Geographically Weighted Logistic Regression Model on Binomial Data to Explore Weather Spatial Non-Stationarity in Covid-19 Cases

Pepi Novianti, Gunardi, Dedi Rosadi*

***Abstract*—Geographically Weighted Logistic Regression (GWLR) creates a logistic regression model by incorporating spatial variation through geographical weighting in parameter estimation. GWLR can be employed for investigating the relationship among a variable and its independent variables when the dependent variable has a Bernoulli distribution. In some studies, not only a single individual or experimental unit is observed but also a group of binary samples that have all been treated in the same manner. As a result, data with n total observations can be considered to have a binomial distribution, and the total availability also indicates that the data have a binomial distribution rather than a Poisson distribution or a negative binomial distribution. In the present research, the GWLR model is modified into Geographically Weighted Binomial Logistic Regression (GWBLR). Unlike the GWLR, which only observes one binary sample as the response variable, the GWBLR allows n binary samples as the response variable. The parameter estimations were calculated by maximum likelihood estimation (MLE), and the optimum bandwidth was determined by a fixed Gaussian function based on the least Cross Validation (CV) result. The likelihood ratio test and the Wald test were used to test hypotheses. We created the model using both simulation and real data. The simulation parameter estimations demonstrated high consistency with the generated data. The GWBLR model was applied to real data to investigate the spatial effects of temperature, sunshine, humidity and precipitation on the number of confirmed Covid-19 cases in Indonesia. Temperature, sunshine, humidity and precipitation all had an important effect on confirmed Covid-19 cases, according to the data.**

***Index Terms*—binomial distribution, geographically weighted binomial logistic regression, geographically weighted logistic regression, maximum likelihood estimation, likelihood ratio test**

I. Introduction

Regression analysis is technique for investigating possible relationships involving predictor factors and variable related to responses statistically. The Ordinary Least Squares (OLS) method, which assumes that all variables are continuous, is a popular estimation method of the parameters in a linear regression model. The normality of the data is also required in classical linear regression. The Generalized Linear Model (GLM) evolved from classical linear regression when normality was not met. GLM assumes that the response variable is of the exponential family, which includes binomial, Poisson, and negative binomial distributions.

Observations in an experimental unit or individual can result in one of two outcomes in various fields of research. Those are generally expressed in terms of ”success” and ”failure”. Binary data have a Bernoulli distribution and has two possible outcomes. In some studies, not only one individual or experimental unit is observed, but also a group of units that have all been treated similarly. The data are then referred to as ”grouped binary data” and show the number of successful observations out of all observations. Given that such a sample has a binomial distribution, data with a Bernoulli distribution are a subset of binomial distributions with a single observation [1]. Hilbe [2] distinguishes two types of logistic regression: binary (”1” or ”0”) response logistic models based on the Bernoulli distribution and proportional (”n/m”) response logistic models based on the binomial probability distribution. The Bernoulli distribution is a binomial model with a response denominator of one (”1”). If there are not too many predictors with binary or categorical outcomes with multiple levels, binary response logistic models can be converted to clustered models.

The analysis for binomial data has been extended. Hall [3] suggested a model for finite count data known as zero-inflated binomial (ZIB), which adapts the model of zero-inflated Poisson (ZIP). Jacups and Cheng [4] proposed using binomial logistic regression in the health sector to describe cases of community-acquired pneumonia (CAP). Furthermore, [5] and [6] used a Bayesian approach to modify the binomial logistic model in spatial analysis and industrial experiments, respectively. Binomial logistic regression is also widely used and implemented in many statistical programs for data with binary or binomial responses [7]. The number of known samples or finite integers is critical in binomial data analysis. The number of samples in the form of upper bounded count data or finite integers must be considered in binomial data.

Spatial analysis is one of the statistical analyses that is currently being developed. The primary principle of spatial analysis is to consider the relationship between areas represented by spatial data. The heterogeneity and spatial dependency tests are used to test for spatial effects. Area approaches, such as Spatial Autoregressive Models, can be used to resolve data with spatial dependency effects [8], [9]. This approach includes conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models. A point approach is used to solve spatial data with heterogeneity effects, such as the Geographically Weighted Regression (GWR) introduced by [10]. [11] developed a Geographically

Manuscript received January 13, 2023; revised June 12, 2023.

This work was supported by the Lembaga Pengelola Dana Pendidikan (LPDP), Ministry of Finance Republic of Indonesia as the scholarship funder.

P. Novianti is a doctoral student in Department of Mathematics, Universitas Gadjah Mada, Indonesia. (e-mail:pepi.novianti@mail.ugm.ac.id).

Gunardi is an Associate Professor in Department of Mathematics, Universitas Gadjah Mada, Indonesia. (e-mail:gunardi@ugm.ac.id).

*D. Rosadi (corresponding author) is a Professor in Department of Mathematics, Universitas Gadjah Mada, Indonesia. (e-mail:dedirosadi@ugm.ac.id).

Weighted Logistic Regression model (GWLR) for binary outcome data with spatial effects that are non-stationary. The GWLR model combines the GWR model [10], which investigates spatial heterogeneity or nonstationarity, and logistic regression with binary response variables. GWLR with binary response variables has been widely used in the health sector, as demonstrated by [12], [13] and [14]. GWR models with discrete and count data, in addition to GWLR, have been developed, including Geographically Weighted Poisson Regression (GWPR) and Geographically Weighted Negative Binomial Regression (GWNBR). Both of these models have been widely used, with the assumption of equidispersion data for GWPR [15] and overdispersion data for GWNBR [16]. In these two models, the entire number of observations is not required, and a given number of binary observations can be an infinite count data.

In some studies, data collected in one location, such as data on confirmed malaria cases, can be in the form of grouped binary data, similar to logistic regression of binomial data. Ssempiira, *et.al.* ([17] investigated the widespread incidence of malaria of kids in the Gambia using Bayesian spatiotemporal negative binomial models, and [18] proposed spatial correlogram models for detecting spatiotemporal trends in the risk of malaria in Uganda. The presence of total observations in binary data sets and the existence of more than one observation indicate that the data has a binomial distribution rather than a Poisson distribution or a negative binomial distribution [3]. The GWLR developed by [11] must be adjusted to the logistic regression of binomial data for these grouped binary data. As a result, GWLR with binomial data remains required to model and analyze data with a binomial distribution that has spatial effects.

A spatial approach based on GWR is commonly used in the health sector. [19] and [20] used several GWR models under the assumption of Gaussian modeling. The GWLR model with a binary response variable was used in [12], [13] and [14]. [21], [22] and [23] employed GWPR on the response variable represented by Poisson data. Furthermore, [24] and [25] conducted the most recent epidemiological research using the GWNBR model to investigate spatial factors that affect Covid-19 incidences.

Almost every country is concerned about the prevalence of Covid-19. The propagation of the COVID-19 pandemic, according to Lamlili [26], is a difficult event to comprehend and anticipate due to the multiple components involved, and it is a new virus with unknown characteristics. As a result, more research into the distribution of Covid-19 instances is required. Some Covid-19 research in Indonesia has been published. Forecasting studies for Covid-19 cases have been conducted, including the use of the Richard Model [27], Singular Spectrum Analysis [28], LSTM Neural Network [29], [30] and multi-state SVIRS model [31]. Furthermore, [32] forecasted using an autocorrelation analysis based on the Hijri calendar. [33], [34] and [35] investigated Covid-19's influence and impact in the social and economic sectors. Moreover, [36] and [37] investigated Covid-19 effects in the financial sector. In contrast, [38] investigated the weather effect on Covid-19 cases in Jakarta. However, [38] did not account for meteorological heterogeneity in Covid-19 infection transmission in Indonesia. Several studies, including [39], [40], [41], [42], [43] and [44], have shown that climatic circumstances can influence the dissemination of Covid-19. As a result of the spatial non-stationarity, it was necessary to conduct study on the impact of weather on Covid-19 cases in Indonesia.

The objective of the present research is to extend a GWLR model to include a dependent variable represented by binomial data, which is recognized as Geographically Weighted Binomial Logistic Regression (GWBLR). The following section will provide a brief explanation of logistic regression for binomial data, GWLR for binomial data, parameter estimation using the Maximum Likelihood Method (MLE), and bandwidth determination using the Cross-Validation (CV) approach. Section 3 discusses the model's application to simulation and real-world data. The last section of this paper is the conclusion.

II. Methods

A. Logistic Regression for Binomial Data

The observed response for binary or binomial data is the ratio y_i/n_i, denoted by p_i, for $i = 1, 2, \ldots, n$. The response for the i^{th} observation is represented by a binomial distribution $B(n_i, p_i)$, in which p_i represents the success probability or response probability and n_i represents the total number of observations. The expected value of the response variable is thus $E(Y_i) = n_i p_i$, and $Var(Y_i) = n_i p_i(1-p_i)$. For binary data specifically, $n_i = 1$ for all i and $y_i = 0$ (failure) or $y_i = 1$ (success). Therefore, $E(Y_i) = p_i$ and $Var(Y_i) = p_i(1-p_i)$ are true for this condition.

The model is examined in logistic regression in terms of the likelihood that the i^{th} observation will be successful, $p_i = E(Y_i/n_i)$, which has a value between 0 and 1. A logistic transformation is used to ensure that the probability is between 0 and 1 when modeling the linear equation. The logit transformation is used to transform the probability of success p_i, which is denoted by $logit(p_i)$ and can be calculated using the formula below [1]:

$$logit(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_d x_{xi} \quad (1)$$

and can be realign into the success probability equation presented below:

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_d x_{xi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_d x_{xi})}. \quad (2)$$

The unknown parameter in equation (1) or (2), $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_d\}$ is estimated using MLE. Likelihood is determined using probability function of binomial distribution given by

$$f(y_i; n_i, p_i) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i}, \text{ for } i = 1, \ldots, n. \quad (3)$$

Parameter estimation is obtained by determining the value of $\hat{\boldsymbol{\beta}}$ which maximizes the likelihood function by determining its first derivative and equating to zero. The likelihood function is maximized in logarithmic form, which is referred to as log-likelihood, as follows:

$$\begin{aligned} L(\beta) &= \log[l(\beta)] = \log\left\{\prod_{i=1}^{n}\binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i}\right\} \\ &= \sum_{i=1}^{n}\left\{\log\binom{n_i}{y_i} + y_i \log\eta_i - n_i\log(1+e^{\eta_i})\right\} \quad (4) \end{aligned}$$

where $\eta_i = \sum_{k=0}^{d} \beta_k x_{ki}$ and $x_{0i} = 1$ for each i. With regard to the $d+1$ unknown parameter, the derivative of this log likelihood function is given by:

$$\frac{\partial L(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} y_i x_{ki} - \sum_{i=1}^{n} n_i x_{ki} e^{\eta_i}(1+e^{\eta_i})^{-1}, j = 0, 1, \ldots, d. \quad (5)$$

These derivatives are evaluated at $\hat{\boldsymbol{\beta}}$ and are then equalized to zero, resulting in a set of $d+1$ nonlinear equations with $\hat{\beta}_j$ unknown parameters that can only be solved numerically.

B. Geographically Weighted Logistic Regression for Binomial Data

Geographically Weighted Logistic Regression (GWLR) is one approach for obtaining regression parameters while accounting for spatial factors. GWLR is a non-stationary parameter approach to GWR that combines categorical data and non-stationary parameters. [11] assumes that the GWLR response variable has a bernoulli distribution with Y between 0 and 1. The GWLR model is expressed as follows:

$$\pi(lon_i, lat_i) = \frac{\exp(\sum_{k=0}^{d} \beta_k(lon_i, lat_i)x_{ki})}{1 + \exp(\sum_{k=0}^{d} \beta_k(lon_i, lat_i)x_{ki})}, \quad (6)$$

where x_i is the observation value of the i response for $i = 1, 2, \ldots, n,$, $\beta_k(lon_i, lat_i)$ is the coefficient at the i observation location or geographic coordinates (*longitude*, *latitude*), and d is the number of predictor variable parameters. By denoting the linear function $\eta_i(lon_i, lat_i) = \sum_{k=0}^{d} \beta_{ki}(lon_i, lat_i)x_{ki}$ with $x_{0i} = 1$, then Equation (6) can be expressed in

$$\pi_i(lon_i, lat_i) = \frac{\exp(\eta_i(lon_i, lat_i))}{1 + \exp(\eta_i(lon_i, lat_i))}, i = 1, 2, \ldots, n. \quad (7)$$

If at one location the interest is the number of successes from n known samples with the same probability of each action, then the data obtained can be in the form of grouped binary data or binomial data. Similarly binomial logistic regression, which has the same model as binary logistic regression, the GWLBR model can also be expressed in equation (6) and (7). Think about that n observations of the dependent variable have a binomial distribution and y_i is the number of successful events of the total n_i events at the i location. The expected value of the variable y_i is $E(Y_i) = n_i \pi_i(lon_i, lat_i)$ where $\pi_i(lon_i, lat_i) = \frac{\exp(\eta_i(lon_i, lat_i))}{1+\exp(\eta_i(lon_i, lat_i))}$ is a probability of each corresponding location.

C. Parameter Estimation of GWBLR Model

To estimate the GWLBR model parameters, the MLE approach is employed. By differentiating the log-likelihood function to $beta_k(lon_i, lat_i)$ and equalizing the outcome to zero, parameter estimation is obtained. The likelihood function for the dependent variable in the GWLBR, which has a binomial distribution, is as follows:

$$\begin{aligned} l(\beta) &= \prod_{i=1}^{n} P(Y = y_i) \\ &= \prod_{i=1}^{n} \binom{n_i}{y_i} \pi_i(lon_i, lat_i)^{y_i}(1 - \pi_i(lon_i, lat_i))^{n_i - y_i}. \end{aligned} \quad (8)$$

and log-likelihood function is as follow:

$$\begin{aligned} L(\beta) &= \log[l(\beta)] \\ &= \log\left\{\prod_{i=1}^{n} \binom{n_i}{y_i} \pi_i(lon_i, lat_i)^{y_i}[1 - \pi_i(lon_i, lat_i)]^{n_i - y_i}\right\} \\ &= \sum_{i=1}^{n}\left\{\log\binom{n_i}{y_i} + y_i \log \pi_i(lon_i, lat_i) + \right. \\ &\quad \left.(n_i - y_i)\log(1 - \pi_i(lon_i, lat_i))\right\} \\ &= \sum_{i=1}^{n}\left\{\log\binom{n_i}{y_i} + y_i \log\left(\frac{\pi_i(lon_i, lat_i)}{1 - \pi_i(lon_i, lat_i)}\right) + \right. \\ &\quad \left. n_i \log(1 - \pi_i(lon_i, lat_i))\right\} \\ &= \sum_{i=1}^{n}\left\{\log\binom{n_i}{y_i} + y_i\left(\sum_{k=0}^{p} \beta_l(lon_i, lat_i)x_{li}\right)\right. \\ &\quad \left. - n_i \log\left(1 + \exp\left(\sum_{k=0}^{p} \beta_l(lon_i, lat_i)x_{li}\right)\right)\right\} \\ &= \sum_{i=1}^{n}\left\{\log\binom{n_i}{y_i} + y_i\eta_i - n_i \log\left(1 + \exp(\eta_i)\right)\right\}. \end{aligned} \quad (9)$$

The weighting factor in the GWBLR model is geographical location, and each region has local parameters with unique values. To produce the GWBLR model, the log-likelihood function is weighted by w_{ij}, which represents the geographically weighted of the j^{th} observation at the i^{th} regression point. In order to estimate the parameters at the i^{th} location, the local log-likelihood function to be employed is

$$\begin{aligned} L^*(\beta(lon_i, lat_i)) = \sum_{j=1}^{m}&\left\{\log\binom{n_j}{y_j} + y_j\eta_j(lon_i, lat_i) - \right. \\ &\left. n_j \log(1 + \exp(\eta_j(lon_i, lat_i)))\right\} w_{ij}. \end{aligned} \quad (10)$$

Parameter estimation $\beta_k(lon_i, lat_i)$ is derived by differentiating (10) to estimate parameter $\beta_k(lon_i, lat_i)$, where $\eta_j = \sum_{k=0}^{d} \beta_k(lon_i, lat_i)x_{kj}$. The following equation provides the derivative of the log likelihood function to the $d+1$ parameter $\beta_k(lon_i, lat_i)$:

$$\begin{aligned} \frac{\partial L^*(\beta(lon_i, lat_i))}{\partial \beta_k(lon_i, lat_i)} &= \sum_{j=1}^{m}\left\{y_j x_{kj} - n_j x_{kj}\left(\frac{\exp(\eta_j)}{1 + \exp(\eta_j)}\right)\right\} w_{ij} \\ &= \sum_{j=1}^{n} w_{ij}\left\{y_j x_{kj} - n_j x_{kj}\pi_j\right\} \\ &= \sum_{j=1}^{n} x_{kj} w_{ij}\left\{y_j - n_j\pi_j\right\} \end{aligned} \quad (11)$$

which can be expressed in the form of a matrix

$$\frac{\partial L^*(\beta(lon_i, lat_i))}{\partial \beta_k(lon_i, lat_i)} = \mathbf{X}^T \boldsymbol{W(lon_i, lat_i)y^*} \quad (12)$$

where, $\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & \cdots & x_{md} \end{pmatrix}$ $\boldsymbol{W(lon_i, lat_i)} = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{im} \end{pmatrix}$ and $\mathbf{y}^* = \begin{pmatrix} y_1 - n_1\hat{\pi}_1 \\ y_2 - n_2\hat{\pi}_2 \\ \vdots \\ y_m - n_m\hat{\pi}_m \end{pmatrix}$ with $\hat{\pi}_j = \frac{\exp(\sum_{k=0}^{d} \hat{\beta}_k(lon_i, lat_i)x_{jk})}{1+\exp(\sum_{k=0}^{d} \hat{\beta}_k(lon_i, lat_i)x_{jk})}, j = 1, 2, \ldots, m$. The first partial derivative is equal to zero to maximize the log-likelihood function, so

$$\frac{\partial L^*(\beta(lon_i, lat_i))}{\partial \beta_k(lon_i, lat_i)} = \mathbf{X}^T \boldsymbol{W(lon_i, lat_i) y^*} = 0. \quad (13)$$

The MLE approach was unable to produce an analytical solution due to the implicit nature of Equation (13). To generate parameter estimates, Fisher's Scoring algorithm, a numerical iteration process, was used. This method was selected because iteration can be used quickly and effectively to estimate parameter values when the information matrix and the gradient vector are known. The information matrix is also the inverse of the maximum likelihood estimate's asymptotic variance-covariance matrix.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{(t+1)}}(lon_i, lat_i) =& \hat{\boldsymbol{\beta}}_{\boldsymbol{(t)}}(lon_i, lat_i) I^{(-1)}(\hat{\boldsymbol{\beta}}_{\boldsymbol{(t)}}(lon_i, lat_i)) \\ & \boldsymbol{g}(\hat{\boldsymbol{\beta}}_{\boldsymbol{(t)}}(lon_i, lat_i)), \quad (14)\end{aligned}$$

where $\hat{\boldsymbol{\beta}}_{\boldsymbol{t}}(lon_i, lat_i)$ is the value of the parameter estimate for the t^{th} iteration at the i^{th} location, $\boldsymbol{g}(\hat{\boldsymbol{\beta}}_{\boldsymbol{(t)}}(lon_i, lat_i))$ is the value of the gradient vector at the t^{th} iteration at the i^{th} location, and $\boldsymbol{I}(\hat{\boldsymbol{\beta}}_{\boldsymbol{(t)}}(lon_i, lat_i))$ is the Information matrix at the i^{th} location for the t^{th} iteration. Elements in $\boldsymbol{g}(\hat{\boldsymbol{\beta}}_{\boldsymbol{(t)}}(lon_i, lat_i))$ are the initial partial derivatives of the log-likelihood function in (10) for each parameter as in (12) and elements in $\boldsymbol{I}(\hat{\boldsymbol{\beta}}_{\boldsymbol{(t)}}(lon_i, lat_i))$ are minus value of the expected value for the second differential of the log-likelihood function.

D. Hypothesis Testing

The parameters of the GWBLR model are available for simultaneous and partial testing. A simultaneous test is utilized to identify at least one predictor that significantly affects the response at the observation location. A partial test is employed to identify the specific predictor that has a significant effect on the response at each observation point.

The following is the simultaneous hypothesis testing:

H_0 : $\beta_1(lon_i, lat_i) = \beta_2(lon_i, lat_i) = \cdots = \beta_d(lon_i, lat_i) = 0$ for $i = 1, \ldots, m$

H_1 : There is at least one $\beta_k(lon_i, lat_i) \neq 0$, for $k = 1, \ldots, d$; and $i = 1, \ldots, m$

The likelihood ratio test (LRT) is applied to calculate the test statistic for this hypothesis [45]. This method compares the most extreme likelihood function corresponding to the parameter set under H_0 to the most extreme value of the likelihood function for the parameter set under the population. The likelihood function is equivalent to the log-likelihood function, and the log transformation is implemented to produce a sample approximation with a chi-square distribution.

$$\begin{aligned} l(\hat{\omega}_{GWBLR}) =& \sum_{i=1}^{n} \left\{ \log \binom{n_i}{y_i} + y_i \log \tilde{\pi}_i(lon_i, lat_i) \right. \\ & \left. + (n_i - y_i) \log(1 - \tilde{\pi}_i(lon_i, lat_i)) \right\} \quad (15)\end{aligned}$$

Next the set of parameters under the population is determined, which is $\hat{\Omega}_{GWBLR} = \{\hat{\beta}_0(lon_i, lat_i), \hat{\beta}_1(lon_i, lat_i), \ldots, \hat{\beta}_d(lon_i, lat_i); i = 1, \ldots, n\}$ and the maximum value of the likelihood function of the parameter set under the population is as follows

$$\begin{aligned} l(\hat{\Omega}_{GWBLR}) =& \sum_{i=1}^{m} \left\{ \log \binom{n_i}{y_i} + y_i \log \hat{\pi}_i(lon_i, lat_i) + \right. \\ & \left. (n_i - y_i) \log(1 - \hat{\pi}_i(lon_i, lat_i)) \right\}. \quad (16)\end{aligned}$$

As a result, the deviance value can be calculated using the formula below:

$$\begin{aligned} D =& -2 \frac{l(\hat{\omega}_{GWBLR})}{l(\hat{\Omega}_{GWBLR})} \\ =& 2 \sum_{i=1}^{n} \left\{ y_i \log \left(\frac{\hat{\pi}_i(lon_i, lat_i)}{\tilde{\pi}_i(lon_i, lat_i)} \right) + \right. \\ & \left. (n_i - y_i) \log \left(\frac{1 - \hat{\pi}_i(lon_i, lat_i)}{1 - \tilde{\pi}_i(lon_i, lat_i)} \right) \right\} \quad (17)\end{aligned}$$

If the number of successes based on the model set of parameters under the population is $\hat{y}_i(lon_i, lat_i) = n_i \hat{\pi}_i$ and the model under H_0 is $\tilde{y}_i(lon_i, lat_i) = n_i \tilde{\pi}_i$, the test statistic for the likelihood ratio test can be written as:

$$\begin{aligned} D =& 2 \sum_{i=1}^{n} \left\{ y_i \log \left(\frac{\hat{y}(lon_i, lat_i)}{\tilde{y}(lon_i, lat_i)} \right) + \right. \\ & \left. (n_i - y_i) \log \left(\frac{(n_i - \hat{y}(lon_i, lat_i))}{(n_i - \tilde{y}(lon_i, lat_i))} \right) \right\} \quad (18)\end{aligned}$$

Test statistic D in (17) has asymptotic property and chi-square distributed with the degrees of freedom equal to the number of effective parameters, which is $v = trace(I)$. The H_0 rejection region of the hypothesis testing for the α significance level was H_0 rejection if the value of $D > \lambda_{(\alpha,v)}$.

Following the simultaneous test, a partial test is performed. Testing the GWBLR model parameters partially is employed to determine the parameters that have a significant effect on the model. The partial hypothesis test for the β_k parameter with a certain value of k and i is

H_0 : $\beta_k(lon_i, lat_i) = 0$

H_1 : $\beta_k(lon_i, lat_i) \neq 0_k$;

for $i = 1, \ldots, n$ and $k = 1, \ldots, p$. The Wald test statistic is employed as the test statistic. Wald's test equals the parameter estimate to its estimated standard error as given by:

$$W = \frac{\hat{\beta}_k(lon_i, lat_i)}{SE(\hat{\beta}_k(lon_i, lat_i))} \quad (19)$$

with $SE(\hat{\beta}_k(lon_i, lat_i)) = \sqrt{Var(\hat{\beta}_k(lon_i, lat_i)}$. $Var(\hat{\beta}_k(lon_i, lat_i)$ is the k diagonal element of the inverse Fisher Information matrix at location (lon_i, lat_i). The test statistic in (19) is close to the standard normal distribution for a large enough sample size n. The critical area is to reject H_0 if the value is $|W| > Z_{\alpha/2}$.

E. Parameter Estimation Interpretation

The interpretation of the GWBLR model was obtained by exponentiating the logit (1), so the odds of success are as follows:

$$\hat{\pi} = \frac{e^{\hat{\beta}_0(lon_i, lat_i)} \left(e^{\hat{\beta}_1(lon_i, lat_i)}\right)^{x_{1i}} \cdots \left(e^{\hat{\beta}_k(lon_i, lat_i)}\right)^{x_{ki}}}{1 + e^{\hat{\beta}_0(lon_i, lat_i)} e^{\hat{\beta}_1(lon_i, lat_i)})^{x_{1i}} \cdots \left(e^{\hat{\beta}_k(lon_i, lat_i)}\right)^{x_{ki}}}. \quad (20)$$

When all independent variables are 0, then

$$\hat{\pi} = \frac{e^{\hat{\beta}_0(lon_i, lat_i)}}{1 + e^{\hat{\beta}_0(lon_i, lat_i)}} \quad (21)$$

According to what a result, if the intercept is negative, the probability of the observed occurrence is less than 0.5. If the intercept, on the other hand, has a positive sign, the probability of the observed occurrence is greater than 0.5.

In contrast to the intercept coefficient, the estimated parameter coefficient of each independent variable in the regression model can be interpreted as each increase in $\hat{\beta}_k$ units of variable X_k. In addition, it can also be interpreted that the negative coefficient reduces log odds and the positive coefficient improves log odds. The odds value is multiplied by $e^{\hat{\beta}_k(lon_i,lat_i)}$ for every 1 unit increase in x_{li}. If $\hat{\beta}_k(lon_i, lat_i) = 0$, then $e^{\hat{\beta}_k(lon_i,lat_i)} = 1$, and it means that the odds do not change when x changes.

F. Optimum bandwidth determination

The significance of weights in GWR cannot be overstated. The distance between the observation points determines the GWR weighting. The weighting matrix's function is to select a different parameter estimator at each point of observation. There are several methods for calculating the weighting value.

One of the locality-based weighting function developments is excluding observations that are more than a certain distance d from the location i^{th}. The following is the weighting function:

$$w_{ij} = \begin{cases} 1, \text{ if } d_{ij} < d \\ 0, \; d_{ij} \text{ otherwise} \end{cases} \tag{22}$$

where d_{ij} is the Euclidean distance between the i and j observation locations, as expressed by the formula

$$d_{ij} = \sqrt{(lon_i - lon_j))^2 + (lat_i - lat_j))^2} \tag{23}$$

The use of this function simplifies model formation because only a portion of the data is used to form the model, i.e. the distance from the observation location to i is less than d. This function, however, has a discontinuity problem, which causes the resulting parameters to change dramatically when the observation location changes.

[10] provides a solution to the weighting discontinuity problem by forming w_{ij} as a continuous function of d_{ij}. The Gaussian weighting function, which is additionally recognized as the fixed kernel Gaussian function, is a common function that is stated as follows:

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \tag{24}$$

where b is a non-negative parameter called bandwidth. Bandwidth is the balance controller between the smoothness of the function and its appropriateness to the data. The weighting value of the data will approach 1 if the distance is close or coincides and will get smaller so it approaches zero if the distance is further away.

To find the best model with the best bandwidth, models with different bandwidths must be evaluated. To find the best bandwidth and model for GWBLR, the Cross Validation (CV) approach was adopted. The CV value is employed to determine whether a model specification fits the data and to provide an alternative method for testing the relevance of the spatial relationship in the spatial regression model without making assumptions. For GWBLR, the model with the lowest CV value is the best. The CV value was calculated using the following formula:

$$CV(b) = \sum_{i=1}^{n}(y_i - \hat{y}_{\neq i}(b))^2 \tag{25}$$

where $\hat{y_i}(b_i)$ is the estimated value of y_i which administers bandwidth b_i and the variable in which the observations at location (lon_i, lat_i) are removed from the estimation process.

III. Simulation and Application

A. Simulation

This section investigates the use of simulation data to determine the optimum bandwidth and estimate parameters based on the formulas in section 2. As a result, describing a computer algorithm for determining optimal bandwidth and estimating parameters is critical. The optimum bandwidth and parameter estimation algorithm are adopted from package **spgwr** in **R** [46], while the Fisher's Scoring algorithm refers to [47]. The algorithms are presented below:

Algorithm 1 Parameter Estimation Algorithm

Require: $N_{n\times 1}$, $y_{n\times 1}$, $x_{n\times p}$, $(lon, lat)_{n\times 2}$, bdw
Ensure: $\hat{\beta}_{n\times p}$
  Function $binom.gwr(\hat{\beta}, N, \pi, x, y, W, V)$
  $H = x^T W V x$
  $g = x^T W(y - (N\pi))$
  $\hat{\beta}_{t+1} = \hat{\beta}_t + H^{-1} g \hat{\beta}_t$
  EndFunction
  Calculating the weighting matrix
  $d_{ij} = \sqrt{(lon_i - lon_j))^2 + (lat_i - lat_j))^2}$ {Euclidean Distance}
  $w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{bdw}\right)^2\right]$ {Gaussian weighting function}
  for $i = 1$ to n **do**
    $W = diag(w[, i])$
    $\hat{\beta}_0 = (0, \ldots, 0)$
    repeat
      $\hat{\beta} = \hat{\beta}_0$
      $\pi = \frac{\exp(x\hat{\beta})}{1+\exp(x\hat{\beta})}$
      $V = diag(N\pi(1 - \pi))$
      $\hat{\beta}_b = binom.gwr(\hat{\beta}, N, \pi, x, y, W, V)$
      $\hat{\beta} = \hat{\beta}_b$
    until $\sum((\hat{\beta}_b - \hat{\beta})^2) < 0.0000000001$
    $\hat{\beta}[i,] = \hat{\beta}$ {Coefficient Estimation in location i^{th}}
  end for

A simulation study was conducted to evaluate the feasibility of the proposed method. As a consequence, before applying the GWBLR model, it is important to generate data. The generated data in this simulation referred to [48] in order to evaluate the performance of GWR and traditional geostatistical models. The generated data are made up of three variables, two independent variables and one dependent variable. The independent variables are X_1, X_2 and dependent variable is Y. Each variable has coefficient functions built into the simulation. Variables X_1 and X_2 were returned by $Uniform(-5, 5)$ and $Uniform(70, 90)$, respectively. The pairs of spatial coordinate (lon_i, lat_i) for a sample of 625 points of data are given below:

$$(u_i, v_i) = (0.5 \times ((i-1) \bmod 25), 0.5 \times ((i-1) \text{ fl } 25)), \tag{26}$$

Algorithm 2 Optimum Bandwidth Algorithm

Require: $N_{n\times 1}$, $y_{n\times 1}$, $x_{n\times p}$, $(lon, lat)_{n\times 2}$
Ensure: bdw
  Function $CV.score(bdw, \hat{\beta}, N, \pi, (lon, lat))$
  $d_{ij} = \sqrt{(lon_i - lon_j))^2 + (lat_i - lat_j))^2}$ {Euclidean Distance}
  $w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{bdw}\right)^2\right]$ {Gaussian weighting function}
  for $i = 1$ to n **do**
    $wt.i = wt[, i]$
    $wt_{noi} = diag(wt.i[-i])$ {Remove element-i^th of $wt.i$}
    $y_{noi} = y[-i]$ {Remove element-i^th of y}
    $x_{noi} = x[-i,]$ {Remove row-i^th of x}
    $N_{noi} = N[-i]$ {Remove row-i^th of N}
    $\hat{\beta}_0 = (0, \ldots, 0)$
    repeat
      $\hat{\beta} = \hat{\beta}_0$
      $\pi = \frac{\exp(x.noi\hat{\beta})}{1+\exp(x_{noi}\hat{\beta})}$
      $V_{noi} = diag(N_{noi}\pi(1-\pi))$
      $\hat{\beta}_b = binom.gwr(\hat{\beta}, N_{noi}, \pi, x_{noi}, y_{noi}, wt_{noi}, V_{noi})$
      $\hat{\beta} = \hat{\beta}_b$
    until $\sum((\hat{\beta}_b - \hat{\beta})^2) < 0.0000000001$
    $\hat{\pi} = \frac{\exp(x_{noi}\hat{\beta})}{1+\exp(x_{noi}\hat{\beta})}$
    $\hat{y}_{noi} = N_{noi}\hat{\pi}$
    $CV[i] = y_{noi} - \hat{y}_{noi}$
  end for
  $cv.score = \sum(CV^2)$
  EndFunction
  $bb = \begin{pmatrix}\min(lon, lat)\\ \max(lon, lat)\end{pmatrix}$
  $up = \sqrt{(bb[1,2] - bb[1,1])^2 + (bb[2,2] - bb[2,1])^2)}$
  $low = up/100$
  $band = (low, \ldots, up)_{100\times 1}$
  for i=1 to n=100 **do**
    $CV[i] = CV.score(band[i], \hat{\beta}, N, \pi, (lon, lat))$
  end for
  $opt.bdw$ {Bandwidth value that minimizes CV}

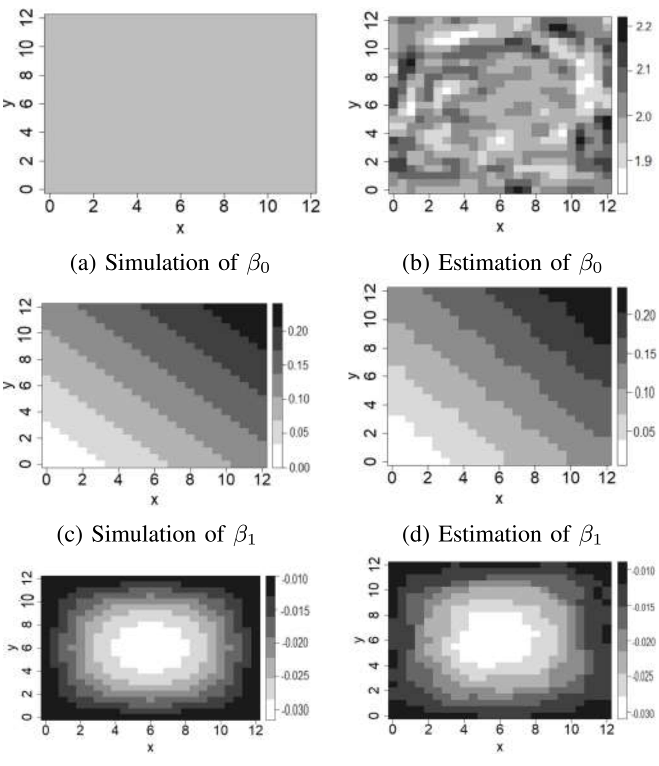
for $i = 1, 2, \ldots, 625$, on which $(a \bmod b)$ is the remaining value of a divided by b and $fl(a)$ is a's floor. Then, as functions of the coordinate pairs (lon, lat), the coefficients $\beta_0(lon_i, lat_i)$, $\beta_1(lon_i, lat_i)$, and $\beta_2(lon_i, lat_i)$ are formulated as follows:

$$\beta_0(lon, lat) = 2$$
$$\beta_1(lon, lat) = sign(lon)0.01lon + sign(lat)0.01lat,$$
$$\beta_2(lon, lat) = -\frac{1 + \left(\frac{(36-(6-lon)^2)(36-(6-lat)^2)}{600}\right)}{100}$$

For each location (lon_i, lat_i), the dependent variable Y is simulated based on a binomial distribution with probability P and total occurrence Y_{total}. The total occurrence at each observation point Y_{total} is generated from $Uniform(500, 1500)$ and the probability P is generated from the logit function as follows:

$$P = \frac{e^{(\beta_0(lon,lat)+\beta_1(lon,lat)x_1+\beta_2(lon,lat)X_2)}}{1 + e^{(\beta_0(lon,lat)+\beta_1(lon,lat)X_1+\beta_2(lon,lat)X_2)}}. \quad (27)$$

The GWBLR model was applied to these data to determine estimated parameters b_0, b_1 and b_2. The geographic weighting is determined before the parameter coefficients are estimated using the fixed kernel Gaussian function. The optimum bandwidth is 0.48 based on the smallest CV value. With a geographical weighting of 0.48, the variables X_1, X_2 and Y were generated by 100 data sets. Each generation data set was used to estimate parameter coefficients b_0, b_1 and b_2, hence, 100 estimated coefficients were obtained at 625 observation points. The average estimate at each sample point is utilized to estimate the GWBLR model of generation data parameter coefficients. The results of the estimated parameter coefficients from the simulation data are compared below:



(a) Simulation of β_0 (b) Estimation of β_0
(c) Simulation of β_1 (d) Estimation of β_1
(e) Simulation of β_2 (f) Estimation of β_2

Fig. 1: Simulation and Estimation of Parameter Coefficients

The simulation result map of β_0 intercept coefficient and its estimation is presented in Figure 1a and 1b. These two images demonstrate slightly different results in which β_0 is simulated with a constant of value 2, whereas the estimated intercept coefficient is varied and lies in the $(1.7, 2.2)$ interval. The estimates for β_1 and β_2 in Figures 1d and 1f have nearly the same shape as the actual simulated values. If the correlation between simulation data and estimation was calculated, for β_1 the Pearson correlation was 0.989 ($p < 0.0000$) and β_2 is 0.999 ($p < 0.0000$). It illustrates a high consistency between the simulation data and its estimation using the GWBLR model.

B. Application

The proposed GWBLR model was implemented in R Program and utilized to model Covid-19 cases for 34 Indonesian provinces in 2020. Based on the availability and completeness of the data, the GWBLR application in this study aims to investigate the spatial variation of the variables temperature, humidity, rainfall, and sunlight on the quantity of positive confirmed Covid-19 cases. The statistics for

Covid-19 instances come from the Indonesia Health Profile 2020, and the meteorological data come from Statistics Indonesia 2021.

First, the optimal spatial distance bandwidth is determined using the CV value before estimating the parameters. The optimal bandwidth is determined to be 9.184 with a CV of 11817676823. The CV value profile is depicted in the figure below:

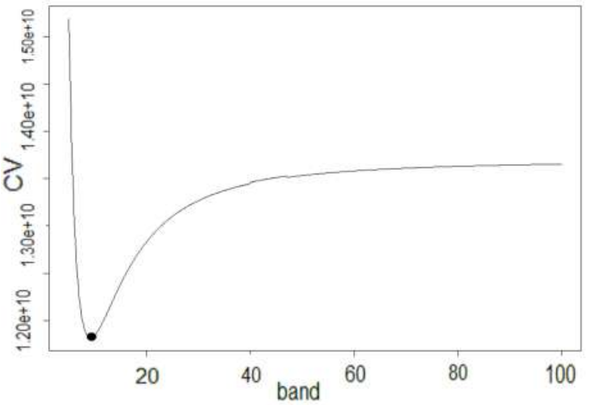


Fig. 2: Cross-Validation Values for the Fixed Kernel Gaussian Function Applied to Covid-19 Data. The black dot indicates the optimum bandwidth with the minimum Cross-Validation Value.

A weighting matrix is applied to form the GWBLR model based on the optimal bandwidth value. The estimated model parameters formed with MLE have the following minimum, maximum, and median values:

TABLE I: Statistics of Estimate Parameters from the GWBLR model

Variables	Min.	1st	Median	3rd	Max.
Intercept	-5.2343	-3.4475	-2.1211	-0.6893	2.1963
Temperature	-0.0929	-0.0406	-0.0170	0.0093	0.0279
Humidity	-0.0019	0.0170	0.0233	0.0279	0.0419
Precipitation	-0.0002	-0.0002	-0.0002	-0.0001	-0.0001
Sunshine	-0.0074	-0.0068	-0.0040	-0.0001	0.0053

Estimates of temperature, humidity, rainfall, and sunshine duration at the provincial level represent the relationship that occurs between verified Covid-19 cases and these parameters in 34 provinces. Table I depicts the relationship that occurs between Covid-19 cases and each independent variable in each province. Each province's variation in the coefficient estimate is a local estimate that can describe the different effects of independent variables on Covid-19 cases.

After obtaining the GWBLR model's parameter estimator, the next step is to perform simultaneous and partial testing. The statistical likelihood ratio test is calculated for simultaneous testing using equation 18, and the value is 1184697, which is greater than the value of $\chi_{(5\%,8.0258)} = 15.5441$. As a result, H_0 is rejected, and it is feasible to conclude that the weather factor is important in the rise of Covid-19 cases in Indonesia.

The hypothesis testing that follows is a preliminary test that seeks to evaluate the influence of each weather factor variable on the addition of Covid-19 instances. The Wald test statistics p-value for 34 provinces is as follows:

TABLE II: *p-value* of Estimate Parameter from GWBLR model

Province	β_1	β_2	β_3	β_4
Aceh	0.0000	0.0172	0.0000	0.0000
West_Sumatra	0.0000	0.0000	0.0000	0.0000
South_Sumatra	0.0000	0.0000	0.0000	0.0000
North_Sumatra	0.0000	0.0000	0.0000	0.0000
Jambi	0.0000	0.0000	0.0000	0.0000
Lampung	0.0000	0.0000	0.0000	0.0000
Riau	0.0000	0.0000	0.0000	0.0000
Bengkulu	0.0000	0.0000	0.0000	0.0000
Riau_Islands	0.0000	0.0000	0.0000	0.0000
Bangka_Islands	0.0000	0.0000	0.0000	0.0000
West_Java	0.0000	0.0000	0.0000	0.0000
DKI_Jakarta	0.0000	0.0000	0.0000	0.0000
Central_Java	0.0000	0.0000	0.0000	0.0000
East_Java	0.0000	0.0000	0.0000	0.0000
DI_Yogyakarta	0.0000	0.0000	0.0000	0.0000
Bali	0.0000	0.0000	0.0000	0.0000
Banten	0.0000	0.0000	0.0000	0.0000
East_Nusa_Tenggara	0.0000	0.0000	0.0000	0.0217
West_Nusa_Tenggara	0.0007	0.0000	0.0000	0.0000
South_Kalimantan	0.0000	0.0000	0.0000	0.0000
West_Kalimantan	0.0000	0.0000	0.0000	0.0000
Central_Kalimantan	0.0000	0.0000	0.0000	0.0000
North_Kalimantan	0.0034	0.0000	0.0000	0.0000
East_Kalimantan	0.0375	0.0000	0.0000	0.0000
North_Maluku	0.0000	0.0000	0.0000	0.0000
Maluku	0.0000	0.0000	0.0000	0.0000
Central_Sulawesi	0.0000	0.0000	0.0000	0.0000
North_Sulawesi	0.0000	0.0000	0.0000	0.0000
Southeast_Sulawesi	0.0000	0.0000	0.0000	0.0000
South_Sulawesi	0.0000	0.0000	0.0000	0.1852
West_Sulawesi	0.0000	0.0000	0.0000	0.3868
Gorontalo	0.0000	0.0000	0.0000	0.0000
Papua	0.0173	0.0000	0.0000	0.0410
West_Papua	0.0000	0.0000	0.0000	0.0000

According to Table II, the $p-value$ of the fourth variable in the provinces of South and West Sulawesi is greater than $\alpha = 5\%$. As a result, it is possible to conclude that the variable duration of solar radiation has no effect on the addition of Covid-19 cases in the province. With the exception of the variable duration of irradiation in the two provinces, all $p-values$ are less than $\alpha = 5\%$, implying that each of the other variables has a substantial role in the addition of Covid-19 cases.

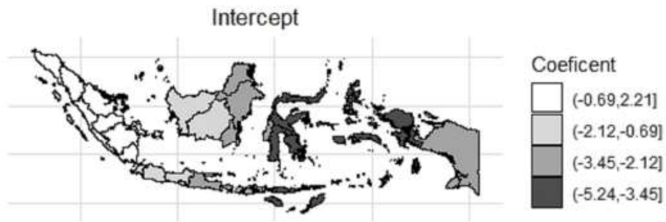


Fig. 3: Intercept Coefficient Estimations

Estimated intercept parameter b_0 from -5.2343 to the third quartile has a negative value, which was -0.6893. If all of the independent variables are equal, the probability of a COVID-19 case is less than 0.5. Meanwhile, an estimated positive value of b_0 exists in the fourth quartile, indicating that the chance of a COVID-19 case is greater than 0.5 if all independent variables are 0. Aceh, West Sumatra, North Sumatra, Jambi, Riau and Bengkulu are the provinces with positive intercept values on Sumatra Island. Figure 3 depicts the estimated b_0 coefficients that differ from the GWLR model parameters.

The provinces of Sumatra Island have a temperature coefficient value in the first quartile and a negative value, indicating that the higher the air temperature, the lower the chance of Covid-19 cases occurring in the region. Meanwhile, except for Papua, provinces in eastern Indonesia have a positive estimate of the coefficient of temperature parameter. This could imply that an increase in air temperature in the eastern region increases the likelihood of an increase in Covid-19 cases. Figure 4 portrays the distribution of the estimated value of the temperature coefficient:

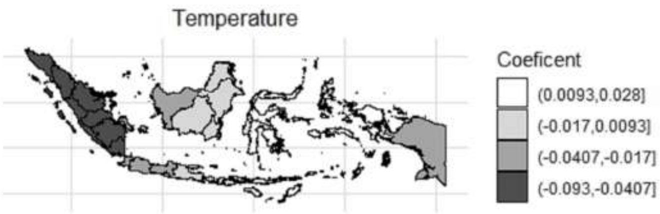


Fig. 4: Average Temperature Coefficient Estimations

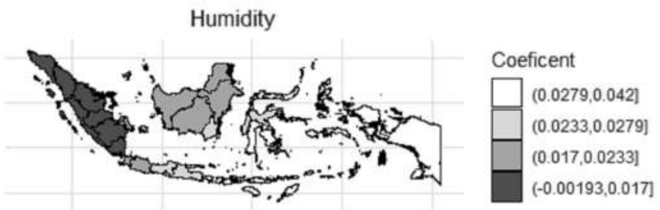


Fig. 5: Humidity Coefficient Estimations

Figure 5 demonstrates the mapping of the estimated coefficient of humidity parameters. Except for Aceh, all provinces in Indonesia have positive coefficient estimates for the Covid-19 cases. This demonstrates that higher humidity levels can increase the probability of Covid-19 by more than 50%. When compared to areas in the western part of Indonesia, provinces in the east have a higher chance of increasing cases if the weather is humid. Except for the provinces of Maluku, North Maluku, West Papua, and Sulawesi, this corresponds to the estimated value of the coefficient on the duration of solar radiation, which is negative. The negative estimation of the sunshine parameter indicates that the longer the duration of solar irradiation, the lower the chance of a rise in the quantity of Covid-19 cases. The mapping of the estimated coefficient of the sunshine parameter parameters is presented in Figure 6

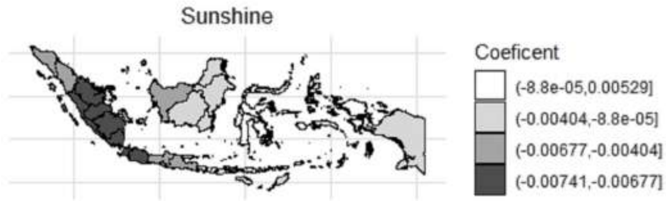


Fig. 6: Duration of Sunshine Coefficient Estimations

Figure 7 depicts a mapping of the estimated coefficient distribution for the rainfall parameter. According to the GWBLR model, all estimates are negative in the range of -0.000221 and -0.000087 or in the interval of increasing probability of $(0.49994, 0.49997)$. With this relatively small increase in the opportunity interval, it is possible to conclude that the effect of the rainfall variable on the rise in Covid-19 cases is nearly identical across all provinces.

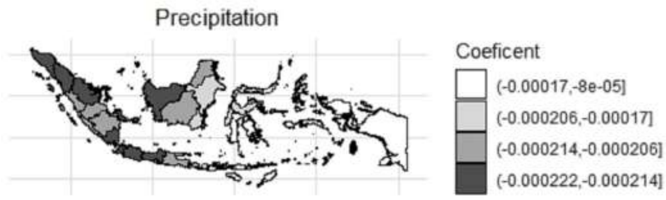


Fig. 7: Precipitation Coefficient Estimations

Furthermore, the GWBLR model and the BLR were compared to determine which model is better suited to modeling Covid-19 cases in Indonesia. Mean Square Error (MSE), R-Squared (R^2), Mean Absolute Percentage Error (MAPE) and Akaike Information Criterion (AIC) are the criteria used. The best model has the highest R^2 value as well as the lowest MSE, AIC, and MAPE. The results of the comparison between the GWBLR and BLR models are shown in the table below:

TABLE III: Comparison of BLR and GWBLR Model

Criterion	BLR	GWBLR
R^2	0.9715	0.9797
MSE	36363296	25838449
AIC	599.9084	588.2907
$MAPE$	30.75%	26.73%

Based on Table III, it is possible to conclude that the GWBLR model outperforms the BLR method using the R^2, MSE, AIC and MAPE criteria. The GWBLR model has been displayed to increase the value of R^2 while decreasing the value of MSE, AIC and MAPE. The GWBLR model can increase the value of R^2 by 0.28% over the BLR model. The MAPE values for both models fall into the reasonable category. However, the MAPE value for the GWBLR model is smaller than the BLR model which indicates the GWBLR model is also more accurate in modelling data than BLR. Comparison charts for visualizing fitted values from the BLR and GWBLR models are depicted in Figure 8. According to Figure 8, the model estimation using the GWBLR method is better suited for calculating positive confirmed Covid-19 cases according to weather-related factors in each region of Indonesia. This is demonstrated by the graph above, indicating in which the GWBLR model is considerably more precise than the BLR model.

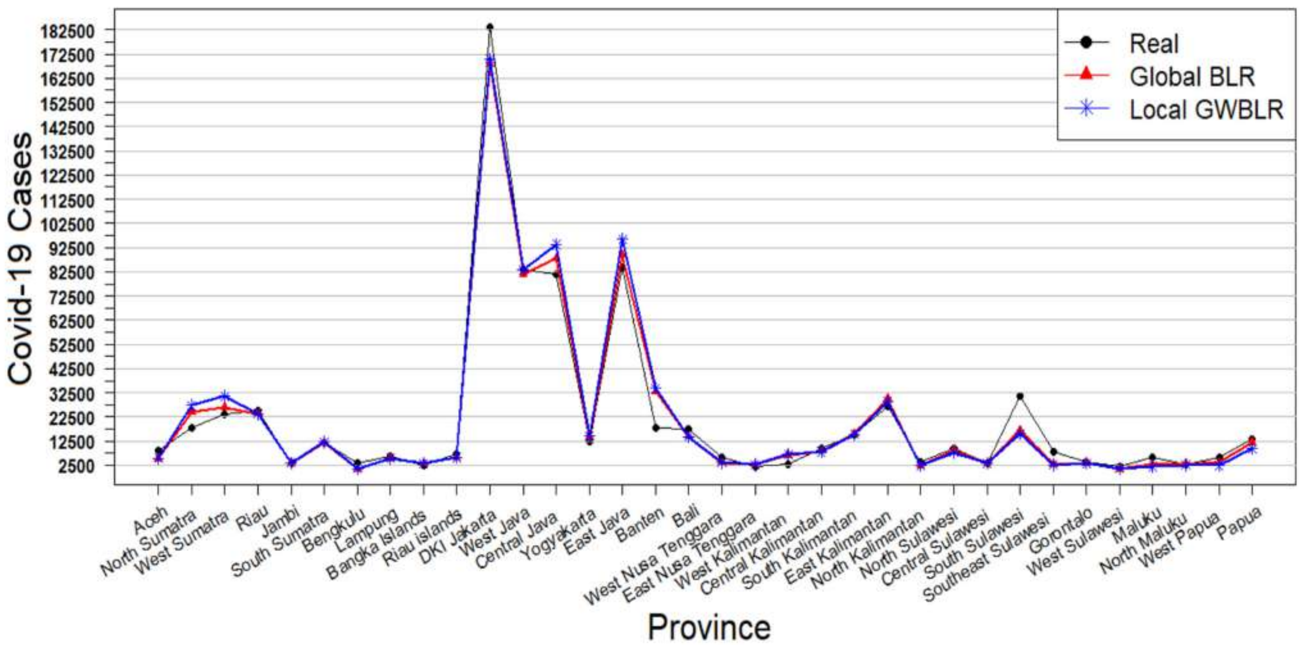


Fig. 8: Comparison of the fitted value of the Global BLR and Local GWBLR model to the observed value

IV. Conclusions

For the GWBLR model with a binomial distribution, the GWLR model with the assumption that the dependent variable has a Bernouli distribution can be developed. MLE is used to estimate parameters by assigning geographical weights to each observation in a different location. The GWBLR model parameter hypothesis can be tested simultaneously using the Likelihood Ratio Test Statistic and partially tested using the Wald Test. The weighting function used is a Gaussian weighting function with the same bandwidth value at each location and the optimum bandwidth determined by the CV value. The parameter estimation of the GWBLR model may demonstrate high consistency to the generated data based on the simulation study. The GWBLR model and parameter coefficient estimations can describe the variety of spatial relationships with regard to independent and dependent variables.

According to the GWBLR model, the spatial relationship between temperature, humidity, rainfall, and sunlight has substantial impact on the quantity of Covid-19 cases in Indonesia in 2020. Local parameter estimation causes parameter estimators to differ across provinces. Parameter estimates in neighboring provinces tend to be in the same quartiles. Aside from the weather, the most critical determinants in Covid-19 transmission are social and demographic factors. As a result, this research can be expanded to examine the relation among social and demographic characteristics and the prevalence of Covid-19 cases in Indonesia.

References

[1] D. Collett, *Modeling Binary Data*. Chapman & Hall, 2014.

[2] J. M. Hilbe, *Logistic regression models*, ser. Texts in statistical science. CRC Press, 2009.

[3] D. B. Hall, "Zero-inflated poisson and binomial regression with random effects: A case study," *Biometrics*, vol. 56, no. 4, pp. 1030–1039, 2000. [Online]. Available: http://www.jstor.org/stable/2677034

[4] S. P. Jacups and A. Cheng, "The epidemiology of community acquired bacteremic pneumonia, due to streptococcus pneumoniae, in the top end of the Northern Territory, Australia—over 22 years," *Vaccine*, vol. 29, no. 33, pp. 5386–5392, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0264410X11008292

[5] A. C. A. Clements, N. J. S. Lwambo, L. Blair, U. Nyandindi, G. Kaatano, S. Kinung'hi, J. P. Webster, A. Fenwick, and S. Brooker, "Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania," *Tropical Medicine and International Health*, vol. 11(4), 2006.

[6] B. P. Weaver and M. S. Hamada, "A bayesian approach to the analysis of industrial experiments: An illustration with binomial count data," *Quality Engineering*, vol. 20, no. 3, pp. 269–280, 2008. [Online]. Available: https://doi.org/10.1080/08982110701810451

[7] A. J. Dobson and A. G. Barnett, *An Introduction to Generalized Linear Models, Fourth Edition*. CRC Press, 2018.

[8] I. Jaya and A. Chadidjah, "Spatial autoregressive in ecological studies: A comparison of the sar and car models," *Engineering Letters*, vol. 29, no. 1, pp. 207–212, 2021.

[9] K. Du, G. Liu, and G. Gu, "Accelerating monte carlo method for pricing multi-asset options under stochastic volatility models," *IAENG International Journal of Applied Mathematics*, vol. 44, no. 2, pp. 62–70, 2014.

[10] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically weighted regression: A method for exploring spatial nonstationarity," *Geographical Analysis*, vol. 28, no. 4, pp. 281–298, 1996.

[11] P. M. Atkinson, S. E. German, D. A. Sear, and M. J. Clark, "Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression," *Geographical Analysis*, vol. 35, no. 1, pp. 58–82, 2003.

[12] L. Wu, F. Deng, Z. Xie, S. Hu, S. Shen, J. Shi, and D. Liu, "Spatial analysis of severe fever with thrombocytopenia syndrome virus in China using a geographically weighted logistic regression model," *International Journal of Environmental Research and Public Health*, vol. 13, no. 11, 2016.

[13] Y. B. Zhou, Q. X. Wang, M. X. Yang, Y. H. Gong, Y. Yang, S. J. Nie, S. Liang, L. Nan, A. Coatsworth, A. H. Yang, and et al., "Geographical variations of risk factors associated with hcv infection in drug users in southwestern china," *Epidemiology and Infection*, vol. 144, no. 6, p. 1291–1300, 2016.

[14] M. C. Wimberly, A. D. Baer, and M. J. Yabsley, "Enhanced spatial models for predicting the geographic distributions of tick-borne pathogens," *Int J Health Geogr*, vol. 7:15, 2008.

[15] T. Nakaya, A. Fotheringham, C. Brunsdon, and M. Charlton, "Geographically weighted poisson regression for disease association mapping," *Statistics in Medicine*, vol. 24, pp. 2695–2717, 2005.

[16] A. R. Da Silva and T. C. V. Rodrigues, "Geographically weighted

negative binomial regression—incorporating overdispersion," *Statistics and Computing*, vol. 24, p. 769–783, 2014.
[17] J. Ssempiira, J. Kissa, B. Nambuusi, E. Mukooyo, J. Opigo, F. Makumbi, S. Kasasa, and P. Vounatsou, "Interactions between climatic changes and intervention effects on malaria spatio-temporal dynamics in Uganda," *Parasite Epidemiology and Control*, vol. 3(3), 2018.
[18] M. Cattelan and C. Varin, "Marginal logistic regression for spatially clustered binary data," *Applied Statistics*, vol. 67(4), 2018.
[19] H. M. Khormi and L. Kumar, "Modeling dengue fever risk based on socioeconomic parameters, nationality and age groups: Gis and remote sensing based case study," *Sci Total Environ*, vol. 409, no. 22, pp. 4713–4719, 2011.
[20] E. Delmelle, M. Hagenlocher, S. Kienberger, and I. Casas, "A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia," *Acta Tropica*, vol. 164, pp. 169–176, 2016.
[21] L. T. Smith, L. E. O. C. Aragão, C. E. Sabel, and T. Nakaya, "Drought impacts on children's respiratory health in the Brazilian Amazon," *Sci Rep*, vol. 4, p. 3726, 2014.
[22] E. M. Y. Cheng, P. M. Atkinson, and A. K. Shahani, "Elucidating the spatially varying relation between cervical cancer and socio-economic conditions in England," *Int J Health Geogr*, vol. 10, no. 51, 2011.
[23] B. Kauhl, J. Heil, C. J. P. A. Hoebe, J. Schweikart, T. Krafft, and N. H. T. M. Dukers-Muijrers, "The spatial distribution of hepatitis c virus infections and associated determinants—an application of a geographically weighted poisson regression for evidence-based screening interventions in hotspots," *PLOS ONE*, vol. 10, no. 9, pp. 1–19, 09 2015.
[24] B. Weinstein, A. R. da Silva, D. E. Kouzoukas, T. Bose, G.-J. Kim, P. A. Correa, S. Pondugula, Y. Lee, J. Kim, and D. O. Carpenter, "Precision mapping of covid-19 vulnerable locales by epidemiological and socioeconomic risk factors, developed using South Korean data," *International Journal of Environmental Research and Public Health*, vol. 18(2), p. 604, 2021.
[25] R. Fitriani and I. G. N. M. Jaya, "Spatial modeling of confirmed covid-19 pandemic in East Java province by geographically weighted negative binomial regression," *Commun. Math. Biol. Neurosci.*, 2020, article ID 58.
[26] E. M. Lamlili, M. Derouich, and A. Boutayeb, "A siar model of covid-19 with control," *Engineering Letters*, vol. 30, no. 1, pp. 98–107, 2022.
[27] F. Zuhairoh and D. Rosadi, "Real-time forecasting of the covid-19 epidemic using the richards model in South Sulawesi, Indonesia," *Indonesian Journal of Science and Technology*, vol. 5(3), pp. 182–189, 2020.
[28] G. Darmawan, D. Rosadi, and B. N. Ruchjana, "Daily forecast for covid-19 during ramadhan by singular spectrum analysis in Indonesia," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 12, no. 6, pp. 818–825, 2020.
[29] Z. E. Rasjida, R. Setiawana, and A. Effendi, "A comparison: Prediction of death and infected covid-19 cases in Indonesia using time series smoothing and lstm neural network," *Procedia Computer Science*, vol. 179, p. 982–988, 2021.
[30] R. A. A. Viadinugroho and D. Rosadi, "A comparative study of lstm, bi-lstm, and dbi-lstm network model in forecasting covid-19 new cases and new deaths in Indonesia," *Presented in the 3rd ISSTEC*, 2021.
[31] F. Zuhairoh, D. Rosadi, and A. R. Effendie, "Multi-state discrete-time markov chain svirs model on the spread of covid-19," *Engineering Letters*, vol. 30(2), pp. 598–608, 2022.
[32] G. Darmawan, D. Rosadi, B. Ruchjana, and Hermansah, "Autocorrelation analysis of COVID-19 based on hijri calendar," *Journal of Physics: Conference Series*, vol. 1918, no. 4, p. 042004, jun 2021.
[33] R. Caraka, Y. Lee, R. Kurniawan, R. Herliansyah, P. Kaban, B. Nasution, P. Gio, R. Chen, T. Toharudin, and B. Pardamean, "Impact of covid-19 large scale restriction on environment and economy in Indonesia," *Global Journal of Environmental Science and Management*, vol. 6, no. Special Issue (Covid-19), pp. 65–84, 2020.
[34] D. K. Sari, R. Amelia, R. Dharmajaya, L. M. Sari, and N. K. Fitri, "Positive correlation between general public knowledge and attitudes regarding covid-19 outbreak one month after first cases reported in Indonesia," *Journal of Community Health*, vol. 46(1), pp. 182–189, 2020.
[35] D. Yuniarti, D. Rosadi, and Abdurakhman, "Inflation of Indonesia during the COVID-19 pandemic," *Journal of Physics: Conference Series*, vol. 1821, no. 1, p. 012039, mar 2021.
[36] R. A. A. Viadinugroho and D. Rosadi, "Long short-term memory neural network model for time series forecasting: Case study of forecasting IHSG during covid-19 outbreak," *Journal of Physics: Conference Series*, vol. 1863, no. 1, p. 012016, mar 2021.
[37] R. Subekti, Abdurakhman, and D. Rosadi, "Toward the black–litterman with shariah-compliant asset pricing model: a case study on the Indonesian stock market during the covid-19 pandemic," *International Journal of Islamic and Middle Eastern Finance and Management*, 2022.
[38] R. Tosepu, J. Gunawan, D. S. Effendy, L. O. A. I. Ahmad, H. Lestari, H. Bahar, and P. Asfian, "Correlation between weather and covid-19 pandemic in Jakarta, Indonesia," *Science of The Total Environment*, vol. 725, p. 138436, 2020.
[39] M. M. Menebo, "Temperature and precipitation associate with covid-19 new daily cases: A correlation study between weather and covid-19 pandemic in Oslo, Norway," *Science of The Total Environment*, vol. 737, p. 139659, 2020.
[40] A. Mollalo, B. Vahedi, and K. M. Rivera, "Gis-based spatial modeling of covid-19 incidence rate in the Continental United States," *Science of The Total Environment*, vol. 728, p. 138884, 2020.
[41] D. K. Rosario, Y. S. Mutz, P. C. Bernardes, and C. A. Conte-Junior, "Relationship between covid-19 and weather: Case study in a tropical country," *International Journal of Hygiene and Environmental Health*, vol. 229, p. 113587, 2020.
[42] J. Wang, K. Tang, K. Feng, X. Lin, W. Lv, K. Chen, and F. Wang, "Impact of temperature and relative humidity on the transmission of covid-19: a modelling study in China and the United States," *BMJ Open*, vol. 11, no. 2, 2021.
[43] M. F. F. Sobral, G. B. Duarte, A. I. G. d. P. Sobral, M. L. M. Marinho, and A. d. S. Melo, "Association between climate variables and global transmission of sars-cov-2," *Science of The Total Environment*, vol. 729, p. 138997, 2020.
[44] M. Ganslmeier, D. Furceri, and J. D. Ostry, "The impact of weather on covid-19 pandemic," *Scientific Report*, vol. 11, p. 22027, 2021.
[45] M. Fathurahman, Purhadi, Sutikno, and V. Ratnasari, "Geographically weighted multivariate logistic regression model and its application," *Abstract and Applied Analysis*, vol. 2020, 2020.
[46] R. Bivand and D. Yu, "Geographically weighted regression," https://github.com/cran/spgwr.git, 2022.
[47] R. Ehlers, "Maximum likelihood estimation procedures for categorical data," Master's thesis, 2002, available from https://repository.up.ac.za/bitstream/handle/2263/26533/Complete.pdf.
[48] P. Harris, A. S. Fotheringham, R. Crespo, and M. Charlton, "The use of geographically weighted regression for spatial prediction: An evaluation of models using simulated data sets," *Mathematical Geosciences*, vol. 42, p. 657–680, 2010.

Pepi Novianti is studying in the Doctoral Mathematics Programme at Universitas Gadjah Mada, Indonesia. She is also a staff in the Department of Statistics at the University of Bengkulu, Indonesia. She completed a Bachelor of Mathematics from the University of Bengkulu in 2007 and a Master of Statistics from the Bogor Agricultural Institute in 2010.

Gunardi is an Associate Professor and Head of the Statistics Laboratory at Universitas Gadjah Mada (UGM), Faculty of Mathematics and Sciences, the Department of Mathematics. He received his undergraduate degree from Statistics at Universitas Gadjah Mada (UGM) in 1989. He graduated the Master degree in 1995 and completed Ph.D. degrees in Mathematics at Universitas Gadjah Mada (UGM) in 2007.

Dedi Rosadi is a professor at Universitas Gadjah Mada (UGM), Faculty of Mathematics and Sciences, the Department of Mathematics. He received his degree from the Statistics Program at Universitas Gadjah Mada in February 1996 and began working as a lecturer at UGM. In August 1997, he began studying for a Master of Science at the University of Twente with minor is Stochastic Modelling and Operation Research , which he finished in June 1999. He was a doctoral student in the Institute of Econometrics and Operation Research at Vienna University of Technology (TU Wien) in Austria from September 2001 to September 2004.