

AN ITERATIVE PROCEDURE FOR OUTLIER DETECTION IN GSTAR(1;1) MODEL

Nur'ainul Miftahul Huda¹, Utriweni Mukhaiyar², Nurfitri Imro'ah^{3*}

¹Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Tanjungpura,
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Indonesia

²Statistics Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung,
Jl. Ganessa 10, Bandung, 40132, Indonesia

³Statistics Department, Faculty of Mathematics and Natural Sciences, Universitas Tanjungpura,
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Indonesia

Corresponding author's e-mail: ^{3*} nurfitriimroah@math.untan.ac.id

Abstract. Outliers are observations that differ significantly from others that can affect the estimation results in the model and reduce the estimator's accuracy. To deal with outliers is to remove outliers from the data. However, sometimes important information is contained in the outlier, so eliminating outliers is a misinterpretation. There are two types of outliers in the time series model, Innovative Outlier (IO) and Additive Outlier (AO). In the GSTAR model, outliers and spatial and time correlations can also be detected. We introduce an iterative procedure for detecting outliers in the GSTAR model. The first step is to form a GSTAR model without outlier factors. Furthermore, the detection of outliers from the model's residuals. If an outlier is detected, add an outlier factor into the initial model and estimate the parameters so that a new GSTAR model and residuals are obtained from the model. The process is repeated by detecting outliers and adding them to the model until a GSTAR model is obtained with no outliers detected. As a result, outliers are not removed or ignored but add an outlier factor to the GSTAR model. This paper presents case studies about Dengue Hemorrhagic Fever cases in five locations in West Kalimantan Province. These are the subject of the GSTAR model with adding outlier factors. The result of this paper is that using an iterative procedure to detect outliers based on the GSTAR residual model provides better accuracy than the regular GSTAR model (without adding outliers to the model). It can be solved without removing outliers from the data by adding outlier factors to the model. This way, the critical information in the outlier id is not lost, and an accurate ore model is obtained.

Keywords: additive, GSTAR, innovative, outliers.

Article info:

Submitted: 27th June 2022

Accepted: 20th August 2022

How to cite this article:

N. M. Huda, U. Mukhaiyar and N. Imro'ah, "AN ITERATIVE PROCEDURE FOR OUTLIER DETECTION IN GSTAR(1;1) MODEL", *BAREKENG: J. Math. & App.*, vol. 16, iss. 3, pp. 975-984, September, 2022.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).
Copyright © 2022 Author(s)

1. INTRODUCTION

Unexpected extraordinary observations that look different from most observations in the data set are often encountered in various kinds of data analysis. Outliers are observations that differ significantly from others. Its presence is unexpected because there are so many factors that can cause its presence. Commonly, outliers are removed. In this paper, outliers were not released, but the way to deal with outliers is by adding an outlier factor to the model. In the time series model, Auto-regressive Integrated Moving Average (ARIMA). Chang et al. have developed the iterative procedure for detecting the outliers and overcoming the presence of outliers by adding the outliers factor to the ARIMA model.

Space-time analysis is no exception, for outliers can be detected. In this paper, we proposed the iterative procedure for detecting the outliers in the space-time model, the generalized Space-Time Autoregressive (GSTAR) model. The GSTAR model is a space-time model used to model and forecast spatial time series data. Nowadays, the GSTAR model is growing in Indonesia. Some development of the GSTAR(1;1) model has been done by some researchers, such as making a new procedure for Generalized STAR modeling using Inverse Auto-covariance Matrix (IACM) approach [1]. In terms of weighting on the GSTAR model. GSTAR has been modeled using the weighted average of fuzz sets concept approach and applied that model to oil palm production [2]. Yundari et al. (2017) researched error assumptions in the GSTAR model [3]. Recently, Yundari et al. (2018) did the research on spatial weight determination of the GSTAR(1;1) model by using the Kernel function [4]. In application, the GSTAR model is rapidly used to forecast Gross Domestic Product (GDP) in West European [5], chili price in Bandung's market [6], and criminality [7]. A combination of GSTAR modeling and variogram of spatial analysis was conducted [8]. However, spatial analysis is an older science than space-time analysis, the development of spatial models is still ongoing with several applications. The bootstrap approach estimated the parameters of the isotropic semivariogram [9]. Moreover, the effect of spatial aggregation on the space-time model was investigated [10].

Dengue fever is transmitted through the *Aedes Aegypti* mosquitos. *Aedes Aegypti* is one of the dengue mosquitoes that likes a warm climate. The case study in this paper is the cases of dengue fever sufferers in six districts in West Kalimantan. Climate change has occurred significantly. This significant increase can indicate outliers detection in the GSTAR model.

This paper proposes the iterative procedure for outlier detection in the GSTAR model and applies it to the study case. This paper is divided into six sections. Section two briefly explains the GSTAR model. The definition and types of outliers are discussed in section three. Then the iterative procedure for outlier detection is discussed in section four. The application of the GSTAR model by adding an outlier factor is discussed in section five. Conclusions and remarks are put forward in section six.

2. RESEARCH METHOD

2.1 GSTAR Model

GSTAR model is the development of the STAR model. In the STAR model, all locations have the same autoregressive parameters, so the locations used are assumed homogenous. It causes that this model only can be used for uniform locations. In reality, we have a heterogeneous location, so a model that can solve this problem is needed. The GSTAR model enables capturing a phenomenon with heterogeneous characteristics locations, and the parameters for each location are different from each other. Consider a random variable $Y_t = (Y_{1,t}, Y_{2,t}, \dots, Y_{n,t})$, Y_t follows the GSTAR($p; \lambda_1, \lambda_2, \dots, \lambda_p$) model. It can be stated as [11]

$$Y_t = \left(\sum_{k=1}^p \sum_{l=0}^{\lambda_k} \Phi_{kl} W^{(l)} Y_{t-k} \right) + e_t$$

where Φ_{kl} is the diagonal matrix of parameters in GSTAR model, W^l is weight matrix, $e_t = (e_{1,t}, e_{2,t}, \dots, e_{n,t})$ is the matrix residuals in GSTAR model, and $e_t \sim N(0, \sigma_e^2)$.

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

$$Y_t = \Phi_0 Y_{t-1} + \Phi_1 W Y_{t-1} + e_t$$

The model in Eq. (1) is called as the GSTAR(1;1) model without outlier factor.

The uniqueness of the GSTAR model is the weight matrix. The weight is defined based on the correlation between the location to one location and another. The following are the weight of the GSTAR model [4]

1. **Uniform Weight** gives the same weight for each location. Therefore, this weight is often used on homogenous data or has the exact distance between locations. This formulation calculates the values of the uniform location weights

$$w_{ij}^{(l)} = \frac{1}{n_i^{(l)}}$$

where $w_{ij}^{(l)}$ is the weight between locations i and j , $n_i^{(l)}$ is the number of locations adjacent to the i – location in spatial lag l .

2. **Binary Weight** only has zero and one value. The correlation between two geographically adjacent cities is defined by $w_{ij} = 1$. Whereas if geographically far apart is defined $w_{ij} = 0$.
3. **Inverse Distance Weight** is based on the actual distance between locations. The weight calculation is obtained from the normalization of the actual inverse distance results. The first step is calculating the actual distance between locations, \mathbf{D} , i.e.

$$\mathbf{D} = [d_{ij}] = \begin{bmatrix} 0 & d_{12} & \dots & d_{1N} \\ d_{21} & 0 & \dots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \dots & 0 \end{bmatrix}$$

where d_{ij} is the distance between locations i and j . \mathbf{D} is a symmetrical matrix.

Then the \mathbf{D} matrix is standardized in \mathbf{W} form with $\sum_{j=1}^N w_{ij}^{(l)} = 1$ where $i \neq j$. Generally the inverse weight of distance for each location is stated by

$$w_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{j=1}^N \frac{1}{d_{ij}}}$$

where $i \neq j$, $\sum_{j=1}^N w_{ij}^{(l)} = 1$, and $\sum_{i=1}^N \sum_{j=1}^N w_{ij}^{(l)} = N$.

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

2.2 ARIMA(1,0,0) and GSTAR(1;1) Model

Let Y_t is the sequence of random variables and follows the ARIMA(1,0,0) model, then Y_t can be defined as

$$Y_t = \phi Y_{t-1} + e_t$$

where ϕ is an autoregressive parameter and e_t is error term at time t .

Then if we have \mathbf{Y}_t where \mathbf{Y}_t is $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$ are the sequence of random variables, then we can define

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ Y_{n,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 & \dots & 0 \\ 0 & \phi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_n \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ \vdots \\ Y_{n,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{n,t} \end{bmatrix}$$

This model is VAR(1) model. Further, if we add the spatial factor to VAR(1) model, then we get the GSTAR(1;1) model (see Eq. 1). The GSTAR(1;1) model can be stated as VAR(1) model with

$$\Phi = \Phi_0 + \Phi_1 \mathbf{W}$$

It can conclude that the structure of AR(1) model is not much different with GSTAR(1;1).

2.3 Outlier

Outliers are inconsistent observation data due to unforeseen events such as turbulent political or economic crises [12]. Outliers can cause unreliable and invalid estimators, so outlier detection needs to be done. Outlier detection was firstly introduced by Fox (1972) [13]. Outliers consist of two types

1. **Additive Outlier** (AO) is an event that affects to time series data at a time. The definition of additive outlier model as follows [13]:

$$Y_t = \begin{cases} u_t & , t \neq T \\ u_t + \omega & , t = T \end{cases}$$

or

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

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

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

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

2. **Innovative Outlier (IO)** gives a more complicated effect compared to the other three types of outliers. The definition of innovative outlier model as follows [13]:

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

If the outlier is AO, the effect given only occurs at the time of the T observation. But if the outlier is IO, the effect given at the entire observation Y_T, Y_{T+1}, \dots

Generally, time series data can contain several different types of outliers. The outlier model in general as follows [12]:

$$Y_t = \sum_{h=1}^H \omega_h v_h(B) I_t^{(T)} + u_t$$

where u_t is ARIMA model without outlier factor such that $u_t = \frac{\theta(B)}{\phi(B)} a_t$ and $v_h(B) = \begin{cases} 1 & , AO \\ \frac{\theta(B)}{\phi(B)} & , IO \end{cases}$

After getting the time series model with the outliers factor, start the iterative procedure for outliers detection based on the time series model with an outliers factor. If no outliers are found, then stop the iterative procedure. Otherwise, the estimation stage is repeated, with the newly identified outliers incorporated into the model (Eq. 2), until no more outliers can be found and all the outlier effects have been simultaneously estimated with the time series parameters.

This paper adopts the types of outliers, iterative procedures of outliers detection, and the GSTAR model with outlier factors from the time series model. In that case, these observations develop into several locations and have spatial correlations among locations. Therefore, outliers may also be detected in the GSTAR model and correlate locations (spatial correlation).

2.4 An Iterative Procedure for Outlier Detection in GSTAR Model

This procedure is developed by an outlier’s iterative procedure for ARIMA model. The procedure begins with modelling 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, i.e.

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

$$\Phi(B) Z_t = e_t$$

where $\Phi(B) = I - \Phi_0 B - \Phi_1 WB$, B is the backshift operator such that

$$B^k Z_t = Z_{t-k}$$

The following is an iterative procedure for outlier detection in GSTAR(1;1) model

1. Compute the residual \hat{e}_t , where $\hat{e}_t = (e_{1,t}, e_{2,t}, \dots, e_{N,t})$ and let $\hat{\sigma}_e^2$ be the estimate of variance residual $\hat{\sigma}_e^2$, such that

$$\hat{e}_t = \hat{Z}_t - Z_t$$

$$\hat{\sigma}_e^2 = \frac{1}{m} \sum_{t=1}^m \hat{e}_t$$

2. Compute $\hat{\lambda}_{1,i,T}$ and $\hat{\lambda}_{2,i,T}$, where

$$\hat{\lambda}_{1,i,T} = \frac{\hat{\omega}_1}{\hat{\sigma}_e} = \frac{\hat{e}_T}{\hat{\sigma}_e} = (\hat{\sigma}_e)^{-1} \hat{e}_T$$

and

$$\hat{\lambda}_{2,i,T} = \frac{\hat{\omega}_A}{\hat{\rho}\hat{\sigma}_e} = \frac{\hat{\rho}^2(\mathbf{I} - \hat{\Pi}_1F - \dots - \hat{\Pi}_{N-T}F^{N-T})\hat{\mathbf{e}}_T}{\hat{\rho}\hat{\sigma}_e}$$

where $\hat{\rho}^2 = (\mathbf{I} + \hat{\Pi}_1^2 + \dots + \hat{\Pi}_{N-T}^2)^{-1}$, $\hat{\Pi}_1 = \Phi_0 + \Phi_1\mathbf{W}$ and $\hat{\Pi}_2, \hat{\Pi}_3, \dots = \mathbf{0}$.

The $\hat{\lambda}_{1,i,T}$ and $\hat{\lambda}_{2,i,T}$ are used for classifying the type of outliers at the next step.

3. Define

$$\eta(t) = \max\{|\hat{\lambda}_{1,i,T}|, |\hat{\lambda}_{2,i,T}|\}$$

for $t = 1, 2, \dots, m$.

If $\max \eta(t) = |\hat{\lambda}_{1,i,T}| > C$ then the IO is detected at time T . The impact of IO at time T , $\hat{\omega}_{iIT}$, can be stated as

$$\hat{\omega}_{iIT} = \hat{\mathbf{e}}_T$$

where i is location, $i = 1, \dots, N$.

While if $\max \eta(t) = |\hat{\lambda}_{2,i,T}| > C$ then the type of outliers detected is AO. The impact of AO at time T , $\hat{\omega}_{iAT}$, can be defined as

$$\hat{\omega}_{iAT} = \hat{\rho}^2\hat{\Pi}(F)\hat{\mathbf{e}}_T$$

Then eliminate the effect by defining a modified residual at time T , $\check{\mathbf{e}}_T$. Due to the two types of outliers, there were also two ways to compute the new residual. For IO, the change is just affected at time T , so that

$$\check{\mathbf{e}}_T = \begin{cases} \hat{\mathbf{e}}_T - \hat{\omega}_{iIT} & , t = T \\ \hat{\mathbf{e}}_t & , \text{otherwise} \end{cases}$$

While for AO, the change is NOT just affected at time T , so that

$$\check{\mathbf{e}}_T = \begin{cases} \hat{\mathbf{e}}_T - \hat{\omega}_{iAT}\hat{\Pi}(B)\mathbf{I}_t^{(T)} & , t \geq T \\ \hat{\mathbf{e}}_t & , 1 \leq t \leq T \end{cases}$$

A new estimator $\check{\sigma}_{ie}^2$, is computed from the modified residuals, i.e.

$$\check{\sigma}_{ie}^2 = \frac{1}{n} \sum_{t=1}^n \check{\mathbf{e}}_{i,t}^2$$

4. If an AO/IO is identified on previous step, recompute $\hat{\lambda}_{1,i,T}$ and $\hat{\lambda}_{2,i,T}$ based on the same initial estimates of the parameters, but using the modified residuals $\check{\mathbf{e}}_T$ and $\check{\sigma}_{ie}^2$.
5. Repeat the third – the fourth step and stop the iteration until no further outlier candidates can be identified

The following flowchart in Fig. 1 illustrates an iterative outlier detection procedure in GSTAR(1,1) model. This output of this procedure is the time when outliers are detected. Generally, space-time data can contain several different types of outliers. To overcome the main problem in this paper, so we need to add the outlier factor, $\omega_h\mathbf{v}_h(B)\mathbf{I}_t^{(T_h)}$ for h is number of outliers and T_h is the time T when outliers are detected for index h , to the GSTAR(1,1) model. The GSTAR model with outlier factors, in general, is as follows:

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

where $\mathbf{u}_t = \frac{\Theta^*(B)}{\Phi^*(B)}\mathbf{e}_t$ is the GSTAR model, $\Phi^*(B) = \mathbf{I} - \Phi_0^*B - \Phi_1^*\mathbf{W}B$, $\mathbf{I}_t^{(T)} = \begin{cases} \mathbf{I} & t \neq T \\ \mathbf{0} & t = T \end{cases}$ is indicator

outlier variable at time T and $\mathbf{v}_h(B) = \begin{cases} \mathbf{I} & AO \\ \frac{\Theta^*(B)}{\Phi^*(B)} & IO \end{cases}$.

After getting the GSTAR model with the outliers factor, restart the iterative procedure for outlier detection based on the GSTAR model with an outlier factor. If no outliers are found, then stop the iterative procedure. Otherwise, the estimation stage is repeated, with the newly identified outliers incorporated into the model (Eq. 5), until no more outliers can be found and all the outlier effects have been simultaneously estimated with the time series parameters. The model (Eq. 5) is called GSTAR(1; 1) with an outliers factor.

3. RESULT AND DISCUSSION

3.1 Descriptive Analysis

The data used in this paper are secondary, i.e., dengue fever cases in Kalimantan Barat. This paper uses five locations, namely A, B, C, D, and E city. Data on dengue fever cases were obtained from the Health Office of West Kalimantan Province from January 2015 to June 2018 (42 data)[14]. Fig. 2 shows time series plot data in each location. Those plots show that in the presence of outliers in DF cases. The red bullets are the outliers detected based on the boxplot. The descriptive statistic, such as mean, first, second, and the third quartile, is also shown in the boxplot. Visually, the possibility of detecting outliers several times in each location can be seen. More outliers were detected from the beginning of 2018 until mid-2018; this case becomes interesting to analyze. One of the assumptions of space-time analysis is the correlation of events between locations. Table 1 shows the correlation of dengue fever cases between locations. The GSTAR model with outlier factors in general is as follows:

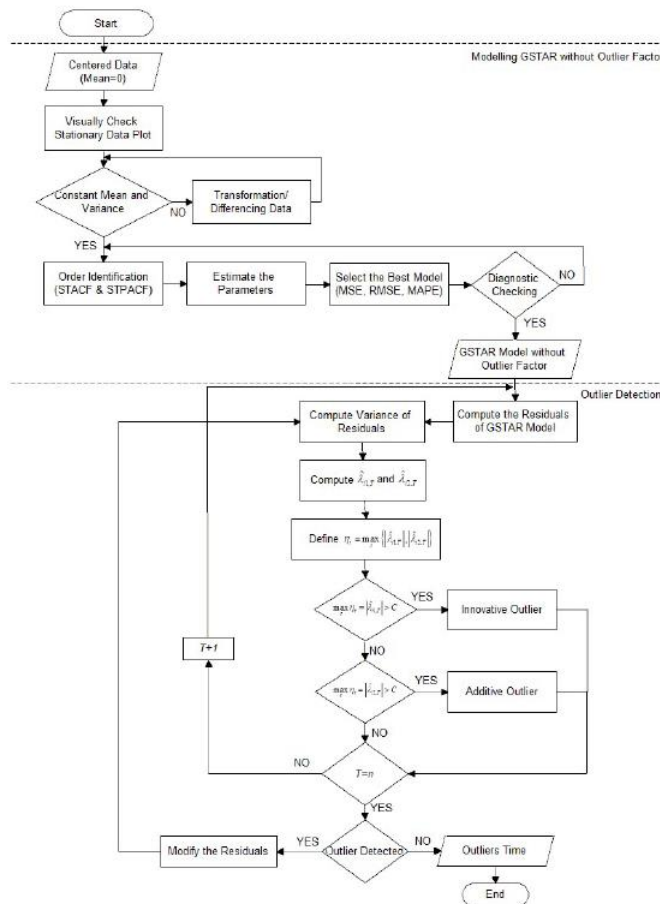
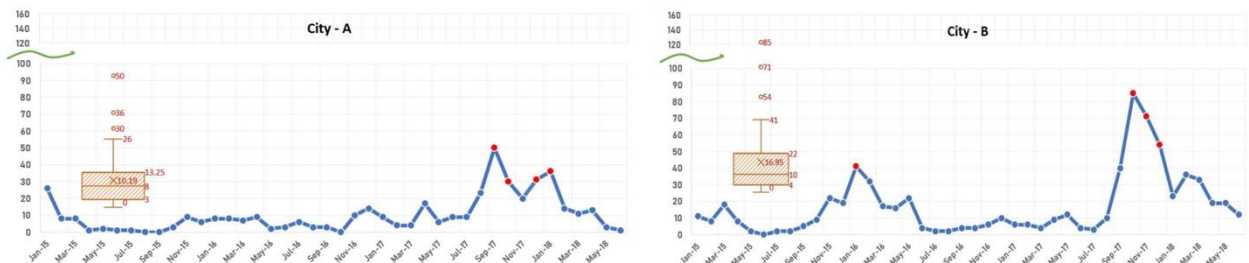


Figure 1. An Iterative Procedure for Outlier Detection in GSTAR(1;1) Model



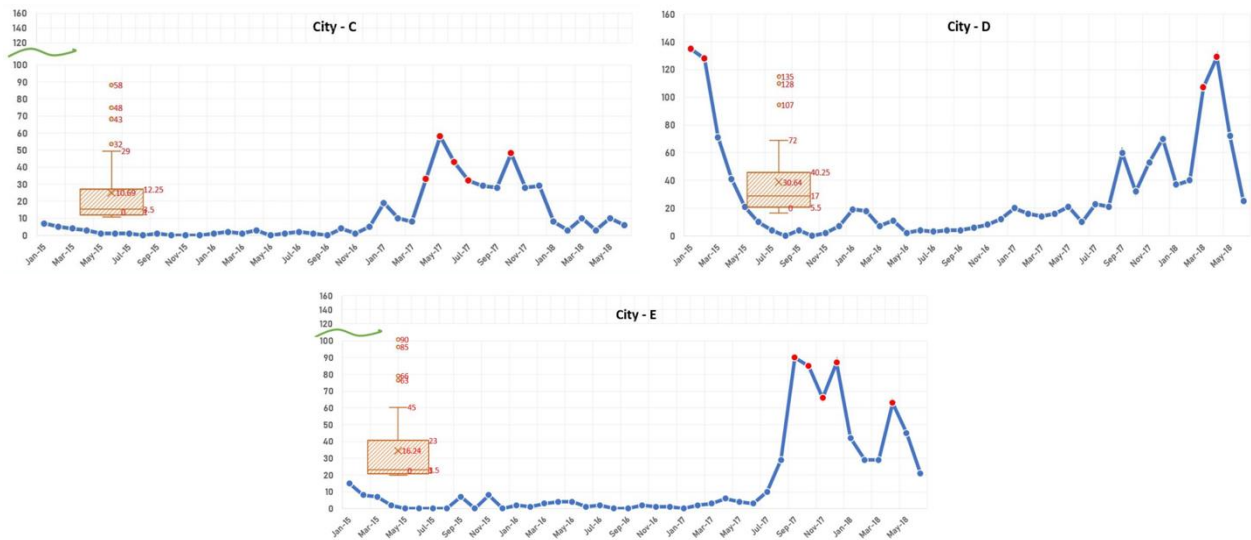


Figure 2. Dengue Fever (DF) Cases Plot. Boxplot of DF Cases (Orange)

Table 1. Correlation of Dengue Fever Cases and Distance among Locations, written as “Corr.(Dist.)”.

	A	B	C	D	E
A	1.00 (0)	0.59 (313)	0.47 (231)	0.39 (456)	0.77 (574)
B	0.59 (313)	1.00 (0)	0.37 (418)	0.28 (571)	0.76 (262)
C	0.47 (231)	0.37 (418)	1.00 (0)	0.09 (632)	0.45 (679)
D	0.39 (456)	0.28 (571)	0.09 (632)	1.00 (0)	0.49 (832)
E	0.77 (574)	0.76 (262)	0.45 (679)	0.49 (832)	1.00 (0)

Based on Table 1, B and E show a strong correlation, i.e., 0.76. In contrast, C and D show a minor correlation compared to other locations. It is due to the closer distance between the two locations than the other three locations. The distance of the location also causes this. The higher the correlation, the stronger the relationship between those cases. The further the distance, the smaller the correlation.

3.2 An Iterative Procedure for Outlier Detection in GSTAR(1;1) Model

The following is the procedure for outlier detection in GSTAR(1;1) model.

1. Modelling the GSTAR(1;1) model by supposing that there is no outlier detected. The weight matrix used is the inverse distance weight matrix. This model assumes that the dengue fever cases in one location are affected by distance or closeness with other locations. Table 1 also shows five locations' actual distance (in km). D and E are locations that have the farthest distance compared to other locations, namely 872 km. In contrast, A and C are locations with the closest distance to other locations, namely 231 km. Then the inverse distance weight matrix can be notified as follows:

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

The parameter estimation of GSTAR(1;1) model was done by using least square method, those are

$$\Phi_0 = \text{diag}(0.54; 0.58; 0.83; 0.66; 0.71)$$

$$\Phi_1 = \text{diag}(0.11; 0.32; -0.02; 0.41; 0.25)$$

2. Compute the residuals $\hat{\epsilon}_t$ and $\hat{\sigma}_\epsilon^2$
3. Do the iterative procedure for $i = 1, 2, \dots, 5$ and $T = 1, 2, \dots, 42$, compute $\hat{\lambda}_{1,i,T}$ and $\hat{\lambda}_{2,i,T}$ (Eq.(3) and (4))
4. Detect the type of the outliers by comparing $\hat{\lambda}_{1,i,T}$ and $\hat{\lambda}_{2,i,T}$. Define $\eta(t) = \max\{|\hat{\lambda}_{1,i,T}|, |\hat{\lambda}_{2,i,T}|\}$ for $t = 1, 2, \dots, m$ and $i = 1, 2, \dots, 5$.
5. According to the iterative procedure for detecting the outliers using the model in Eq. (1), the procedure correctly identifies the time when the outlier has happened. We have four models: GSTAR(1;1) model

without outlier factor, first, second, and third. Afterward, do the parameter estimation outlier factor based on residual in GSTAR(1;1) model without outlier factor, one iteration on the second model (GSTAR(1;1) model with outlier factor based on residual in the first model) and two iterations on the third model (GSTAR(1;1) model with outlier factor based on residual in the second model). Table 2 shows the time identified in the iterative procedure.

Table 2. The Time when The Outliers Detected.

City	Outliers	Model		
		1 st	2 nd	3 rd
A	IO	[33,-,-,-]	[(1,2)]	[-,-]
	AO	[-,1,-,-]	[-]	[-,-]
B	IO	[34,-,-,13,33]	[37]	[-,-]
	AO	[-,-,-,-]	[-]	[-,-]
C	IO	[(29,34),28,37,-,35]	[-]	[-,-]
	AO	[-,-,-,25,-]	[-]	[-,-]
D	IO	[1,39,-,33,-]	[-]	[-,-]
	AO	[-,-,40,-,2]	[-]	[-,-]
E	IO	[33,(36,40),-,-,-]	[37]	[42,32]
	AO	[-,-,-,-]	[-]	[-,-]

The first model shows many outliers detected. It caused by the residual used is residual based on GSTAR(1;1) model without outlier factor. So that, the iterative procedure captures all the origin outliers.

No more outliers detected on that model after getting the third model. The Auto-regressive's parameters in the GSTAR(1;1) model with outlier factor are

$$\Phi_0^* = \text{diag}(0.42; 0.72; 0.77; 0.46; 0.89)$$

$$\Phi_1^* = \text{diag}(0.17; 0.05; 0.07; 0.44; -0.09)$$

While the outlier's parameters are shown in Table 3.

Table 3. The Outlier's Parameter in GSTAR(1;1) Model with Outlier Factor.

Parameter	A	B	C	D	E
ω_{1j}	[15.9,-]	[-,-]	[-,-]	[105.8,-]	[-,-]
ω_{2j}	[-11.6,-]	[-,-]	[-,-]	[-,-]	[-,-]
ω_{13j}	[-,-]	[123.2,-]	[-,-]	[-,-]	[-,-]
ω_{25j}	[-,-]	[-,-]	[-,13.4]	[-,-]	[-,-]
ω_{28j}	[-,-]	[-,-]	[25.1,-]	[-,-]	[-,-]
ω_{29j}	[-,-]	[-,-]	[29.9,-]	[-,-]	[-,-]
ω_{32j}	[-,-]	[-,-]	[-,-]	[-,-]	[18.4,-]
ω_{33j}	[33.6,-]	[27.9,-]	[-,-]	[32.3,-]	[62.7,-]
ω_{34j}	[-,-]	[49.9,-]	[21.7,-]	[-,-]	[-,-]
ω_{35j}	[-,-]	[-,-]	[-14.7,-]	[-,-42.9]	[-,-]
ω_{36j}	[-,-]	[-,-]	[-,-]	[-,-]	[29.1,-]
ω_{37j}	[-,-]	[-22.5,-]	[-19.4,-]	[-,-]	[-34.5,-]
ω_{39j}	[-,-]	[-,-]	[-,-]	[70.1,-]	[-,-]
ω_{40j}	[-,-]	[-,-]	[-,-]	[-,-54.9]	[37.1,-]
ω_{42j}	[-,-]	[-,-]	[-,-]	[-,-]	[-19.9,-]

These parameters gotten by solve the 3rd model using Least Square. The 3rd model is chosen cause no more outlier detected from that residuals.

Compute the Mean Square Error (MSE) for getting the best model. By adding the outlier factor to the model can increase the accuracy of the model. The analysis produces the result as presented as Table 4. Therefore, the best model is the third model.

Table 4. Mean Square Error (MSE) of GSTAR(1;1) Model with Outlier Factor.

MSE	A	B	C	D	E
Without O.F	78.262	130.243	76.453	781.063	210.291
1st Model	54.724	46.349	9.272	136.851	56.018
2nd Model	30.569	34.621	9.272	130.409	36.214
3rd Model	30.569	34.621	9.272	130.409	19.223

3.3 Diagnostic Checking Model

A diagnostic checking model examines whether the model assumptions are fulfilled. The basic assumption of the model is residual white noise. The basic assumption of the model is residual white noise [15]. The residual's white noise can be detected by checking that the normality of residuals is independent of the lag of times. Visually, residuals are independent if there is no correlation between lag times in the ACF plot of the residuals model (see Fig. 3). The plot from left to right connected by arrows shows the ACF of GSTAR(1;1) model without outlier factor → The first model → Second model → Third model (The best model).

While the normality residuals is done by Kolmogorov Smirnov test. The significance level (α) used in this test is 0.05. The p-value of the KS test sequentially for the city- A, B, C, D, and E are 1.253;0.655;0.517;0.717;0.850. We can conclude that the residuals are normally distributed. Fig. 4 gives the plot of the fitted values based on the GSTAR(1;1) model with the outlier factor versus the observations. The plot shows that the outliers can be reached by using this model. Indirectly, this can help reduce the error model in approaching the actual data without eliminating important information from the outliers.

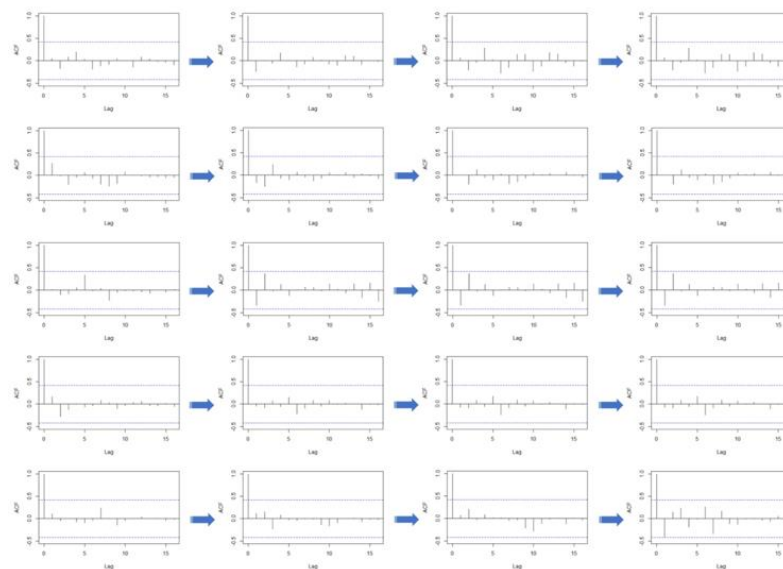
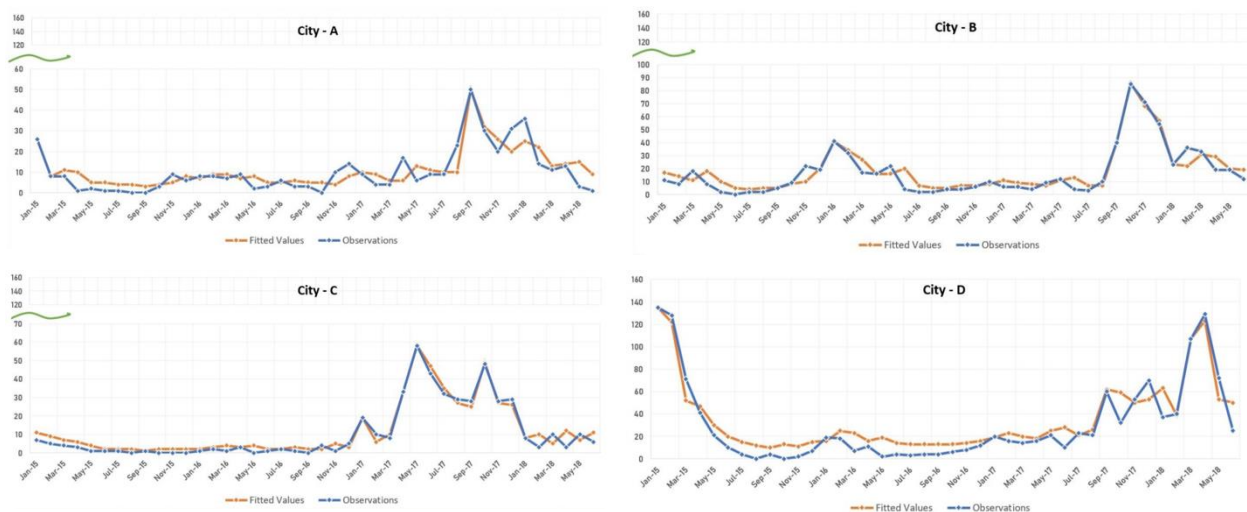


Figure 3. ACF of Residuals in GSTAR(1;1) Model with Outlier Factor.



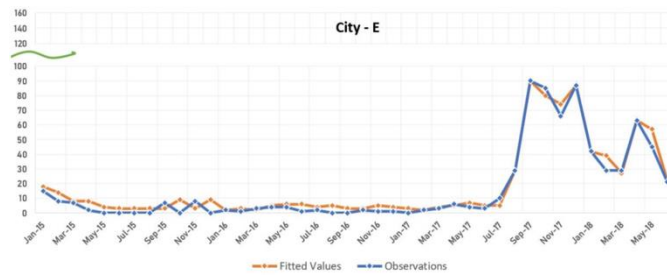


Figure 4. Fitted Values vs Observations based on The GSTAR(1;1) Model with Outlier Factor.

4. CONCLUSIONS

Based on iterative procedures in detecting outliers on the GSTAR(1;1) model, 11 times for IO and three times AO are detected. By adding the outlier factor to the GSTAR(1;1) model, the GSTAR(1;1) model with an outlier factor is obtained better than the GSTAR(1;1) model without adding the outlier factor. It means that the influence of outliers is very significant, so if the outlier is ignored, it can produce a biased model. If the data containing the outlier are removed, it will reduce the information from the data. Therefore, the right solution to handle data containing outliers is adding the outlier factor to the model.

AKNOWLEDGEMENT

The authors thank to the Health Office of West Kalimantan Province for the data.

REFERENCES

- [1] U. Mukhaiyar and U. S. Pasaribu, "A New Procedure of Generalized STAR Modeling using IAcM Approach," *ITB J. Sci.*, vol. 44, no. 2, pp. 179–192, 2012, doi: 10.5614/itbj.sci.2012.44.2.7.
- [2] R. F. Nugraha, S. Setiyowati, U. Mukhaiyar, and A. Yuliawati, "Prediction of oil palm production using the weighted average of fuzzy sets concept approach," 2015, p. 020012, doi: 10.1063/1.4936440.
- [3] Y. Yundari, U. S. Pasaribu, and U. Mukhaiyar, "Error Assumptions on Generalized STAR Model," *J. Math. Fundam. Sci.*, vol. 49, no. 2, p. 136, Oct. 2017, doi: 10.5614/j.math.fund.sci.2017.49.2.4.
- [4] Yundari, U. S. Pasaribu, U. Mukhaiyar, and M. N. Heriawan, "Spatial Weight Determination of GSTAR(1;1) Model by Using Kernel Function," *J. Phys. Conf. Ser.*, vol. 1028, p. 012223, Jun. 2018, doi: 10.1088/1742-6596/1028/1/012223.
- [5] N. Nurhayati, U. S. Pasaribu, and O. Neswan, "Application of Generalized Space-Time Autoregressive Model on GDP Data in West European Countries," *J. Probab. Stat.*, vol. 2012, pp. 1–16, 2012, doi: 10.1155/2012/867056.
- [6] N. F. F. I., U. Mukhaiyar, and F. Fahmi, "The generalized STAR(1,1) modeling with time correlated errors to red-chili weekly prices of some traditional markets in Bandung, West Java," 2015, p. 020014, doi: 10.1063/1.4936442.
- [7] D. Masteriana and U. Mukhaiyar, "Monte Carlo Simulation of Error Assumptions in Generalized STAR(1; 1) Model," in *Proceedings on the Jangjeon Mathematical Society*, 2019, pp. 43–45.
- [8] U. Mukhaiyar, "The goodness of generalized STAR in spatial dependency observations modeling," 2015, p. 020008, doi: 10.1063/1.4936436.
- [9] K. N. Sari, U. S. Pasaribu, O. Neswan, and A. K. Permadi, "Estimation of the parameters of isotropic semivariogram model through bootstrap," *Appl. Math. Sci.*, vol. 9, pp. 5123–5137, 2015, doi: 10.12988/ams.2015.54293.
- [10] J.-H. Jeon, S.-Y. Kho, J. J. Park, and D.-K. Kim, "Effects of spatial aggregation level on an urban transportation planning model," *KSCE J. Civ. Eng.*, vol. 16, no. 5, pp. 835–844, Jul. 2012, doi: 10.1007/s12205-012-1400-4.
- [11] U. Mukhaiyar, U. S. Pasaribu, W. S. Budhi, and K. Syuhada, "The stationarity of generalized STAR(2; λ_1, λ_2) process through the invers of autocovariance matrix," 2014, pp. 484–487, doi: 10.1063/1.4868849.
- [12] I. Chang, G. C. Tiao, and C. Chen, "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, vol. 30, no. 2, p. 193, May 1988, doi: 10.2307/1270165.
- [13] W.W.S. Wei, *Time Series Analysis Univariate and Multivariate Analysis*. Boston: Pearson Education Inc, 2006.
- [14] Dinas Kesehatan Provinsi Kalimantan Barat, "Data Kasus Demam Berdarah Dengue," Pontianak, 2018.
- [15] C. D. Jonathan and C. Kung-Sik, *Time Series Analysis: With Application in R*. USA: Springer Science and Bussiness Media, 2008.