

THE BEST GLOBAL AND LOCAL VARIABLES OF THE MIXED GEOGRAPHICALLY AND TEMPORALLY WEIGHTED REGRESSION MODEL *

Nuramaliyah¹, Asep Saefuddin^{2‡}, and Muhammad Nur Aidi³

¹Department of Statistics, IPB University, Indonesia, nuramaliyah12@gmail.com

²Department of Statistics, IPB University, Indonesia, asaefuddin@gmail.com

³Department of Statistics, IPB University, Indonesia, nuraidi@yahoo.com

‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 3 No 3 (2019), 320 - 330

Copyright © 2019 Nuramaliyah, Asep Saefuddin, and Muhammad Nur Aidi. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Geographically and temporally weighted regression (GTWR) is a method used when there is spatial and temporal diversity in an observation. GTWR model just consider the local influences of spatial-temporal independent variables on dependent variable. In some cases, the model not only about local influences but there are the global influences of spatial-temporal variables too, so that mixed geographically and temporally weighted regression (MGTWR) model more suitable to use. This study aimed to determine the best global and local variables in MGTWR and to determine the model to be used in North Sumatra's poverty cases in 2010 to 2015. The result show that the Unemployment rate and labor force participation rates are global variables. Whereas the variable literacy rate, school enrollment rates and households buying rice for poor (raskin) are local variables. Furthermore, based on Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC) showed that MGTWR better than GTWR when it used in North Sumatra's poverty cases.

Keywords: global and local spatial-temporal variable, mixed geographically and temporally weighted regression, poverty.

* Received Sep 2019; Accepted Oct 2019; Published online on Oct 2019

1. Pendahuluan

Linear regression is a method that has a linear effect between the independent and response variable. In linear regression, the coefficients are the same for all regions, whereas each region sometimes has different influencing factors, consequently, spatial diversity does exist. Spatial diversity occurs when one location to another has a different relationship between variables (Fotheringham et al., 2002). Therefore, Geographically Weighted Regression (GWR) model is more suitable for this condition than linear regression

GWR is a method that used for exploring spatial nonstationarity (Leung et al., 2000). GWR constructs a regression model that has local effects at each observation location (Rahmawati & Djuraidah, 2010). GWR is a development of linear regression model at each location by adding weight in the model (Brunsdon et al., 1999). In 2002, Fotheringham et al. said that the regression parameters in the GWR model are assumed to vary spatially so that it provides different interpretations for each observation. For example, in the case of poverty, the factors that affect poverty in each region are different, so poverty is more accurately approached using the GWR model.

In its application, the GWR model can only control the spatial effect without controlling the time effect (Ma et al., 2018). In fact, when there is a time effect, the factors that influence will also different. therefore, the GWR model is not appropriate for the case with time effect. The addition of time effect can be solved by applying the geographically and temporally weighted regression (GTWR) method. This method generates a model that is local for each location and for each time so that the result is more representative (Huang et al., 2010).

Each region has diverse causes of poverty, so an approach that can determine explanatory variables that affect the model, both locally and globally is needed. In dealing with these problems, the GTWR method is not appropriate anymore, that is why the mixed geographically and temporally weighted regression (MGTWR) method is developed to seeking out an influential explanatory variable, both locally according to location and time and globally for all locations and times.

Study on MGTWR has been widely carried out. In the environmental field, Winarso et al., (2014) modeled the air pollution standards index in Surabaya with MGWR and MGTWR, revealed that the best model for the effect of air pollution in Surabaya is MGTWR. In 2015, Utama (2015) conducted research on poverty cases in East Java using the MGTWR method using the Gaussian kernel and Bisquare kernel functions. The results of this study are that MGTWR with a Gaussian kernel function is better used in modeling poverty cases in East Java Province. On the other hand, modeling of poverty data with MGTWR was also carried out by Yasin et al. (2015). Liu et al. (2017) comparing the MGWR, GTWR, and MGTWR methods based on the exploration of temporal-spatial variations from a global and local perspective.

Based on these, this study aims to find out the best model with the best global and local variables on the poverty rate in North Sumatra Province in 2010-2015 using the MGTWR.

2. Methodology

2.1. Research Materials

(i) Breusch-Pagan Test

The Breusch-Pagan Test (BP) is used to test spatial heterogeneity related to differences in environmental and geographical characteristics between locations of observation. Thus, each observation may have different variations or there is an effect of explanatory variables on the response variables for each observation location (Anselin, 1988).

$$BP = \frac{1}{2} \mathbf{f}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{f} \sim \chi^2_{(p)} \quad (1)$$

and vector \mathbf{f} is:

$$f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1 \right) \quad (2)$$

with \mathbf{Z} is a matrix with dimension $n \times (p + 1)$ containing a standardized vector.

(ii) Mixed Geographically and Temporally Weighted Regression (MGTWR)

Mixed geographically and temporally weighted regression (MGTWR) model is the development of the MGWR model that involves the time effect. Similar to the MGWR model, the MGTWR model contains several coefficients of explanatory variables, assumed to be constant for all observation while other coefficients vary according to the location of the observation. The MGTWR model can be expressed as follows (Liu et al., 2017):

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{g=1}^p \beta_g x_{ig} + \sum_{l=1}^q \beta_l(u_i, v_i, t_i) x_{il} + \varepsilon_i \quad (3)$$

Parameters of MGTWR model $\hat{\beta}_l(u_i, v_i, t_i)$ are obtained using the weighted least squares method by giving different weights for each location and time. The steps in estimating the MGTWR model are the same as those in estimating the GTWR model (Fotheringham et al., 2002), can be expressed as follows:

$$\hat{\beta}_g = [X'_g(I - S_l)'(I - S_l)X_g]^{-1} X'_g(I - S_l)'(I - S_l)Y' \quad (4)$$

$$\hat{\beta}_l(u_i, v_i, t_i) = [X'_l W(u_i, v_i, t_i) X_l]^{-1} X'_l W(u_i, v_i, t_i) (Y - X_g \hat{\beta}_g) \quad (5)$$

with \mathbf{W} is

$$W(u_i, v_i, t_i) = \text{diag}[w_1(u_i, v_i, t_i), w_2(u_i, v_i, t_i), \dots, w_n(u_i, v_i, t_i)] \quad (6)$$

and

$$S_l = \begin{bmatrix} X'_{l1}[X'_l W(u_1, v_1, t_1)X_l]^{-1}X'_l W(u_1, v_1, t_1) \\ X'_{l2}[X'_l W(u_2, v_2, t_2)X_l]^{-1}X'_l W(u_2, v_2, t_2) \\ \vdots \\ X'_{ln}[X'_l W(u_n, v_n, t_n)X_l]^{-1}X'_l W(u_n, v_n, t_n) \end{bmatrix} \quad (7)$$

The definition and measurement of spatial and temporal distances occur because the estimation of the model parameters of each observation location located near the *i*-th location is more influential. Determination of the spatial and temporal approach is based on the euclidean distance function can be written as follows:

$$(d_{ij}^S)^2 = (u_i - u_j)^2 + (v_i - v_j)^2 \quad (8)$$

$$(d_{ij}^T)^2 = (t_i - t_j)^2 \quad (9)$$

Because of the differences in scale spatially and temporally, the spatial-temporal distance function (d^{ST}) was formed through a combination of the spatial distance function (d^S) and the temporal distance function (d^T). Spatial-temporal distance functions can be written as follows:

$$(d_{ij}^{ST})^2 = \lambda(d_{ij}^S)^2 + \mu(d_{ij}^T)^2$$

$$(d_{ij}^{ST})^2 = \lambda [(u_i - u_j)^2 + (v_i - v_j)^2] + \mu(t_i - t_j)^2 \quad (10)$$

with λ and μ are balancing parameter to the effect of the scale difference between location and time on spatial-temporal distance measurement (Huang et al., 2010).

Based on Liu et al. (2017), For example τ is a parameter ratio of $\tau = \mu/\lambda$ with $\lambda \neq 0$, we get the following equation:

$$\frac{(d_{ij}^{ST})^2}{\lambda} = (u_i - u_j)^2 + (v_i - v_j)^2 + \tau(t_i - t_j)^2 \quad (11)$$

τ is obtained from the minimum cross validation (CV) criteria through initial initialization of the τ value, can be expressed as follows:

$$CV(\tau) = \sum_{i=1}^n [y_i - \hat{y}_{\neq i}(\tau)]^2 \quad (12)$$

2.2. Data and Analysis Method

(i) Data

Data used in this study is secondary data of poverty in North Sumatra of each district/cities and factors that affect it in 2010-2015. Data were obtained from the publications of "Provinsi Sumatera Utara dalam Angka 2010-2015" and "Statistik Kesejahteraan Rakyat Provinsi Sumatera Utara 2010-2015" published by Central Bureau of Statistics (BPS, 2015a, 2015b). There are 33 district/cities in this study with variables as presented in Table 1.

Table 1: The variables used this the study.

Variables	Note	Unit Variable
Y	Poverty rate	%
X ₁	Literacy rate	%
X ₂	Unemployment rate	%
X ₃	School enrollment rate	%
X ₄	Labor force participation rate	%
X ₅	Households buying rice for the poor	%

(ii) Analysis Procedure

Steps of data analysis using R Studio are carried out as follows:

1. Explore the data.
2. Test the spatial and time heteroscedastic using the Breusch-Pagan (BP) test.

a. Determine the hypothesis of BP test as follows (Breusch & Pagan, 1979):

H₀: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma = 0$ (homoscedastic disturbances)

H₁: At least there is one i where $\sigma_i^2 \neq \sigma$ (heteroscedastic disturbances)

b. Calculate the BP test statistic as follows:

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

with

$$f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1 \right)$$

3. Determine the best global and local variables using the BP test.

a. Determine the function of x variables as follows:

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

b. Determine the hypothesis BP test as follows:

H₀: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma = 0$ (Global Variable)

H₁: At least there is one i where $\sigma_i^2 \neq \sigma$ (Local Variable)

c. Compare BP value and $\chi^2_{(k-1)}$. If $BP > \chi^2_{(k-1)}$, then that variables are local variables. While If $BP < \chi^2_{(k-1)}$, then that variables are global variables.

4. MGTWR analysis model

a. Determine spatial-temporal parameter (τ), spatial parameter (λ), temporal parameter (μ) using the minimum CV.

$$CV(\tau) = \sum_{i=1}^n [y_{\neq i} - \hat{y}_{\neq i}(\tau)]^2$$

b. Determine the optimum bandwidth (h_{sT}) using the minimum CV.

$$CV(h_{sT}) = \sum_{i=1}^n [y_{\neq i} - \hat{y}_{\neq i}(h_{sT})]^2$$

c. Calculate the esimated value of global ($\hat{\beta}_g$) and local ($\hat{\beta}_l$) parameters.

$$\hat{\beta}_g = [X'_g(I - S_l)'(I - S_l)X_g]^{-1} X'_g(I - S_l)'(I - S_l)Y'$$

$$\hat{\beta}_l(u_i, v_i, t_i) = [X'_l W(u_i, v_i, t_i) X_l]^{-1} X'_l W(u_i, v_i, t_i) (Y - X_g \hat{\beta}_g)$$

- d. Test the global parameters $\hat{\beta}_g$ and the local parameters $\hat{\beta}_l$ of MGTWR model.
5. Find the best model for poverty in North Sumatra Province in 2010-2015 using root mean square error (RMSE) and akaike information criterion (AIC).
6. Establish the distribution map of poverty in Nort Sumtra Province based on significant explanatory variables.

3. Results

3.1 Data Exploration

Poverty is a major problem that is complex and multidimensional. Poverty can be a benchmark of socio-economic conditions in assessing the success of development carried out by local governments. Poverty is influenced by several factors namely education, employment, also health and other factors. North Sumatra Province is one of provinces with a high poverty rate, even have the highest poverty rate outside of Java Island. North Sumatra Province is a large province consisting of 25 districts and 8 cities spread across various regions.

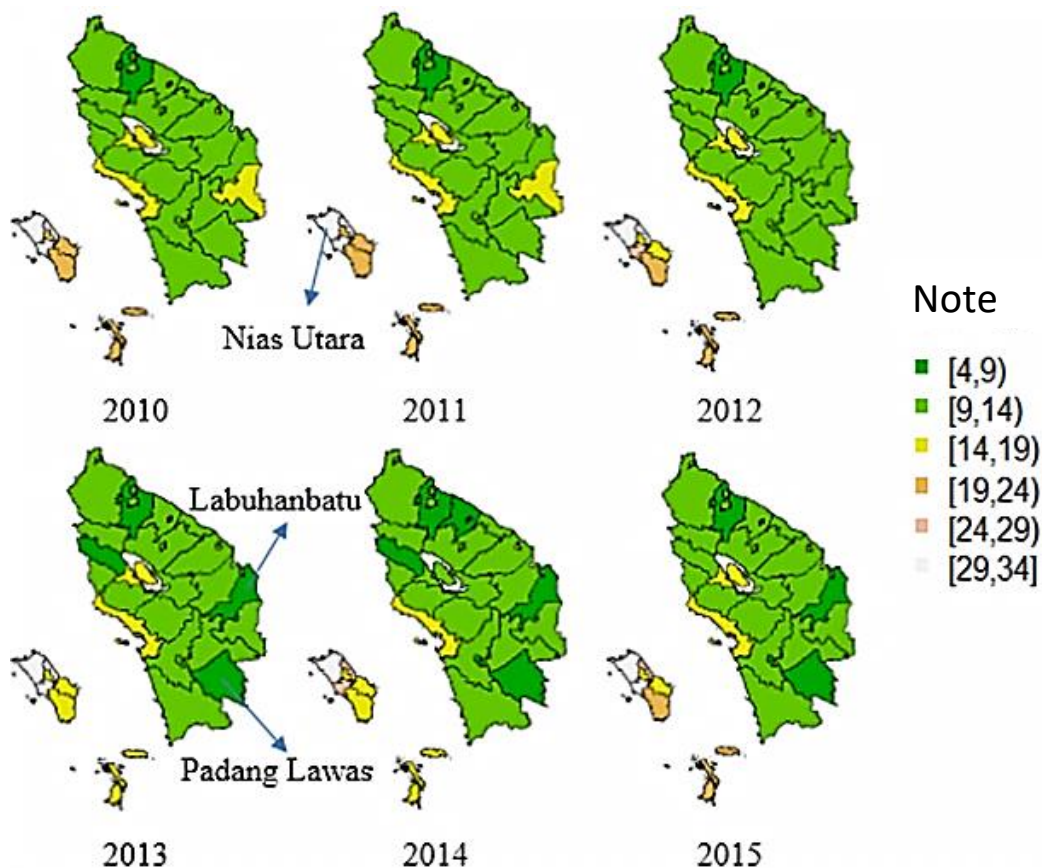


Figure 1: Distribution map of the North Sumatra's poverty rate in 2010-2015.

Figure 1 shows that districts/cities in North Sumatra have poverty levels that are almost similar or do not differ greatly between regions. In addition, Figure 1 also

describes that Nias Utara Regency has the highest poverty rate of all. In the period 2010 to 2015, it appears that there was a decline in poverty levels in Labuhanbatu Regency, Padang Lawas Regency, and in several other regions.

3.2 Spatial Heteroscedastic

Testing for spatial heteroscedastic using the Breusch-Pagan (BP) test. This test is conducted to determine whether there is diversity due to spatial effect. The test is conducted annually and simultaneously on 33 districts/cities in North Sumatra province in 2010 to 2015 with a significance level of 5% and 10%. Results of the Breusch-Pagan test are presented in Table 2.

Table 2: Breusch-Pagan test.

Year	BP statistic value	P-value
2010	12.167	0.03256**
2011	13.069	0.02274**
2012	11.064	0.05013*
2013	12.856	0.02477**
2014	13.055	0.02286**
2015	6.970	0.22290

Note : * significant at $\alpha = 10\%$, ** significant at $\alpha = 5\%$

Based on Table 2, it is known that observations for 2010 to 2014 showed spatial heteroscedastic, but no spatial heteroscedastic occurred in poverty rate of North Sumatra in 2015. As a result, simultaneous testing of spatial heteroscedastic showed that there was spatial diversity in poverty levels in North Sumatra. Therefore, a simple regression method is not appropriate to be used. Thus, the spatial regression method using geographically weighted regression is more appropriate for the poverty rate in North Sumatra.

3.3 Mixed Geographically and Temporally Weighted Regression (MGTWR)

GTWR Model can only obtain local models, whereas if there are global and local models, the use of the MGTWR model is needed. The MGTWR model consists of two variables, namely global and local variables. The determination of this variable is based on the Breusch-pagan test value obtained from the combination of the response variable (Y) with all the independent variables ($X_1, X_2, X_3, X_4,$ and X_5). The results of the determination of these variables are 31 combinations of Y with all X and some major combinations taken between Y and X are presented in Table 3. Table 3 shows that X_2 and X_4 are global variables because the the p-value is greater than the significance level of 0.05. Whereas X_1, X_3 and X_5 are local variables because the p-value is less than the significance level 0.05.

Table 3: The Breusch-Pagan Test combination for the determination of variables.

Variable	BP test value	P-value	Note
Y dengan X_1	49.693	1.798×10^{-12} *	Local
Y dengan X_2	3.5657	0.058	Global
Y dengan X_3	21.234	4.064×10^{-12} *	Local
Y dengan X_4	1.749	0.186	Global
Y dengan X_5	17.302	0.000*	Local

Note: * significant at $\alpha = 5\%$

MGTWR modeling works with the selected Exponential kernel function based on the smallest Cross-Validation (CV) value. Based on the kernel function, the spatial distance (λ) is 1.38611, temporal distance parameter (μ) is 0.00129, spatial-temporal distance parameter (τ) is 0.00088, and optimum bandwidth (h_{ST}) is 0.058. The distance parameters are obtained based on the smallest result of cross-validation. MGTWR modeling produces modeling as much as the location and time studied. In this study, there are 33 locations \times 6 times = 198 models. Each location has 6 models that are distinguished by year, 2010 to 2015.

3.4 Estimated Parameter of MGTWR Model

The MGTWR model has different parameter estimates at each location and at each time. Table 4 and Table 5 show the summary results of global and local estimated parameters of the MGTWR model.

Table 4: Global estimated parameter of the MGTWR model in North Sumatra.

Parameter	Estimation Value
$\hat{\beta}_2$	0.02101
$\hat{\beta}_4$	-0.00273

The parameters $\hat{\beta}_2$ and $\hat{\beta}_4$ are parameters for the unemployment rate (X_2) and the labor force participation rate (X_4), which are global variables, respectively. Parameters $\hat{\beta}_2$ and $\hat{\beta}_4$ had no significant effect on the poverty 0.02101 rate between locations in North Sumatra.

Table 5: Summary of Local estimated parameter in MGTWR model in North Sumatra.

Parameter	Minimum	Maximum	Average	Standard Deviation
$\hat{\beta}_0$	-179.85606	225.64410	51.46219	64.98797
$\hat{\beta}_1$	-2.24232	1.90538	-0.35423	0.70314
$\hat{\beta}_3$	-1.90062	1.15481	-0.13367	0.51396
$\hat{\beta}_5$	-0.15880	0.32157	0.01170	0.08371

The parameters $\hat{\beta}_1, \hat{\beta}_3$ and $\hat{\beta}_5$ are parameters for the literacy rate (X_1), the school participation (X_3) and the households buying rice for the poor (X_5) which are local variables. Parameters $\hat{\beta}_1, \hat{\beta}_3$ and $\hat{\beta}_5$ have significant influences on poverty rate between each locations in North Sumatra. In 2010 to 2015, the parameters $\hat{\beta}_1, \hat{\beta}_3$ and $\hat{\beta}_5$ gave positive and negative values to district/cities in North Sumatra.

The MGTWR model can be formed from the all of $\hat{\beta}$ values. All of $\hat{\beta}$ have different value according to lokasi and time. So that, in this method there are 198 models from 33 locations x 6 times. Based on Table 4 and Table 5, the MGTWR model of poverty in North Sumatra Province in 2010-2015 can be written as follows:

$$\hat{Y}_i = \hat{\beta}_0 - \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} - \hat{\beta}_3 X_{i3} - \hat{\beta}_4 X_{i4} + \hat{\beta}_5 X_{i5}$$

$$\hat{Y}_i = 51.46219 - 0.35423 X_{i1} + 0.02101 X_{i2} - 0.35423 X_{i3} - 0.00273 X_{i4} + 0.01170 X_{i5}$$

The MGTWR model of poverty in North Sumatra Province in 2010-2015 contains average value of $\hat{\beta}_1, \hat{\beta}_3$, and $\hat{\beta}_5$ from 198 models. So that, the model is only an average of 198 models. The model explains that if the literacy rate rise by one percent then the poverty rate decrease by 0.35423 percent assuming other variables are constant.

3.5 Parameter Testing of MGTWR Model

The explanatory variables in the MGTWR model have different effects in each region and each time so that it can be said that the poverty rate in regions is influenced by different factors. If these factors are grouped according to several categories with similar characteristics, then there are 8 categories, as presented in Table 6.

Table 6: The grouping of regions is based on the similarity of significant variables.

Significant variables	Region	Group
-	South Tapanuli, Simalungun	1
X_1	North Tapanuli, Toba Samosir, Dairi, Karo, Humbang Hasundutan, Samosir, Padang Lawas, North Labuhanbatu, Pematangsiantar, Padangsidempuan	2
X_3	Serdang Bedagai	3
X_5	South Nias	4
X_1, X_3	Mandailing Natal, Central Tapanuli, Labuhanbatu, Asahan, Nort Nias, Sibolga, Binjai, Gunungsitoli	5
X_1, X_5	Deli Serdang, Pakpak Bharat, North Padang Lawas, South Labuhanbatu, West Nias,	6
X_3, X_5	Batu Bara, Tanjungbalai, Tebing Tinggi	7
X_1, X_3, X_5	Nias, Langkat, Medan	8

3.6 Comparison Goodness of Model

The goodness of fit can be carried out by Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC), as presented in Table 7. It shows that the MGTWR has a smaller RMSE and AIC than the GTWR. This indicates that the variation produced by the MGTWR model is almost the same as the variation in the observed data. In other words, the MGTWR is more accurate than the GTWR.

Table 7: The comparison of goodness of model.

Model	RMSE	AIC
GTWR	1.10771	1553.968
MGTWR	0.63279	662.279

4. Conclusion

Based on the results and the discussion conducted it can be concluded that the variables that have global influences on the MGTWR model are the open unemployment rate (X_2) and the labor force participation rate (X_4). Variable literacy rate (X_1), school participation rate (X_3) and households buying rice for the poor (X_5) are local influences. In addition, the poverty rate in North Sumatera 2010 to 2015, the MGTWR is better compared to the GTWR model.

Daftar Pustaka

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. New York (US): Springer Science & Business Media.
- [BPS] Badan Pusat Statistik Sumatera Utara. (2015a). *Provinsi Sumatera Utara dalam angka 2010-2015*. Medan (ID): Badan Pusat Statistik Sumatera Utara.
- [BPS] Badan Pusat Statistik Sumatera Utara. (2015b). *Statistik kesejahteraan rakyat provinsi Sumatera Utara 2010-2015*. Medan (ID): Badan Pusat Statistik Sumatera Utara.
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, **47**(5): 1287–1294.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, **39**(3): 497–524.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester, UK: John Wiley & Sons.
- Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, **24**(3): 383–401.

- Leung, Y., Mei, C.-L., & Zhang, W.-X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, **32**(1): 9–32.
- Liu, J., Zhao, Y., Yang, Y., Xu, S., Zhang, F., Zhang, X., ... Qiu, A. (2017). A mixed geographically and temporally weighted regression: Exploring spatial-temporal variations from global and local perspectives. *Entropy*, **19**(2): 53.
- Ma, X., Zhang, J., Ding, C., & Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Computers, Environment and Urban Systems*, **70**: 113–124.
- Rahmawati, R., & Djuraidah, A. (2010). Regresi Terboboti Geografis dengan Pembobot Kernel Kuadrat Ganda untuk Data Kemiskinan di Kabupaten Jember. *Forum Statistika Dan Komputasi*, **15**(2): 32–37.
- Utama, N. (2015). *Pemodelan mixed geographically temporally weighted regression (mgtwr) pada kasus kemiskinan di provinsi jawa timur* (Skripsi). Universitas Brawijaya, Malang (ID).
- Winarso, K., Notobroto, H. B., & Fatmawati. (2014). Development of air polluter model for the carbon monoxide (CO) element based on mixed geographically temporal weighted regression (MGTWR) Kriging. *Applied Mathematical Sciences*, **8**(118): 5863–5873.
- Yasin, H., Sugito, & Prahutama, A. (2015). Analisis data kemiskinan di Jawa Tengah menggunakan metode mixed geographically and temporally weighted regressions (MGTWR). *BIAStatistics*, **9**(1): 15–23.