another (random area effects). Testing spatial heterogeneity in the regression model is important because ignoring it will lead to inefficient estimates and inferred conclusions. Spatial heterogeneity testing was performed using the Breusch-Pagan Test test statistic [7]. Statistical Test p-value (0.3722) > α (0.05) then Failed Reject H_0 . By using 95% confidence level obtained decision for this analysis is rejected H_0 so it can be concluded that the residual of the model is homoscedasticity, as shown in Figure 4.

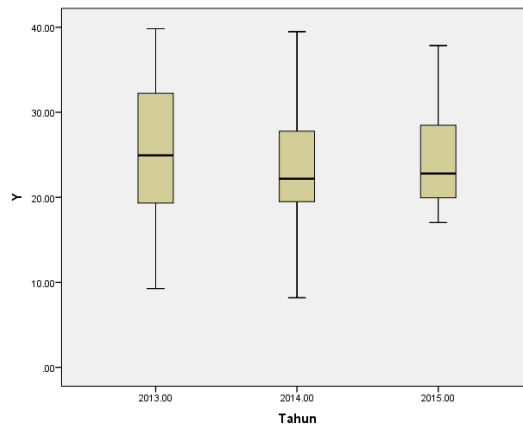


Figure 4 Heterogenitas Temporal Analisis

3.6 Modeling Using GTWR

Estimation of parameters that have significant effect on Underfive Mortality Rate on modeling of Underfive Mortality rate in Central Java using GTWR method. In modeling using GTWR we use different weighting matrices with GWR because of the influence of time (temporal). The main thing to do to get the euclidean distance matrix is to estimate the parameter τ . By using software R 3.2.5 in the Appendix obtained parameter estimation value τ iteratively with initialization $\tau_0 = 150$ with the final result of 1 at the minimum CV of 3601.164. Furthermore, parameter estimation μ and λ can be obtained iteratively based on result of estimation τ which yields CV minimum so that obtained value of μ and λ respectively are 3.9 and 3. After got value of μ and λ hence obtained value of spatial-temporal (hst) bandwidth equal to 131.9085 which can be used to generate a weighted matrix. In modeling using GTWR we use different weighting matrices with GWR because of the influence of time (temporal).

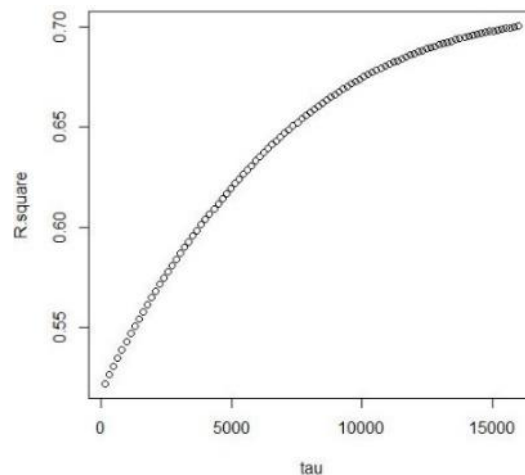


Figure 5 Iterated Results parameters

Figure 5 shows the result of the optimum parameter iteration estimation adalah is 16000 with R^2 value of 0.7003. The value of τ is the ratio between λ and μ . So then do the parameter estimation process λ and μ iteratively with initial value μ of 0.026 and λ of 0.02.

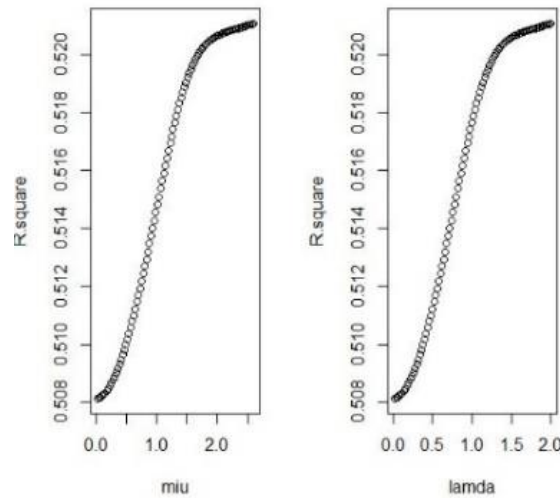


Figure 6 Iterations Estimation Parameters μ and λ

Figure 6 shows the result of the optimum parameter estimate μ and λ is 2.6 and 2 with R^2 value of 0.5210. After obtaining the parameter estimation value λ and μ , then got the value of spatial-temporal bandwidth (hSt) equal to 131.9058. The next step is to calculate the weighted matrix. Then parameter estimation β . There are 105 models obtained from 105 observation data. The model equation of GTWR location of Kebumen Regency in 2013 is as follows.

$$y = 13.95856 + 1.573918X_2 - 0.10022X_{12} + 1.882818X_{15} - 0.18495X_{16}$$

The equation explains that, if the independent variable is fixed and Low Birth Weight increases one unit then the Underfive Mortality Rate will increase by 1.573918. Inappropriate Sanitation has increased one unit then Infant Mortality Rate will decrease by 1.882818. Women with Elementary Grade have an increase of one unit then Mortality Rate of Under Fives will increase by 1.882818, Puskesmas facilities have increased one unit then Infant Mortality Rate will decrease equal to 0.18495.

3.7 Comparison of Multiple Linear Regression Model and GTWR Model

A larger parameter value λ indicates that the temporal heterogeneous effect has a greater effect on the modeling. The value of determination coefficient (R^2) in the attachment obtained based on GTWR model is 52.10%. This value indicates that the obtained GTWR model is able to explain the Infant Mortality Rate in Central Java of 52.10%, while the remaining 47.9% is influenced by other predictor variables that have not been included in the model. Further comparison is done to decide the best modeling of Underfive Mortality Rate in Central Java.

Table 4 Comparison of Multiple Linear Regression Model and GTWR

Method	R^2	MSE
Linear Regression	0.5080	27.137
GTWR	0.5210	1.084

To compare the best model between global regression and GTWR can use the easiest step, that is comparing the value of R^2 and MSE. The best model is obtained by the criteria of the larger R^2 value and the smaller MSE value. Table 4 shows that modeling of under-five mortality rate in Central Java using GTWR gives better result than modeling using global regression. These results show that global regression gives only R^2 value 0,5080 and MSE 27,137 while GTWR model is able to maximize R^2 value equal to 0.5210 with MSE value smaller than global regression that is 1.084.

4. Conclusion

Based on the results and discussion in the previous chapter, can be obtained some conclusions, including: (1) Factors that significantly affect Underfive Mortality Rate in Central Java Province based on global regression model are. Global regression model obtained is: $y = 14.8245 + 1.592254x_2 - 0.103176x_{12} + 0.139785x_{14} + 1.953588x_{15} - 0.181876x_{16}$; (2) There are 105 models obtained from 105 observation data on districts / cities in Central Java Province in 2013-2015 in Appendix 1. The model equation of GTWR location of Kebumen Regency in 2013 is as follows: $y = 13.95856 + 1.573918X_2 - 0.10022X_{12} + 1.882818X_{15} - 0.18495X_{16}$; (3) After comparison, it is found that open unemployment modeling in Central Java using GTWR is better than using global regression. This is because the GTWR model is able to maximize the R^2 value of 0.5210 with the MSE value smaller than the global regression of 1,084. While the global regression model gives only R^2 of 0.5080 and MSE of 27,137.

5. Reference

- [1] BPS 2016 Pengertian Angka Kematian Balita Retrieved from <https://sirusa.bps.go.id/index.php?r=indikator/view&id=78>
- [2] Yeni K 2014 Pemodelan Proporsi Penduduk Miskin Kabupaten dan Kota di Provinsi Jawa Tengah Menggunakan Geographically and Temporally Weighted Regression *Jurnal Statistika Universitas Diponegoro*
- [3] Aisyah K 2014 Pemodelan Konsentrasi Partikel Debu (PM10) pada Pencemaran Udara di Kota Surabaya dengan Metode Geographically-Temporally Weighted Regression *Jurnal Statistika Institut Teknologi Sepuluh November*
- [4] Anselin L & Rey S J 2010 Perspectives on Spatial Data Analysis (New York. Springer)
- [5] Wang P 2006 Exploring spatial effects on housing price: the case study of the city of Calgary *Master dissertation* (Canada: University of Calgary)
- [6] Choiruddin A & Sutikno 2013 Pemodelan Indikator Pencemar Biological Oxygen Demand di Kali Surabaya Menggunakan Pendekatan Spatial-Temporal Weighted Regression (STWR) *Jurnal Statistika Institut Teknologi Sepuluh November*
- [7] Dinas Kesehatan Jawa Tengah 2015 Profil Kesehatan Jawa Tengah tahun 2015 Retrieved from http://www.depkes.go.id/resources/download/profil/PROFIL_KES_PROVINSI_2015/13_Jateng_2015.pdf
- [8] Dinas Kesehatan Jawa Tengah 2014 Profil Kesehatan Jawa Tengah tahun 2014 Retrieved from www.depkes.go.id/resources/download/profil/PROFIL_KES.../13_Jateng_2014.pdf.
- [9] Dinas Kesehatan Jawa Tengah 2013 Profil Kesehatan Jawa Tengah tahun 2013 Retrieved from www.depkes.go.id/resources/download/profil/...2013/13_Prov_Jateng_2013.pdf