JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 7, No. 4, October 2023, pp. 961-975

Geographically Weighted Panel Regression Modelling of Dengue Hemorrhagic Fever Data Using Exponential Kernel Function

Risti Raihani[1], Sifriyani[2*], Surya Prangga[3]
[1,2,3]Statistics, Department of Mathematics, Mulawarman University, Indonesia
sifriyani@fmipa.unmul.ac.id[2]

Article History:
Received : 02-07-2023
Revised : 22-09-2023
Accepted : 27-09-2023
Online : 07-10-2023

Keywords:
Adaptive Exponential;
Panel Regression;
Fixed Effect Model;
Geographically Weighted Panel Regression;
Dengue Hemorrhagic Fever.

ABSTRACT

Geographically Weighted Panel Regression (GWPR) model is a panel regression model applied to spatial data. This research takes the Fixed Effect Model (FEM) panel regression as the global model and GWPR as the local model for dengue hemorrhagic fever (DHF) in East Kalimantan Province data over the years 2018-2020. DHF is a disease that has the potential to become an extraordinary event which is accompanied by death. In comparison to Indonesia, East Kalimantan Province's DHF Incident Rate (IR) was high in 2020. East Kalimantan's IR is 60.6 per 100,000 population, compared to Indonesia's IR of 40.0 per 100,000 population. This research aims to obtain the GWPR model, as well as to acquire factors that affect DHF in East Kalimantan Province over the years 2018-2020 based GWPR model. The parameter of the GWPR model was estimated on each observation location using the Weighted Least Square (WLS) method, which is an Ordinary Least Square (OLS) with the addition of spatial weighting. The spatial weighting on the GWPR model was determined by the best weighting function between fixed exponential and adaptive exponential. The optimum weighting function with a minimum cross-validation (CV) value of 1.7317×10^{6} is adaptive exponential. Based on GWPR parameter testing, factors that affect DHF are local and diverse in each 10 regencies/municipalities in East Kalimantan Province. These factors are population density, number of health facilities, percentage of proper sanitation use in the household, percentage of household with qualified drinking water sources, and percentage of health services. The coefficient of determination of the GWPR model obtains a higher value than the FEM, which is 95.33%. Based on the measurement of goodness using the coefficient of determination value, it can be concluded that GWPR is the best method to model the DHF data rather than the FEM.

Crossref
https://doi.org/10.31764/jtam.v7i4.16235

——————————— ◆ ———————————

A. INTRODUCTION

Regression analysis is a statistical method using an equation to express the relationship between a variable of interest and a set of related predictor variables (Montgomery et al., 2021). The commonly used regression model is linear regression. The parameters of the linear regression model are estimated using Ordinary Least Square (OLS) by minimizing the sum of squared errors. The assumptions of the linear regression model are normality, multicollinearity, homoscedasticity, and non-autocorrelation (Gujarati, 2022). The data used in linear regression analysis is generally cross-sectional data. In fact, much of the available data is not only obtained in the form of cross-sections, but also in the form of panel data.

Panel data is a combination of cross section data and time series data where the same cross section unit is measured at different times. The right method for modeling panel data is panel regression. There are three models in panel regression, which are Common Effect Model (CEM), Random Effect Model (REM) and Fixed Effect Model (FEM). The panel regression model in this study is limited to the FEM model (Yusra et al., 2019). Panel data is mostly found in the form of spatial panel data. Spatial panel data is a combination of panel data and spatial data. Spatial data contains attribute information, location, and there is a relation between data and geographic location. Panel data containing spatial heterogeneity cannot be modeled using panel regression. The right model for spatial panel data is Geographically Weighted Panel Regression (GWPR) (Bruna & Yu, 2013).

According to previous research conducted by Qur'ani (2014) shows that the GWPR model provides great results due to the GWPR model produces a high coefficient of determination value of 0.9447. Other research by Meutuah, Yasin, and Maruddani (2017) shows that the GWPR model with fixed Exponential weighting function produces a coefficient of determination value of 0.9227. GWPR model in this study was applied to Dengue Hemorrhagic Fever (DHF) data. DHF is a disease with the potential for extraordinary events which is accompanied by death. At the beginning of 2020, World Health Organization listed dengue fever as a global health threat among 10 other diseases (WHO, 2020). Until now, DHF is still not well controlled, as evidenced by the significant increase in the incidence of DHF throughout the world and the outbreaks that occur every year in Indonesia. Bhatt et al. (2013) estimated that 390 million dengue infections occur every year and 96 million of them have clinical manifestations with varying levels of disease severity.

According to Ministry of Health of the Republic of Indonesia, DHF cases occur equally in females (49%) and males (51%). Most dengue cases occur in the 15-44 year age group (39%). This pattern is different from deaths due to DHF, which are more dominant in females (55%) and in the younger age group, specifically 5-14 years (45%). Widoyono (2011) stated factors that are thought to affect the incidence of DHF include health factors, environmental factors and community behavior as well as social environmental factors. Understanding the elements that have a substantial impact on the rise in DHF cases is one strategy to stop the spread of DHF. One of the previous studies conducted by Fatati, Wijayanto and Soleh (2017) states that the amount of health facilities, population density, the percentage of households with an adequate source of drinking water, and the quantity of protected water springs all affect the number of DHF incidents in Central Java. Other research by Boleng, Ginting and Ariyanto (2022) states that population density have a significant effect against dengue cases in East Nusa Tenggara Province.

According to the Indonesian Ministry of Health, reported cases of DHF in Indonesia in 2020 were recorded as many as 95,893 cases. In 2020, there were 661 deaths caused by DHF (Dinas Kesehatan, 2020). According to BPS, the DHF morbidity rate (IR) in East Kalimantan Province in 2020 is quite high compared to DHF IR in Indonesia. East Kalimantan has an IR of 60.6 per 100,000 population, while Indonesia's IR in 2020 is 40.0 per 100,000 population (BPS, 2020). One of the efforts to reduce DHF cases in East Kalimantan Province statistically is to find out the factors that have a significant effect on DHF. The purpose of this study is to compare the FEM and GWPR models and to identify the variables that affect DHF in East Kalimantan using the GWPR model.

B. METHODS

The research data is secondary data obtained from the publication of East Kalimantan Central Statistics Agency (BPS). The sampling technique used in this research is total sampling. The research variables consist of dependent variables and independent variables from 2018 – 2020 due to limited data in the latest publications for the variable number of DHF incidents. The variables are number of DHF incidents, population density, number of health facilities, percentage of used proper sanitation in households, percentage of households with an adequate source of drinking water, and percentage of health services as shown in Table 1.

Table 1. Variable Data

Variable	Variable Name	Definition
y	Number of DHF incidents	Number of dengue fever incidents in each regency/city obtained from records by health workers with a clear address.
x_1	Population density	Comparison of population numbers by area.
x_2	Number of health facilities	Number of public health-related facilities in a region.
x_3	Percentage of used proper sanitation in households	Comparison between the number of households that have access regarding sanitation services meet health requirements to the number of total households.
x_4	Percentage of households with an adequate source of drinking water	Comparison between the number of households that have access to adequate drinking water sources which are less than 10 metres from waste disposal to the number of total households.
x_5	Percentage of health services	Comparison between the number of neonates with complications handled according to standards by trained health personnel in specific region and time period.

The method used in this study are FEM and GWPR panel regression. The steps of analysis using R and Octave are as follows:

1. FEM Modelling

FEM assumed that each individual has a different intercept coefficient (Hsiao, 2007) Prior to treatment, parallel trajectories or trends must exist for fixed effects to be a reliable estimating technique (Jones & Lewis, 2015). The general form of the FEM model is:

$$y_{it} = \beta_{0_i} + \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_p x_{itp} + \varepsilon_{it} \quad (1)$$

Parameter estimation of FEM is done by transforming β_{0_i} using *within estimator*. Within estimator is formed by subtracting the actual data from the time series average (Wooldridge, 2002). The average time series for each $t = 1, 2, \ldots, T$ is as follows:

$$\bar{y}_i = \beta_{0_i} + \beta_1 \bar{x}_{i1} + \beta_2 \bar{x}_{i2} + \ldots + \beta_p \bar{x}_{ip} + \bar{\varepsilon}_i \tag{2}$$

FEM with *within the estimator* is formed by reducing equation (1) with equation (2).

$$y_{it}^* = x_{it1}^* \beta_1 + x_{it2}^* \beta_2 + \cdots + x_{itp}^* \beta_p + \varepsilon_{it}^* \tag{3}$$

Parameter $\boldsymbol{\beta}$ is estimated using the OLS method (Moon & Weidner, 2017) that obtained:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_F^{*T} \mathbf{X}_F^*)^{-1} \mathbf{X}_F^{*T} \mathbf{y}_F^* \tag{4}$$

2. FEM Parameter Significance Test

Parameter significance test for FEM is done simultaneously and partially. The purpose of simultan test is to determine whether the independent variables and the dependent variable affect (Alita et al., 2021). The hypothesis is as follows:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1 : There is at least one $\beta_k \neq 0,\ k = 1, 2, 3, \ldots, p$

With the following test statistics:

$$F_1 = \frac{KTR_p}{KTG_p} \tag{5}$$

KTR_p is the mean square regression of the FEM model and KTG_p is the mean square error of the FEM model. Null hypothesis rejected if $F_1 > F_{(\alpha\ ;\ p\ ;\ nT-n-p)}$ or $pvalue < \alpha$ (Setiawan & Kusrini, 2010). The purpose of partial test is to verify whether the independent variables individually affects dependent variable (Montgomery et al., 2021). The hypothesis is as follows:

$H_0 : \beta_k = 0, k = 1, 2, 3, \ldots, p$

$H_1 : \beta_k \neq 0, k = 1, 2, 3, \ldots, p$

With the following test statistics:

$$T_1 = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \tag{6}$$

with $SE(\hat{\beta}_k)$ is the standard deviation of the estimator $\hat{\beta}_k$, that is:

$$SE(\hat{\beta}_k) = s\sqrt{g_{kk}} \tag{7}$$

$$s = \frac{JKG_p}{nT - p},\ \ g_{kk} = diag\begin{bmatrix} g_{11} & g_{22} & \cdots & g_{pp} \end{bmatrix} \tag{8}$$

JKG_p is the sum of square error of FEM model and g_{kk} adalah elemen matriks $\left(\mathbf{X}_F^{*T}\mathbf{X}_F^{*}\right)^{-1}$. Null hypothesis rejected if $|T_1| \geq t_{\left(\frac{\alpha}{2};\, nT-p\right)}$ or $pvalue < \alpha$ (Setiawan & Kusrini, 2010).

3. Multicollinearity Test

Multicollinearity means there is a strong linear correlation between the independent variables in the regression model. The method used to detect multicollinearity is Variance Inflation Factor (VIF) value (Lin et al., 2019). Multicollinearity occurs when the VIF value is greater than 10. The VIF value can be obtained by:

$$VIF = \frac{1}{1-R^2} \tag{9}$$

With R^2 is the coefficient of determination of the independent variable x_k regression model on other independent variables (Shrestha, 2020).

4. Homoscedasticity Test

The error variance that is not constant will cause inefficient estimation of model parameters and the conclusions obtained are not valid. Homoscedasticity testing can be done using Glejser test (Obabire et al., 2020) with the following hypothesis:

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

H_1 : There is at least one $\sigma_i^2 \neq \sigma^2$; $i = 1, 2, \ldots, n$

With the following test statistics:

$$F_2 = \frac{(\hat{\boldsymbol{\varphi}}^T \mathbf{X}^T \boldsymbol{\varepsilon} - n\bar{\varepsilon}^2) / p}{(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} - \hat{\boldsymbol{\varphi}}^T \mathbf{X}^T \boldsymbol{\varepsilon}) / (n-p-1)} \tag{10}$$

n is the number of observations and p is the number of independent variables. Null hypothesis rejected if $F_2 > F_{(\alpha;p;n-p-1)}$ or $pvalue < \alpha$ (Gujarati, 2022).

5. Spasial Weighting Function

The GWPR model's weighting is identical to the GWR model's weighting, which is based on the separation between points of observation (Ningrum et al., 2020). Typically, a distance-decaying kernel function is used to approximate the spatial weighting system (Cai et al., 2014). There are two types of kernel functions, namely fixed kernel functions and adaptive kernel functions (Sifriyani et al., 2022). The weighting function used in this study is the exponential kernel function because it does not require specific conditions for euclidean distance and bandwidth. Spatial weighting using fixed exponential can be calculated by the following equation:

$$w_{ij} = exp\left(\frac{-d_{ij}}{h}\right) \tag{11}$$

Spatial weighting using adaptive exponential can be calculated by the following equation:

$$w_{ij} = exp\left(\frac{-d_{ij}}{h_i}\right) \tag{12}$$

With d_{ij} is the Euclidean distance or the distance between location *i* and location *j* which can be calculated using the following equation:

$$d_{ij}^2 = \left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2 \tag{13}$$

While h is a non-negative parameter known as the bandwidth or smoothing parameter. One of the methods to get the optimum bandwidth is to use the Cross Validation (CV) approach. The optimum bandwidth will produce the minimum CV value. CV value can be calculated by the following equation:

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(h_i)]^2 \tag{14}$$

With $\hat{y}_{\neq i}(h)$ is the estimated value for y_i by eliminating the observation at point *i* from the parameter testing process (Fotheringham et al., 2003).

6. *Geographically Weighted Panel Regression* (GWPR)

GWPR is a local form of FEM *within estimator* using spatial panel data. This technique was created to get around the problem with traditional linear regression's parameter estimation, which yields a regression coefficient that is supposed to be universally applicable to the entire observation unit (Chotimah et al., 2019). GWPR model is as follows (Mar'ah, 2023):

$$y_{it}^* = \beta_1(u_i, v_i)x_{it1}^* + \beta_2(u_i, v_i)x_{it2}^* + \cdots + \beta_p(u_i, v_i)x_{itp}^* + \varepsilon_{it}^* \tag{15}$$

GWPR model parameter estimation was carried out using the WLS method, which is the OLS model by giving different weights to each observation location. The GWPR model parameter estimator is:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \left(\mathbf{X}^{*T}\mathbf{W}(u_i, v_i)\mathbf{X}^*\right)^{-1}\mathbf{X}^{*T}\mathbf{W}(u_i, v_i)\mathbf{y}^* \tag{16}$$

The predicted value for $\mathbf{y}_{it}^*$ can be obtained by the following equation.

$$\hat{\mathbf{y}}_i^* = \mathbf{X}_{it}^{*T}\hat{\boldsymbol{\beta}}(u_i, v_i) = \mathbf{X}_{it}^{*T}\left[\mathbf{X}^{*T}\mathbf{W}(u_i, v_i)\mathbf{X}^*\right]^{-1}\mathbf{X}^{*T}\mathbf{W}(u_i, v_i)\mathbf{y}^* \tag{17}$$

7. GWPR Suitability Testing

Testing the suitability of the GWPR model is carried out by simultaneously testing the suitability of the parameters with the following hypothesis.

$H_0 : \beta_k(u_i, v_i) = \beta_k,\ k = 1, 2, \ldots, p\ ;\ i = 1, 2, \ldots, n$

H_1 : There is at least one $\beta_k(u_i, v_i) \neq \beta_k$

With the following test statistics:

$$F_3 = \frac{JKG_G(H_0) / db_1}{JKG_G(H_1) / db_2} \tag{18}$$

$JKG_G(H_0)$ is according to the FEM model and is according to GWPR model. The null hypothesis is rejected if $F_3 > F_{\alpha;db_1;db_2}$ or $pvalue < \alpha$, where $db_1 = \delta_1^2 / \delta_2$ and $db_2 = (nT - p - 1)$ (Sifriyani et al., 2022). The value of $\delta_1 = tr\left((\mathbf{I} - \mathbf{L}^*)^T(\mathbf{I} - \mathbf{L}^*)\right)$ and $\delta_2 = tr\left((\mathbf{I} - \mathbf{L}^*)^T(\mathbf{I} - \mathbf{L}^*)\right)^2$. $\mathbf{I}$ is an identity matrix of size $nT \times nT$ and $\mathbf{L}^*$ is a matrix of size $nT \times nT$ (Leung et al., 2000).

8. GWPR Parameter Significance Test

Testing the partial parameters of the GWPR model aims to determine certain independent variables that affect the *i*-th location. The hypothesis are as follows:

$H_0 : \beta_k(u_i, v_i) = 0,\ k = 1, 2, \ldots, p\ ; i = 1, 2, \ldots, n$

$H_1 : \beta_k(u_i, v_i) \neq 0,\ k = 1, 2, \ldots, p\ ; i = 1, 2, \ldots, n$

With the following test statistics:

$$T_2 = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{c_{kk}}} \tag{19}$$

Null hypothesis is rejected if $|T_2| > t_{\left(\alpha/2\ ;\ \delta_1^2/\delta_2\right)}$ or $pvalue < \alpha$ (Leung et al., 2000).

9. Goodness of Fit

Goodness of fit measures for FEM and GWPR model will calculated using coefficient of determination. The coefficient of determination is a measure of the goodness of the model which shows how strong the independent variables affect the dependent variable. The coefficient of determination is shown in the following equation:

$$R^2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{y}_{it}^{\ *} - \bar{y}^*\right)^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it}^* - \bar{y}^*\right)^2} \tag{20}$$

The range of the coefficient of determination is in the range $0 \le R \le 1$ (Piepho, 2019).

C. RESULT AND DISCUSSION

1. Descriptive Analysis

The descriptive statistics of the research data include mean, maximum value, minimum value and standard deviation which are shown in Table 2.

Table 2. Descriptive Statistics

Variable	Mean	*Max*	*Min*	Standard Deviation
y	448.20	1838.00	40.00	470.26
x_1	371.72	1343.00	1.35	547.21
x_2	469.00	1627.00	56.00	411.27
x_3	82.89	95.98	41.61	12.04
x_4	21.10	33.87	5.58	7.72
x_5	85.11	99.60	56.30	11.74

Based on Table 2, it is known that the average number of DHF cases is 448 cases with a standard deviation of 470.26. The highest DHF cases occurred in Balikpapan City in 2019 with 1838 cases and the lowest DHF cases occurred in Mahakam Ulu Regency in 2020 with 40 cases.

2. Multicollinearity Test

Multicollinearity detection using VIF value criteria based on the equation (9) is presented in Table 3.

Table 3. VIF Value

Variable	VIF_k
x_1	1.7867
x_2	1.8487
x_3	1.4432
x_4	1.2560
x_5	1.2941

Based on Table 3, it can be seen that the VIF value of each independent variable is less than 10. Therefore, it can be indicated that there is no multicollinearity between the independent variables in the regression model.

3. FEM Modelling

Based on the FEM general model in equation (3), the FEM general model for DHF data with 5 independent variables is:

$$y_{it}^{*} = \beta_1 x_{it1}^{*} + \beta_2 x_{it2}^{*} + \beta_3 x_{it3}^{*} + \beta_4 x_{it4}^{*} + \beta_5 x_{it5}^{*} + \varepsilon_{it}^{*} \quad (21)$$

FEM parameter estimation uses the OLS method and the estimation results are shown in Table 4.

Table 4. FEM Parameter Estimation

Parameter	Estimation
β_1	7.7959
β_2	-4.7039
β_3	-6.1990
β_4	-13.4286
β_5	6.7491

Based on the estimated parameter values in Table 4, the FEM model formed is:

$$\hat{y}_{it}^{*} = 7.7959 x_{it1}^{*} - 4.7039 x_{it2}^{*} - 6.199 x_{it3}^{*} - 13.4286 x_{it4}^{*} + 6.7491 x_{it5}^{*} \quad (22)$$

To determine the suitability of the model, parameter significance testing is carried out simultaneously with the following hypothesis.

H_0 : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

(Simultaneously, the independent variables have no effect on the incidence of DHF in East Kalimantan Province).

H_1 : There is at least one $\beta_k \neq 0,\ k = 1, 2, \ldots, 5$

(At least one of the independent variables has an effect on the incidence of DHF in East Kalimantan Province). Statistic value F_1 and *pvalue* are shown in Table 5.

Table 5. F Test for FEM

F_1	$F_{(0,05;5;15)}$	*pvalue*	Decision
5.8006	2.9013	0.003555	H_0 rejected

Based on Table 5, the conclusion obtained is that at least one of the independent variables has an effect on the incidence of DHF in East Kalimantan Province. Furthermore, a partial parameter significance test was carried out to determine the effect of each independent variable on the dependent variable with the following hypothesis. H_0 is $\beta_k = 0$ (There is no effect of the independent variables x_k on the incidence of DHF in the districts/cities of East Kalimantan Province). H_1 is $\beta_k \neq 0$ (There is an effect of the independent variables x_k on the incidence of DHF in the districts/cities of East Kalimantan Province). Statistic value T_1 and *pvalue* are shown in Table 6.

Table 6. FEM Partial Significance Test

Variable	$\lvert T_1 \rvert$	*pvalue*	Decision
x_1	2.5177	0.0237	H_0 rejected
x_2	-4.2590	0.0007	H_0 rejected
x_3	-1.1389	0.2726	H_0 not rejected
x_4	-0.9767	0.3442	H_0 not rejected
x_5	0.9929	0.3365	H_0 not rejected

Based on Table 6, the results are obtained variable x_1 and x_2 partially affect the number of DHF incidents in the districts/cities of East Kalimantan Province. This is indicated by the statistic value $\lvert T_1 \rvert$ of each variables is more than $t_{(0,025;25)} = 2.060$ or $pvalue < \alpha = 0.05$. Based on Table 6, the results are also obtained variable partially did not affect the number of DHF incidents in the districts/cities of East Kalimantan Province. This is indicated by the statistic value $\lvert T_1 \rvert$ of each variables is less than $t_{(0,025;25)} = 2.060$ or $pvalue < \alpha = 0.05$.

4. Homoscedasticity Test for FEM

The homoscedasticity test examines at whether the variance of the error is constant across all observation sites with hypotheses as follows: H_0 is $\sigma_{1,1}^2 = \sigma_{2,1}^2 = \ldots = \sigma_{10,3}^2 = \sigma^2$ (The error variance is constant across all observation areas); H_1 is There is at least one $\sigma_{i,t}^2 \neq \sigma^2$; $i = 1, 2, \ldots, 10$; $t = 1, 2, 3$ (The error variance is not constant across all observation areas). Statistic value F_2 and *pvalue* are shown in Table 7.

Table 7. FEM Homoscedasticity Test

F_2	$F_{(0,05;5;15)}$	*pvalue*	Decision
2.9186	2.9013	0.0491	H_0 rejected

According to Table 7, the assumption of homoscedasticity is not met since the error variance is not constant across all observation locations. As a result, modeling using GWPR was done.

5. GWPR Modelling

The general GWPR model (Sifriyani et al., 2022) for DHF data with 5 independent variables is based on equation (15) is:

$$y_{it}^* = \beta_1(u_i,v_i)x_{it1}^* + \beta_2(u_i,v_i)x_{it2}^* + \beta_3(u_i,v_i)x_{it3}^* + \beta_4(u_i,v_i)x_{it4}^* + \beta_5(u_i,v_i)x_{it5}^* + \varepsilon_{it}^* \quad (23)$$

The technique is described as the repeated estimation of a local regression at each location in space using a subsample of cross-sectional data correctly weighted according to their proximity to each regression point. Calculating the Euclidean distance using equation (13) is the first step in the parameter estimation stage. The adaptive exponential has the best weighting function when determining the optimal bandwidth using the CV criteria based on equation (14). The GWPR model formed is 10 models. One of the GWPR models for Paser District is as follows:

$$\hat{y}_{1t}^{*} = 6,5703x_{1t1}^{*} - 4,4890x_{1t2}^{*} - 7,4917x_{1t3}^{*} - 13,1600x_{1t4}^{*} + 8,9340x_{1t5}^{*} \; ; \; t = 1,2,3 \quad (24)$$

6. GWPR Suitability Testing

The hypothesis are as follows: H_0 is $\beta_k(u_i, v_i) = \beta_k,\ k = 1,2,...,5\ ;\ i = 1,2,...,10$ (There is no significant difference between the FEM model and the GWPR model); H_1 is There is at least one $\beta_k(u_i, v_i) \neq \beta_k,\ k = 1,2,...,5\ ;\ i = 1,2,...,10$ (There is significant differences between the FEM model and the GWPR model). Statistic value F_3 and $pvalue$ are shown in Table 8.

Table 8. GWPR Suitaibility Testing

F_3	$F_{(0,05;7;23)}$	$pvalue$	**Decision**
52.183	2.4422	1.2779×10^{-12}	H_0 rejected

According to Table 8, it can be concluded that there is a significant difference between the FEM model and the GWPR model.

7. GWPR Partial Parameter Significance Test

The hypothesis are as follows: H_0 : $\beta_k(u_i, v_i) = 0,\ k = 1,2,...,5\ ; i = 1,2,...,10$ (The independent variable x_k has no impact on the DHF incidence in East Kalimantan Province) H_1 is $\beta_k(u_i, v_i) \neq 0,\ k = 1,2,...,5\ ; i = 1,2,...,10$ (The independent variable x_k has an impact on the DHF incidence in East Kalimantan Province); H_0 is rejected at the significance level $\alpha = 0,05$ if $|T_2| > t_{(0,025;7)} = 2.365$ or $pvalue < \alpha = 0.05$. For example, Table 9 displays one of the partial test results for Paser District.

Table 9. GWPR Partial Test for Paser District

Parameter	T_2	$pvalue$
β_1	4.5175	0.0029*
β_2	-8.8054	0.0001*
β_3	-2.5884	0.0365*
β_4	-1.7913	0.1171
β_5	2.2303	0.0616

According to Table 9, it can be concluded that the factors that influence the DHF incidence in Paser District are population density, number of health facilities and the percentage of households using proper sanitation. Table 10 shows the GWPR model grouping for all observation location based on significant variables.

Table 10. GWPR Model Grouping

Group	Significant Variable	Location
1	x_1, x_2 and x_3	Paser and Kutai Timur
2	x_2	Kutai Barat, Mahakam Ulu and Balikpapan
3	x_1 , x_2 , x_4 and x_5	Kutai Kartanegara
4	x_4	Berau
5	x_3	Penajam Paser Utara
6	x_1 , x_2 , x_3 and x_5	Samarinda
7	x_2 and x_3	Bontang

Table 10 can be used to group the GWPR model depending on the significant variables, which can be shown in Figure 1 as a distribution map.

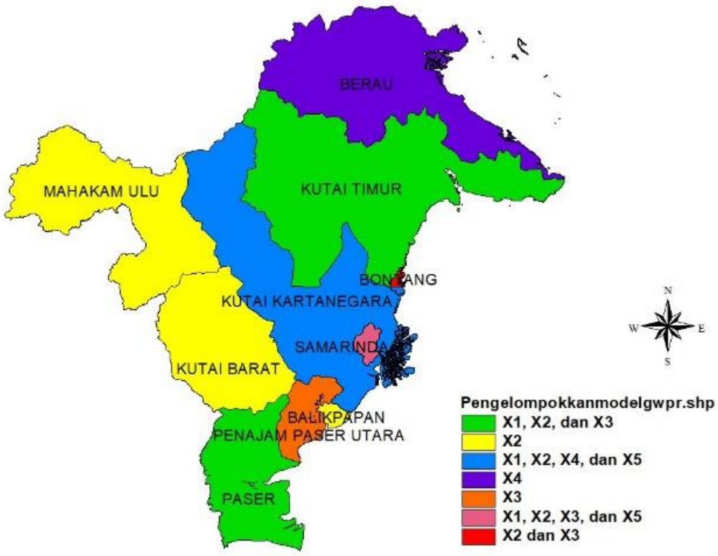

Figure 1. Distribution Map of GWPR Model Grouping

According to Table 10 and Figure 1, the regencies and cities in East Kalimantan Province can be divided into seven groups based on the variables that affect the number of DHF incidents. The first group of GWPR models are Paser Regency and Kutai Timur Regency. Factors that affect the number of DHF incidents at the observation location are population density, number of health facilities, and percentage of used proper sanitation in households. The second group are Kutai Barat Regency, Mahakam Ulu Regency, and Balikpapan City. Factor that affects the number of DHF incidents at the observation location is number of health facilities. The third group is Kutai Kartanegara Regency. Factors that affect the number of DHF incidents at the observation location are population density, number of health facilities, percentage of households with an adequate source of drinking water, and percentage of health services. The fourth group is Berau Regency. Factor that affects the number of DHF incidents at the observation location is percentage of households with an adequate source of drinking water. The fifth group is Penajam Paser Utara

Regency. Factor that affects the number of DHF incidents at the observation location is percentage of used proper sanitation in households. The sixth group is Samarinda City. Factors that affect the number of DHF incidents at the observation location are population density, number of health facilities, percentage of used proper sanitation in households, and percentage of health services. The seventh group is Bontang City. Factors that affect the number of DHF incidents at the observation location are number of health facilities and percentage of used proper sanitation in households.

8. GWPR Homoscedasticity Test

The Glejser test is used for homoscedasticity testing with the following hypothesis: H_0 is $\sigma_{1,1}^2 = \sigma_{2,1}^2 = \ldots = \sigma_{10,3}^2 = \sigma^2$ (The error variance is constant across all observation areas); H_1 : There is at least one $\sigma_{i,t}^2 \neq \sigma^2$; $i = 1, 2, \ldots, 10$; $t = 1, 2, 3$ (The error variance is not constant across all observation areas). Statistic values and *pvalue* are shown in Table 11.

Table 11. GWPR Homoscedasticity Test

F_2	$F_{(0,05;5;24)}$	*pvalue*	**Decision**
0.3271	2.6207	0.8916	H_0 not rejected

Table 11 indicates that the error variance is constant across all observation locations.

9. Goodness of Fit Measures

The coefficient of determination was used to compute the goodness of fit measures. The goodness of fit measures values are shown in Table 12.

Table 12. The Goodness of Fit Measures Value

Model	R^2
FEM	0.6591
GWPR	0.9533

The GWPR model performs better at simulating DHF cases than the FEM model, according to the good measure of the coefficient of determination. This is demonstrated by the fact that the GWPR model's coefficient of determination is higher than the FEM model's coefficient of determination. Besides, FEM in this study cannot simulate the incidence of DHF in East Kalimantan, considering the assumption of homoscedasticity is not met since the error variance is not constant across all observation locations.

D. CONCLUSION AND SUGGESTIONS

The GWPR model and the variables affecting the incidence of DHF vary depending on the observation site. Factors that affect the number of DHF incidents in Paser Regency and Kutai Timur Regency are population density, number of health facilities, and percentage of used proper sanitation in households. Factor that affects the number of DHF incidents in Kutai Barat Regency, Mahakam Ulu Regency, and Balikpapan City is number of health facilities. Factors that affect the number of DHF incidents in Kutai Kartanegara Regency are population density, number of health facilities, percentage of households with an adequate source of drinking water, and percentage of health services. Factor that affects the number of DHF incidents in Berau Regency is percentage of households with an adequate source of drinking water. Factor that affects the number of DHF incidents in Penajam Paser Utara Regency is percentage of used proper sanitation in households. Factors that affect the number of DHF incidents in Samarinda City are population density, number of health facilities, percentage of used proper sanitation in households, and percentage of health services. Factors that affect the number of DHF incidents in Bontang City are number of health facilities and percentage of used proper sanitation in households. The GWPR model's coefficient of determination, which is 0.9533 or 95.33%, shows that it is more effective in simulating DHF cases in East Kalimantan Province compared to FEM which only has a coefficient of determination value of 0.6591 or 65.91%. Future studies could make use of more time series units and different bandwidth-selection methods, such as the Akaike Information Criterion (AIC).

REFERENCES

Alita, D., Putra, A. D., & Darwis, D. (2021). Analysis of Classic Assumption Test and Multiple Linear Regression Coefficient Test for Employee Structural Office Recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(3), 295–306. https://doi.org/10.22146/ijccs.65586

BPS. (2020). *Provinsi Kalimantan Timur Dalam Angka 2020*. BPS Provinsi Kalimantan Timur. https://kaltim.bps.go.id/

Bruna, F., & Yu, D. (2013). Geographically Weighted Panel Regression. *In XI Congreso Galego de Estatística e Investigación de Operacións. http://Xisgapeio. Udc. Es.*, 1–23.

Cai, R., Yu, D., & Oppenheimer, M. (2014). Estimating the Spatially Varying Responses of Corn Yields to Weather Variations using Geographically Weighted Panel Regression. *Journal of Agricultural and Resource Economics*, *39*(2), 230–252. http://www.jstor.org/stable/44131327

Chotimah, C., Sutikno, & Setiawan. (2019). Modelling of Income Inequality in East Java Using Geographically Weighted Panel Regression. *IOP Conference Series: Materials Science and Engineering*, *546*(5), 0–10. https://doi.org/10.1088/1757-899X/546/5/052019

Dinas Kesehatan. (2020). *Profil Kesehatan Provinsi Kalimantan Timur 2020*. Dinas Kesehatan Provinsi Kalimantan Timur. https://dinkes.kaltimprov.go.id/

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.

Musella, G., & Rivieccio, G. (2022). Evaluating the Determinants of Innovation from a Spatio-Temporal Perspective. The GWPR approach. *Scientific Meeting of the Italian Statistical Society*, 354–365. https://hdl.handle.net/11573/1680228.

Gujarati, D. N. (2022). *Basic Econometric Fifth Edition*. McGraw-Hill.

Hsiao, C. (2007). Panel Data Analysis—Advantages and Challenges. *TEST*, *16*(1), 1–22. https://doi.org/10.1007/s11749-007-0046-x

Jones, K. W., & Lewis, D. J. (2015). Estimating the Counterfactual Impact of Conservation Programs on Land Cover Outcomes: The Role of Matching and Panel Regression Techniques. *PLoS ONE*, *10*(10), 1–22. https://doi.org/10.1371/journal.pone.0141380

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. https://doi.org/10.1068/a3162

Lin, X., Su, Y. C., Shang, J., Sha, J., Li, X., Sun, Y. Y., Ji, J., & Jin, B. (2019). Geographically Weighted Regression Effects on Soil Zinc Content Hyperspectral Modeling by Applying The Fractional-Order Differential. *Remote Sensing*, *11*(6). https://doi.org/10.3390/RS11060636

Mar'ah, Z., Sifriyani. (2023). (GWPR) For Covid-19 Case In Indonesia. *Barekeng, 17*(2), 879–886.

https://doi.org/10.30598/barekengvol17iss2pp0879-0886.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Moon, H. R., & Weidner, M. (2017). Dynamic Linear Panel Regression Models with Interactive Fixed Effects. *Econometric Theory*, *33*(1), 158–195. https://doi.org/10.1017/S0266466615000328.

Murakami, D., & Lu, B. (2021). Stable Geographically Weighted Poisson Regression for Count Data. *UC Santa Barbara*. https://doi.org/10.25436/E2X59B

Ningrum, A. S., Rusgiyono, A., & Prahutama, A. (2020). Village Classification Index Prediction Using Geographically Weighted Panel Regression. *Journal of Physics: Conference Series*, *1524*(1), 0–8. https://doi.org/10.1088/1742-6596/1524/1/012040

Obabire, A., Agboola, J. O., Ajao Isaac, O., & Adegbilero-Iwari, O. E. (2020). Comparison of Different Tests for Detecting Heteroscedasticity in Datasets. *Annals. Computer Science Series*, *18(2)*, 78–85.

Piepho, H. (2019). A Coefficient of Determination (R2) for Generalized Linear Mixed Models. *Biometrical Journal*, *61*(4), 860–872. https://doi.org/10.1002/bimj.201800270.

Setiawan, K., & Kusrini, E. (2010). *Ekonometrika*. Yogyakarta: Andi Offset.

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1.

Sifriyani, S., Mandang I., Amijaya, F. D. T., & Ruslan, R. (2022) Developing Geographically Weighted Panel Regression Model For Spatio-Temporal Analysis Of Covid-19 Positive Cases In Kalimantan, Indonesia. *Journal of Southwest Jiaotong University*, *57*(3). https://doi.org/10.35741/issn.0258-2724.57.3.10.

Sifriyani, S., Mandang, I., & Amijaya, F. D. T. (2022). Geographically Temporally Weighted Regression Model for GIS Mapping of Influence COVID-19 in East Kalimantan. *AIP Conference Proceedings, 2668*(1). AIP Publishing. https://doi.org/10.1063/5.0111808.

Sifriyani, Mandang, I., Deny, F., Amijaya, T., Sholihin, M., Tri, A., & Dani, R. (2022). A Spatio-Temporal Description of Covid-19 Cases in East Borneo Using Improved Geographically and Temporally Weighted Regression (I-GTWR). *Commun. in Mathematical Biology and Neuroscience,* 1–22. https://doi.org/10.28919/cmbn/7572.

Sifriyani, S., Rasjid, M., Rosadi, D., Anwar, S., Wahyuni, R. D., & Jalaluddin, S. (2022). Spatial-Temporal Epidemiology of COVID-19 Using a Geographically and Temporally Weighted Regression Model. *Symmetry*, *14*(4). https://doi.org/10.3390/sym14040742.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press

Yusra, I., Hadya, R., Begawati, N., Istiqomah, L., & Kurniasih, N. (2019). Panel Data Model Estimation: The Effect of Managerial Ownership, Capital Structure, and Company Size on Corporate Value. *Journal of Physics: Conference Series*, *1175*(1), 12285. https://doi.org/10.1088/1742-6596/1175/1/012285.