Available online at http://scik.org
J. Math. Comput. Sci. 2022, 12:164
https://doi.org/10.28919/jmcs/7427
ISSN: 1927-5307

# GEOGRAPHICALLY AND TEMPORALLY WEIGHTED REGRESSION MODELING IN STATISTICAL DOWNSCALING MODELING FOR THE ESTIMATION OF MONTHLY RAINFALL

AAN KARDIANA[1], AJI HAMIM WIGENA[2,*], ANIK DJURAIDAH[2], AGUS MOHAMAD SOLEH[2]

[1]Department of Informatics Engineering, YARSI University, Jakarta 10510, Indonesia

[2]Department of Statistics, IPB University, Bogor 16680, Indonesia



**Abstract:** Rainfall estimation was carried out using various Statistical Downscaling (SD) models namely Projection Pursuit, Quantile Regression, Multiple Linear Regression, Partial Least Squares Regression, Clustered Linear Regression and Two-Stage Modeling, and Clusterwise Regression. Global regression cannot handle the relationship between response variables and predictor variables in data containing spatial and temporal variability. The Geographically and Temporally Weighted Regression (GTWR) model can be used to overcome this. This study will perform SD modeling for the estimation of monthly rainfall using the Weighted Least Squares method. The response variables are monthly rainfall data from 35 stations in West Java Province from January 1983 to December 2012 and the predictor variables are temperature and monthly precipitation from the General Circulation Model from the National Centers for Environmental Prediction in the form of a Climate Forecast System Reanalysis model. The results of the study show that the GTWR method, which employs the Exponential kernel function and a fixed bandwidth to model monthly rainfall, outperforms the variable selection method, by giving the value of $R^2 = 70.62\%$ and the Root

---

*Corresponding author
E-mail address: aji_hw@apps.ipb.ac.id
Received April 13, 2022

Mean Square Error (RMSE) = 84.25 while the variable selection method gives the value of $R^2$ = 31.21% and RMSE = 128.91. The combination of the GTWR method and Spline interpolation method is the best method for estimating the monthly rainfall value in an unobserved location.

**Keywords:** statistical downscaling; regression; spatial; temporal; GTWR; kriging and spline interpolation.

**2010 AMS Subject Classification:** 62P12.

## 1. INTRODUCTION

Various Statistical Downscaling (SD) models were used to estimate rainfall. [1] states that SD is a way of interpolating regional-scale atmospheric predictor variables to smaller-scale variables. SD technique is a process of transforming data from a grid with large-scale units into data on a grid with smaller-scale units.

The General Circulation Model (GCM) is one of the remote sensing technology products that can be used to estimate rainfall. GCM is a computer-based model that consists of a number of numerical and deterministic equations that are integrated and follow physics rules. GCM is able to simulate large-scale climatic conditions and by using downscaling techniques can produce important information with higher resolution such as at local scale stations [2]. Various SD models were used, including: Projection Pursuit [2], [3], [4], and [5]; Clustered Linear Regression and Two-Stage Modeling [6] and Clusterwise Regression [7].

[2] stated that in general, the rainfall data are nonlinear and do not spread normally, while the GCM output data have the character of curse of dimensionality and multicollinearity, therefore a particular step is needed in SD modeling to reduce the dimensions of the data. The PPR (Projection Pursuit Regression) method can perform dimension reduction and develop a nonparametric and data-driven regression model. The results of the analysis study show that the estimation of monthly rainfall using the PPR model is more accurate and the pattern of the estimated values is closer to the actual data pattern than the PCR (Principal Component Regression) model, especially for historical data lengths that are more than or equal to 20 years.

[5] suggested that while GCM can simulate climate parameters on a global scale, it cannot

produce local scale climate parameter data such as rain gauge stations. The research was conducted through classification modeling using ordinal logistic regression to estimate local rainfall groups based on their quartiles associated with GCM outputs, and then estimating rainfall from each local rainfall group using PCR and Partial Least Squares Regression,

The relationship between response variables and predictor variables in data containing spatial diversity can use local regression models, including Geographically Weighted Regression (GWR). The heterogeneous data was overcome by including the location as a weight in the estimation of model parameters that varied by location. [8] stated that the GWR has the advantage of investigating the non-stationary characteristics and the scale dependence of the relationship between the response variable and the predictor variable. Based on the assumption that the relationship between rainfall and environmental variables varies spatially, [9] and [10] proposed the GWR model to obtain a higher resolution rainfall data set.

The new GWR method handles the diversity of data from the standpoint of location. This method must be developed to handle the diversity of data from the temporal side. [11] has developed GWR into GWR and Temporal (GTWR) so as to provide different model parameter values at each location $(u_i, v_i)$ and time $t_i$.

According to [12], rainfall in Indonesia varies greatly both spatially and temporally and the pattern is influenced by the monsoon pattern which is characterized by the presence of rainy and dry seasons. Rainfall data collection in the field is carried out using a rain gauge. Rainfall measurement also utilizes remote sensing technology via satellite so that it becomes an important reference for measuring rainfall in an area with a wide coverage, particularly in areas that are difficult to reach.

Based on the literature study, there is no GTWR model estimation in SD modeling for estimating monthly rainfall using the Weighted Least Squares (WLS) method which involves the predictor variables of temperature and monthly precipitation of GCM outputs from NCEP (National Centers for Environmental Prediction) in the form of a CFSR (Climate Forecast System Reanalysis) model. For this reason, it is necessary to estimate monthly rainfall using the WLS model in order to obtain information about the goodness of the model and the best predictor variables, temperature or precipitation.

The goal of this study is to use the GTWR model in SD modeling for estimating monthly rainfall using the WLS method and estimating monthly rainfall values at unobserved locations. The response variables are monthly rainfall data from 35 stations in West Java Province from January 1983 to December 2012 and the predictor variables are monthly GCM output temperature and precipitation from NCEP in the form of the CSFR model. The results of the study will provide information about the goodness of the model as well as the best predictor variables, temperature or precipitation.

## 2. LITERATURE REVIEW

### 2.1. Geographically Weighted Regression

[13] suggested that GWR considers location effects in the estimation of model parameters. This weighting process follows Tobler's First Law of Geography [14], which states that locations closer to location *i* have a greater influence in estimating parameters at location *i* than locations farther away.

The GWR model is [15]:

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

$$i = 1, 2, 3, \ldots, n; \; k = 1, 2, \ldots, p.$$

where: $y_i$ is the value of the response variable at the location $(u_i, v_i)$, $\beta_0(u_i, v_i)$ is the intercept at the observation location $(u_i, v_i)$, $\beta_k(u_i, v_i)$ is the regression coefficient of the sequence $k$ predictor variable at the location$(u_i, v_i)$, $x_{ik}$ is the value of the sequence $k$ predictor variable at location $(u_i, v_i)$ dan $\varepsilon_i$ is the sequence $i$ location error, assumed to be identical, independent, and normally distributed with zero mean and variance $\sigma^2$.

Equation (1) in matrix notation becomes:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

where:

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \boldsymbol{\beta}(u_i, v_i) = \begin{pmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_k(u_i, v_i) \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

The parameter estimators are [16]:

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \boldsymbol{W}(u_i, v_i) \boldsymbol{y} \quad (3)$$

$$\boldsymbol{W}(u_i, v_i) = diag(w_{i1}, w_{i2}, \ldots, w_{in}) \quad (4)$$

$\boldsymbol{W}(u_i, v_i)$ is the weighting matrix at location $(u_i, v_i)$ and the Euclidean distance between location $(u_i, v_i)$ and location $(u_j, v_j)$ is:

$$(d_{ij})^2 = (u_i - u_j)^2 + (v_i - v_j)^2 \quad (5)$$

The weighting for the GWR model at each location is determined using several weighting functions [13], including:

a. Gaussian Kernel function:

$$w_{ij}(u_i, v_i) = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \quad (6)$$

$d_{ij}$ is the spatial distance between a point at location *i* and location *j*, and h is a non-negative parameter known as bandwidth or smoothing parameter.

b. Exponential Kernel:

$$w_{ij}(u_i, v_i) = \sqrt{\exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right)} \quad (7)$$

*c.* Bisquare Kernel function:

$$w_{ij}(u_i, v_i) = \begin{cases} \sqrt{1 - \left(\frac{d_{ij}}{h}\right)^2}, & for\ d_{ij} \le h \\ 0, & for\ d_{ij} > h \end{cases} \quad (8)$$

The bandwidth value is determined using the Cross Validation (CV) procedure as follows [15]:

$$\mathrm{CV} = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(h)]^2 \quad (9)$$

$\hat{y}_{\neq i}$ is the approximate value for $y_i$ excluding location *i*. This approach tests the model only with examples close to location $(u_i, v_i)$ other than at location *i*. The optimal *h* value will be obtained at the minimum CV value.

**2.2. Geographically and Temporally Weighted Regression**

The GTWR model is a development of the GWR model to handle the diversity of data from the spatial and temporal side simultaneously [17] and [18]. The weighting matrix is a combination of spatial and temporal information in identifying spatial and temporal diversity. The model is:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^{p} \boldsymbol{\beta}_k(u_i, v_i, t_i)\boldsymbol{x}_{ik} + \varepsilon_i \qquad (10)$$

$$i=1, 2, 3, \ldots, n;\ k=1, 2, \ldots, p.$$

where $y_i$ is the observed value of the response variable at the observation location $(u_i, v_i)$ and time $t_i$, $\beta_0(u_i, v_i, t_i)$ is the intercept at the observation location and time sequence $i$, $\boldsymbol{\beta}_k(u_i, v_i, t_i)$ is the regression coefficient of the variable the sequence $k$ predictor at the observation location $(u_i, v_i)$ and time $t_i$ and $\boldsymbol{x}_{ik}$ are the observed values of the sequence $k$ predictor variable at the observation location $(u_i, v_i)$ and time $t_i$ and $\varepsilon_i$ are the errors of the sequence $i$ observation, assumed to be identical, independent, and $\varepsilon_i \sim N(0, \sigma^2)$.

The coefficient $\boldsymbol{\beta}_k(u_i, v_i, t_i)$ for each $k$ variable at the sequence $i$ point is obtained using the weighted least squares method by giving different weights for each location and time. The estimated coefficients are:

$$\widehat{\boldsymbol{\beta}}(u_i, v_i, t_i) = [\mathbf{X}'\mathbf{W}(u_i, v_i, t_i)\mathbf{X}]^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i, v_i, t_i)\boldsymbol{y} \qquad (11)$$

$$\boldsymbol{W}(u_i, v_i, t_i) = diag(w_{i1}, w_{i2}, \ldots, w_{in}) \qquad (12)$$

$\boldsymbol{W}(u_i, v_i, t_i)$ is the weighting matrix for the observation ($u_i$,) and sequence $i$ time.

The spatial-temporal distance function is a combination of the spatial distance function and the temporal distance function, namely:

$$\left(d_{ij}^{ST}\right)^2 = \varphi^S[(u_i - u_j)^2 + (v_i - v_j)^2] + \varphi^T(t_i - t_j)^2 \qquad (13)$$

$\varphi^S$ and $\varphi^T$ are the balancing parameters for the effect of unit differences between location and time on temporal spatial distance measurements ([18], [19], [20]).

$$\frac{\left(d_{ij}^{ST}\right)^2}{\varphi^S} = (u_i - u_j)^2 + (v_i - v_j)^2 + \tau(t_i - t_i)^2 \qquad (14)$$

where $\tau$ is the parameter to increase or decrease the ratio of the temporal distance to the spatial distance obtained from the minimum CV:

$$\mathrm{CV}(\tau) = \sum_{i=1}^{n}\left(y_i - \hat{y}_{\neq i}(\tau)\right)^2 \tag{15}$$

where $y_i$ is the value of the response variable at the location $(u_i, v_i)$ and $\hat{y}_{\neq i}$ is the estimated value of the response variable at the location $(u_i, v_i)$ without including $y_i$. The initial assignment of $\tau$ and the use of iterative methods will produce estimators for $\varphi^T$ and $\varphi^S$.

### 2.3. Estimation of Data Values at Unobserved Locations

In the field, it has been discovered that some locations are not observed because they lack measuring instruments; a natural disaster or social unrest occurs at the measurement location; the measuring device is damaged; cultural factors; safety conditions; or the high cost of constructing measurement facilities. These locations require an estimate of the value of the response variable. The estimation methods used in this study were the Kriging interpolation and the Spline interpolation method.

#### 2.3.1. Kriging Interpolation

Kriging is a geostatistical method used to estimate the value of the response variable at a location that is not observed, as a linear combination of the values of the response variable around the location to be estimated. This estimation involves a Kriging spatial weighting function derived from the estimation of the minimum variance using the semivariogram model [21].

Variogram consists of experimental variogram and theoretical variogram. The experimental variogram was obtained from observational data. This variogram is formulated as [21]:

$$2\gamma(h) = \frac{1}{N(h)}\sum[Z(s_i) - Z(s_i + h)]^2 \tag{16}$$

where $s_i$ is the location of the sequence $i$ sample point, $Z(s_i)$ is the observation value at location $s_i$, $N(h)$ is the number of data pairs that have a distance of $h$, $\varepsilon_i$ is the location error $(u_i, v_i)$ and time $t_i$, assumed to be identical, independent, and normal spread with zero mean and variance $\sigma^2$.

The theoretical semivariogram model that will be used as a comparison in this research is [21]:

a. Spherical Model

$$\gamma(\mathrm{h}) = \begin{cases} C_0 + C\left[\left(\frac{3h}{2a}\right) - 0.5\left(\frac{h}{a}\right)^3\right], \text{for h} \le \text{a} \\ C_0 + C \qquad\qquad\qquad\quad , \text{for h} > \text{a} \end{cases} \tag{17}$$

where $\mathrm{C}_0 + \mathrm{C}$ is sill, a is range.

b. Gaussian Model

$$\gamma(h) = C_0 + C\left[1 - exp\left(-\frac{3h^2}{a^2}\right)\right] \tag{18}$$

The Kriging method consists of Simple Kriging (SK), Ordinary Kriging (OK), and Universal Kriging (UK). SK assumes the population mean is known and has a constant value, OK assumes the population mean is unknown and the data does not contain trends and outliers, while UK can be used on data that contains trends and is not stationary.

The OK estimator is a linear combination of known sample variables [22]. Mathematically stated as follows:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} w_i Z(s_i) \tag{19}$$

where $\hat{Z}(s_0)$ is the estimated response value at unobserved locations, $w_i$ is the weight coefficient of $Z(s_i)$, with $\sum_{i=1}^{n} w_i = 1$, $Z(s_i)$ is the response value at observed locations and *n* is the number of examples.

### 2.3.2. Spline Interpolation

The Spline interpolation is a method to estimate the value of the response variable at a certain point by using a mathematical function that minimizes the overall surface curvature that passes through the input point correctly [23]. The sample points are extruded to the height of their magnitude. It fits a mathematical function to a specified number of nearest input points while passing through the sample points [24].

The spline interpolation equation is:

$$S(x, y) = T(x, y) + \sum_{j=i}^{n} \lambda_j R(r_j) \tag{20}$$

where j= 1, 2, …, n, n is the number of points, $\lambda_j$ is the coefficient found from the system of linear equations, $r_j$ is the distance between point (x,y) to point j, $T(x, y)$ and and $R(r_j)$ is defined differently, based on the selection method (regularized spline or tension spline) [24].

## 3. MATERIALS AND METHODS

### 3.1. Data

This research used a response variable in the form of monthly rainfall data (mm/month) from 35 rain observation stations in West Java Province which are located at -7.78°S to -6.28°S and 108.40°E to 107.87°E, from January 1983 to June 2021 (126 months), which was obtained from BMKG. The predictor variables were temperature and monthly precipitation, from January 1984 to June 2020 the GCM output from NCEP in the form of a CSFR model downloaded from http://rda.ncar.edu [25] using a grid size of 2.5° $\times$ 2.5° [26].

### 3.2. Best Model Selection Criteria

The measure of the goodness of the model used is the coefficient of determination ($R^2$ and $R^2_{Adjusted}$), Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC), namely:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{21}$$

$$R^2_{Adjusted} = 1 - \frac{(1-R)(n-1)}{n-k-1} \tag{22}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}(y_i-\hat{y}_i)^2\right)} \tag{23}$$

$$\text{AIC} = -n\,log\left(\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}\right) + 2p \tag{24}$$

## 4. RESULTS AND DISCUSSION

### 4.1. Estimated Monthly Rainfall Using GTWR

Based on the boxplot of the monthly rainfall from 1984 to 2012 in Figure 1, it can be seen that the highest monthly rainfall value was achieved in January and then decreased to the lowest in August, then increased again until it reached its highest value again in January. This is in accordance with Tukidi [27] that the monsoon pattern is influenced by sea breezes on a very wide scale. This type of rain is distinguished by a distinct difference between the rainy and dry seasons in a year, as well as a single maximum monthly rainfall in a year. This type of rainfall is most common on the Indonesian islands of Java, Bali, and Nusa Tenggara.

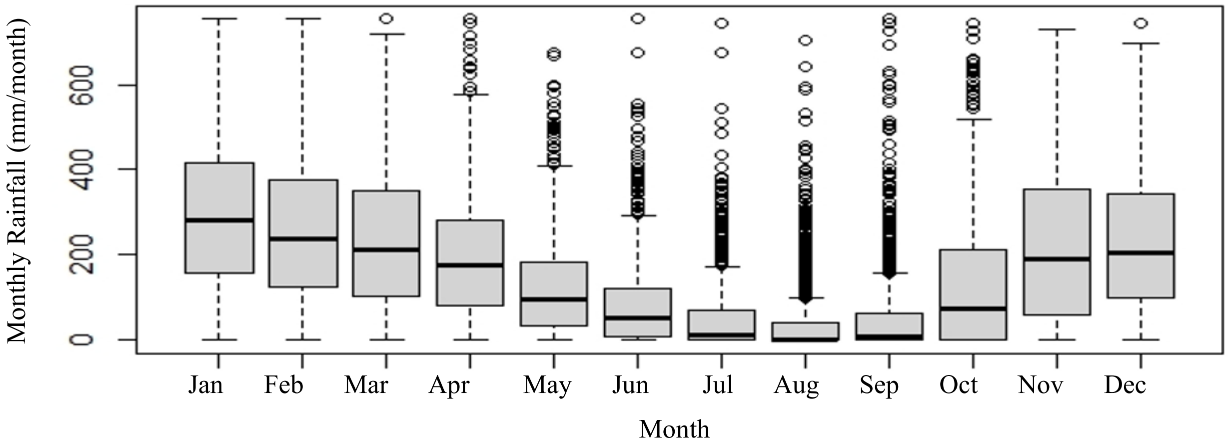


Figure 1. Monthly Rainfall Boxplot per Month

The distribution of monthly rainfall at the first six stations in West Java Province is shown in Figure 2. In general, the histogram of monthly rainfall for each station is not symmetrical, tends to skew to the right. The low-value monthly rainfall has a fairly large frequency, while the high-value monthly rainfall has a low frequency.

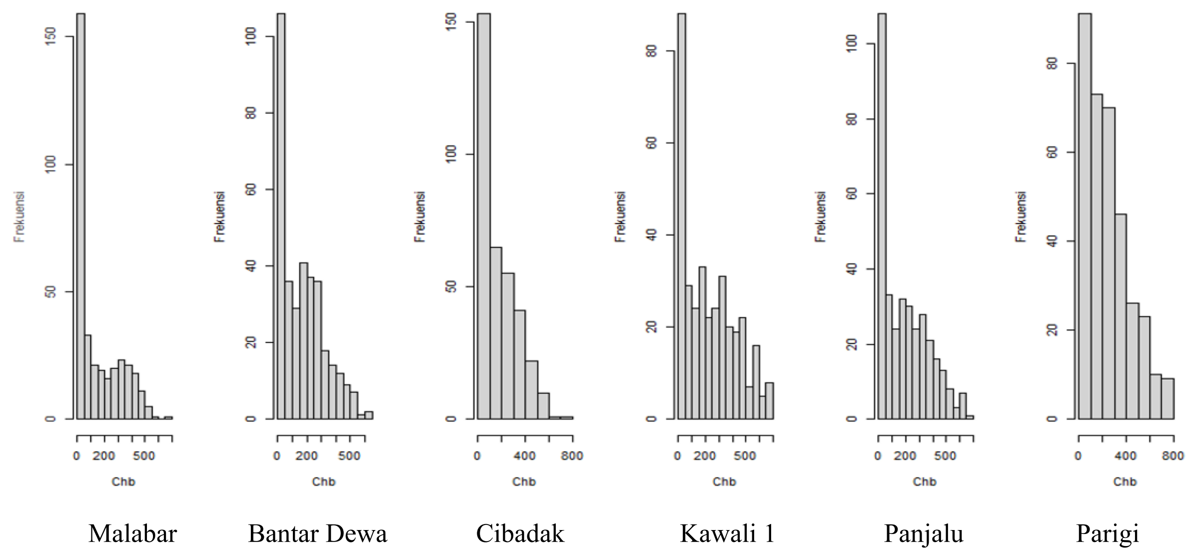


Figure 2. Monthly Rainfall Histogram of the $1^{st}$ to $6^{th}$ Station.

The next step is to estimate the Multiple Linear Regression parameter by selecting the significant variables using the Stepwise method. Furthermore, spatial and temporal heterogeneity was examined every month. The period which mostly contains spatial and temporal heterogeneity is used as the basis for examining the spatial and temporal model.

The response variable used is monthly rainfall from 1984 to 2012 from 35 stations in West Java province. The predictor variables were: monthly rainfall in the previous year ($y_0$);

precipitation and GCM output temperature in the 5×8 domain with a resolution of 2.50×2.50 (x1, x2, …, x40), the dummy variables of monthly rainfall in a month are listed in Table 1:

Table 1. The Dummy Variables of Monthly Rainfall

| Month | $d_1$ | $d_4$ | $d_{10}$ |
|---|---|---|---|
| December/January/February | 1 | 0 | 0 |
| March, April, May | 0 | 1 | 0 |
| September/October/November | 0 | 0 | 1 |
| June/July/August | 0 | 0 | 0 |

and the dummy variables of station altitude status are listed in Table 2:

Tabel 2. The Dummy Variables of Station Altitude Status

| Altitude Status | $s_1$ | $s_2$ |
|---|---|---|
| Low | 1 | 0 |
| Medium | 0 | 1 |
| High | 0 | 0 |

The values of $R^2$, $R^2_{Adj}$, RMSE and AIC from the Stepwise Regression model in 2010 to 2012, with the predictor variables of precipitation and temperature listed in Table 3. Based on the value of the goodness of the model, $R^2$ and $R^2_{Adj}$are the largest and the value The smallest RMSE and AIC, Stepwise Regression model with precipitation predictor variables were selected as the best model to proceed to the spatial temporal modeling.

Table 3. Values of $R^2$, $R^2_{Adj}$, RMSE and AIC from The Stepwise Regression Model with Predictor Variables of Precipitation and Temperature from 2010 to 2012.

| Predictor Variable | $R^2$ | $R^2_{Adj}$ | RMSE | AIC |
|---|---|---|---|---|
| Precipitation | 0.312 | 0.302 | 128.908 | 15862.66 |
| Temperature | 0.308 | 0.294 | 129.252 | 15881.36 |

### 4.2. Spatial Diversity Analysis

The Breusch Pagan test is used to identify the spatial effect on the data set. The diversity test is carried out every month so that the temporal effect can also be known. If there is spatial and temporal diversity in the data, then proceed to temporal spatial modeling. The results of the examination showed that the period that the majority had spatial diversity was from January 2010 to December 2012. Table 4 shows the spatial diversity values for each month.

Table 4. Spatial Diversity Test Results.

| Month | 2010 | | 2011 | | 2012 | |
|---|---|---|---|---|---|---|
| | BP Value | p-Value | BP value | p-Value | BP Value | p-Value |
| January | 13.607 | 0.0035 | 3.948 | 0.2671 | 9.051 | 0.0286 |
| February | 4.968 | 0.1742 | 1.942 | 0.5844 | 12.857 | 0.0049 |
| March | 6.608 | 0.0857 | 5.518 | 0.1376 | 3.061 | 0.3824 |
| April | 5.450 | 0.1416 | 1.960 | 0.5808 | 1.937 | 0.5857 |
| May | 12.250 | 0.0066 | 0.457 | 0.9283 | 11.473 | 0.0094 |
| June | 17.679 | 0.0005 | 3.447 | 0.3277 | 8.871 | 0.0311 |
| July | 19.438 | 0.0002 | 8.035 | 0.0453 | 16.987 | 0.0007 |
| August | 2.120 | 0.5478 | 0.545 | 0.9089 | 0.221 | 0.9742 |
| September | 15.039 | 0.0018 | 4.768 | 0.1896 | 6.384 | 0.0942 |
| October | 8.139 | 0.0432 | 19.398 | 0.0002 | 1.624 | 0.6540 |
| November | 8.312 | 0.0399 | 7.106 | 0.0686 | 5.619 | 0.1317 |
| December | 4.516 | 0.2109 | 9.782 | 0.0205 | 1.057 | 0.7874 |

### 4.3. Temporal Diversity Analysis

Furthermore, an examination is carried out whether there is a temporal variation in the selected. The boxplots of the residuals in each month from 2010 to 2012 are listed in Figure 3, Figure 4, and Figure 5.

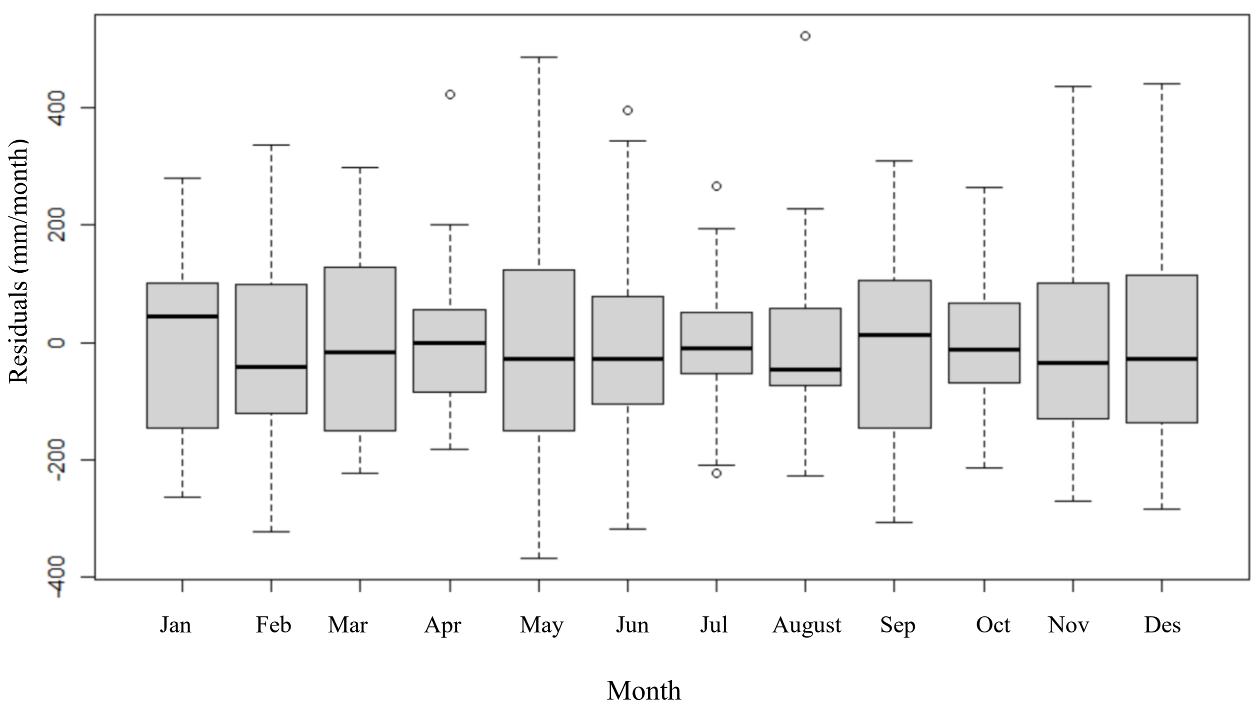


Figure 3. Boxplot of Residuals in 2010

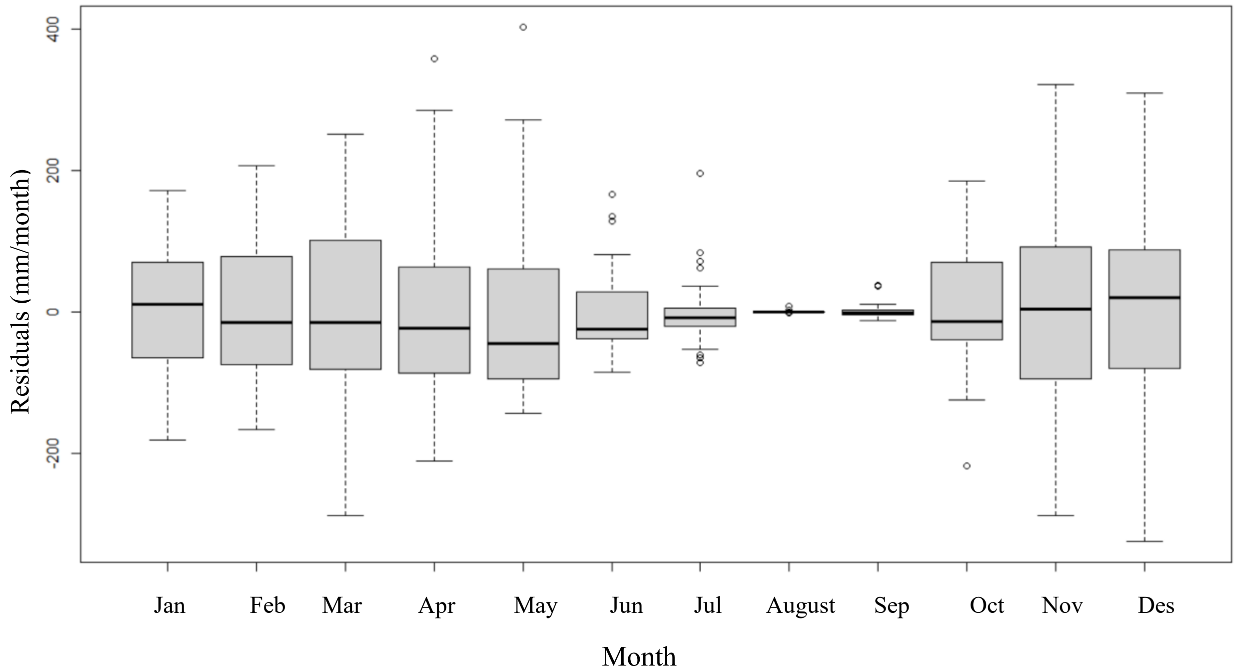


Figure 4. Boxplot of Residuals in 2011

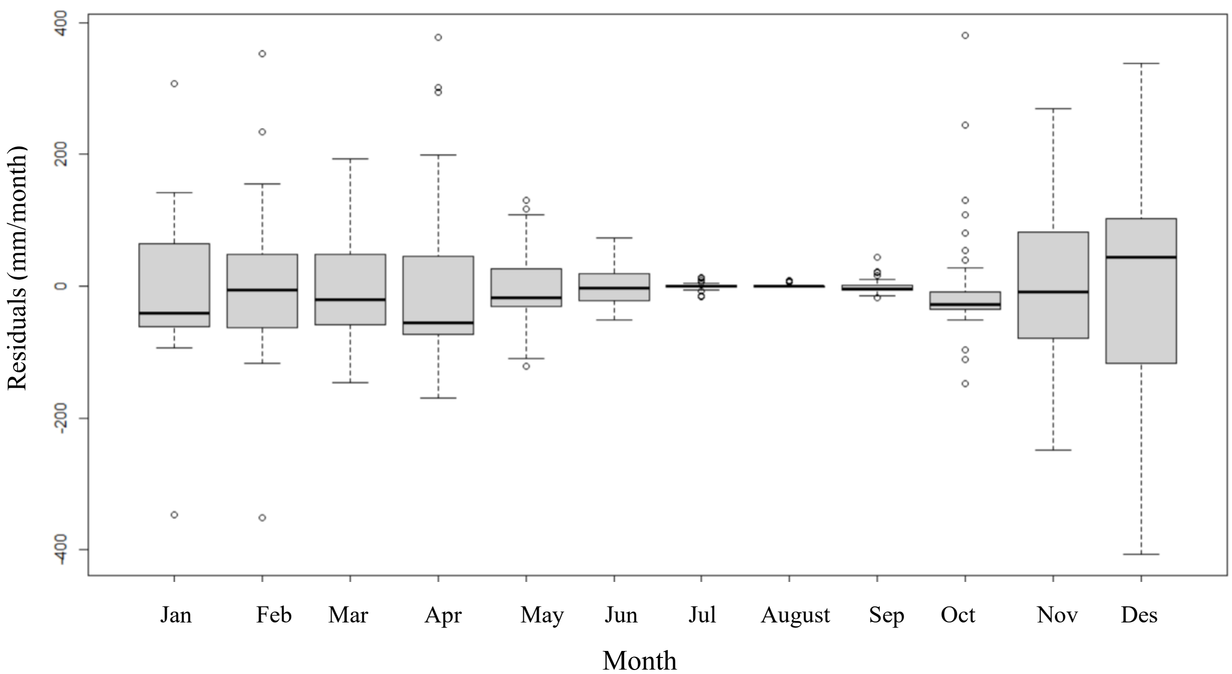


Figure 5. Boxplot of Residuals in 2012

Based on the distance between quartiles in the three figures, it can be seen that the residuals of the global regression model are not homogeneous. The height of the box represents the diversity of the data. It can be concluded that there are spatial and temporal variations in the monthly rainfall data. This variability will then be addressed by applying spatial-temporal regression.

### 4.4. GTWR Modeling

GTWR modeling used response variables and predictor variables of precipitation and temperature with fixed bandwidth, the results are shown in Table 5.

Table 5. Values of $R^2$, $R^2_{Adj}$, RMSE and AIC from Local Models with Predictor Variables of Precipitation and Temperature and Fixed Bandwidth.

| Kernel Function | Predictor | $R^2$ | $R^2_{Adj}$ | RMSE | AIC |
|---|---|---|---|---|---|
| Gaussian | Precipitation | 0.543 | 0.475 | 104.957 | 15430.330 |
| | Temperature | 0.460 | 0.416 | 114.196 | 15590.750 |
| Exponential | Precipitation | 0.706 | 0.568 | 84.248 | 15031.520 |
| | Temperature | 0.672 | 0.546 | 88.964 | 15130.130 |
| Bisquare | Precipitation | 0.391 | 0.366 | 121.341 | 15710.200 |
| | Temperature | 0.370 | 0.347 | 123.391 | 15749.740 |

Based on the criteria for the largest $R^2$ and $R^2_{Adj}$ values and the smallest RMSE and AIC in Table 5, it is found that the best model is the GTWR model with precipitation predictor variables, using Exponential kernel functions and fixed bandwidth with $R^2$ = 70.6% and RMSE = 84.25. This value is higher than the Stepwise regression model with $R^2$ = 31.2% and RMSE = 128.91. The predictor variables selected were: y0, x106, x108, x109, x111, x113, x131, x132, x134, x135, x156, x157, x158 and x159.

The GTWR model generates a model for each station and for each month. Consequently, 1260 models were produced for 35 stations for 12 months in 2010, 2011 and 2012. Each station has 36 models from January 2010 to December 2012. Figure 6 shows the Boxplot of the GTWR coefficient estimator.

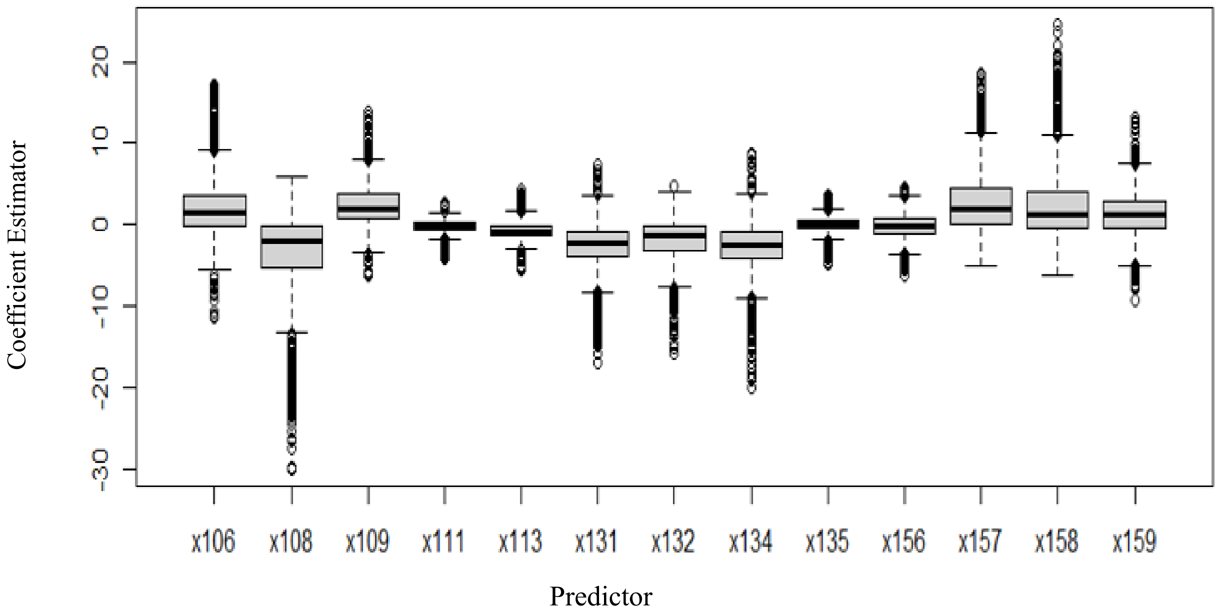


Figure 6. Boxplot of GTWR Coefficient Estimator

The value of the estimator coefficient of the predictor variable parameter varies between stations and months. The predictor variable x111 has the smallest inter-quartile distance of 0.82 while the largest inter-quartile distance is 5.27 for the predictor variable x108. This means that when there is a change in the value of the predictor variable by one unit, it will decrease the value of monthly rainfall for those whose estimating coefficient is negative or increase the value of monthly rainfall for those whose estimator coefficient is positive. The smallest GTWR coefficient estimator is -24.51 for the predictor variable x108 while the largest is 29.98 for the predictor variable x158.

### 4.5. Estimation of Data Values at Unobserved Locations

The results of GTWR modeling can be used to estimate the monthly rainfall value at a station for which there is no data. This study compares the Kriging (Gauss and Spherical) interpolation and Spline interpolation methods. The performance results of the three methods are shown in Table 6.

Table 6. The RMSE Value Derived from The Method of Estimating Monthly Rainfall in January and July of 2010-2012

| Month | Year | Kriging Gauss | Kriging Spherical | Spline |
|---|---|---|---|---|
| January | 2010 | 73.088 | 69.806 | 6.827 E-10 |
| | 2011 | 74.321 | 75.550 | 5.590 E-10 |
| | 2012 | 85.905 | 86.624 | 1.341 E-10 |
| July | 2010 | 191.621 | 77.187 | 4.652 E-10 |
| | 2011 | 41.281 | 38.891 | 6.018 E-10 |
| | 2012 | 24.553 | 25.622 | 1.522 E-10 |

The Spline interpolation method is the best method for estimating data at unobserved locations compared to the Kriging Gauss and Kriging Spherical interpolation methods. This method always gives the smallest RMSE value in January 2010 to July 2012, the smallest value is 1.341E-10 in January 2012 and the largest is 6.827E-10 in January 2010. Contour map of monthly rainfall values for January 2010 until July 2012 can be seen in Figure 7, Figure 8, and Figure 9.

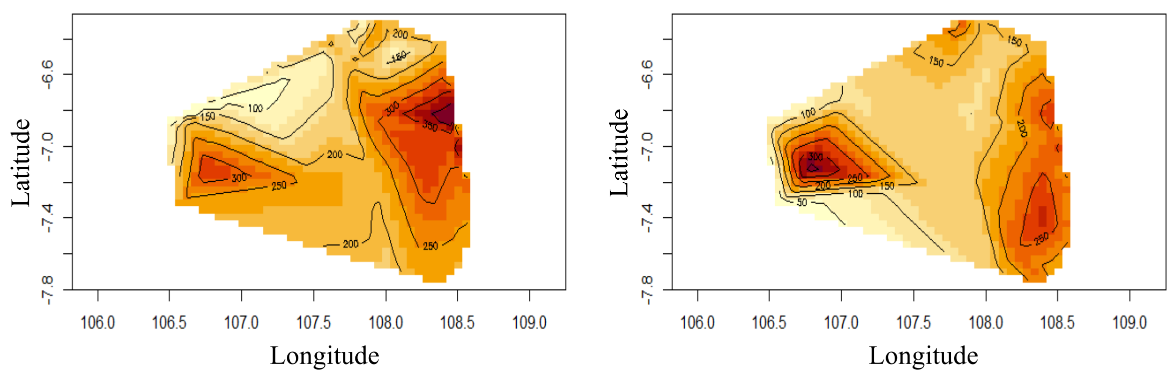


Figure 7. Contour Map of Monthly Rainfall Values for January 2010 and January 2011

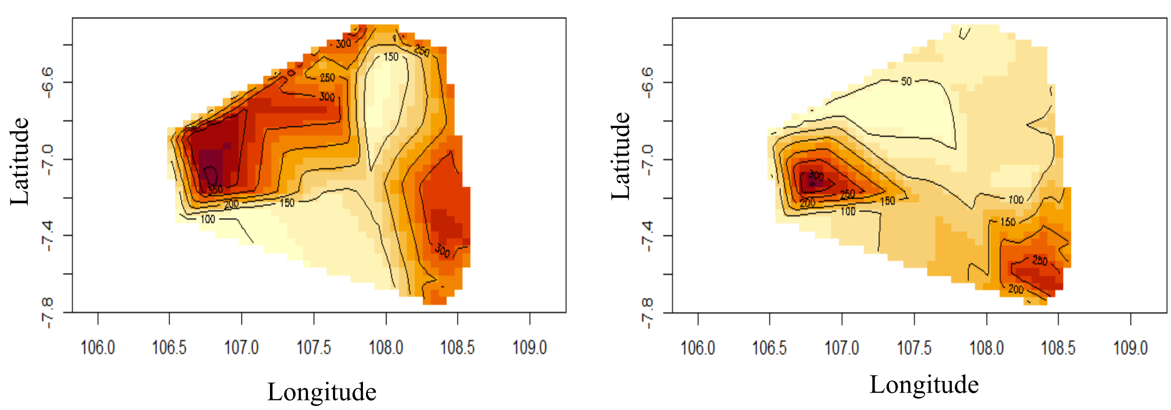


Figure 8. Contour Map of Monthly Rainfall Values for January 2012 and July 2010

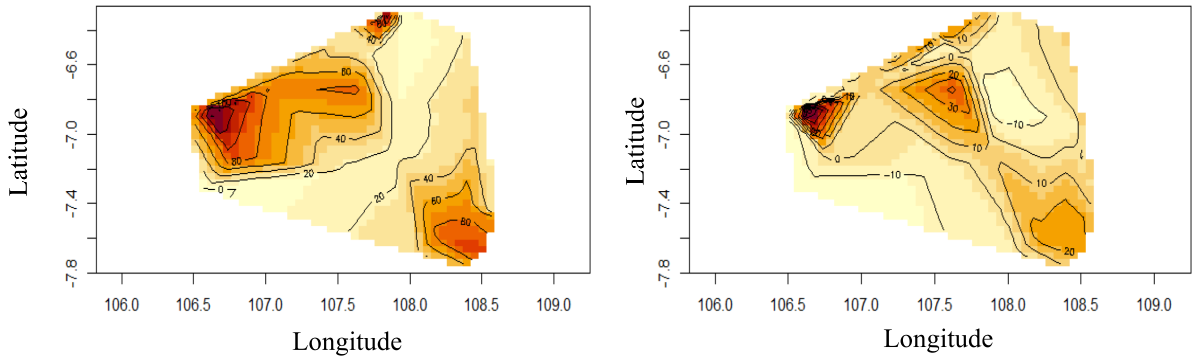


Figure 9. Contour Map of Monthly Rainfall Values for July 2011 and July 2012

Based on the contour maps of monthly rainfall values in Figure 7, Figure 8, and Figure 9, it can be seen that in general the monthly rainfall values in the rainy season are generally high in the west and southeast, while in the southwest it tends to be low. Meanwhile, during the dry season, monthly rainfall values tend to be high in the west and southeast and low in the southwest and northeast. This can provide us and stakeholders with preliminary information about the estimated value of rainfall in locations whose measurement data values are not known. If the longitude and latitude values of a location are provided, the estimated monthly rainfall value can be obtained by referring to the corresponding monthly rainfall contour map generated by the Spline interpolation method.

## 5. CONCLUSIONS

Based on the description provided, the following conclusions were reached:

1. The GTWR method uses the Exponential kernel function and fixed bandwidth to model monthly rainfall better than the Stepwise Regression method, by giving $R^2$ = 70.62% and RMSE = 84.25, while the Stepwise Regression gives $R^2$ = 31.21% and RMSE = 128.91.
2. The combination of GTWR and Spline interpolation methods is the best method for estimating monthly rainfall values in locations that are not observed.

## CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

## REFERENCES

[1] R.L. Wilby, T.M.L.Wigley, Downscaling general circulation model output: A review of methods and limitations, Progress Phys. Geography. 214 (1997), 530-548

[2] A.H.Wigena, Pemodelan Statistical Downscaling dengan Regresi Projection Pursuit untuk Peramalan Curah Hujan Bulanan. Kasus Curah Hujan Bulanan di Indramayu [disertasi]. Bogor: Sekolah Pascasarjana Institut Pertanian Bogor, 2006.

[3] A. Djuraidah, A.H. Wigena, Regresi kuantil untuk eksplorasi pola curah hujan di Kabupaten Indramayu. Jurnal Ilmu Dasar. 12 (2011), 50-56.

[4] S.F. Jia, W.B. Zhu, A.F. Lű et al. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China. Remote Sens. Environ. 115 (2011), 3069–3079.

[5] Khairunnisa, A.M. Soleh, U.D.Syafitri, Prediksi Curah Hujan Bulanan dengan Teknik Statistical Downscaling Melalui Pemodelan Dua Tahap. Journal of Computer Engineering System and Science, 2019.

[6] Nadya AR. Pemodelan Statistical Downscaling untuk Memprediksi Curah Hujan degan Regresi Linear Gerombol dan Pemodelan Dua Tahap [tesis]. Bogor (ID): Institut Pertanian Bogor, 2018.

[7] V. Butar-butar, Pemodelan clusterwise Regression pada Statistical Downscaling Untuk Prediksi Curah Hujan Bulanan [tesis]. Bogor: Sekolah Pascasarjana Institut Pertanian Bogor, 2019.

[8] G.M. Foody, Geographical weighting as a further refinement to regression modelling: an example focused on the NDVI–rainfall relationship. Remote Sens. Environ. 88 (2003), 283–293.

[9] F.R. Chen, Y. Liu, Q. Liu et al. Spatial downscaling of TRMM 3B43 precipitation considering spatial heterogeneity, Int. J. Remote Sens. 35 (2014), 3074–3093.

[10] S.G. Xu, C.Y. Wu, L. Wang et al. A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics. Remote Sens. Environ. 162 (2015), 119–140.

[11] A.S. Fotheringham, R. Crespo, J. Yao, Geographical and temporal weighted regression (GTWR). Geograph. Anal. 47(2015), 431-452.

[12] T.A. Nuraini, D.E. Nuryanto, K.E. Komalasari, et al. Pengembangan Model HyBMG 2.07 untuk Prediksi Iklim Di Indonesia Dengan Menggunakan Data Tropical Rainfall Measuring Mission (TRMM). Jurnal Meteorologi

dan Geofisika (2019), 7-12. Pusat Penelitian dan Pengembangan Badan Meteorologi Klimatologi dan Geofisika (BMKG) http://puslitbang.bmkg.go.id/jmg.

[13] A.S. Fotheringham, C. Brundson, M.E.Charlton, Geographically weighted regression: analysis of spatially varying relationship. John Wiley and Sons Ltd: England, 2002.

[14] J.M. Harvey, Tobler's first law and spatial analysis, Ann. Assoc. Amer. Geograph. 94 (2004), 284-289.

[15] B. Huang, B. Wu, M.Barry, Geographically and temporally weighted regression for modeling spatio-temporal variataion in houses prices, Int. J. Geograph. Inform. Sci. 24 (2010), 383-401.

[16] C. Brudson, A.S. Fotheringham, M. Charton, Geographically weighted regression: A method for exploring spatial nonstationarity, Geograph. Anal. 28 (1996), 281-298.

[17] P. Wang, Exploring spatial effects on housing price: the case study of the city of calgary [dissertation]. Canada: University of Calgary, 2006.

[18] G. Erda, Prediksi Model Regresi Terboboti Geografis dan Temporal Kekar Menggunakan Penduga–M (Studi Kasus: Jumlah Peserta KB Aktif Jawa Timur Tahun 2009-2018) [tesis]. Bogor (ID): Institut Pertanian Bogor, 2018.

[19] Z. Mar'ah, A.H. Wigena, A. Djuraidah, Semi parametric geographically weighted regression modelling using liear model of coregionalization, Int. J. Sci., Basic Appl. Res. 34 (2017), 178-186.

[20] M. Sholihin, Pengembangan Regresi Terboboti Geografis dan Temporal Menggunakan Interaksi Jarak Spasial-Temporal (Studi Kasus: Pertumbuhan Ekonomi di Jawa Tengah Tahun 2011-2015) [tesis]. Bogor (ID): Institut Pertanian Bogor, 2018.

[21] N.A.C. Cresie, Statistics for spatial data, John Wiley and Sons, Inc. New York, 1993.

[22] Isaaks, Srivastava, An introduction to applied geostatistics, Oxford University Press, New York, 1989, 278p.

[23] J. McCoy, K. Johnston, Using ArcGIS Spatial Analyst. ESRI Press. Redland. CA. USA, 2001.

[24] L. Mitas, H. Mitasova, General variational approach to the interpolation problem, Computer Math. Appl. 16 (1988), 983-992.

[25] S. Saha, Q. Liu, J. Meng, et al. NCEP Climate Forecast System Reanalysis (CFSR) Monthly Products, January 1979 to December 2010. Research Data Archive at the 16 National Center for Atmospheric Research, Computational and Information Systems Laboratory. 2010. https://doi.org/10.5065/D6DN438J.

[26] N. Fadhli, Penentuan Domain Global Climate Model dengan LASSO untuk Peningkatan Akurasi Prediksi Curah Hujan di Jawa Barat [tesis]. Bogor (ID): Institut Pertanian Bogor, 2019.

[27] Tukidi, Karakter Curah Hujan di Indonesia, Jurusan Geografi FIS UNNES, 7 (2010), 136-145.