Geographically Weighted Regression (GWR) Modelling with Weighted Fixed Gaussian Kernel and Queen Contiguity for Dengue Fever Case Data

Grissila Yustisia[1], Waego Hadi Nugroho [1], Maria Bernadetha Theresia Mitakda [1]

[1]Department of Statistics, Brawijaya University, Malang

Email: y_grissila@yahoo.com

ABSTRACT

Regression analysis is a method for determining the effect of the response and predictor variables, yet simple regression does not consider the different properties in each location. Methods Geographically Weighted Regression (GWR) is a technique point of approach to a simple regression model be weighted regression model. The purpose of this study is to establish a model using Geographically Weighted Regression (GWR) with a weighted Fixed Gaussian Kernel and Queen Contiguity in cases of dengue fever patients and to determine the best weighting between the weighted Euclidean distance as well as the Queen Contiguity based on the value of R2. Results from the study showed that the modeling Geographically Weighted Regression (GWR) with a weighted Fixed Gaussian Kernel showed that all predictor variables affect the number of dengue fever patients, whereas the weighted Queen Contiguity, not all predictor variables affect the dengue fever patients. Based on the value of R2 is known that a weighted Fixed Gaussian Kernel is better used.

Keywords: geographically weighted regression, fixed gaussian kernel, queen contiguity

INTRODUCTION

Spatial data is the measurement data containing location information. Methods Geographically Weighted Regression (GWR) is a technique point of approach to a simple regression model be weighted regression model [1]. Spatial weighting matrix is used to determine the closeness of the relationship between the regions. GWR weighting role model is very important because it represents a weighted value of research data layout. Weighted grouped by distance (distance) and region (contiguity) [2]. In GWR models often use a weighting based on the distance (distance) without considering other weighting associated with the area (contiguity).

In GWR models often use a weighting based on the distance (distance) without considering other weighting associated with the area (contiguity). In general, dengue fever clustered in specific locations. Things to note about the location other than the one with the distance (distance), which is the state of the location or about the area (contiguity). Therefore, in this study will consider the distance (distance) and region (contiguity) as weighting to search for the best model in dengue cases.

METHODS

2.1 Testing the Effect of Spatial

Testing of spatial heterogeneity can be done by using the test statistic Breusch-Pagan (BP), is based on the hypothesis

H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_j^2 = \sigma^2$

H_1: there are at least one j where $\sigma_j^2 \neq \sigma^2$

If H_0 true test statistic:

$$\text{BP} = \left(\frac{1}{2}\right) \boldsymbol{f}_{(1xn)}^T \boldsymbol{Z}_{(nxn)} \left(\boldsymbol{Z}_{(nxn)}^T \boldsymbol{Z}_{(nxn)}\right)^{-1} \boldsymbol{Z}_{(nxn)}^T \boldsymbol{f}_{(nx1)} + \left(\frac{1}{T}\right)\left[\frac{e_{(1xn)}^T \boldsymbol{W}_{(nxn)} e_{(nx1)}}{\sigma^2}\right]^2 \sim \chi_{(p+1)}^2 \qquad (1)$$

2.2 Fixed Gaussian Kernel

Fixed Gaussian kernel is the weighting matrix based on the proximity of the location of the observation point i and point to another location. Fixed weighted Gaussian kernel as follows [1]:

$$w_{ij} = \exp[-((\text{d}_{ij}/h)/2)^2] \qquad (2)$$

If the location of the i located at coordinates (u_i, v_i) it will obtain the Euclidean distance (d_{ij}) between all locations i and j are:

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \qquad (3)$$

Bandwidth is a circle of radius from the center point location. Methods to determine the bandwidth is Cross Validation (CV) [1]):

$$\text{CV} = \sum_{i=1}^{n}\left(y_i - \hat{y}_{\neq i}(h)\right)^2 \qquad (4)$$

2.3 Queen Contiguity

Observations on the adjacent locations tend to be similar compared to locations far apart, because they relate to weighted location [3].

$$w_{ij} = \begin{cases} 1, & \text{if regions i and j intersect,} \quad \text{where} \quad i \neq j \\ 0, & \text{another} \end{cases}$$

$$w_{ij} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \overbrace{\begin{matrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{matrix}}^{w_i} \qquad (5)$$

$$w_i = \sum W_{Q(nxn)}$$

$$w_{ij} = \frac{W_{Q(nxn)}}{w_i}$$

where:

w_i : Sum total row

w_{ij} : The weighting matrix row i column j

2.4 Geographically Weighted Regression

Spatial data is the measurement data containing location information. Methods Geographically Weighted Regression (GWR) is a technique point of approach to a simple regression model be weighted regression model [1]. According to Yasin [4] model of GWR is:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i)x_{ij} + \varepsilon_i$$

(u_i, v_i)= Coordinates (longitude, latitude) point i to a geographical location.

2.5 Testing Geographically Weighted Regression Model Parameters

Testing Geographically Weighted Regression model parameters is done simultaneously and partially [5]:

1. Simultaneous testing to determine the effect of predictor variables together against the response variable.

   If H_0 true test statistic:

   $$\frac{JK_{sisa}(H_1)\Big/\frac{\delta_1^2}{\delta_2}}{JK_{sisa}(H_0)\Big/n-(p+1)} \sim F^*_{(\frac{\delta_1^2}{\delta_2},n-(p+1))} \tag{6}$$

   $\delta_1 = trace\{(I-L)^T(I-L)\}$

   $\delta_2 = trace\{(I-L)^T(I-L)\}^2$

   where $JK_{sisa}(H_0) = Y^T_{(1xn)}\big(I_{(nxn)} - L_{(nxn)}\big)^T\big(I_{(nxn)} - L_{(nxn)}\big)Y_{(nx1)}$

   $$JK_{sisa}(H_1) = Y^T_{(1xn)}\big(I_{(nxn)} - H_{(nxn)}\big)Y_{(nx1)}$$
   $$H = X_{(nxp)}(X^T_{(pxn)}X_{(nxp)})^{-1}X^T_{(pxn)}$$
   $$L_{(nxn)} = \begin{pmatrix} x_1^T[X^TW(u_1,v_1)X]^{-1}X^TW(u_1,v_1) \\ x_2^T[X^TW(u_2,v_2)X]^{-1}X^TW(u_2,v_2) \\ \vdots \\ x_n^T[X^TW(u_n,v_n)X]^{-1}X^TW(u_n,v_n) \end{pmatrix}$$

   I = identity matrix of order n

   using criteria reject H_0 if the test statistic $|F| > F^*_{(\alpha,\frac{\delta_1^2}{\delta_2},n-(p+1))}$

2. Partial testing to determine which predictor variables that influence the response variable for each observation location, using the t test statistic is based on the hypothesis:

   H_0: $\beta_j(u_i,v_i) = 0$

   H_1: $\beta_j(u_i,v_i) \neq 0$, j = 1, 2, $\cdots$, p

   t test statistic can be written as follows:

   $$\frac{\hat{\beta}_j(u_i,v_i)}{\hat{\sigma}\sqrt{c_{jj}}} \sim t_{(n-(p+1))} \tag{7}$$

   $$\sigma^2 = \frac{JK_{sisa}}{db_{galat}} = \frac{e'e}{n-(p+1)}$$

   where c_{jj} a diagonal matrix element $\mathbf{CC^T}$,

   $$C = (X^T_{(pxn)}W(u_i,v_i)_{(nxn)}X_{(nxp)})^{-1} \tag{8}$$

   Reject if the test statistic $|t| > t_{(\frac{\alpha}{2}, n-(p+1))}$

2.5 Testing Assessment Geographically Weighted Regression Model

The coefficient of determination can describe the magnitude of the response variable diversity can be explained by the predictor variables. GWR R^2 value obtained by the following mathematical equation [1]:

$$R^2(u_i,v_i) = \frac{JKR_W}{JKT_W} = \frac{\sum_{j=1}^p W_{ij}(y_i-\hat{y}_i)^2}{\sum_{j=1}^p W_{ij}(y_i-\bar{y}_i)^2} \tag{9}$$

RESULTS AND DISCUSSION

The results of statistical calculations Breusch-Pagan test with both weighting are presented in Table 1.

Table 1. Testing Result *Breusch-Pagan*

Weighting	*Breusch-pagan*
Fixed Gaussian Kernel	13.341
Queen Contiguity	18.089

Table 1 shows that the critical point test χ^2 with error level $\alpha = 0,05$ and degrees of freedom (p+1) is 11,707 then reject H_0 so the conclusion is that there is spatial heterogeneity in the data dengue cases.

Further Testing GWR model parameters simultaneously, the test results are presented in Table 2.

Table 2. Model Parameter Testing Results GWR Simultaneous

Weighting	*F*
Fixed Gaussian Kernel	4.397
Queen Contiguity	0.21

Table 2 shows that the predictor variables with Gaussian kernel weighting Fixed effecting simultaneously the response variable for F > (3,305)$F_{0.05(2,31)}$, where the weighted predictor variables Queen Contiguity no effect along the response variable for F > (2,922) $F_{0.05(3,30)}$.

The partial test shows throughout predictor variables with Fixed Gaussian kernel weighting affect the response variable at each location and weighted Queen Contiguity affect the response variable at each location.

Comparison of methods performed to determine the best weighting. Criteria for selection of the best weighting by using R^2 are presented in Table 3. Best weighting is weighted with the largest value of R^2.

Table 3. Comparison of R^2 at GWR Model

Weighting	R^2
Fixed Gaussian Kernel	0.42
Queen Contiguity	0.06

Rated R^2 for a model with Gaussian kernel weighting Fixed bigger than Queen contiguity weighted, so that it can be concluded that the weighting Fixed weighting Gaussian kernel is better used for data dengue cases in this study.

CONCLUSION

The GWR method with the weighted *Fixed Gaussian Kernel* yields a R^2 greater than the weighted *Queen Contiguity*. This result indicates that the weighted *Fixed Gaussian Kernel* use for dengue fever case data in this study.

REFERENCES

[1] H. Yasin, "Pemilihan Variabel pada Model Geographically Weighted Regression," *Jurnal Media Statistika,* vol. IV, no. 2, pp. 63-72, 2011.

[2] M. Ilham and I. Dwi, "Pemodelan Data Kemiskinan di Provinsi Sumatera Barat dengan Metode Geographically Weighted Regression," *Jurnal Media Statistika,* vol. VI, no. 1, pp. 37-49, 2013.

[3] A. S. Fotheringham, C. Brundson and M. Charlthon, Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, 2002.

[4] L. Anselin, I. Syabri and K. Youngihn, GeoDa: An Introduction to Spatial Data Analysis, Urbana: University of Illinois, 2004.

[5] J. LeSage and R. K. Pace, Introduction to Spatial Econometrics, Boca Ration: CRC Press, 2009.