Modeling Dengue Fever Cases by Using GSTAR(1;1) Model with Outlier Factor

Utriweni Mukhaiyar, Nur’ainul Miftahul Huda, RR Kurnia Novita Sari and Udjianna Sekteria Pasaribu
Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia

Abstract. Dengue fever is an endemic disease transmitted through the Aedes Aegypti mosquitoes. Dengue virus can be transmitted from human hosts who have been infected by the virus to the mosquitoes to be transmitted back to other humans. So that, it is possible for the virus to be transmitted to several surrounding locations. Aedes Aegypti is one of the dengue mosquitoes that likes a warm climate and not too wet or dry. In addition, many un-expected factors can cause a significant increase in the number of dengue fever cases. So that the number of dengue fever cases can increase significantly far different from other data. An observation data that has different characteristics from others is called outlier. The existence of outliers can indicate individuals or groups that have very different behavior from the most of the individuals of the dataset. Outlier data in a data set are often encountered in various kinds of data analysis. Frequently, outliers are removed to improve accuracy of the estimators. But sometimes the presence of an outlier has a certain meaning, which explanation can be lost if the outlier is removed. In this paper, modeling dengue fever cases using GSTAR(1;1) with outlier factors was firstly proposed.

1. Introduction
Modeling time series data that have correlation in time and spatial is called space time data. Generalized Space Time Autoregressive (GSTAR) model is one of space time models used to modeling and forecasting spatial time series data. Nowadays, the GSTAR model is growing rapidly in Indonesia. Some development of GSTAR (1;1) model have been done by some researcher such as [1], making a new procedure for Generalized STAR modeling using IAcM (Inverse Autocovariance Matrix) approach. This model was applied to the monthly tea production of some plantations in West Java, Indonesia. In terms of weighting on the GSTAR model, [2] modeled GSTAR using the weighted average of fuzzy sets concept approach and applied that model to oil palm production. [3] conducted research on error assumption on GSTAR model. Recently [4] did a research about Spatial Weight Determination of GSTAR(1;1) Model by Using Kernel Function. This research made weight matrix construction was less subjective. In application, GSTAR model is rapidly used to forecast Gross Domestic Product (GDP) West European [5], chili price in Bandung’s market [6] and criminality [7].

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
The combination of GSTAR modeling and variogram of spatial analysis was conducted by [8]. Though the spatial analysis is an older science than space-time analysis, the development of spatial models still on going with several application. [9] use bootstrap approach to estimate the parameters of isotropic semivariogram and [10] used spatial weighting approach to disaggregate MDGs indicators. Furthermore the effect of spatial aggregation to space time model was investigated by [11].

Unexpected extraordinary observations that look differently far from most observations in a data set are often encountered in various kinds of data analysis. Space time analysis is no exception. Outlier is an observation data that has different characteristics from others. Its presence is unexpected because there are so many factors that can cause its presence. The existence of outliers can indicate individuals or groups that have very different behavior from the most of the individuals of the dataset. Frequently, outliers are removed to improve accuracy of the estimators. But sometimes the presence of an outlier has a certain meaning, which explanation can be lost if the outlier is removed. In this paper the presence of outliers was not removed, but the way to deal with outliers is by adding an outlier factor to the model.

In this paper, the data used to apply the GSTAR model with outlier factors is dengue fever cases data in six regencies in West Kalimantan from January 2015 – June 2018. Dengue fever is an endemic disease transmitted through the Aedes *Aegypti* mosquitoes. Aedes *Aegypti* is one of the dengue mosquitoes that likes a warm climate and not too wet or dry with optimum rainfall ranging from 200-300 mm / month. Climate change that has occurred in recent years has caused dengue cases increase significantly. This significant increase can be an indication of outliers detection in the GSTAR model. Deterministically, modeling of dengue fever cases has been examined by [12]. The study discussed about estimation of the basic reproductive ratio for dengue fever at the take-off period of dengue infection.

The purpose of this study are to analysis the algorithm for outlier detection in GSTAR model and to apply it in data about dengue fever cases in 6 regencies in West Kalimantan. This paper is divided into six sections. Section 2 briefly explains GSTAR model. Section 3 discusses the definition and types of outliers. Then the algorithm for outlier detection is discussed in Section 4. Application of GSTAR model by adding an outlier factor is discussed in Section 5. Conclusions and remarks are put forward in Section 6.

2. GSTAR Model
GSTAR model is the development of STAR model. In STAR model, all locations have same autoregressive parameters, so that the locations used is assumed homogen. Consequently its model only can be used for uniform location. Whereas in reality, we have heterogeneous location. GSTAR model enable to capture a phenomenon with heterogeneous characteristics locations and the parameters for each location are different from each other. Consider a random variable $Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{n,t})$, $Y_t$ follows GSTAR($p; \lambda_1, \lambda_2, \ldots, \lambda_p$) model. It can be stated as [14]:

$$Y_t = \left( \sum_{k=1}^{p} \lambda_k \Phi_{k,t} W^{(t)} Y_{t-k} \right) + a_t \tag{1}$$

where $\Phi_{k,t}$ is the diagonal matrix of parameters in GSTAR model, $W^{(t)}$ is weight matrix, $a_t = (a_{1,t}, a_{2,t}, \ldots, a_{n,t})$ is the matrix residuals in GSTAR model, and $a_t \sim N(0, \sigma_a^2)$.

For example, GSTAR model with both autoregressive’s order and spatial’s order are 1, GSTAR(1; 1), could be written as

$$Y_t = \Phi_0 Y_{t-1} + \Phi_1 W Y_{t-1} + a_t \tag{2}$$
The model (2) is called as GSTAR(1; 1) model without outlier factor. The following are the weight on the GSTAR model

(i) Uniform Weight
Uniform weight give the same weight for each locations. The values of uniform location weights are calculated by
\[ w_{ij}^{(\ell)} = \frac{1}{n_i^{(\ell)}}, \quad \text{where} \quad w_{ij}^{(\ell)} \text{ is the weight between locations } i \text{ and } j \]
and \( n_i^{(\ell)} \) is the number of locations adjacent to the \( i \)-location in spatial lag \( \ell \)

(ii) Binary Weight
The value of binary weight is zero and one on elements other than the main diagonal. The relationship between two geographically adjacent cities is defined by \( w_{ij} = 1 \). Whereas if geographically far apart is defined \( w_{ij} = 0 \).

(iii) Inverse Distance Weight
The determination of this weight is based on the actual distance between locations. The calculation of weights is obtained from the normalization of the actual inverse distance results. The first step is calculating the actual distance between locations. The shape of the distance matrix formed is
\[
D = \begin{cases} 
0, & \text{for } t = T \\
\frac{1}{d_{ij}}, & \text{for } t \neq T 
\end{cases}
\]
where \( d_{ij} \) is the distance between locations \( i \) and \( j \). \( D \) is a symmetrical matrix.

Then the \( D \) matrix is standardized in \( W \) form with \( \sum_{j=1}^{N} w_{ij}^{(\ell)} = 1 \) where \( i \neq j \). Assuming that close distances have a strong relationship among locations, generally the inverse weight of distance for each location is stated by
\[
W = \begin{cases} 
0, & \text{for } t = T \\
\frac{1}{\sum_{j=1}^{N} \frac{1}{d_{ij}}}, & \text{for } t \neq T 
\end{cases}
\]
where \( \sum_{j=1}^{N} w_{ij}^{(\ell)} = 1, \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^{(\ell)} = N \)

The diagonal matrix of inverse weight \( w_{ij} \) is zero, because a location has no distance with itself. The inverse weight form distance \( W \) is not a symmetrical matrix.

3. Outlier
Outliers in a time series data are an inconsistent observation data as a result of unexpected and unconscious extraordinary events such as turbulent political or economic crises. These observations are commonly known in time series in the form of outliers [13]. Outliers can cause the results of data analysis to be unreliable and valid, so that outlier detection needs to be done to eliminate the outlier effect.

Outlier detection was firstly introduced by [14]. Outliers consist of several types, namely Additive Outlier (AO), Innovative Outlier (IO), Level Shift (LS) and Temporary Change (TC). In this study, the outliers discussed were AO and IO.

(i) Additive Outlier (AO)
Additive Outlier (AO) is an event that affects to time series data at a time. Wei (2006) defines the additive outlier model as follows:
\[
Y_t = \begin{cases} 
\xi_t, & \text{for } t \neq T \\
\xi_t + \omega, & \text{for } t = T 
\end{cases}
\]
or

\[ Y_t = u_t + \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} u_t + \omega I_t^{(T)} \]  

(6)

where \( u_t \) is ARIMA model without outlier factor, \( B \) is the backshift operator such that \( B^k Y_t = Y_{t-k} \) and \( I_t^{(T)} \) is indicator outlier variable at time \( T \), such that

\[ I_t^{(T)} = \begin{cases} 
1, & \text{for } t \neq T \\
0, & \text{for } t = T 
\end{cases} \]  

(7)

(ii) Innovative Outlier (IO)

Time series data that contains Innovative Outlier (IO) gives a more complicated effect compared to the other three types of outliers. Wei (2006) defines the IO model as follows:

\[ Y_t = u_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} (a_t + \omega I_t^{(T)}) \]  

(8)

If in AO, the effect given only occurs at the time of the \( T \) observation, then the IO affects the entire observation \( Y_T, Y_{T+1}, \ldots \).

Generally, time series data can contain several different types of outliers. The outlier model in general as follows (Chang et al, 1988):

\[ Y_t = \sum_{h=1}^{H} \omega_h v_h(B) I_t^{(T_h)} + u_t \]  

(9)

where \( v_h(B) = \begin{cases} 
1, & \text{for AO} \\
\frac{\theta(B)}{\phi(B)}, & \text{for IO} 
\end{cases} \).

In this paper, the types of outliers, the algorithm of outlier detection and GSTAR model with outlier factors are adopted from the time series model. If in the time series model, we see observations only occur in one location, then in the GSTAR model these observations develop into several locations and have spatial correlations between locations. Therefore in terms of outliers, outliers may also be detected on the GSTAR model and have a correlation between locations (spatial correlation).

4. Outlier Detection in GSTAR Model

The procedure begins with modeling the original series \( Z_t \) by supposing that there is no outlier. Estimated parameters of the GSTAR model without outlier assumptions. Let \( Z_t \) be a stochastic process following an GSTAR(1;1) model, that is

\[ Z_t = \Phi_0 Z_{t-1} + \Phi_1 W Z_{t-1} + e_t \]

\[ \Phi(B)Z_t = e_t \]

where \( \Phi(B) = I - \Phi_0 B - \Phi_1 W B \), \( B \) is the backshift operator such that \( B^k Z_t = Z_{t-k} \).

The following is the algorithm for outlier detection in GSTAR(1;1) model:

(i) Compute the residual \( \hat{e}_t \) with \( \hat{e}_t = (e_{1,t}, e_{2,t}, \ldots, e_{n,t}) \) and let \( \hat{\sigma}_e^2 \) be the estimate of variance residual \( \sigma_e^2 \), such that \( \hat{e}_{i,t} = Z_{i,t} - \hat{Z}_{i,t} \) and \( \hat{\sigma}_e^2 = \frac{1}{m} \sum_{t=1}^{m} \hat{e}_{i,t}^2 \).
(ii) Compute $\hat{\lambda}_{1,i,T}$ and $\hat{\lambda}_{2,i,T}$, which are $\hat{\lambda}_{1,i,T} = \frac{\hat{\Omega}_I}{\hat{\sigma}_e}$ and $\hat{\lambda}_{2,i,T} = \frac{\hat{\Omega}_A}{\hat{\rho}_2}$ where $\hat{\rho}_2 = (1 + \hat{\Pi}_1 + \ldots + \hat{\Pi}_{n-T})^{-1}$, $\Pi_1 = \Phi_0 + \Phi_1 \mathbf{W}$, $\Pi_2, \Pi_3, \ldots = 0$ and $\hat{\Omega} = \text{diag}(\hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_n)$.

(iii) Define $\eta_t = \max \{ |\hat{\lambda}_{1,T}|, |\hat{\lambda}_{2,T}| \}$ for $t = 1, \ldots, m$ and $i = 1, \ldots, n$. If $\max \eta_t = |\hat{\lambda}_{1,T}| > C$ then there is the possibility of Innovative Outlier detected at time $T$, so that the impact $\omega$ of this possible IO is estimated by

$$\hat{\omega}_{iIT} = \hat{e}_{i,T}$$

Then eliminate its effect by defining a new residual at time $T$, that is $\hat{\omega}_{iIT} = \hat{e}_{i,T} - \hat{\omega}_{iIT} = 0$.

If $\max \eta_t = |\hat{\lambda}_{2,T}| > C$ then there is the possibility of Additive Outlier detected at time $T$, so that the impact $\omega$ of this possible AO is estimated by

$$\hat{\Omega}_{AT} = \hat{\rho}_2^2 \hat{\Pi}(P)\hat{\sigma}_e$$

Then eliminate its effect by defining a residual at time $t \geq T$, that is $\hat{e}_T = \hat{e}_T - \hat{\Omega}_{AT} \hat{\Pi}(B)\hat{I}_{T-1}^{(T)}$ and residual at time $1 \leq t < T$, that is $\hat{e}_T = \hat{e}_t$. Then, a new estimator $\hat{\sigma}_{ic}^2$, is computed from the modified residuals, that is $\hat{\sigma}_{ic}^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{e}_{i,t}^2$.

(iv) If an AO/IO is identified on previous step, recompute $\hat{\lambda}_{1,T}$ and $\hat{\lambda}_{2,T}$ based on the same initial estimates of time series parameters, but using the modified residuals $\hat{e}_T$ and the estimate $\hat{\sigma}_{ic}^2$.

(v) Repeat the third – the fourth step and stop the iteration until no further outlier candidates can be identified.

For more details, detecting the outlier in GSTAR(1;1) model is illustrated by the following flowchart in Figure 1. Generally, space time data can contain several different types of outliers. The GSTAR model with outlier factor in general as follows:

$$\mathbf{Y}_t = \sum_{h=1}^{H} \Omega_h \mathbf{v}_h(B)\hat{I}_t^{(T_h)} + \mathbf{u}_t$$

where $\mathbf{u}_t = \Theta(B)\hat{e}_t$ is GSTAR model, $\Omega = \text{diag}(\omega_1, \omega_2, \ldots, \omega_n)$ is outlier factor, $\hat{I}_t^{(T)} = \begin{cases} \mathbf{I}, & \text{for } t \neq T \\ 0, & \text{for } t = T \end{cases}$ is indicator outlier variable at time $T$ and $\mathbf{v}_h(B) = \begin{cases} \mathbf{I}, & \text{for AO} \\ \Theta(B), & \text{for IO} \end{cases}$.

After getting the GSTAR model with outlier factor, start detecting the outlier again based on the GSTAR model with outlier factor. If no outliers are found, then stop the algorithm. Otherwise, the estimation stage is repeated, with the newly identified outliers incorporated into model (12), until no more outliers can be found and all the outlier effects have been simultaneously estimated with the time series parameters. The model (12) is called as GSTAR(1;1) with outlier factor.
5. Data Analysis
5.1. Descriptive Analysis
The data used in this paper are secondary data, that is dengue fever cases in Kalimantan Barat. There are five location used in this paper, namely Pontianak, Sintang, Sambas, Ketapang and Kapuas Hulu. Data on Dengue Fever (DF) cases was obtained from the Health Office of West Kalimantan Province. The data size used were 42 data from January 2015 to June 2018. Figure 2 shows time series plot data in each locations. Based on Figure 2, visually it can be seen the possibility of detecting outliers at several times in all of the locations. This case becomes interesting to analyze. One of the assumptions of space time analysis is the correlation of events between locations. A strong correlation between locations indicates a strong relationship between these locations. Table 1 shows correlation of dengue fever cases between locations. Based on Table 1, Kapuas Hulu and Sintang show a strong correlation, that is 0.768. this is due to the closer distance between the two locations compared to the other three locations. Whereas Ketapang and Sambas show the smallest correlation compared to other locations. This is also caused by the distance of the location.

5.2. Outlier Detection in GSTAR Model
The following is the procedure for outlier detection in GSTAR(1;1) model.
Table 2. Distance among locations.

|            | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|------------|-----------|---------|--------|----------|-------------|
| Pontianak  | 0         | 313     | 231    | 456      | 574         |
| Sintang    | 313       | 0       | 418    | 571      | 262         |
| Sambas     | 231       | 418     | 0      | 632      | 679         |
| Ketapang   | 456       | 571     | 632    | 0        | 832         |
| Kapuas Hulu | 574       | 262     | 679    | 832      | 0           |

(i) Modelling the GSTAR(1;1) model by supposing that there is no outlier detected.

In this paper, we use inverse distance weight matrix as a weight matrix. Table 2 shows the real distance (in km) in five locations. Based on the distance of each of these locations, then the inverse distance weight matrix can be notified as follows:

\[
W = \begin{bmatrix}
0 & 0.28 & 0.38 & 0.19 & 0.15 \\
0.29 & 0 & 0.21 & 0.16 & 0.34 \\
0.44 & 0.24 & 0 & 0.16 & 0.15 \\
0.33 & 0.26 & 0.24 & 0 & 0.18 \\
0.21 & 0.46 & 0.18 & 0.15 & 0
\end{bmatrix}
\]

Based on the weight matrix above, the parameter estimation of GSTAR(1; 1) model was done by using least square method. The result of parameter estimation as follows:

\[
\Phi_0 = \text{diag}[0.54, 0.58, 0.82, 0.66, 0.71] \quad \text{and} \quad \Phi_1 = \text{diag}[0.11, 0.32, -0.02, 0.41, 0.25]
\]

After getting the best model of GSTAR(1; 1), the next step is computing the residuals.

(ii) Compute the residuals \( \hat{e}_t \) and let \( \hat{\sigma}_e^2 \)

(iii) Do the iteration for \( i = 1, \ldots, 5 \) and \( T = 1, \ldots, 42 \). Compute \( \hat{\lambda}_{1,i,T} \) and \( \hat{\lambda}_{2,i,T} \)

(iv) Detect the type of the outliers by comparing \( \hat{\lambda}_{1,i,T} \) and \( \hat{\lambda}_{2,i,T} \). Define

\[
\eta_t = \max\{ |\hat{\lambda}_{1,T}|, |\hat{\lambda}_{2,T}| \} \quad \text{for} \quad t = 1, \ldots, 42 \quad \text{and} \quad i = 1, \ldots, 5
\]

(v) According to the algorithm on detecting the outliers using the Model (2), the procedure correctly identifies the time when the outlier is happened. There are five iterations on the first model, one iteration on the second model and two iterations on the third model. Table 3 shows the outlier time identified.

After getting it, do the parameter estimation based on the GSTAR(1; 1) model by adding the outlier factor which that time is identified on the iterations. Table 4 shows the result of parameter estimation in GSTAR(1;1) with outlier factor. After getting the parameter on the model, compute the Mean Square Error (MSE) for getting the best model. The analysis produces the result as presented as Table 5. Based on Table 5, adding the outlier factor to the model can increase the accuracy of the model. Therefore the best model is the third model, that is

\[
\hat{Y}_t = \hat{\Phi}_0 Y_{t-1} + \hat{\Phi}_1 W Y_{t-1} + \hat{e}_t + \hat{\Omega}_1 I_t^{(1)} + \hat{\Omega}_2 I_t^{(2)} + \hat{\Omega}_3 I_t^{(13)} + \hat{\Omega}_4 I_t^{(28)} + \hat{\Omega}_5 I_t^{(29)} + \hat{\Omega}_6 I_t^{(32)} + \hat{\Omega}_7 I_t^{(33)} + \hat{\Omega}_8 I_t^{(34)} + \hat{\Omega}_9 I_t^{(35)} + \hat{\Omega}_{10} I_t^{(36)} + \hat{\Omega}_{11} I_t^{(37)} + \hat{\Omega}_{12} I_t^{(39)} + \hat{\Omega}_{13} I_t^{(40)} + \hat{\Omega}_{14} I_t^{(42)} + \hat{\Omega}_{15} \hat{\Phi}(B) I_t^{(41)} + \hat{\Omega}_{16} \hat{\Phi}(B) I_t^{(40)}
\]

Figure 3 shows the time series plot in each locations based on the GSTAR(1; 1) model with outlier factor versus observation data.
Table 3. Detected outliers.

| Model | Iteration | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|-------|-----------|-----------|---------|--------|----------|------------|
|       |           | IO       | AO      | IO     | AO       | IO         | AO         |
| 1     | 3         | 33       | -       | 34     | -        | 29,34      | -          | 1          | -         | 33        | -          |
| 2     | -         | 1        | -       | -      | 28       | -          | 39         | -          | 36,40     | -          |
| 3     | 3         | -        | -       | -      | 37       | -          | 40         | -          | -         | -          |
| 4     | -         | 13       | -       | -      | 25       | 33         | -          | -         | -         | -          |
| 5     | -         | 33       | -       | 35     | -        | 2          | -          | -         | -         | -          |
| 2     | 1         | 1,2      | -       | 37     | -        | -          | -          | -         | -         | 37        |
| 3     | 1         | -        | -       | -      | -        | -          | -          | -         | -         | 42,32     |

Table 4. Parameter estimation of GSTAR model with outlier factor.

| Parameter | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|-----------|-----------|---------|--------|----------|------------|
|           | IO | AO | IO | AO | IO | AO | IO | AO | IO | AO |
| $\phi_{0j}$ | 0.4188 | 0.7215 | 0.7736 | 0.4584 | 0.8943 |
| $\phi_{1j}$ | 0.1693 | 0.0534 | 0.0709 | 0.4408 | -0.0893 |
| $\omega_{1j}$ | 15.9333 | - | - | - | 105.8444 | - | - | - | - |
| $\omega_{2j}$ | -11.5882 | - | - | - | - | - | - | - | - |
| $\omega_{3j}$ | - | 123.2036 | - | - | - | - | - | - | - |
| $\omega_{4j}$ | - | - | - | 25.0358 | - | - | - | - | - |
| $\omega_{5j}$ | - | - | - | - | 29.8679 | - | - | - | - |
| $\omega_{6j}$ | - | - | - | - | 18.3682 | - | - | - | - |
| $\omega_{7j}$ | 33.6417 | - | 27.8673 | - | - | 32.3056 | - | 62.6731 | - |
| $\omega_{8j}$ | - | - | 49.9030 | - | 21.7552 | - | - | - | - |
| $\omega_{9j}$ | - | - | - | - | -42.9161 | - | - | - | - |
| $\omega_{10j}$ | - | - | - | - | 29.1173 | - | - | - | - |
| $\omega_{11j}$ | - | - | - | - | -34.4772 | - | - | - | - |
| $\omega_{12j}$ | - | - | - | - | 70.1285 | - | - | - | - |
| $\omega_{13j}$ | - | - | - | - | -19.9641 | - | - | - | - |
| $\omega_{14j}$ | - | - | - | - | -54.9705 | 37.0742 | - | - | - |

Table 5. MSE of GSTAR(1; 1) model with outlier factor

| Model | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|-------|-----------|---------|--------|----------|------------|
|       | IO | AO | IO | AO | IO | AO | IO | AO | IO | AO |
| GSTAR model without outlier | 78.2619 | 130.2432 | 76.4535 | 781.0634 | 210.2904 |
| 1st Model | 54.7239 | 46.3485 | 9.2715 | 136.8514 | 56.0178 |
| 2nd Model | 30.5691 | 34.6208 | 9.2715 | 130.4089 | 36.2143 |
| 3rd Model | 30.5691 | 34.6208 | 9.2715 | 130.4089 | 19.2234 |
Figure 3. Time series plot in each locations using GSTAR(1; 1) model with outlier factor versus observation data

Figure 4. ACF residuals in each locations based on GSTAR(1; 1) model with outlier factor

6. Diagnostic Checking Model
Diagnostic checking model is an examination of whether the model assumptions fulfilled. The basic assumption of the model is residual white noise that can be done by checking the normality of residuals and independent for lag of times. Usually, Figure 4 shows that the ACF of residuals GSTAR(1; 1) model with outlier factor have no autocorrelation or independent. While the normality test of residuals is done by Kolmogorov Smirnov test, which the null hypothesis is the residuals are normally distributed. The null hypothesis is accepted if the KS p-value is greater than significance level. By using the significance level (α) is 0.05, the result of this test given in Table 6. Based on Table 6, the residuals are normally distributed.
Table 6. Kolmogorov-Smirnov test for normality of residual

| $\alpha$ | n  | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|---------|----|-----------|---------|--------|----------|-------------|
| 0.05    | 42 | 0.0870    | 0.7840  | 0.9520 | 0.6900   | 0.4650      |

7. Conclusion

Based on the algorithm in detecting outliers on GSTAR(1; 1) model, 11 times Innovative Outliers (IO) and 3 times Additive Outliers (AO) are detected. By adding an outlier factor to the GSTAR(1; 1) model, the GSTAR (1; 1) model with outlier factor is obtained better than the GSTAR (1; 1) model without adding an outlier factor. It means that the influence of outliers is very significant so that if the outlier is ignored, it can produce a biased model. Whereas if the data containing the outlier is removed, it will reduce the information from the data. Therefore, the right solution to handle data containing outliers is adding an outlier factor to the model. It means that to get the best model in modeling dengue fever cases is by using GSTAR(1; 1) model with outlier factor.

8. Acknowledgement

The authors thank to Hibah RISTEK DIKTI 2018 – 2019 for supporting funds. We also thank to the Health Office of West Kalimantan Province for the data.

9. References

[1] Mukhaiyar U and Pasaribu US 2012 ITB J. Sci. 44A(2) 179-192
[2] Nugraha RF, Setyowati S and Mukhaiyar U 2015 Prediction of oil palm production using the weighted average of Fuzzy sets concept approach. AIP Conf. Proc. 1692
[3] Yundari, Pasaribu US and Mukhaiyar U 2017 J of Mathematical and Fundamental Sciences 49(2) 136-55
[4] Yundari, Pasaribu US, Mukhaiyar U and Heriawan MN 2018 Journal of Physics:Conference Series 1028 1-8
[5] Nurhayati N, Pasaribu US and Neswan O 2012 Journal of Probability and Statistics
[6] Fadlilah NFI, Mukhaiyar U and Fahmi F 2015 The generalized STAR(1;1) modeling with time correlated errors to red-chili weekly prices of some traditional markets in Bandung, West Java. AIP Conf. Proc.
[7] Masteriana D and Mukhaiyar U 2019 Monte Carlo simulation of error assumptions in generalized STAR(1;1) model. Proceedings on the Jangjeon Mathematical Society 22(1) 43-50.
[8] Mukhaiyar U 2015 The goodness of generalized STAR in spatial dependency observations modeling. AIP Conf. Proc. 1692
[9] Sari RKN, Pasaribu US and Neswan O 2015 Applied Mathematical Sciences 9(103) 5123-37
[10] Permadi SD, Mukhaiyar U, Satyaning NLP, Soleh M and Aini Q 2018 IOP Conference Series 332
[11] Gehman AJ 2016 The effects of spatial aggregation on spatial time series modeling and forecasting, Master Thesis
[12] Jafaruddin, Indratno SW, Nuraini N, Supriatna AK and Soewono E 2015 Computational and Mathematical Methods in Medicine 1-14
[13] Chang I, Tiao GC and Chen C 1988 J Technometrics 30 193-204
[14] Wei WWS 2006 Time series analysis univariate and multivariate analysis 2nd Edition (Boston: Pearson Education)