Jurnal Varian
Vol. 6, No. 2, April 2023, Hal. 105–118
e-ISSN: 2581-2017

Robust Spatial-Temporal Analysis of Toddler Pneumonia Cases and its Influencing Factors

Musdalifah[1], Siswanto[2], Nirwan Ilyas[3]
[1,2,3]Hasanuddin University, Indonesia

Article Info

Article history:

Received : 12-22-2022
Revised : 05-06-2023
Accepted : 05-16-2023

Keywords:

Outliers;
Pneumonia;
Regression;
Robust Geographically and Temporally Weighted;
M-Estimator.

ABSTRACT

Pneumonia is a disease that causes inflammation of the lungs and is one of the most common diseases infecting toddlers. As a directly infectious disease, there is a possibility of the influence of location diversity on the number of pneumonia sufferers. This study was conducted to model pneumonia sufferers under five and to find out the factors that significantly affect the number of sufferers in each observation. Robust Geographically and Temporally Weighted Regression (RGTWR) is a method used to model data by considering the heterogeneity of location and time and to overcome outliers in the data. The data used is the number of pneumonia sufferers aged under five and the factors that are thought to influence it, namely the number of health centers, population density, percentage of children under five with complete basic immunizations, percentage of children under five who are exclusively breastfed 0-6 months, and percentage of poor people. RGTWR produces an optimal model with an R^2 value of 99.9997%, a Mean Absolute Deviation of 21.6852, and a Median Absolute Deviation of 6.9661 compared to Geographically and Temporally Weighted Regression model. Variables number of puskesmas, percentage of infants with complete basic immunization, and percentage of poor population are factors that influence the number of pneumonia sufferers under five in most locations in 34 provinces and 5 years of observation.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v6i1.2599

Corresponding Author:

Siswanto,
Department of Statistics, Hasanuddin University, Indonesia
Email: siswanto@unhas.ac.id

A. INTRODUCTION

Pneumonia is one type of direct infectious disease. The Ministry of Health of the Republic of Indonesia through the publication of Profil Kesehatan Indonesia stated that the handling of pneumonia has so far been prioritized on controlling pneumonia under five (Kementerian Kesehatan Republik Indonesia, 2020). The number of children aged 0-4 years in Indonesia in 2018 is estimated to be around 23,729,583 people (Kementerian Kesehatan Republik Indonesia, 2019). Global estimates show that in that year, every hour 71 children in Indonesia contracted pneumonia and more than 19,000 children under five died from pneumonia, or more than 2 children under five every hour (UNICEF, 2019). The importance of preventing pneumonia in children under 5 years of age should be considered, including considering the factors that influence on increasing the number of pneumonia sufferers in children under five.

Spatial analysis techniques have proven useful in understanding the epidemiology of infectious diseases. For example, it can identify new risk factors or areas of increased risk that could benefit from targeted public health interventions (Blain et al., 2014). Referring to previous research, pneumonia as an infectious disease causes a possible spatial effect on the number of cases (Alwi et al., 2022; Nadya, 2017; Renika et al., 2021; Ristiani et al., 2021). For example, research conducted by (Nadya, 2017), shows that in general the factors that influence the number of pneumonia sufferers in toddlers in West Java are the number of severely malnourished toddlers, the percentage of infants who are fully immunized, and the number of health centers. Research by (Ristiani et al., 2021) shows the results that the number of poor people, population density, malnutrition, and water suitable for consumption are the significant factors that affect the number of toddlers with pneumonia in West Java. Spatio-temporal modelling is the process of extracting hidden and useful knowledge from large spatio-temporal datasets, finds wide-ranging applications in the geospatial domain (Liu et al., 2017; Yin et al., 2016). One method that can be used to analyse the factors that influence a variable by considering different locations or the presence of spatial diversity is the Geographically Weighted Regression (GWR) model (Fotheringham et al., 2003). GWR is a popular local regression method used in various disciplines (Wu et al., 2019). Geographically Weighted model is suitable when the data is not well explained by the global model (Gollini et al., 2013). The GWR models produce local parameter estimators for each observation location, contrast to the linear regression model (Yasin et al., 2012). Then, to get more accurate parameter estimates, a time component should be added to the GWR model (Fotheringham et al., 2015). Adding the time element in GWR can be done using the Geographically and Temporally Weighted Regression (GTWR) method. This explains the spatial and temporal heterogeneity (Guo et al., 2017; Huang et al., 2010; Ma et al., 2018; Wu et al., 2018). Unlike standard GWR models, GTWR combines temporal and spatial information in a weighting matrix to identify the presence of spatial and temporal heterogeneity (Conita and Purwaningsih, 2020).

GTWR is not robust to outliers. Therefore, outliers in the data create biased coefficient estimates and errors in completing regression relationships (Zhang and Mei, 2011). One method that accommodates outliers is Robust Geographically and Temporally Weighted Regression (RGTWR) using M-Estimator. Robust regression is a regression method that is used when the distribution of the residuals is not normal and/or when there are a large number of outliers that influence the model (Olive, 2008). The M-Estimator is a robust regression method that is commonly used which was introduced by Huber (Huber, 1965). The parameter estimation in this method uses an iterative process called iteratively reweighted least squares (IRLS), and used to build outlier-resistant models (Siswanto et al., 2017). Several previous research results show that RGTWR model produces a better model than GTWR model in explaining data conditions (Erda et al., 2019; Haryanto et al., 2019; Putra, 2019). For example, (Putra, 2019) modelled the RGTWR using the S-Estimator in cases of crime rates in East Java. This research produced a GTWR model with a high coefficient of determination, which is 97.5%, but there are many outliers in the model error. RGTWR modeling with the S-Estimator then produces a higher coefficient of determination, amounting to 98.2%, and the value of the goodness models which is smaller than the GTWR model.

Based on the description above, the purpose of this study was to model the RGTWR with the M-Estimator in pneumonia patients under five years of age in Indonesia and obtain the factors that significantly influence the number of cases. Research related to factors that affect the number of under-five pneumonia sufferers previously used many spatial, temporal, or focused aspects to overcome outliers with robust methods, but no one has used all three. In this study, these three aspects will be used, so that the spatial and temporal aspects of the factors thought to influence changes in the number of pneumonia sufferers under five can be accommodated, as well as outlier problems that can reduce the accuracy of the analysis results can be overcome. The data used in this study is the number of under-five pneumonia sufferers in Indonesia during the period 2016 to 2020 and the factors that can influence it.

Aside from the various data and research locations, the novel aspect of this study is the use of a robust M estimator in estimating the regression parameters, as well as the use of two robust objective functions in the estimation process, namely the Tukey and Huber objective functions, with the Tukey objective function being the objective function used because it produces the highest coefficient of determination.

1. Breusch-Pagan Test

The Breusch-Pagan test statistic can be used to test for spatial heterogeneity with the following hypothesis:

H_0 : $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$ (No spatial heterogeneity)

H_1 : At least one $\sigma_i^2 \neq \sigma^2; i = 1, 2, \ldots, n$ (Have spatial heterogeneity)

The test statistics used are as follows.

$$BP = \left(\frac{1}{2}\right) f^T Z \left(Z^T Z\right)^{-1} Z^T f \tag{1}$$

where the element vector f is:

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

with:

e_i : Residual at observation-i

Z : A Matrix of size $n \times (p+1)$ containing the standardized vectors for each observation. H_0 will be rejected if $BP\ value > \chi^2_{(\alpha;p)}$ (Breusch and Pagan, 1980).

DOI: https://doi.org/10.30812/varian.v6i1.2599

2. Geographically and Temporally Weighted Regression

Geographically and temporally weighted regression (GTWR) is a method developed for GWR that accommodates for spatial and temporal heterogeneity (Huang et al., 2010). The general equation for the GTWR model of the response variable p with the explanatory variable y_i at the location (u_i, v_i, t_i) is as follows:

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

with:

y_i	:	The observed value of the response variable for the-i observation location
$\beta_0(u_i, v_i, t_i)$	:	Intercept at observation location (u_i, v_i) and time-t_i
$\beta_k(u_i, v_i, t_i)$	:	Coefficient of k-th explanatory variable at the location of the-(u_i, v_i) and the-t_i time
x_{ik}	:	Observation value of the k-th explanatory variable at the i-th observation location with $k = 1, 2, \ldots, p$
ε_i	:	Observation residual at location-i

Parameter estimation is written as follows.

$$\widehat{\beta}(u_i, v_i, t_i) = \left(X^T W(u_i, v_i, t_i) X\right)^{-1} x^T W(u_i, v_i, t_i) Y \tag{3}$$

with

$$W(u_i, v_i, t_i) = diag\left(w_{i1}, w_{i2}, \ldots, w_{in}\right)$$

with:

n	:	Number of observation data
$W(u_i, v_i, t_i)$	:	Weighting matrix on observations (u_i, v_i) and time t_i.

3. Robust Geographically and Temporally Weighted Regression

A Robust regression for GTWR is done by adding weights to Equation 2. Robust Geographically and Temporally Weighted Regression (RGTWR) model for the i-th location and time with outliers is:

$$\rho(y_i) = \rho\left[\beta_0(u_i, v_i, t_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i, t_i)x_{ik} + \varepsilon_i\right]$$

with: $i = 1, 2, \ldots, n;\;\; x_i = x_{i1}, x_{i2}, \ldots, x_{ip}$

RGTWR using M-Estimator is done by minimizing the error with equation:

$$\widehat{\beta}_m = \min_{\beta} \sum_{i=1}^{n} \rho(\varepsilon_i) = \min_{\beta} \sum_{i=1}^{n} \rho(y_i - X_i^T \beta) = \sum_{i=1}^{n} \left(\rho(y_i - X_i^T \beta(u_i, v_i, t_i))\right)$$

The M-Estimator is performed using the Iteratively Reweighted Least Square (IRLS) method, or weighted least squares iteration. In this iteration, the value of w_i changes its value each iteration, so that:

$$\widehat{\beta}(u_i, v_i, t_i)^m = \left(X_i' W^{m-1} X_i\right)^{-1} X_i' W^{m-1} y_i \tag{4}$$

with m is the number of iterations.

At the given weight W^m we get the estimator:

$$\widehat{\beta}(u_i, v_i, t_i)^{m+1} = \left(X_i' W^m X_i\right)^{-1} X_i' W^m y_i \tag{5}$$

The computation of the above equation is repeated until a convergence estimator is obtained, that is, until the difference between the values of $(u_i, v_i, t_i)^{m+1}$ and $\beta(u_i, v_i, t_i)^m$ approaches zero (Erda et al., 2019).

B. RESEARCH METHOD

1. Data

The data used in this study is secondary data on pneumonia sufferers aged under five in 34 provinces of Indonesia in 2016-2020 obtained through the publication of Badan Pusat Statistik (BPS) and also the publication of the Minister of Health of the

DOI: https://doi.org/10.30812/varian.v6i1.2599

Republic of Indonesia through the Profil Kesehatan Indonesia. Table 1 shows the independent and dependent variables used in this study.

Table 1. Variable Used

Variable	Description
Y	Number of pneumonia sufferers in children under five
X_1	Number of PUSKESMAS
X_2	Population density
X_3	Percentage of infants with complete basic immunization
X_4	Percentage of infants exclusively breastfed 0-6 months
X_5	Percentage of poor people

2. Analysis Step

The research flow is explained through the flowchart according to Figure 1, and explained in detail through the following steps:

1. Exploring data on the number of pneumonia sufferers in toddlers
2. Perform linear regression modelling and test residual assumptions
3. Checking the spatial diversity and time diversity of the data using a boxplot and the Breusch-Pagan test using Equation 1
4. GTWR model analysis
   (a) Determination of temporal spatial ratio parameter (τ)
   (b) Determination of spatial parameters (λ) and temporal parameters (μ)
   (c) Determination of optimum bandwidth h_{ST}
   (d) Perform parameter calculations $\beta(u_i, v_i, t_i)$ using Equation 3.
5. Detect outlier data using a boxplot
6. Analyze RGTWR model using M-Estimator for each observation in location and time
   (a) Calculate the value of $\widehat{y}_i$ using the formula:
   $$\widehat{y}_i = \chi_i^T \widehat{\beta}(u_i, v_i, t_i)^0$$
   where $\widehat{\beta}(u_i, v_i, t_i)^0$ is the value of the GTWR model parameter for the i-th location that has been obtained previously
   (b) Calculating the weighted value w_i using the Tukey weighting function
   (c) Calculate the value $\widehat{\beta}(u_i, v_i, t_i)^0$ where m is the number of iterations using Equation 4.
   (d) Repeat steps (a)-(c) until convergent parameter estimator are obtained, that is, the value of $\widehat{\beta}(u_i, v_i, t_i)^{m+1}$ dan $\widehat{\beta}(u_i, v_i, t_i)^m$ approaches zero.
7. Partial test for each parameter RGTWR
8. Selection of the best model based on the value of R^2, Median Absolute Deviation (MdAD), and Mean Absolute Deviation (MeAD).

DOI: https://doi.org/10.30812/varian.v6i1.2599

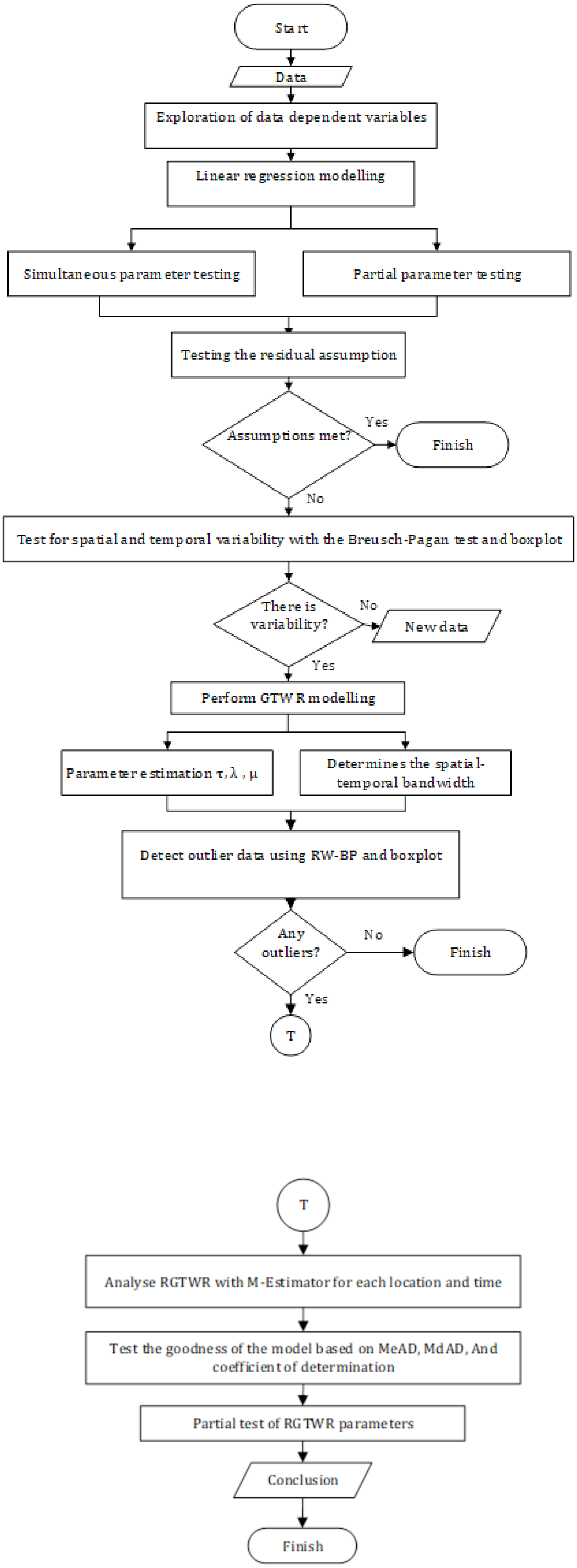

Figure 1. Flowchart of research analysis steps

C. RESULTS AND DISCUSSION

1. Description of the number of pneumonia sufferers under five in Indonesia

Table 2. Description of the number of pneumonia sufferers under five in Indonesia

Year	Mean	Standard Deviation	Maximum	Minimum	Median
2016	16710	33974.7544	174612	0	5550
2017	15042	30004.9988	145135	53	5411
2018	14863	28222.9192	131382	427	4956
2019	13770	24462.2179	104866	17	4684
2020	9112.9	18449.1810	76929	0	2173

From Table 2, every year the median and average pneumonia sufferers aged under five in Indonesia are decreasing, and a large decline occurred in 2020. One of the reasons for this large decline was the stigma of COVID-19 sufferers which affected the decrease in the number of visits. toddler cough or difficulty breathing at the Puskesmas. In 2019 the number of visits for toddlers with coughing or difficulty breathing was 7,047,834 visits, in 2020 it was 4,972,553 visits, a decrease from visits in 2019 which ultimately resulted in the discovery of pneumonia under five (Kementerian Kesehatan Republik Indonesia, 2020)

2. Multiple Linear Regression Modelling

Linear regression analysis is a method of linearly describing the relationship between response and predictor variables (Draper & Smith, 1998). Multiple linear regression modelling was carried out on the data on the number of pneumonia patients aged under five. In the simultaneous test of the model, the results are as shown in Table 3 with the hypothesis used as follows.

H_0 : $\beta_0 = \beta_1 = \ldots = \beta_5 = 0$

H_1 : At least one $\beta_k \neq 0$; $k = 0, 1, 2, \ldots, 5$

Table 3. Analysis of variance multiple linear regression model

Source of variation	Sum of squares	Degrees of Freedom	Mean of squares	F	p-*value*
Regression	92183197785.235	5	18436639557.047	88.911	0.000
Residual	34007257237.053	164	207361324.616		
Total	126190455022.288	169			

Obtained $p - value = 0.000 < \alpha = 0.05$, then reject H_0, there is at least 1 parameter that affects the number of pneumonia patients aged under five. The independent variables simultaneously affect the number of pneumonia sufferers under five. Subsequently, a partial test of the parameters of the linear regression model was carried out as shown in Table 4 with the following hypothesis.

H_0 : $\beta_k = 0$

H_1 : $\beta_k \neq 0$; $k = 0, 1, 2, \ldots, 5$

Table 4. Partial test of multiple linear regression model

Variable	$\widehat{\beta}_k$	$SE\left(\widehat{\beta}_k\right)$	$\lvert t_{test}\rvert$	$t_{\frac{\alpha}{2},(n-p-1)}$	Decision
Intercept	-31760.768	8927.444	3.558	2.2622	Significant
X_1	88.048	4.769	18.462	2.2622	Significant
X_2	1.357	0.435	3.117	2.2622	Significant
X_3	291.034	88.523	3.288	2.2622	Significant
X_4	2.695	59.685	0.045	2.2622	Not significant
X_5	-559.474	211.616	-2.644	2.2622	Significant

Variables X_1, X_2, X_3, and X_5 significantly affect the number of pneumonia sufferers under five, while the other 3 variables did not significantly affect. The following is a multiple linear regression model obtained.

$$\widehat{y} = -31760.768 + 88.0480X_1 + 1.357X_2 + 291.034X_3 - 559.474X_5$$

1. Normality Test

The Kolmogorov-Smirnov Normality Test was conducted to determine whether the residuals from the model were normally distributed or not. The Kolmogorov-Smirnov test results yield a value of 0.52353 and a $p - value < 2.2 \times 10^{-16}$, which is smaller than the significant level of 0.05, so the decision taken is that the residual is not normally distributed.

2. Independent Test

   The autocorrelation test was conducted to determine the residuals of the multiple linear regression models were independent of each other or not (Draper and Smith, 1998). Based on the results of the Durbin Watson test, the value of $d = 1.9959$ was obtained. Conclusions are drawn by looking at the value of the Durbin Watson table. The Durbin Watson table for $n = 170$ and $k = 5$ is 1.6890 for the d_L value and 1.8100 for the d_U value. Based on the decisionmaking criteria in the Durbin Watson test, it was found:

$$dU = 1.8100 < d = 1.9959 < 4 - dL = 2.311$$

   it is concluded that there is no correlation between residuals, and the independent assumptions of the model are met.

3. Identic Test

   Identical assumption (homoscedasticity) means that the variance on the residuals is the same or identical. The opposite is heteroscedasticity, which is a condition if the residual variance is not identical (Gujarati, 2003). The Glejser test was carried out to see whether there was no heteroscedasticity problem in the model, the F_{test} value was 15.683 and the p-value was 0.000. Based on the decision criteria for the Glejser test, because the value $F_{test} = 15.683 > F_{(\alpha,p,(n-p-1))} = 2.2693$ and $p-value = 0.000 < \alpha = 0.05$, the identical assumption is not met, there is a heteroscedasticity problem in the residual model. One of the possible causes of the data not meeting the assumption of homoscedasticity is the presence of spatial heterogeneity, which will be tested using the Breusch-Pagan test.

4. Multicollinearity Test

   A Multicollinearity test was performed to determine if there was a correlation between the independent variables. A VIF score > 10 indicates the presence of multicollinearity (Laksana, 2018).

Table 5. VIF value of independent variables

Variable	VIP
X_1	1.0814
X_2	1.0865
X_3	1.1897
X_4	1.0434
X_5	1.2176

Based on the test results as shown in Table 5, it is found that the VIF value for all independent variables is less than 10, so it can be concluded that there is no correlation between the independent variables.

3. Spatial and Temporal Heterogeneity Test

The Breuch-Pagan test was conducted to determine the presence or absence of spatial heterogeneity in the data. After testing, the value of $= 39.3860$ and $p - value = 1.985 \times 10^{-7}$. BP value $= 39.3860 > \chi^2_{0.05,7} = 11.0705$ and $p - value = 1.985 \times 10^{-7} < \alpha = 0.05$, it can be concluded that there is spatial heterogeneity in the data. Furthermore, spatial heterogeneity testing was carried out using a boxplot diagram.

DOI: https://doi.org/10.30812/varian.v6i1.2599

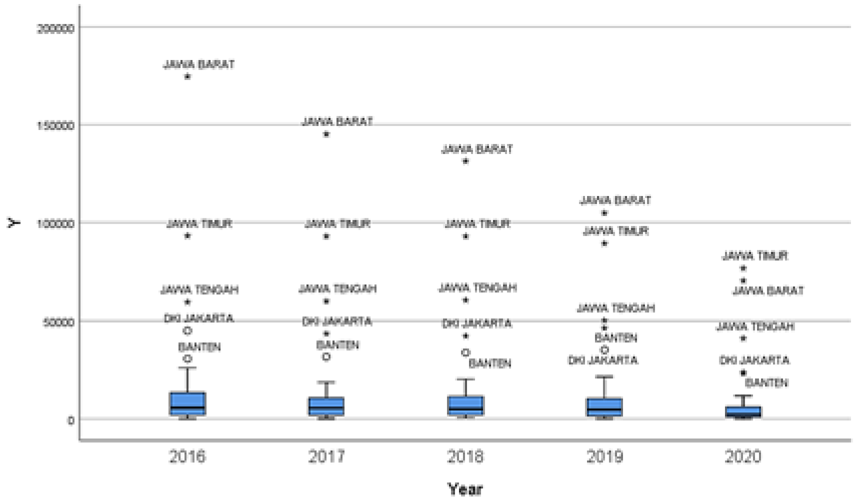

Figure 2. Boxplot Distribution of Toddler Age Pneumonia Patients

Based on Figure 2, there is a change in the pattern in the form of a decreasing median for each year. This pattern of decline indicates that there is temporal heterogeneity in the data.

4. Modelling Geographically and Temporally Weighted Regression

The initial step taken in GTWR modelling is to determine the kernel function that will be used for weighting, with the selection criteria being the kernel function with the optimum bandwidth that produces the smallest Cross Validation value, which is the best kernel function used for GTWR modelling.

Table 6. VIF value of independent variables

Kernel Function	Bandwidth GWR	CV GWR	Bandwidth GTWR	CV GTWR
Gaussian	1.5793	17501752416	2.5260	6365076914
Exponential	0.8222	16513907206	0.9875	5087110056
Bisquare	3.1719	28222712605	13.1878	27749690452
Kernel Function	**Bandwidth GWR**	**CV GWR**	**Bandwidth GTWR**	**CV GTWR**
Tricube	22.7745	40545798638	0.0118	156595363195

Table 6 shows that the Exponential kernel function produces the smallest CV value compared to the optimal bandwidths of the other kernel functions, so the Exponential kernel function is chosen for weighting the GTWR model. Next, determine the bandwidth value and calculate the parameters that will be used in GTWR modelling.

Table 7. VIF value of independent variables

h_S	h_{ST}	τ	μ	λ
0.8222	0.9875	1.7957	22.1395	12.3293

Table 7 shows the obtained GTWR model parameter values. The spatial temporal bandwidth value of 0.9875 indicates that the area that is less than or equal to 0.9875 has a large influence on the i-th observation point. GTWR modelling produced 170 models from 34 provinces and 5 years of observation. The following shows an example of the GTWR model obtained, namely the model for Aceh Province in 2016.

$$\widehat{y} = -0.3302 + 4.9019X_1 + 34.2047X_2 + 3.8284X_3 - 50.2079X_4 - 7.3022X_5$$

DOI: https://doi.org/10.30812/varian.v6i1.2599

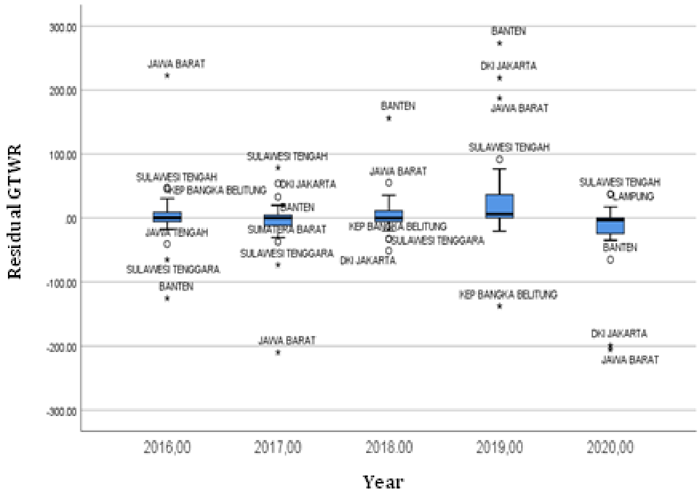

Figure 3. GTWR Model Residual Boxplot

Detection of outliers using the boxplot in Figure 3 resulted in 27 provinces indicated as outliers for five years of observation. Modeling RGTWR using M-Estimator is then carried out to overcome outliers in the GTWR model. After 34 iterations, the parameter $\beta_k(u_i, v_i, t_i)$ converges. The following is a thematic map of the distribution of parameters $\beta_k(u_i, v_i, t_i)$ for 5 years of observation.

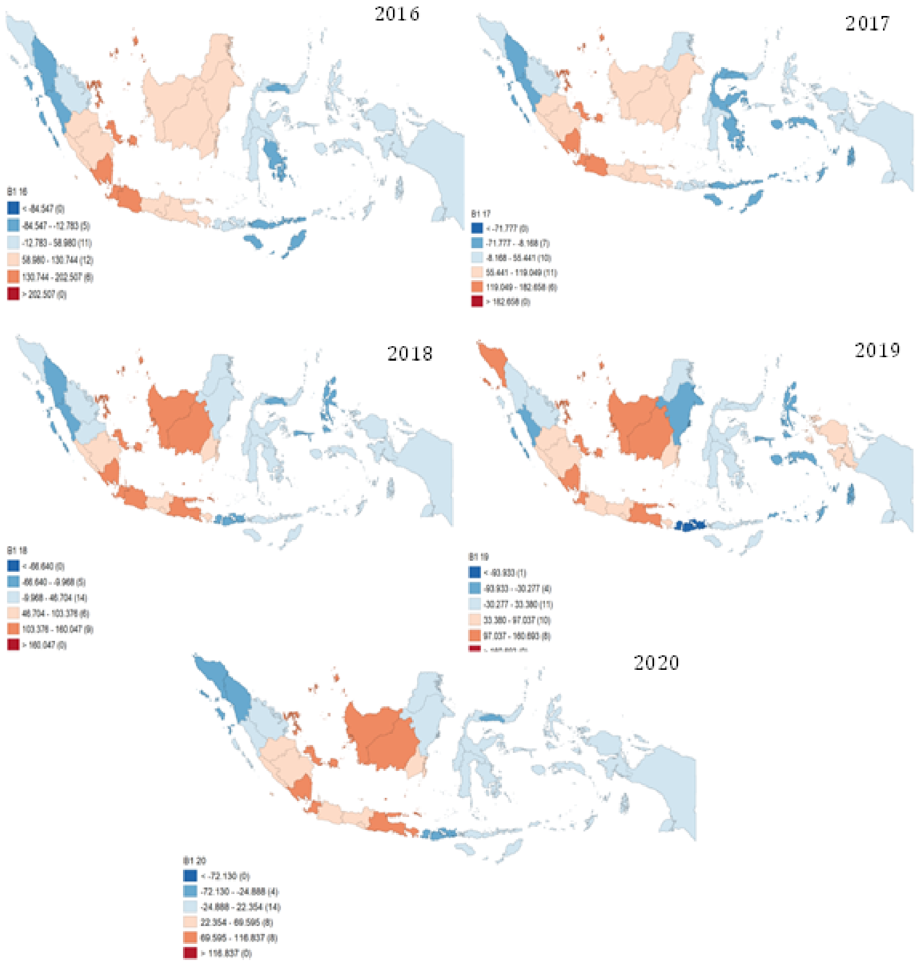

Figure 4. Distribution of parameters β_1

DOI: https://doi.org/10.30812/varian.v6i1.2599

Figure 4 shows the distribution of the parameter β_1, the number of puskesmas. From Figure 4 it can be seen that most of the parameters have positive values. This mean that an increase in the number of puskesmas will also increase the number of pneumonia sufferers found.

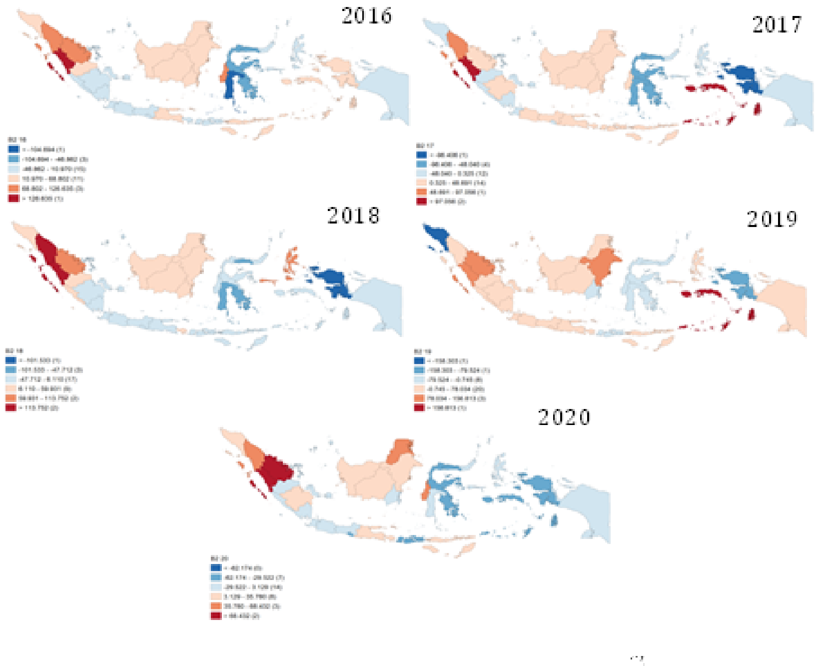

Figure 5. Distribution of parameters β_2

Figure 5 Shows the distribution of the parameter β_2, namely population density. From Figure 5 it can be seen that most of the parameters show positive values, meaning that the greater population density will also increase the number of pneumonia sufferers. West Sumatra Province is the province that has the highest parameter value, it means that in that province, population density has a big effect when compared to other observations.

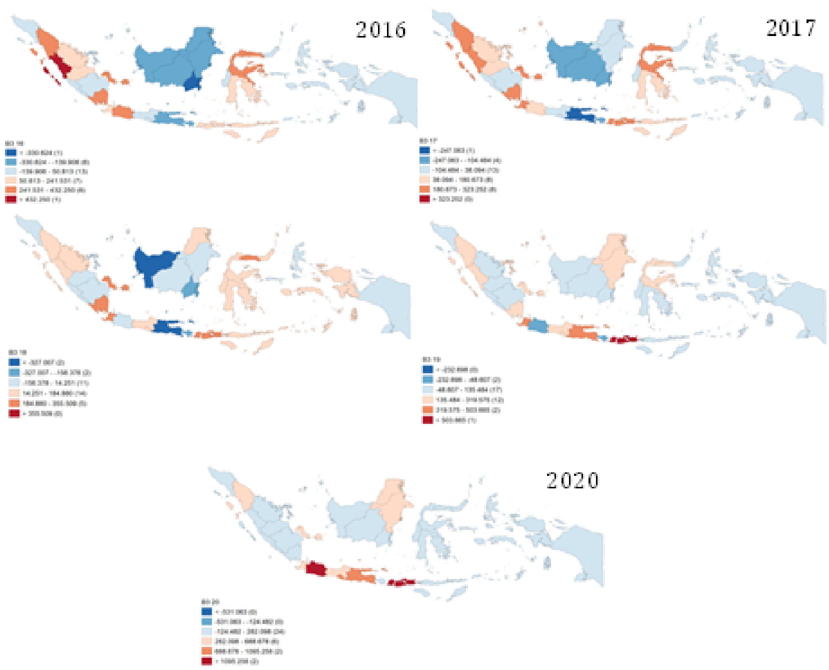

Figure 6. Distribution of parameters β_3

Figure 6 shows the distribution of the parameter β_3 the percentage of infants receiving complete basic immunization. Parameters for observations have various values, but most of them are positive. This mean that the higher the percentage of children under five who received complete basic immunization, the higher the number of pneumonia patients cases found.

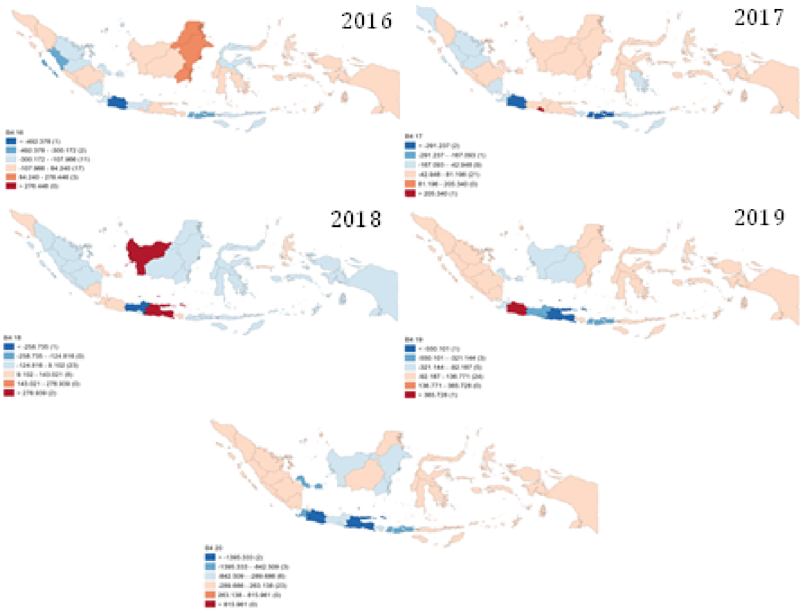

Figure 7. Distribution of parameters β_4

Figure 7 shows the distribution of the parameter β_4 the percentage of infants receiving exclusive breastfeeding for 0-6 months. Most of the parameters are negative, which means that an increase in the percentage of infants receiving exclusive breastfeeding will reduce the number of pneumonia sufferers in toddlers, and vice versa.

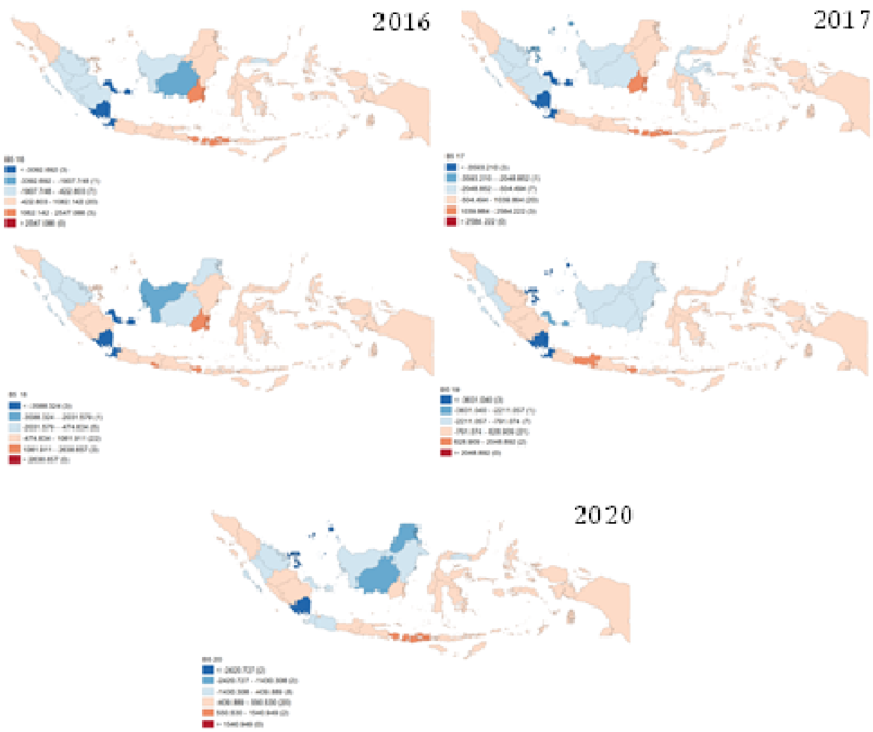

Figure 8. Distribution of parameters β_5

Figure 8 shows the distribution of the parameter β_5, the percentage of poor people. Most of the parameters are negative, meaning that the higher the percentage of poor people, the number of pneumonia sufferers found will decrease. The following are some of the RGTWR models obtained.

Aceh Province in 2016:

$$\widehat{y} = -0.2677 + 6.8796X_1 + 30.1988X_2 + 0.3026X_3 - 51.9513X_4 - 5.7336X_5$$

North Sumatra Province in 2016:

$$\widehat{y} = 3.6458 + 6.8796X_1 + 30.1988X_2 + 0.3026X_3 - 51.9513X_4 + 15.9813X_5$$

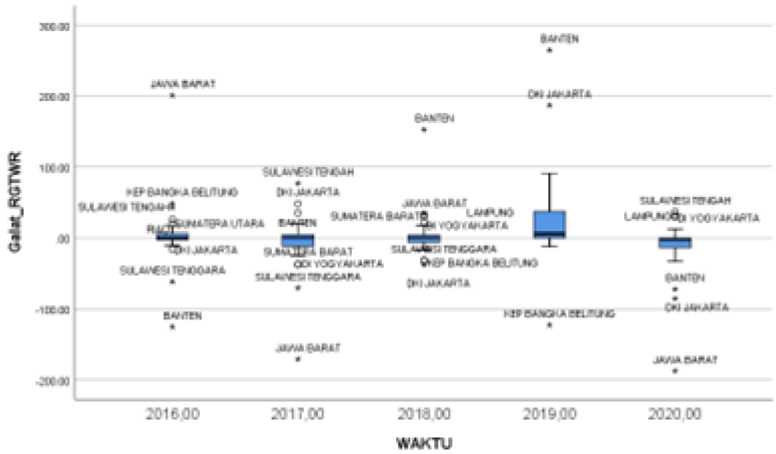

Figure 9. Boxplot Residual Model RGTWR

Figure 9 shows the detection of outliers using a boxplot on the RGTWR model which resulted in 32 observations whose model residuals were detected as outliers. Next, the best model was selected using the criteria of Mean Absolute Deviation (MeAD), Median Absolute Deviation (MdAD) and R^2.

Table 8. VIF value of independent variables

Model	R^2	MdAD	MeAD
GTWR	99.9996%	8.3474	25.1833
RGTWR	99.9997%	6.9661	21.6852

Table 8 shows that the RGTWR model succeeded in reducing the MdAD and MeAD values from the GTWR modelling. This means that the RGTWR modelling with the M-Estimator has an estimated value that is closer to the actual value. The R^2 value of the RGTWR model is also higher than the GTWR model, so it can be said that the RGTWR model is able to explain the diversity of models better than the GTWR model. Compared with previous research, for example, research by (Febrianti et al., 2023) modeling the incidence of pneumonia in toddlers in Bandung using the GTWR method, obtained a coefficient of determination of 87.83%. Research conducted by (Renika et al., 2021) who modeled the extrinsic risk factors for pneumonia in toddlers on Java Island in 2018 using GWR obtained a coefficient of determination of 46.2617%, higher than modeling with multiple linear regression of 35.9%. This study used the RGTWR by taking into account spatial, temporal, and outlier factors, which resulted in a model with a high coefficient of determination, indicating that the method used was correct compared to the GTWR method or multiple linear regression. Based on this, it can be concluded that the RGTWR model with the M-Estimator is better used in modelling pneumonia patients aged under five in Indonesia.

The results of the RGTWR modeling with the M-Estimator show that the variability variable has a significant effect on the number of cases of pneumonia under five each year, both in terms of the magnitude of the influence and the direction of the influence of the variable, namely positive or negative effect. In general, the number of puskesmas, the proportion of infants receiving complete basic gymnastics, and the proportion of poor people are variables that significantly influence the number of pneumonia sufferers under five in most of the observations. Compared to previous studies (Nadya, 2017; Renika et al., 2021;

Ristiani et al., 2021), the factors that influence the number of pneumonia sufferers under five show mixed results, but there are several significant variables that are the same. In line with research conducted by Erda (Erda et al., 2019), the RGTWR method produces a model that is better at describing data conditions, compared to the GTWR method and multiple linear regression.

D. CONCLUSION AND SUGGESTION

The number of cases of pneumonia patients under five in Indonesia can be modelled using the RGTWR method with the M-Estimator. This is due to the spatial and temporal heterogeneity of the data, and the outliers in the GTWR model so that the RGTWR method is appropriate. The RGTWR model with M-Estimator produces a MeAD value of 21.6852, an MdAD of 6.9661 and an R^2 value of 99,9997%. Variable number of puskesmas, percentage of infants with complete basic immunization, and percentage of poor population are variables that affect the number of pneumonia sufferers under five in Indonesia in most of the observations in 34 provinces and 5 years of observation.

To produce a more optimal model for all observations, it is advised to conduct future research in an area with fewer islands. Alternatively, adaptive bandwidth, which has a different bandwidth value for each observation, can be used in conjunction with other robust estimators in RGTWR modeling to produce the best model.

REFERENCES

Alwi, W., Sauddin, A., and Islamiyah, N. I. (2022). Pemodelan generalized poisson regression pada faktor-faktor yang mempengaruhi kasus pneumonia pada balita di provinsi sulawesi selatan 2018. *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, 10(1):9–14.

Blain, A., Thomas, M., Shirley, M., Simmister, C., Elemraid, M., Gorton, R., Pearce, M., Clark, J., Rushton, S., and Spencer, D. (2014). Spatial variation in the risk of hospitalization with childhood pneumonia and empyema in the north of england. *Epidemiology & Infection*, 142(2):388–398.

Breusch, T. S. and Pagan, A. R. (1980). Lagrange multiplier test and to model applications specification in econometrics. *The Review of Economic Studies*, 47(1):239–253.

Conita, C. and Purwaningsih, T. (2020). Under-five mortality rate modeling using geographically and temporal weighted regression (gtwr). In *Ahmad Dahlan International Conference on Mathematics and Mathematics Education*, volume 1, pages 59–67.

Draper, N. R. and Smith, H. (1998). *Applied regression analysis*, volume 326. John Wiley & Sons.

Erda, G., Djuraidah, A., et al. (2019). Outlier handling of robust geographically and temporally weighted regression. In *Journal of Physics: Conference Series*, volume 1175, page 012041. IOP Publishing.

Febrianti, L., Andriyana, Y., and Faidah, D. Y. (2023). Pemodelan dan pemetaan kejadian pneumonia pada balita di kota bandung menggunakan metode geographically and temporally weighted regression (gtwr). *E-Journal BIAStatistics— Departemen Statistika FMIPA Universitas Padjadjaran*, 2023(1):277–295.

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

Fotheringham, A. S., Crespo, R., and Yao, J. (2015). Geographical and temporal weighted regression (gtwr). *Geographical Analysis*, 47(4):431–452.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., and Harris, P. (2013). Gwmodel: an r package for exploring spatial heterogeneity using geographically weighted models. *arXiv preprint arXiv:1306.0413*.

Gujarati, D. (2003). *Basic Econometrics*. Erlangga.

Guo, Y., Tang, Q., Gong, D.-Y., and Zhang, Z. (2017). Estimating ground-level pm2. 5 concentrations in beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment*, 198:140–149.

Haryanto, S., Aidi, M. N., and Djuraidah, A. (2019). Modelling of grdp the construction sector in java island using robust geographically and temporally weighted regression (rgtwr). *Repositories-Dept. of Statistics, IPB University*, pages 165–174.

Huang, B., Wu, B., and 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.

Huber, P. J. (1965). A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, pages 1753–1758.

Kementerian Kesehatan Republik Indonesia (2019). Profil Kesehatan Indonesia 2018. *Kementerian Kesehatan Republik Indonesia*. KementerianKesehatanRepublikIndonesia.https://www.kemkes.go.id/folder/view/01/structure-publikasi-pusdatin-profil-kesehatan.html.

Kementerian Kesehatan Republik Indonesia (2020). Profil Kesehatan Indonesia 2019. *Kementerian Kesehatan Republik Indonesia*. https://www.kemkes.go.id/folder/view/01/structure-publikasi-pusdatin-profil-kesehatan.html.

Laksana, S. E. A. (2018). Pemodelan Kebakaran Hutan di Provinsi Riau dengan Metode Geographically-Temporally Weighted Regression. https://repository.its.ac.id/56605/1/06211440000034-Undergraduate_Theses.pdf.

Liu, J., Zhao, Y., Yang, Y., Xu, S., Zhang, F., Zhang, X., Shi, L., and 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., and 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.

Nadya, M. (2017). *Analisis Geographically Weighted Regression (Gwr) Pada Kasus Pneumonia Balita Di Provinsi Jawa Barat*. PhD thesis, UNIVERSITAS NEGERI JAKARTA.

Olive, D. J. (2008). Applied robust statistics. *Preprint M-02-006*.

Putra, Z. (2019). *Model Regresi Terboboti Geografis Temporal Kekar untuk Tingkat Kriminalitas di Provinsi Jawa Tengah dan Provinsi Jawa Timur*. PhD thesis, IPB University.

Renika, I., Amin, C., et al. (2021). *Identifikasi Faktor Risiko Ekstrinsik Pneumonia pada Balita di Pulau Jawa Tahun 2018 Menggunakan Geographically Weighted Regression*. PhD thesis, Universitas Muhammadiyah Surakarta.

Ristiani, D. A. et al. (2021). Analisis faktor-faktor yang mempengaruhi jumlah balita yang terkena penyakit pneumonia di provinsi jawa barat dengan regresi terboboti geografis. B.S. thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.

Siswanto, S., Raupong, R., and Anisa, A. (2017). Estimasi regresi robust m pada faktorial rancangan acak lengkap yang mengandung outlier. *Jurnal Matematika, Statistika dan Komputasi*, 13(2):171–181.

UNICEF (2019). *Lembaga Kesehatan dan Anak Memeringatkan Satu Anak Meninggal akibat Pneumonia setiap 39 Detik*. https://www.unicef.org/indonesia/id/press-releases/lembaga-kesehatan-dan-anak-memeringatkan-satu-anak-meninggal-akibat-pneumonia-setiap.

Wu, C., Ren, F., Hu, W., and Du, Q. (2019). Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, 33(3):489–511.

Wu, C., Ye, X., Ren, F., and Du, Q. (2018). Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in shenzhen, china. *Cities*, 77:104–116.

Yasin, H. et al. (2012). Mixed geographically weighted regression model (case study: The percentage of poor households in mojokerto 2008). *European Journal of Scientific Research*, 69(2):188–196.

Yin, J., Gao, Y., Du, Z., and Wang, S. (2016). Exploring multi-scale spatiotemporal twitter user mobility patterns with a visual-analytics approach. *ISPRS International Journal of Geo-Information*, 5(10):187.

Zhang, H. and Mei, C. (2011). Local least absolute deviation estimation of spatially varying coefficient models: robust geographically weighted regression approaches. *International Journal of Geographical Information Science*, 25(9):1467–1489.