Analysis of Generalized Space Time Autoregressive with Exogenous Variable (GSTARX) Model with Outlier Factor

U Mukhaiyar, N M Huda*, K N Sari and U S Pasaribu
Statistics Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesha 10 Bandung 40132, Indonesia
E-mail: nurainulmifatulhuda@students.itb.ac.id

Abstract. The outlier is an observation data that has different characteristics from others. Frequently, outliers are removed to improve the accuracy of the estimators. But sometimes the presence of an outlier has a specific meaning, which explanation can be lost if the outlier is removed. There are two exceptional cases from types of outliers, Innovative Outlier (IO) and Additive Outlier (AO). The presence of an outlier in the space-time model is no exception. Space-time model, not only influenced by previous observations at the same location and previous observations in a different location, or there are not only time and location dependencies, but also there are some other things that affect, which can be expressed as an exogenous variable. GSTARX is a model that combines not only time and location but also involves exogenous variables. In the GSTARX model, the presence of outliers may also be detected and may have spatial correlation at a time. In this paper, the iterative procedure in detecting outliers in the GSTARX model was introduced. Therefore data containing outliers is not deleted or ignored but still involves the outlier data by adding an outlier factor to the GSTARX model. The power of the procedure in detecting outliers is investigated by simulation experiments. The result is a GSTARX model with outlier factors that maintain the outlier factor.

1. Introduction

The Space-Time AR model, STAR, is a model combining time and location elements in space-time data. STAR model is a development of the Vector AR model, VAR, which connecting linkages to spatial relationships [1]. One of the weaknesses in the STAR model is less flexible when confronted at locations that have various characteristics. Therefore, the Generalized Space-Time AR model, GSTAR, a capable model representing heterogeneous locations. The GSTAR model was firstly introduced by Pfeifer and Deutsch [2]. This model is widely used in space-time analysis, which is to model and predict space-time data. Practically, in predicting events both in cases, univariate and multivariate often involve other variables or exogenous variables [3]. Some models are ARIMAX, VARIMAX, VARX, and VMAX. The space-time model is not only influenced by past observations at the same and different locations or dependencies among observations are not only based on time and location but also obtained from another factor referred to as the exogenous variable. In space-time analysis, it can involve exogenous variables. When the model GSTAR involves exogenous variables, then the model used is the GSTARX model [4].
The outlier is observational data that has different characteristics from other data [5]. The presence of an outlier is unpredictable due to various factors. An outlier can have a significant influence on the identification and parameter estimation that can cause bias or inaccuracy of predictions on the model [6]. The outlier is a natural occurrence and often occurs within data analysis, including in space-time data. In space-time data, outliers also have the possibility of spatial correlation in other locations. Outlier detection becomes an important thing that also needs to be done in space-time data to overcome the effects of these outliers without removing the outlier from data.

The purposes of this paper are to analyze the GSTAR model with exogenous variables, the algorithm for detecting the outlier in the GSTAR model, the GSTARX model with outlier factors, and to apply the GSTARX model with outlier factor to dengue fever cases in West Kalimantan as a study case. There are six sections in this paper. The first section is about the introduction of space-time. The second section explains the GSTAR with exogenous variables, while the explanation about the outlier and the type of outliers is in section three. In section four, we propose the GSTARX model with the outlier factor. The application of GSTARX model with the outlier factor, applied to dengue fever cases in West Kalimantan, is explained in section five. The last section is conclusions and remarks.

2. Generalized Space Time Autoregressive with Exogenous Variable (GSTARX)

GSTARX model is the development of the GSTAR model by adding the exogenous variable to the model. It is caused by the influence of other factors which called as exogenous variable.

Consider random \(Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{N,t})\), \(Y_t\) follows GSTAR\((1; 1)\) model.

\[
Y_t = \Phi_0 Y_{t-1} + \Phi_1 W Y_{t-1} + \Gamma_0 X_{t-1} + \Gamma_1 W X_{t-1} + e_t
\]

where \(Y_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{N,t})^T\), \(\Phi_0 = diag(\phi_{10}, \phi_{20}, \ldots, \phi_{N0})\) and \(\Phi_1 = diag(\phi_{11}, \phi_{21}, \ldots, \phi_{N1})\) are autoregressive parameters for response variable \(Y\), \(\Gamma_0 = diag(\gamma_{10}, \gamma_{20}, \ldots, \gamma_{N0})\) and \(\Gamma_1 = diag(\gamma_{11}, \gamma_{21}, \ldots, \gamma_{N1})\) are autoregressive parameters for exogenous variable \(X\) and \(W\) is the weight matrix, \(e_t \sim iid N(0, \sigma^2)\) and \(i, t \in N\).

If the exogenous variable in Equation (1) is removed, then that formula will be the GSTAR model. Some developments of the GSTAR model have been done by some researchers such as Mukhaiyar and Pasaribu (2012) make procedures for the model GSTAR use the IAcM (Inverse Autocovariance Matrix) approach. It applies to the monthly tea production data in West Java, Indonesia, from several tea plantations [1]. Nugraha et al. (2015) have the GSTAR model using the weighted average of the fuzzy approach in terms of weighting the GSTAR model [9]. Yundari et al (2017) also conducted related research related to error assumptions on the GSTAR model [10]. Also, Yundari, et al (2018) also conducted related research determining the spatial weight of the GSTAR model (1; 1) using the kernel function [11]. In its application, the GSTAR model was developed to predict Domestic Products Western European Gross Regional (GRDP) [12], chili prices on the market Bandung [13], and criminality [14].

Furthermore, the GSTAR model (1;1) becomes the STAR model (1;1) if all locations have the same autoregressive parameters [7]. It means the location used is assumed to be non-heterogenous. The space-time model involves several locations that assumed to have a strong correlation between locations with observed events. It means that time series models actually can be seen as a time series vector with elements in the vector represent the locations. The differences between space-time model and vector time series model are spatial correlations indicated by adding the weight matrix in the space-time model. Therefore, when the spatial factor of the space-time model in Equation (1) is removed, then that model changes to the VAR(1) model. In the VAR(1) model, \(Y\) as a vector in which the element is the locations that are considered to have no spatial correlation. Then if only one location was observed, so the VAR (1) model changes to the AR model (1).
3. Outlier

The outlier is an inconsistent observation from the data because of unpredicted and extraordinary events like natural disasters or economic crises [5]. Space-time data also have the possibility of an outlier. If in time series data, outliers are only possible in one location, then in space-time data, outliers also have the chance of spatial correlation in other locations. The emergence of outliers in space-time analysis also has the same effect as in time series analysis. One effect of the emergence of outliers is unreliable and invalid results, so outlier detection becomes an important thing that also needs to be done in space-time data to overcome the effects of these outliers. Fox (1972) in Wei (2006) firstly introduced the outlier [5]. There are several types of outliers, namely Innovative Outlier (IO), Additive Outlier (AO), Level Shift (LS), and Temporary Change (TC). AO and IO are the types of outliers we discuss.

3.1. Innovative Outlier

IO model can be expressed as [5]

$$Y_t = u_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T)} = \theta(B) \left( a_t + \omega I_t^{(T)} \right)$$

where $u_t$ is ARIMA model without outlier factor, $\omega$ is outlier parameter, $e_t \sim iid N(0, \sigma^2)$ and

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

The presence of outlier in IO affects all $Y_T, Y_{T+1}, \ldots$ pass at time $T$ along the memory of system given by $\frac{\theta(B)}{\phi(B)}$.

3.2. Additive Outlier

AO model can be expressed as [5]

$$Y_t = u_t + \omega I_t^{(T)} = \theta(B) a_t + \omega I_t^{(T)}$$

The presence of outlier in AO affects just at time $T$ [15].

The algorithm of outlier detection in the GSTARX model is adopted based on the algorithm outlier detection in the time series model introduced by Fox (1972) and the algorithm outlier detection in the GSTAR model [9]. The outlier detection algorithm is an iterative process. The iteration process will stop when all the outlier has detected. Here is the algorithm of outlier detection in the GSTARX model.

(i) Modeling GSTAR without outlier factor

Let $Y_t$ follows GSTARX$(1; 1)$

$$\Phi(B)Y_t + \Gamma(B)X_t = e_t$$

where $\Phi(B) = I - \Phi_0 B - \Phi_1 WB$ and $\Gamma(B) = -\Gamma_0 B - \Gamma_1 WB$

(ii) Compute the residuals $e_t$ and variance of residuals $\sigma^2_{ie}$, such that $\sigma^2_{ie} = \frac{1}{n} \sum_{t=1}^{n} \hat{e}_{i,t}^2$

(iii) Compute $\hat{\lambda}_{1,T}$ and $\hat{\lambda}_{2,T}$, such that

$$\hat{\lambda}_{1,T} = \frac{\bar{\epsilon}_T}{\sigma_{ie}} = \hat{e}_T$$

and

$$\hat{\lambda}_{2,T} = (I + \hat{\Pi}_1^2 + \ldots + \hat{\Pi}_{n-T}^2)^{-1} (1 - \hat{\Pi}_1 F - \ldots - \hat{\Pi}_{n-T} F^{n-T}) \hat{e}_T (\hat{\sigma}_a)^{-1} (1 + \hat{\Pi}_1^2 + \ldots + \hat{\Pi}_{n-T}^2)^{-1/2}$$

where $F$ is forward operator, $\Pi(B) = \frac{\Phi B}{\Theta(B)}$. 


(iv) Decide the type of outliers either IO or AO by defining $\eta_t = \max\{|\hat{\lambda}_{1,T}|, |\hat{\lambda}_{2,T}|\}$. If $\max\eta_t = |\hat{\lambda}_{1,T}| > C$ then IO detected at time $T$, otherwise if $\max\eta_t = |\hat{\lambda}_{2,T}| > C$ then AO detected at time $T$.

(v) Modify the residuals at the time when outlier detected by $\tilde{e}_{i,T} = \hat{e}_{i,T} - \hat{\omega}_{IT} = 0$ for IO and $\tilde{e}_T = \hat{e}_T - \hat{\omega}_{AT}\hat{\Pi}(B)\hat{I}_T^{(T)}$ for AO.

(vi) Recompute $\hat{\lambda}_{1,T}$ and $\hat{\lambda}_{2,T}$ by using modified residuals $\tilde{e}_T$ and $\tilde{\sigma}_{ie}^2$.

(vii) Do the iterative process from the third – the sixth step. The iteration will stop until no outlier identified in the model.

After getting all the time when outliers detected, add the outlier factor based on the time detected to the GSTARX(1;1) model. Do the parameter estimation based on GSTARX(1;1) model. Repeat the algorithm of outlier detection in the GSTARX model by using the new residuals from the GSTARX model with an outlier. The following flowchart in Figure 1 illustrates the algorithm of detecting the outlier in the GSTARX(1;1) model.

4. GSTARX Model with Outlier Factor

Let $T_1, T_2, ..., T_k$ are the time when the outlier detected either AO or IO. The following equation shows the formula of GSTARX model with outlier factor.

$$Y_t = \sum_{j=1}^{k} \Omega_j L_j(B) I_{T_j}^{(T)} + u_t = \Phi(B) \sum_{j=1}^{k} \Omega_j L_j(B) I_{T_j}^{(T)} + \Theta(B) e_t$$

where $u_t = \frac{\Theta(B)}{\Phi(B)} e_t$ is GSTARX model without outlier factor.

Let $Y_t$ follows GSTARX(1;1), then

$$Y_t = \Phi(0) Y_{t-1} + \Phi_1 W Y_{t-1} + \Gamma_0 X_{t-1} + \Phi_1 W X_{t-1} + \Phi(B) \sum_{j=1}^{k} \Omega_j L_j(B) I_{T_j}^{(T)} + e_t$$

The parameter of outlier factor and autoregressive from GSTARX model with outlier factor will be estimated simultaneously using Least Square method.

5. Data Analysis

5.1. Descriptive Statistics

The data used is secondary data about Dengue Fever (DF) cases and the monthly average temperature in five locations in West Kalimantan Province, those are Pontianak, Sintang, Sambas, Ketapang, and Kapuas Hulu. The distance among the locations was calculated by using Google Maps, which can be seen in Table 1. The farthest distance is 832 km, which is between Ketapang and Kapuas Hulu, while the closest distance is between Sintang and Kapuas Hulu, which is 231 km. The data about DF cases were obtained from Healthy Department in West Kalimantan Province, while the data about monthly average temperatures were obtained from the National Oceanic and Atmospheric Administration (NOAA). The data size used was 48 data from January 2015 until December 2018. In this paper, 45 observations, January 2015 – September 2018, used as in-sample data and the three others were used as out-sample data for validating in the forecasting stage. The response variable in this paper is DF cases, while the exogenous variable is the monthly average temperature in Celsius.

The primary assumption in space-time analysis is the spatial correlation of events among locations. The stronger the correlation of the events, the stronger the relationship between
**Figure 1.** The algorithm of detecting the outlier in the GSTARX(1;1) model. This algorithm is an iterative procedure. The iteration will stop if no outlier detected based on the model.

**Table 1.** The distance among the five locations (km). It is used for indicating the strong correlation between locations and calculating the weight matrix.

|        | 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           |
Table 2. The correlation of DF cases between locations. One of assumptions in space time analysis is strong correlation between locations.

|       | Pontianak | Sintang | Sambas  | Ketapang | Kapuas Hulu |
|-------|-----------|---------|---------|----------|-------------|
| Pontianak | 1        | 0.5957  | 0.4730  | 0.3980   | 0.7734      |
| Sintang   | 0.5957   | 1       | 0.3772  | 0.2854   | 0.7683      |
| Sambas    | 0.4730   | 0.3772  | 1       | 0.0944   | 0.4513      |
| Ketapang  | 0.3980   | 0.2854  | 0.0944  | 1        | 0.4980      |
| Kapuas Hulu| 0.7734   | 0.7683  | 0.4513  | 0.4980   | 1           |

those locations. Table 2 shows the correlation of DF cases among the locations. Based on Table 2, Kapuas Hulu and Sintang have a strong correlation, which is 0.78. It is because of the distance between those two locations compared to the other three locations, show the smallest correlation. The distance of the location also causes this. Nevertheless, the correlation between Pontianak and Kapuas Hulu also show a strong correlation, 0.77. Pontianak causes it as a city town of West Kalimantan. Therefore the health center is in Pontianak. Some of the infected humans will be recommended to get better treatment in Pontianak so that the probability of transmitting the virus is more significant than if the infected human gets the treatment n Kapuas Hulu.

Figure 2 (a) and (b) shows the plot of DF cases and the monthly average temperature in five locations. Figure 2 (c) shows the boxplot of DF cases. It is used to pretest the presence of outlier in DF cases. Five locations have extreme values (outliers). The existence of those outliers can cause biased results on the parameters of the GSTARX model, so the action is needed to overcome those outliers. In this case, the existence of outliers is maintained without ignoring and removing it from the data by adding the outlier factor to the GSTARX model.

5.2. GSTARX Model

Modeling the GSTARX with outlier factor is started by modeling GSTARX without outlier factor. The following is the procedure in modeling GSTARX without outlier factor:

(i) Compute the weight matrix

In this paper, the weight matrix used is the inverse distance matrix. Based on Table 1, the weight matrix can be expressed as follows

\[
W^{(1)} = \begin{bmatrix}
0 & 0.28 & 0.38 & 0.19 & 0.15 \\
0.29 & 0 & 0.21 & 0.16 & 0.34 \\
0.44 & 0.24 & 0 & 0.16 & 0.15 \\
0.33 & 0.26 & 0.24 & 0 & 0.18 \\
0.21 & 0.46 & 0.18 & 0.15 & 0
\end{bmatrix}
\]

(ii) Order Identification

The order is identified by using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Figure 3 shows the STACF and STPACF plot of dengue fever cases. This figure shows instationarity. One of the causes is the existence of outliers. Therefore, the outlier factor was added to the GSTARX model. One possible model is GSTARX(1;1) model. Therefore, the GSTARX model can be written as

\[
Y_t = (\Phi_0 + \Phi_1 W)Y_{t-1} + (\Gamma_0 + \Gamma_1 W)X_{t-1} + e_t
\]
Figure 2. Dengue Fever (DF) cases plot (left). Those plot show the possible presence of outliers. Monthly temperature plot (middle). Boxplot of DF cases (right). Those plot show that in the presence of outliers in DF cases.

(iii) Parameter Estimation
The parameters were estimated by using Least Square method. The results are

$$\Phi_0 = diag(0.54; 0.58; 0.82; 0.66; 0.71); \Phi_1 = diag(0.47; 0.58; 0.81; 0.55; 0.77)$$

$$\Gamma_0 = diag(6.62; 6.34; -2.07; 6.50; -19.65); \Gamma_1 = diag(-3.93; -7.54; 4.01; -18.89; 18.69)$$
5.3. GSTARX Model with Outlier Factor

If the parameter of GSTARX model without outlier factor is gotten, then do the iterative procedure in modeling GSTARX with outlier factor.

(i) Compute the residuals $\hat{e}_t$ and variance of residuals $\sigma^2_e$. Based on the GSTARX model without outlier factor, the matrix of deviation standard of residuals is $\sigma_e = diag(7.86; 10.85; 8.76; 23.18; 13.81)$

(ii) Do the iterative procedure for $i = 1, 2, \ldots, 5$ and $T = 1, 2, \ldots, 45$. Compute $\hat{\lambda}_{1,T}$ and $\hat{\lambda}_{2,T}$

(iii) Decide the type of outliers, either AO or IO by defining $\eta_t = max\{|\hat{\lambda}_{1,T}|, |\hat{\lambda}_{2,T}|\}$

(iv) Modify the residuals at the time which outliers are detected.

(v) The procedure correctly identifies the time when the outlier happens. Based on Table 3, there are three models gotten with each iteration on that model. There are five iterations on the first model, one iteration on the second model and two iterations on the third model.

After the outliers are identified, then add the outlier factors to the GSTARX model before and do the parameter estimation to the GSTARX model with outlier factor. Then, do the iterative procedure for detecting the outlier based on the new GSTARX model with outlier factor. The iteration will stop if we get the GSTARX model with outlier factor that free from outlier. Table 3 shows the result of outliers time identified. In this paper, there are four models, those are GSTARX model without outlier factor, first model (GSTARX model with outlier factor based on residual in GSTARX model without outlier factor), second model (GSTARX model with outlier factor based on residual in 1st model) and third model (GSTARX model with outlier factor based on residual in 2nd model). The first model consists of five iterations; the second model consists of three iterations, while the third model consists of one iteration. Table 4 shows the parameter’s result in GSTARX(1;1) model with the outlier factor based on the third model. The standard measure of prediction accuracy in forecasting method in statistics are the Mean Square Error (MSE), Root MSE (RMSE), and Mean Absolute Percentage Error (MAPE). After getting the parameter, compute that measurement from all models to compare and to decide which one is the best model. The smaller, the better. Table 5 shows the measurement from all models.

GSTARX Model is GSTARX model without outlier factor, first model is GSTARX model based on residuals in GSTARX model without outlier factor, second model is GSTARX model with outlier factor based on residuals in the first model, and the third model is the GSTARX model based on residuals in the second model.
Table 3. The time when outliers detected. The time on the first model gotten based on the residual on the GSTARX model without outlier factor. While the second model gotten based on the residual on the first model and also on the third model gotten based on the second model.

| Model       | Iteration | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|-------------|-----------|-----------|---------|--------|----------|-------------|
|             |           | IO AO     | IO AO   | IO AO  | IO AO    | IO AO       |
| First Model | 1         | 33 -      | 34 -    | 29 -   | 1 -      | 33 -        |
|             | 2         | - -       | 13 -    | 28 -   | 39 -     | - -         |
|             | 3         | - -       | 37 -    | 37 -   | 40 -     | 40 -        |
|             | 4         | - -       | 33 -    | - 44   | 33 -     | - -         |
|             | 5         | - -       | - -     | 25 -   | - -      | - -         |
| Second Model| 1         | - - -     | 45 -    | 37 -   | 42,45 -  | -           |
|             | 2         | - - -     | - -     | 2 -    | 32 -     | -           |
|             | 3         | - 1 -     | - -     | - -    | 42 -     | - -         |
| Third Model | 1         | - - -     | - -     | 34 -   | - -      | - -         |

Table 4. Parameter of GSTARX model with outlier factor. The parameters are estimated by using Least Square method based on the third model.

| Parameter | Pontianak ($i = 1$) | Sintang ($i = 2$) | Sambas ($i = 3$) | Ketapang ($i = 4$) | Kapuas Hulu ($i = 5$) |
|-----------|---------------------|-------------------|------------------|---------------------|-----------------------|
|           | IO AO               | IO AO             | IO AO            | IO AO               | IO AO                 |
| $\phi_{i0}$ | 0.34               | 0.67              | 0.73             | 0.65               | 0.92                  |
| $\phi_{i1}$ | 0.19               | 0.13              | 0.03             | 0.64               | -0.16                 |
| $\gamma_{i0}$ | 2.99              | 3.55              | -1.54            | -1.23              | -1.17                 |
| $\gamma_{i1}$ | -3.15              | -5.57             | 3.53             | -4.68              | 2.18                  |
| $\omega_{i33}$ | 33.96              | 25.62             | -29.72           | -61.73             | -                     |
| $\omega_{i41}$ | -15.93             | -105.84           | -23.31           | -40.81             | -                     |
| $\omega_{i34}$ | -46.82             | -23.31            | -17.12           | -42.19             | -                     |
| $\omega_{i13}$ | -24.89             | -                 | -17.12           | -42.19             | -                     |
| $\omega_{i37}$ | -22.74             | -17.12            | -31.35           | -                   | -                     |
| $\omega_{i29}$ | -16.68             | -17.12            | -24.58           | -                   | -                     |
| $\omega_{i38}$ | -                   | -17.12            | -23.31           | -6.71              | -                     |
| $\omega_{i44}$ | -16.68             | -12.69            | -8.71            | -                   | -                     |
| $\omega_{i25}$ | -                   | -                 | 22.48            | -                   | -                     |
| $\omega_{i45}$ | -                   | -                 | -32.99           | 42.87              | 38.03                 |
| $\omega_{i39}$ | -                   | -                 | 63.97            | -                   | -                     |
| $\omega_{i40}$ | -                   | -                 | -42.87           | 38.39              | -                     |
| $\omega_{i2}$ | -                   | -                 | -21.35           | -                   | -                     |
| $\omega_{i42}$ | -                   | -                 | -33.51           | -21.37             | -                     |
| $\omega_{i36}$ | -                   | -                 | -                | -30.36             | -                     |
| $\omega_{i32}$ | -                   | -                 | -                | 17.98              | -                     |
Table 5. The measurement of model accuracy. It is used for comparing the models based on the residuals of each model and finding the best model.

| Model          | The Measurement | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|----------------|-----------------|-----------|---------|--------|----------|-------------|
|                | MSE             | 61.83     | 117.76  | 76.73  | 537.21   | 190.77      |
|                | RMSE            | 7.86      | 10.85   | 8.76   | 23.18    | 13.81       |
| GSTARX Model   | MAPE            | 6.14      | 1.29    | 1.07   | 0.89     | 0.58        |
| First Model    | MSE             | 37.53     | 33.95   | 14.01  | 115.84   | 46.31       |
|                | RMSE            | 6.10      | 5.83    | 3.74   | 10.76    | 6.81        |
|                | MAPE            | 5.33      | 1.27    | 0.63   | 0.67     | 0.41        |
| Second Model   | MSE             | 30.21     | 33.95   | 10.92  | 45.52    | 19.19       |
|                | RMSE            | 5.49      | 5.83    | 3.30   | 6.75     | 4.38        |
|                | MAPE            | 5.15      | 1.28    | 0.61   | 0.44     | 0.23        |
| Third Model    | MSE             | 30.21     | 33.95   | 10.92  | 45.52    | 19.19       |
|                | RMSE            | 5.49      | 5.83    | 3.30   | 5.94     | 4.38        |
|                | MAPE            | 5.15      | 1.28    | 0.61   | 0.27     | 0.23        |

Table 6. The comparison time when the outliers detected. The sum of outliers detected in boxplot is more than in AO and IO. More complex the model, less outliers detected.

| Location       | Boxplot       | Outlier detection based on GSTARX model |
|----------------|---------------|----------------------------------------|
|                |               | AO                                     |
| Pontianak      | 1,33,34,36,37 | 33                                     |
|                | 34,35,36      | 1                                      |
| Sintang        | 28, 29, 30, 31, 32, 34, 36 | 25,34 |
| Ketapang       | 1, 2, 39, 40 | 2,40                                   |
| Kapuas Hulu    | 33, 34, 36   | 36                                     |

Based on the Table 5, the third model shows the smallest MSE, RMSE and MAPE. Therefore the third model is the best model, that is

\[
\hat{Y}_t = \hat{\Phi}_0 Y_{t-1} + \hat{\Phi}_1 WY_{t-1} + \hat{\Gamma}_0 X_{t-1} + e_t + \Omega_1 I_t^{(33)} + \Omega_2 I_t^{(34)} + \Omega_3 I_t^{(13)} + \Omega_4 I_t^{(37)} + \Omega_5 I_t^{(29)} + \Omega_6 I_t^{(28)} + \Omega_7 I_t^{(11)} + \Omega_8 I_t^{(39)} + \Omega_9 I_t^{(40)} + \Omega_{10} \hat{\Phi}(B) I_t^{(34)} + \Omega_{11} \hat{\Phi}(B) I_t^{(44)} + \Omega_{12} \hat{\Phi}(B) I_t^{(25)} + \Omega_{13} \hat{\Phi}(B) I_t^{(33)} + \Omega_{14} \hat{\Phi}(B) I_t^{(36)} + \Omega_{15} I_t^{(45)} + \Omega_{16} I_t^{(42)} + \Omega_{17} I_t^{(32)} + \Omega_{18} \hat{\Phi}(B) I_t^{(2)}
\]

(8)

There are two ways to detect the outlier in this paper, using boxplot and the algorithm of outlier detection based on the GSTARX model. Using boxplot means ignoring the spatial dependencies in data, whereas utilizing the algorithm of outlier detection based on the GSTARX model covers the spatial dependencies in data. Table 6 shows the comparison time of detected outliers using two different methods.
Table 7. Diagnostic test for normality test. GSTARX model with outlier factor shows that the residuals normally distributed.

| Location       | Without Outlier | With Outlier |
|----------------|-----------------|--------------|
| Pontianak      | 0.039           | 0.175        |
| Sintang        | 0.039           | 0.548        |
| Sambas         | 0.008           | 0.079        |
| Ketapang       | 0.013           | 0.256        |
| Kapuas Hulu    | 0.026           | 0.609        |

Table 8. The result of forecasting the number of dengue fever cases. There are two model used for forecast

| Time           | Method            | Pontianak | Sintang | Sambas | Ketapang | Kapuas Hulu |
|----------------|-------------------|-----------|---------|--------|----------|-------------|
| October 2018   | Without outlier factor | 11        | 27      | 22     | 26       | 52          |
|                | With outlier factor | 13        | 19      | 21     | 5        | 45          |
| November 2018  | Without outlier factor | 8         | 33      | 20     | 41       | 48          |
|                | With outlier factor | 18        | 31      | 27     | 31       | 50          |
| December 2018  | Without outlier factor | 8         | 37      | 19     | 47       | 52          |
|                | With outlier factor | 22        | 40      | 32     | 52       | 53          |

5.4. Diagnostic Test
The diagnostic test has become a standard tool for model identification before forecasting the data. The underlying assumptions which have fulfilled by the model are residual white noise (normality test). The normality test can be evaluated by the Kolmogorov Smirnov test. The null hypothesis is that the residuals are normally distributed. If the K-S p-value is greater than the significance level, then the null hypothesis is accepted. Table 7 shows the K-S p-value for both GSTARX(1;1) model without outlier factor and with outlier factor.

By using the significance level is 0.05, the residuals of GSTARX model without outlier factor are not normally distributed, while the residuals of GSTARX model with outlier factor are normally distributed. So that, the GSTARX(1;1) model with outlier factor fulfill the assumptions in diagnostic test.

5.5. Forecasting
Forecasting is the primary goal of space-time modeling. The resulting model is expected to represent historical data and produce good forecasting. The principle of forecasting is based on historical patterns owned. In the case of DF cases, forecasting is made for the next three times based on the model in Equation (7). Therefore the results in Table 8 are obtained. Table 8 also shows the comparison forecasting DF cases without and with the outlier factor.

6. Conclusions
There are three conclusions based on the result above. The first is the outlier detection algorithm in the GSTARX model is an iterative process. This algorithm was adopted from the outlier
detection algorithm in the time series model and GSTAR model, so the stages used are still the same. The difference between both of them is the exogenous variable X in the model, so the process of outlier detection also involves these variables.

The second is the parameters of GSTARX(1;1) model with outlier factor are estimated using Least Square (LS) method. Addition of exogenous variables X and outlier factor in the GSTAR(1;1) model does not change the linearization structure in the LS method. Therefore, the LS method is still appropriate in the estimation of model parameters GSTARX(1;1) with the outlier factor.

The third is in the process of detecting the outliers. Most outliers are detected after 28th month (April 2017), also in the 33rd month (September 2017), outliers were detected in all locations that were the object of research. Ketapang is the location with the highest number of outliers. While Pontianak is the location with the least amount of time, an outlier is detected. A large amount when the outlier is detected indicates the number of causing phenomena DF cases can increase or decrease extremely.

7. Acknowledgement
The authors thank to the RISTEK-DIKTI grant, ”Penelitian Dasar Unggulan Perguruan Tinggi (PDUPT)” 2019, The ITB’s program of Research, Community Service and Innovation (P3MI ITB) and ITB Research Program for supporting funds. We also thank to the Health Office of West Kalimantan Province for the data.

8. References
[1] Mukhaiyar U and Pasaribu U S 2012 A new procedure for generalized STAR modeling using IAcM approach. *ITB J. Sci.*
[2] Pfeifer P E and Deusch S J 1980 A three-stage iterative approach for space-time modeling. *Technometrics* 22 397-408.
[3] Suhartono and Atok R 2006 Pemilihan bobot lokasi yang optimal pada model GSTAR *Presented at National Mathematics Conference XIII*, Semarang, Universitas Negeri Semarang.
[4] Astuti D, Ruchjana B N and Soemartini 2012 Analisis data ekspsor Crude Palm Oil (CPO) dengan pendekatan model space-time *The Asian Mathematical Conference*.
[5] Wei W W S 2006 *Time series analysis univariate and multivariate analysis ed 2* (Boston: Pearson Education, Inc).
[6] Tsay R S 1986 Time series model specification in the presence of outliers *J. of American Statistical Association* 132-41.
[7] Ruchjana B N, Borovkova S A, Lopuhaa H P 2012. Least squares estimation of Generalized Space Time Autoregressive (GSTAR) model and its properties The 5th International Conference on Research and Education in Mathematics *AIP Conf. Proc.* 1450: 61-4.
[8] Fox A J 1972 Outliers in time series *J. of the Royal Statistical Society* 350-363.
[9] Nugraha R F, Setyowati S and Mukhaiyar U 2015 Prediction of oil palm production using the weighted average of Fuzzy sets concept approach. *AIP Conf. Proc.* 1692.
[10] Yundari, Pasaribu U S and Mukhaiyar U 2017 Error assumptions on generalized STAR model. *J. of Mathematical and Fundamental Sciences* 49(2) 136-55.
[11] Yundari, Pasaribu U S, Mukhaiyar U and Heriawan M N 2018 Spatial weight determination of GSTAR(1;1) model by using kernel function *J.Phys.:Conf. Series* 1028 1-8.
[12] Nurhayati N, Pasaribu U S and Neswan O 2012 Application of generalized STAR model on GDP data in West European Countries. *J. of Probability and Statistics* 2012.
[13] Fadilah N F I, 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.* 1692.
[14] 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.
[15] Masteriana D, Riani M I and Mukhaiyar U 2019 Spatial weight determination of GSTAR(1;1) model by using kernel function *J.Phys.:Conf. Series* 1245 012046 1-9.