

Geographically and Temporally Weighted Regression Modeling in Analyzing Factors Affecting the Spread of Dengue Fever in Malang

A. Fahmi Indrayani^{1*}, Henny Pramodyo², Atiek Iriany²

¹Master Program of Mathematics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

²Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

Abstract

Geographically and Temporally weighted regression (GTWR) modeling has been developed to evaluate spatial heterogeneity and temporal heterogeneity in factors influencing the spread of dengue fever in Malang city. By using the monthly data in 2012-2015 as the temporal unit of each urban village in Malang and village is considered as a spatial unit. GTWR model is compared with the GWR model using several statistical criteria. GTWR model shows that the relationship between dengue incidence with population density and monthly average temperature significantly affects each Village in Malang.

Keywords: dengue hemorrhagic fever, Geographically and Temporally weighted regression, spatiotemporal pattern.

INTRODUCTION

Dengue hemorrhagic fever (DHF) is a disease caused by the dengue virus that is found in tropical and subtropical areas. The number of sufferers and the extent of their spreading areas increases with increasing mobility and population density. In Indonesia, the disease transmitted by *Aedes aegypti* mosquitoes was first discovered in Surabaya and Jakarta in 1968 [1].

For the last three years, patients with dengue fever in Malang continue to increase (Health Office of City Government of Malang, 2016, unpublished). Various programs have been implemented for disease prevention and control, but DHF in Malang remains high. High dengue cases require serious efforts to prevent the spread of DHF endemic disease by analyzing the risk factors associated with dengue virus infection. Many research on the factors that cause the spread of dengue virus in the community were conducted [2].

Previous study used the Geographically Weighted Regression (GWR) model in its analysis; where GWR statistical methods adjusted the global framework with local regression models allowing estimation regression parameters for each spatial unit [3]. And it is reported that rainfall significantly affects the spread of DHF. However, this research is limited to spatial effects and rainfall factors, where the relationship between

dengue fever and rainfall varies greatly in each location of observation. As a result, its conclusions only look at the spatial effects and one factor, not paying attention to the influence of time and other factors that can also affect the spread of dengue in the area.

In many cases, data structures are not only owned by geographic location but also related to the time factor. One method for examining spatial and temporal influences is geographically and temporally weighted regression (GTWR). This statistic method is developed from GWR. It is expected that this model can accommodate spatial influence, and also able to analyze the temporal influence which present in the data simultaneously [4].

The purpose of this study is to extend the GTWR modeling framework in analyzing factors causing dengue extensively by observing the presence of spatial heterogeneity and temporal heterogeneity in the modeling. The GTWR model obtained is compared with the GWR model using several statistical tests. The results of this study will practically be useful for the authorities in regulating the prevention program of DHF infection locally and at certain times.

MATERIAL AND METHOD

Study Area

Location of this research is Malang City, East Java, which consists of 57 urban villages. Total area of Malang is 252.10 km² which is located on Java island about 7°06' - 8°07' latitude and 112°06' - 112°07' longitude [5]. Village is used as a spatial unit in this study, there are 57 villages and the month as a temporal unit, consisting of 48 months.

* Correspondence address:

A. Fahmi Indrayani

Email : fahmi.indrayani@gmail.com

Address : Dept. Mathematics, Faculty of Mathematics and Natural Science, University of Brawijaya, Veteran Malang, Malang 65145.

Data Collection

Data on the number of DHF patients in each Village and each month during the year 2012 - 2015 was obtained from the Health Office of Malang city government. Monthly data as a temporal unit in each Village, one Village is defined as one spatial unit. For population density data, the number of *Puskesmas* (health center) in each Village, as well as the height of the place in each Village obtained from Malang Central Agency on Statistics publication [5-8]. Data of rainfall and average temperature obtained from Karangploso Agency for Meteorological, Climatological and Geophysics.

Number of DHF patients selected as response variable (Y), population density (X_1), Number of health center auxiliary (X_2), rainfall (X_3), average temperature (X_4) and altitude of place from sea level (X_5) as predictors variable.

Statistical Methods

The relationship between DHF incidence and factors considered to affect the spread of DHF was analyzed by comparing the results of GWR model and GTWR model. If GTWR model is better than GWR model, the relation between respondent variable and the predictor variables is spatially and temporally different for each Village.

GWR model is a simple model and is usually used in research. The spatial heterogeneity of the relationship between the occurrence of DHF and factors influencing it was analyzed by the GWR model allowing estimation of different regression parameters in each Village [3].

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \dots\dots 1$$

Where $\beta_0(u_i, v_i)$ and $\beta_k(u_i, v_i)$ is the coefficient of GWR in each Village i ; (u_i, v_i) is the location point of the Village is the definition of the longitude and latitude coordinates ; and ε_i is an error that is assumed to be normal distribution with mean 0 and its variant known.

The GTWR model is formed by using the spatial weight function and the temporal weighting function in estimating its parameters [4].

$$Y_{it} = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^p \beta_k(u_i, v_i, t_i) X_{ik} + \varepsilon_i \dots\dots 2$$

Where (u_i, v_i, t_i) is the location of Village i^{th} and the observation time for the observation location. Longitude and Latitude coordinate defined by (u_i, v_i, t_i) ; Y_{it} is a response variable, X_{ik} is a factor that may affect and ε_i is an error of

observation at location i^{th} . The first step for GTWR modeling to define the coordinates of Longitude and Latitude for each Village at t time. Geographical coordinates are used to determine the euclidean distance between villages observed at t time

$$d_{ij} = \sqrt{\lambda[(u_i - u_j)^2 + (v_i - v_j)^2] + \mu(t_i - t_j)^2} \dots\dots 3$$

λ and μ are scale factors used to measure location and time difference. Distance in equation (3) as a basis in weighting data to estimate GTWR model parameters. The closer the distance between the expected points each village, the greater the weight of the data during the estimation. In this study, data weighting is done by using the adaptive Gaussian kernel function

$$w_{ij} = \exp(-(d_{ij} / h)^2) \dots\dots\dots 4$$

Where, h is the constant bandwidth performed by the Incremental Spatial Autocorrelation (ISA) method [9]. The Adaptive Gaussian kernel function is used in forming a weighted matrix.

$$W(u_i, v_i, t_i) = \text{diag}(w_1(u_i, v_i, t_i), \dots, w_n(u_i, v_i, t_i)) \dots\dots\dots 5$$

Each observed data has one weighted matrix $W(u_i, v_i, t_i)$ in estimating the parameters. Using the algebraic matrix approach, and using the Weighted least square (WLS) method the parameter estimation can be written [4]

$$\hat{\beta}(u_i, v_i, t_i) = (X'W(u_i, v_i, t_i)X)^{-1} X'W(u_i, v_i, t_i)Y \dots\dots 6$$

Significant test of GTWR model is given on [10]. Where the null hypothesis of test indicates the function $\beta_k(u_i, v_i)$ for each coordinate (u_i, v_i) of the research location. If the null hypothesis is rejected, then the GTWR model is accepted to describe the data. Meanwhile, tests for the variation of each set of β_k for the entire study site will be used in F_3 statistics [10]. The identification of the GTWR model will also be shown by coefficient determination (R^2). The comparison of the GWR model and the GTWR model will be shown with Root mean square error. According to RMSE the regression model that has the smallest RMSE value is defined as the best model.

RESULT AND DISCUSSION

Result of GWR Model

Using the spatial weighting, the parameter estimation ratios (regression coefficients) for the obtained GWR model are shown in Table 1. In

this model the value of obtained $R^2 = 51.1\%$ and $RMSE = 13.13$. Further analysis of the model indicates that the error approaches normal distribution.

Table 1. Summary for Coefficient of GWR Model

Parameter	Minimum	Median	Maximum
θ_0	0.9138619705260	0.9138767081168	0.9138948801913
θ_1	0.0000010683522	0.0000010684401	0.0000010685115
θ_2	-0.2121692197872	-0.2121678107572	-0.2121659143796
θ_3	0.0004221212435	0.0004221221469	0.0004221234641
θ_4	2.0687979109140	2.0687998269883	2.0688021917249
θ_5	282.8140409500290	282.8165895808340	282.8184768167160

Result of GTWR Model

By using ISA [9], GTWR modeling has bandwidth Optimum $h = 3947.54$. Table 2 shows the estimated estimation of the GTWR model parameters. The GTWR model has a value of $RMSE = 1.22$, which is smaller than the $RMSE$ value of the GWR model. And the coefficient of determination ($R^2 = 69.13\%$) is higher than the GWR model.

Positive value on coefficient for koefisien β_1 , β_2 , β_3 , β_4 and β_5 shows that predictor variables of population density, number of puskesmas / puskesmas auxiliaries, rainfall, temperature and

altitude of the sea surface contribute positively to the response variable, this indicates that if these variables increase, then the number of dengue fever patients every month in Village it will increase.

To determine the factors that influence the incidence of DHF in each Village, conducted statistical calculation of GTWR model parameters test for each urban village in Malang. Value of t compared with $t_{table} (0.05; 2730) = 2.24264$. If $|t_{cal}| > t_{table}$ then it is decided reject H_0 which means the significant k^{th} parameter of i^{th} location.

Table 2. Summary for coefficient of GTWR Model

Parameter	Minimum	Median	Maximum
θ_0	0.1511224153094	0.1511827937156	0.1512318270934
θ_1	0.0000823500371	0.0000823518679	0.0000823541227
θ_2	0.0019856156327	0.1985769641700	0.0019859593066
θ_3	0.0000149726542	0.0001497344800	0.0000149740928
θ_4	0.0208599179654	0.0208606080060	0.0208614577425
θ_5	0.0004859141711	0.0048598685785	0.0004860763600

CONCLUSION

In this paper, a discrete fractional-order Population density and temperature are significant variables on the number of dengue fever occurrences in Malang. Residents living in densely populated locations have a higher risk of dengue transmission. So also if there is an increase in temperature then the risk of transmission of dengue virus will also be greater.

ACKNOWLEDGEMENT

We thank to Health Office of Malang city government, Malang Central Agency on Statistics, and Karangpulo Agency for Meteorological, Climatological and Geophysics for supporting data.

REFERENCES

- [1] Aiken, S., C. Leigh. 1978. Dengue Hemorrhagic fever in South-East Asia. *Trans. Inst. Br. Geogr.* 3(4). 476-497.
- [2] Baharuddin, Suharningsih, B. Ulama. 2014. Geographically Weighted Regression Modeling for analyzing spatial heterogeneity on relationship between Dengue Hemorrhagic Fever incidence and rainfall in Surabaya, Indonesia. *Mod. Appl. Sci.* 8(3). 85-91.
- [3] McMillen, D. P. 2002. Geographically Weighted Regression - the analysis of Spatially Varying Relationships. *Am. J. Agric. Econ.* 86(2). 554-556.
- [4] Huang, B., B. Wu, M. Barry. 2010. Geographically and Temporally Weighted Regression for modelling spatio-temporal

- variation in house prices. *Int. J. Geogr. Inf. Sci.* 24(3). 383-401.
- [5] Central Agency on Statistics. 2013. Malang in figures 2013. Central Agency on Statistics. Malang.
- [6] Central Agency on Statistics. 2014. Malang in figures 2014. Central Agency on Statistics. Malang.
- [7] Central Agency on Statistics. 2015. Malang in figures 2015. Central Agency on Statistics. Malang.
- [8] Central Agency on Statistics. 2016. Malang in figures 2016. Central Agency on Statistics. Malang.
- [9] Ge, L., Y. Zhao, Z. Sheng, N. Wang, K. Zhou, X. Mu, L. Guo, T. Wang, Z. Yang, X. Huo. 2016. Contruction of Seasonal Difference-Geographically and Temporally Weighted Regression (SD-GTWR) model and comparative analysis with GWR-Based Models for Hemorrhagic Fever with Renal Syndrome (HFRS) in Hubei Province (China). *Int. J. Environ. Res. Public Health.* 13(11). pii: E1062.
- [10] Xuan, H., S. Li, M. Amin. 2015. Statistical inference of Geographically and Temporally Weighted Regression Model. *Pakistan J. Statist.* 31(3). 307-325.
- [11] Hii, Y. L., H. Zhu, N. Ng, L.C. Ng, J. Rocklov. 2012. Forecast of Dengue Incidence Using Temperature and Rainfall. *PLOS Neglected Tropical Diseases*. Available at: <https://doi.org/10.1371/journal.pntd.0001908>.