PREDICTION OF PM_{10} POLLUTANT IN SURABAYA USING GENERALIZED SPACE-TIME AUTOREGRESSIVE MOVING AVERAGE

Muhammad Sjahid*, Akbar* Setiawan*2, Suhartono*3, Budi Nurani Ruchjana**4, Dedy Dwi Prastyo*5
* Department of Statistics, Faculty of Science and Data Analytics. Institut Teknologi Sepuluh Nopember, 60111, Indonesia
** Department of Mathematics, Faculty of Mathematics and Natural Sciences, Padjadjaran University, Bandung, 45363, Indonesia.

ABSTRACT
In this paper, we are modeling the concentration of PM_{10} pollutants using the Generalized Space-Time Autoregressive Moving Average (GSTARMA) model. The GSTARMA model is a development of the GSTAR model by remodeling the time effect on the GSTAR residual model. The purpose of this research is to compare the performance of the GSTARMA and the GSTAR model. The two models are applied to model the concentration of PM_{10} pollutants located at three Air Monitoring Stations (AMS) in the city of Surabaya – Indonesia. The estimation methods employed are OLS and SUR. The results of this study show that the GSTARMA model produces better accuracy than the GSTAR model in modeling the PM_{10} data in Surabaya. Moreover, the GSTARMA model can predict well for the two days ahead.

KEY WORDS: GSTARMA, GSTAR, OLS, SUR, PM_{10}.

MSC: 62M10, 91B72 , 37M10 , 46N30 .

RESUMEN
En este documento, estamos modelando la concentración de contaminantes PM10 utilizando el modelo de media móvil autorregresiva espacio-tiempo generalizado (GSTARMA). El modelo GSTARMA es un desarrollo del modelo GSTAR mediante la remodelación del efecto de tiempo en el modelo residual GSTAR. El propósito de esta investigación es comparar el rendimiento del modelo GSTARMA y el modelo GSTAR. Los dos modelos se aplican para modelar la concentración de contaminantes PM10 ubicados en tres estaciones de monitoreo de aire (AMS) en la ciudad de Surabaya, Indonesia. Los métodos de estimación empleados son OLS y SUR. Los resultados de este estudio muestran que el modelo GSTARMA produce una mejor precisión que el modelo GSTAR al modelar los datos de PM_{10} en Surabaya. Además, el modelo GSTARMA puede predecir bien para los próximos dos días.

PALABRAS CLAVE: GSTARMA, GSTAR, OLS, SUR, PM_{10}.

1. INTRODUCTION

Time series data forecasting has now developed by involving the existence of location factors, then referred to as spatio-temporal data forecasting. Several methods can be used to predict spatio-temporal data, one of which is Generalized Space-Time Autoregressive (GSTAR). The GSTAR method is the development of the Space-Time Autoregressive (STAR) model introduced by Cliff & Ord and later developed by Pfeifer & Deutsch [1,2]. The STAR model assumes that each location is homogeneous because the time and location parameter values for all variables have the same amount. This assumption is one of the weaknesses of the STAR model, which was subsequently developed by the GSTAR model by Ruchjana [3]. The GSTAR model is a generalization of the STAR model that assumes the value of the autoregressive parameters has a different value (heterogeneous) at each location. The research using the GSTAR method has been widely used, including Suhartono & Subanar [4], Ruchjana, Borovkova, & Lopuhaa [5], Suhartono, Wahyuningrum, Setiawan, & Akbar [6], and Akbar, Setiawan, Suhartono, Ruchjana, & Riyadi [7]. The GSTARMA model is an extension of the STARMA model, which is characterized by the influence of previous time on the autoregressive and moving average models. Besides the impact of time, there are also influences of other locations, as indicated by spatial autocorrelation [2]. The research using the GSTARMA model include Min,

Hu, & Zhang [8], Nisak [9], and Andayani, Sumertajaya, Ruchjana, & Aidi [10]. These researches not only involve time in the autoregressive and moving average models but also patterns of non-stationarity of data.
The GSTARMA model has been applied in several fields, including in the economic sector, i.e., the price of rice [10], the export of chocolate commodities [11], climatology, i.e., rainfall [9], and transportation [8]. The data used in the research contained interdependent patterns of location. Besides these data, air pollution data also has a habit of dependencies between locations, where air conditions in an area are influenced by air conditions in the surrounding area [12]. Air conditions in an area are affected by several pollutants, including particulate matter (PM_{10}), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2), and ozone (O3) [13]. The toxicology test gives the result that PM_{10} that is sucked directly into the lungs and settles in the alveoli can harm the respiratory system [14]. Rahim & Yeremiah observed that 25% of sufferers of respiratory disorders in Indonesia were caused by PM_{10} [15]. PM_{10} is one of the pollutants that are often the main focus in controlling air quality. Some research on the prediction of pollutant PM_{10} has been done as the early warning on air quality, including by Chrisdayanti, and Suharsono [12] and Suhartono, Prabowo, Prastyo, and Lee [16]. Both kinds of research use a univariate method approach without involving location dependencies. So that in this research, the modeling of PM_{10} data was developed by involving location factors using the GSTARMA method. PM_{10} data used are daily PM_{10} data in the city of Surabaya by using three Air Monitoring Stations (AMS) namely the AMS1, AMS2, and AMS3. In this research, the best modeling will be obtained on PM_{10} data using a combination of different parameter estimates and location weights. The selection of the best model is determined based on the smallest RMSE value.

2. MATERIALS AND METHODS

2.1. Source Data

The data used in this research are PM_{10} daily concentration data in the city of Surabaya. Data were collected from three main Air Monitoring Stations (AMS) in Surabaya, namely AMS1 (Taman Prestasi), AMS2 (Wonorejo), and AMS3 (Kebonsari). The PM_{10} data is secondary data obtained from the Air Quality Monitoring System (AQMS) Surabaya. The period of the data is from January 1st to December 31st, 2018. The data is divided into two parts, namely training data from January 1st to December 24th, 2018, and testing data from December 25th to December 31st, 2018. The data is analyzed using SAS software.

2.2. Steps of Analysis

The analysis steps used in this research as follows.

1. Conduct descriptive analysis of PM_{10} data using time series plots.
2. Calculate the location weights for GSTARMA modeling.
3. Eliminate the pattern of non-stationarity in PM_{10} data by modeling time series regression using a predictor variable in the form of the intervention effect
4. Identify the GSTAR and GSTARMA model order on time series regression residual data using CCF and PCCF.
5. Estimate the parameters of the GSTAR and GSTARMA models using the OLS and GLS methods.
6. The best model selection method for training data uses the RMSE criteria.
7. Forecast the data testing by using the best model in training data.

2.3. GSTARMA Model Identification

The determination of the AR (p) order and MA(q) order on the GSTARMA model can be done using the Cross-Correlation Function (CCF) and Partial Cross-Correlation Function (PCCF). CCF serves to measure the magnitude and direction of the correlation between two random variables [17]. CCF is used to identify the order of the MA model, while the PCCF is the order of the AR model. CCF equation between x_t and y_t is given in equation (1).

$$\hat{\rho}_{xy}(k)=\frac{\sum_{t=1}^{n-k}(x_t-\bar{x})(y_{t+k}-\bar{y})}{\sqrt{\sum_{t=1}^{n}((x_t-\bar{x}))^2\sum_{t=1}^{n}(y_{t+k}-\bar{y})^2}}$$

(1)

The PCCF equation in lag s is like an equation (2).

$$\mathbf{P}(s)=[\mathbf{D}_v(s)]^{-1}\mathbf{V}_{vu}(s)[\mathbf{D}_u(s)]^{-1}$$

(2)

Where $\mathbf{D}_v(s)$ is matrix diagonal with the ith-diagonal element is the ith-diagonal element's root of $\mathbf{V}_v(s)$ and $\mathbf{D}_u(s)$ is defined equal to $\mathbf{V}_u(s)$. $\mathbf{V}_u(s)$ similar to $\operatorname{var}(\boldsymbol{U}_{s-1,t+s})$, $\boldsymbol{V}_{\boldsymbol{V}(s)}$ equal to $\operatorname{var}(\boldsymbol{V}_{s-1,t+s})$, and $\boldsymbol{V}_{\boldsymbol{VU}(s)}$ equal to $\operatorname{cov}(\boldsymbol{V}_{s-1,t},\boldsymbol{U}_{s-1,t+s})$, and

$$\boldsymbol{U}_{s-1,t+s}=\begin{cases}\mathbf{Y}_{t+s}-\sum_{k=1}^{s-1}\boldsymbol{\alpha}_{s-1,k}\mathbf{Y}_{t+s-k}, & s\geq 2\\ \mathbf{Y}_{t+1}, & s=1\end{cases}$$

and

$$\boldsymbol{V}_{s-1,t}=\begin{cases}\mathbf{Y}_{t}-\sum_{k=1}^{s-1}\boldsymbol{\beta}_{s-1,k}\mathbf{Y}_{t+k}, & s\geq 2\\ \mathbf{Y}_{t}, & s=1\end{cases}$$

2.4. Generalized Space-Time Autoregressive Moving Average (GSTARMA)

The GSTARMA model is a development of the GSTAR model by remodeling the residual GSTAR model. In general, the GSTARMA model is given in equation (3) [18].

$$\mathbf{y}_t=\sum_{s=1}^{p}\left(\mathbf{\Phi}_{s0}+\sum_{k=1}^{\lambda_s}\mathbf{\Phi}_{sk}\mathbf{W}^{(k)}\right)\mathbf{y}_{t-s}-\sum_{s=1}^{q}\left(\mathbf{\Theta}_{s0}+\sum_{k=1}^{r_s}\mathbf{\Theta}_{sk}\mathbf{W}^{(k)}\right)\mathbf{e}_{t-s}+\mathbf{u}_t \quad (3)$$

Where $\mathbf{\Phi}_{s0}$ = diag($\phi_{s0}^{(i)}$) = diag($\phi_{s0}^{(1)}$,L ,$\phi_{s0}^{(N)}$), $\mathbf{\Phi}_{sk}$ = diag($\phi_{sk}^{(i)}$) = diag($\phi_{sk}^{(1)}$,L ,$\phi_{sk}^{(N)}$), then $\mathbf{\Phi}_{s0}$ and $\mathbf{\Phi}_{sk}$ as an autoregressive parameter at the ith-location, the sth-time order, and the kth-spatial order. $\mathbf{W}^{(k)}$ is weight matrix $(N\times N)$ the kth spatial order (where k = 0,1,..., λ_s), $\mathbf{w}_{ii}=0$, $\sum_{i\neq j}\mathbf{w}_{ij}^{(k)}=1$, $\mathbf{\Theta}_{s0}$ = diag($\theta_{s0}^{(i)}$) = diag($\theta_{s0}^{(1)}$,L ,$\theta_{s0}^{(N)}$), $\mathbf{\Theta}_{sk}$ = diag($\theta_{sk}^{(i)}$) = diag($\theta_{sk}^{(1)}$,L ,$\theta_{sk}^{(N)}$), whereas Θ_{s0} and Θ_{sk} as moving average parameter at the ith location, the sth time order, and the kth spatial order. $\mathbf{u}_t$ is residual vector following a normal distribution with an average of zero and covariance $\mathbf{\Sigma}$. This study uses uniform weights and inverse distances.

a. Uniform Weights

Uniform weights assume that the locations used in the research are homogeneous, so the weight values used are the same for each location. The formula for uniform location weights is formulated in equation (4).

$$w_{ij}=\frac{1}{s_i}, \quad (4)$$

where s_i is the number of locations adjacent to the ith-location. In this research, there are three locations used, so the matrix for uniform weights is

$$\mathbf{W} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

b. Inverse Distance Weights

The use of inverse distance weights is based on the actual distance between locations. Weight calculation using the inverse distance method is obtained from the normalization of the actual inverse distance result. The formula used to calculate inverse distance weights is given in equation (5).

$$w_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{j=1}^{s} \frac{1}{d_{ij}}}, \quad j \neq i. \tag{5}$$

Table 1 is the distance of the location of each AMS in Surabaya.

Table 1. Distance between the three AMS (in km)

AMS	1	2	3
1	0		
2	13.1	0	
3	9.9	19.5	0

Calculation of distance inverse matrix using the data in Table 1 is reported below.

$$\mathbf{W} = \begin{bmatrix} 0 & 0.432 & 0.568 \\ 0.598 & 0 & 0.402 \\ 0.662 & 0.338 & 0 \end{bmatrix}.$$

2.5. Estimation of GSTARMA Model Parameters

GSTARMA model with order one, for spatial and time at three locations (GSTARMA (k = 1, p = 1, q = 1)) as in the following equation (6).

$$\mathbf{y}_t = \mathbf{\Phi}_{10}\ \mathbf{y}_{t-1} + \mathbf{\Phi}_{11}\ \mathbf{W}^{(1)}\mathbf{y}_{t-1} - \mathbf{\Theta}_{10}\mathbf{e}_{t-1} - \mathbf{\Theta}_{11}\mathbf{W}^{(1)}\mathbf{e}_{t-1} + \mathbf{u}_t \tag{6}$$

where

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_{1,t} \\ \mathbf{y}_{2,t} \\ \mathbf{y}_{3,t} \end{pmatrix}, \quad \mathbf{y}_{t-1} = \begin{pmatrix} \mathbf{y}_{1,t-1} \\ \mathbf{y}_{2,t-1} \\ \mathbf{y}_{3,,t-1} \end{pmatrix}, \mathbf{e}_{t-1} = \begin{pmatrix} \mathbf{e}_{1,t-1} \\ \mathbf{e}_{2,t-1} \\ \mathbf{e}_{3,t-1} \end{pmatrix}, \quad \boldsymbol{u}(t) = \begin{pmatrix} \boldsymbol{u}_1(t) \\ \boldsymbol{u}_2(t) \\ \boldsymbol{u}_3(t) \end{pmatrix}, \quad \mathbf{W}^{(1)} = \begin{pmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{pmatrix},$$

$$\mathbf{\Phi}_{10} = \mathrm{diag}\left(\phi_{10}^{(1)}, \phi_{10}^{(2)}, \phi_{10}^{(3)}\right), \mathbf{\Theta}_{10} = \mathrm{diag}\left(\theta_{10}^{(1)}, \theta_{10}^{(2)}, \theta_{10}^{(3)}\right), \mathbf{\Phi}_{11} = \mathrm{diag}\left(\phi_{11}^{(1)}, \phi_{11}^{(2)}, \phi_{11}^{(3)}\right), \text{ dan } \mathbf{\Phi}_{11} = \mathrm{diag}\left(\theta_{11}^{(1)}, \theta_{11}^{(2)}, \theta_{11}^{(3)}\right).$$

The GSTARMA model in equation (6), if given for example $\mathbf{v}_i = \sum_{j=1}^{N} w_{ij}\mathbf{y}_j$ and $\mathbf{r}_i = \sum_{j=1}^{N} w_{ij}\mathbf{e}_j$, then it will be equation (7).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{7}$$

where $\mathbf{X} = (\mathbf{A} \quad \mathbf{B})$ and $\boldsymbol{\beta}' = (\mathbf{f}' \quad \mathbf{g}')$,

$$\mathbf{A} = \begin{pmatrix} \mathbf{y}_{1,t-1} & 0 & 0 & \mathbf{v}_{1,t-1} & 0 & 0 \\ 0 & \mathbf{y}_{2,t-1} & 0 & 0 & \mathbf{v}_{2,t-1} & 0 \\ 0 & 0 & \mathbf{y}_{3,t-1} & 0 & 0 & \mathbf{v}_{3,t-1} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} \mathbf{e}_{1,t-1} & 0 & 0 & \mathbf{r}_{1,t-1} & 0 & 0 \\ 0 & \mathbf{e}_{2,t-1} & 0 & 0 & \mathbf{r}_{2,t-1} & 0 \\ 0 & 0 & \mathbf{e}_{3,t-1} & 0 & 0 & \mathbf{r}_{3,t-1} \end{pmatrix}$$

$\mathbf{f}' = \left(\phi_{10}^{(1)} \quad \phi_{10}^{(2)} \quad \phi_{10}^{(3)} \quad \phi_{11}^{(1)} \quad \phi_{11}^{(2)} \quad \phi_{11}^{(3)}\right)$

$\mathbf{g}' = \left(\theta_{10}^{(1)} \quad \theta_{10}^{(2)} \quad \theta_{10}^{(3)} \quad \theta_{11}^{(1)} \quad \theta_{11}^{(2)} \quad \theta_{11}^{(3)}\right)$

$\mathbf{u}' = \left(\mathbf{u}'_{1,t} \quad \mathbf{u}'_{2,t} \quad \mathbf{u}'_{3,t}\right)$ and $\mathbf{u} : N\left(\mathbf{0}, \mathbf{\Sigma}\right)$.

The model parameters in equation (7) can be estimated by the Ordinary Least Square (OLS) method with the solution formulated in equation (8) below.

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{8}$$

The solution to the estimation of GSTARMA parameters using the Generalized Least Square (GLS) method is given in equation (9).

$$\hat{\boldsymbol{\beta}}_{\text{SUR}} = \left(\mathbf{X}'\left(\mathbf{\Sigma}\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{\Sigma}\right)^{-1}\boldsymbol{Y} \tag{9}$$

where $\mathrm{E}\left(\mathbf{u}\right) = 0,\ \mathbf{\Sigma}^{*} = \mathrm{E}\left(\mathbf{uu}'\right),$ and $\mathbf{\Sigma} = \mathbf{\Sigma} \otimes \mathbf{I}.$

2.6. Steps to Obtain the Best GSTARMA Model

The steps of building the GSTARMA model until the best model is obtained are as follows:

1. Assume the data has been stationary in the mean and variance.
2. Calculates uniform weights (like equation (4)), and inverse distance weights (like equation (5)) used as weights in the GSTAR and GSTARMA model.
3. The data is divided into two parts, i.e., training data used as modeling and testing data used as model validation.
4. Determine the autoregressive order (p) and the moving average order (q) to identify the GSTAR and GSTARMA models using CCF (like equation (1)) and MPCCF (like equation (2)).
5. Estimate the parameters of the GSTAR and GSTARMA models using the OLS (as in equation (8)) and SUR (as in equation (9)) method.
6. Calculate predictions from each model obtained from step 5.
7. The best model is obtained by selecting the smallest RMSE value in the testing data of each model.

3. RESULT AND DISCUSSION

The characteristics of PM_{10} data from three AMS in Surabaya, namely AMS1, AMS2, and AMS3 starting from January 1st to December 31st, 2018, are displayed through time series plots in Figure 1. Figure 1 shows that the pattern of PM_{10} data tends to be not stationary, as indicated by very high fluctuations at some point. PM_{10} data are divided into two parts, training and testing data. The training data is used from January 1st to December 24th, 2018. The testing data has period from December 25th to December 31st, 2018.

PM_{10} data used in GSTAR and GSTARMA modeling need to be stationary in the mean, so the data must be modeled by time-series regression first. The time-series regression modeling uses predictor variables in the form of daily intervention events. The purpose of this time series regression modeling is to obtain stationary data patterns in the mean for GSTAR and GSTARMA modeling. The time series plot of the residual of the time series regression model in Figure 2 has shown a stationary pattern in the mean so that it can proceed to the next modeling, GSTAR, and GSTARMA.

The GSTAR and GSTARMA modeling are using two weights, uniform, and inverse distance weights. Besides, a combination of parameter estimation is also used in OLS and SUR. The determination of the autoregressive order in the GSTAR model can be identified through the PCCF plot in Figure 4. The PCCF plot in Figure 4 shows that there is a tendency to lag that correlates among the three AMS at lag 4, so the order of the GSTAR model is GSTAR($[4]_1$). Next, to identify moving average order through CCF. In the CCF plot based on Figure 3, the residual of the time series regression of PM_{10} at the AMS3 correlates significantly with AMS1 at lag 3. The residual at AMS1 correlates with AMS3. The residual at AMS2 correlate with AMS1 at lag 7. The relationship between locations of PM_{10} data has been represented by lag 4 in the PCCF plot as well as by lag 3 and lag 7 in the CCF plot.

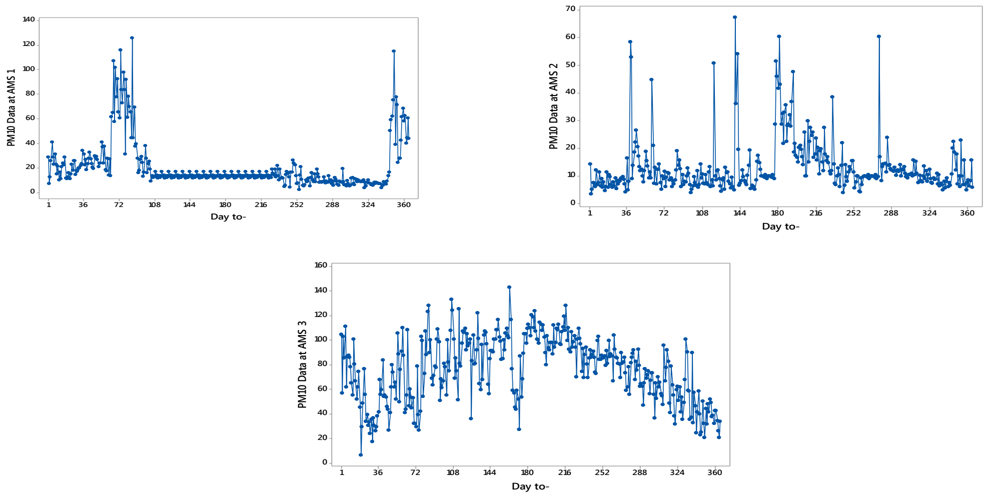

Figure 1. Time Series Plot of PM_{10} Data at AMS1, AMS2, and AMS3

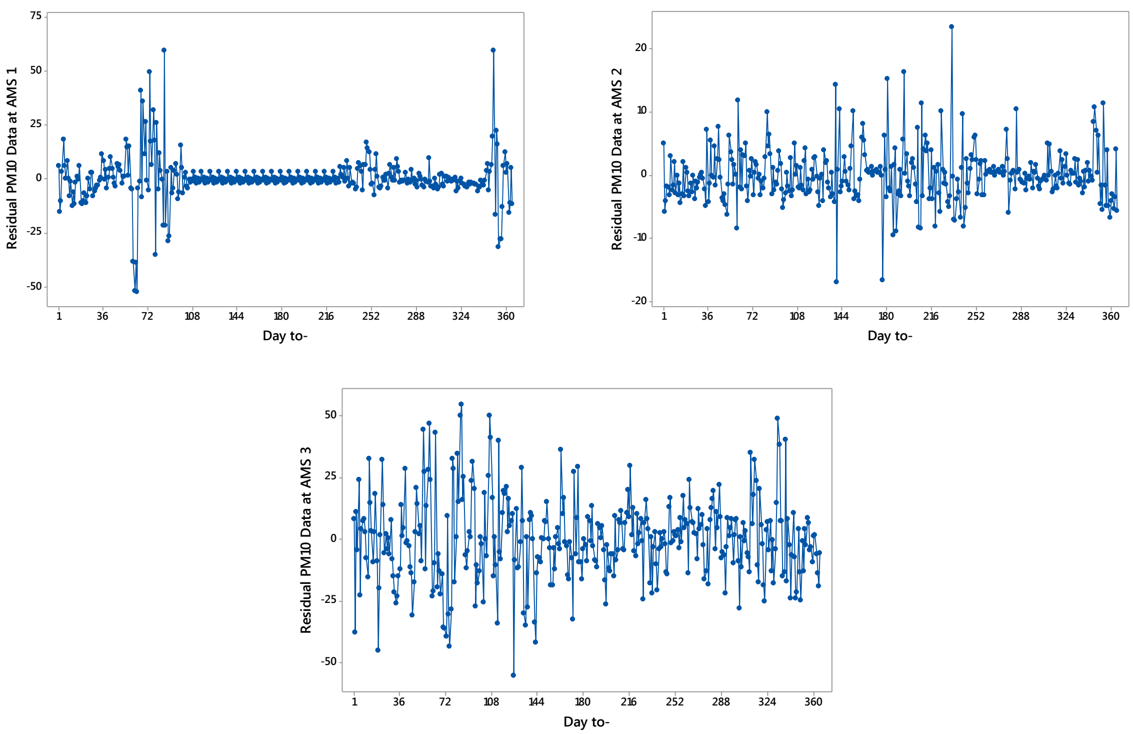

Figure 2. Time Series Plot the Residual of Time Series Regression of PM_{10} Data at AMS1, AMS2, and AMS3

Schematic Representation of Cross Correlations

Variable/ Lag	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
AMS1	+. .	+. .	+. .	. . .	. . .	- . .	. . .	- . +	..+	. .+	- . +	. .+	-.+	-.+	. . .	. . .	. . .
AMS2	.++	. . .	. . .	. . .	. - .	. . .	. . .	- . .	.+.	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
AMS3	.++	. . +	. . .	- . .	. . .	- . .	-. .	. . .	. . .	. . .	. . .	. . .	. . -	. . .	. . .	+. .	. . -

Variable/

```
Lag     17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32
AMS1    ... ... ... ... ... ... ... ... ... ... ... +.. ... ... ... ...
AMS2    ... -.. .+. ... ... ... ... ... ... .-. ... ... ... ... ... ...
AMS3    +.- +.. ... ... ... ... -.. ... ... .-. ... ... ... ... ... ...

          + is > 2*std error,  - is < -2*std error,  . is between
```

Figure 3. CCF Plot of Residual Time Series Regression of PM_{10} Data at AMS1, AMS2, and AMS3

```
          Schematic Representation of Partial Cross Correlations

Variable/
Lag     1   2   3   4   5   6   7   8   9   10  11  12  13  14  15  16
AMS1    +.. +.. ... -.. -.. ... ..+ ..+ ... -.. ... ..+ ... ... ... ...
AMS2    ... ... ... .-. ... .-. ... .+. ... ... ... ... ... ... ... ...
AMS3    ..+ ... -.. .-. -.. ... ... ... ... ... -.. ... ... ... ... ...


Variable/
Lag     17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32
AMS1    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
AMS2    ... -.. .+. ... ... ... ... ... ... ... ... ... ... ... ... ...
AMS3    ... ... ... ... ... ... -.. ... ... .-. ... ... ... ... +.. ...

          + is > 2*std error,  - is < -2*std error,  . is between
```

Figure 4. PCCF Plot of Residual Time Series Regression of PM_{10} Data at AMS1, AMS2, and AMS3

The GSTARMA modeling used the GSTARMA($[4]_1$, $[3,7]_1$) model using the estimated OLS parameters and uniform weights. For other GSTAR and GSTARMA modeling, using OLS parameter estimation with distance inverse weight and using SUR parameter estimation with the uniform weight and distance inverse have the same order, namely GSTAR($[4]_1$) and GSTARMA($[4]_1$,$[3,7]_1$). The selection of the best method for predicting PM_{10} data is using the smallest RMSE values in the testing data. Figure 5 shows the RMSE values of the three AMS of PM_{10} testing data. Based on Figure 5, it can be seen that the best model in the PM_{10} data at AMS1 is GSTARMA-OLS ($[4]_1$, $[3,7]_1$) using uniform weights. While in AMS2, it is known that the best model is GSTARMA-SUR ($[4]_1$, $[3,7]_1$) with uniform weights. In AMS3, the best model that can be applied is GSTARMA-SUR($[4]_1$, $[3,7]_1$) with inverse distance weight. Thus, the equation GSTARMA to model PM_{10} data in the three AMS as in Table 2.

Table 2. The Best GSTARMA Model for Modeling PM_{10} Data at AMS1, AMS2, and AMS3

AMS	Model	Weights	Model equation
1	GSTARMA-OLS($[4]_1$,$[3,7]_1$)	Uniform	$Y_t^{(1)} = -0{,}076Y_{t-4}^{(1)} - 0{,}014Y_{t-4}^{(2)} - 0{,}014Y_{t-4}^{(3)} + 0{,}099a_{t-3}^{(1)} - 0{,}110a_{t-3}^{(2)} - 0{,}110a_{t-3}^{(3)} - 0{,}185a_{t-7}^{(1)} - 0{,}077a_{t-7}^{(2)} - 0{,}077a_{t-7}^{(3)}$
2	GSTARMA-SUR($[4]_1$,$[3,7]_1$)	Uniform	$Y_t^{(2)} = -0{,}021Y_{t-4}^{(1)} - 0{,}146Y_{t-4}^{(2)} - 0{,}021Y_{t-4}^{(3)} + 0{,}021a_{t-3}^{(1)} + 0{,}070a_{t-3}^{(2)} + 0{,}021a_{t-3}^{(3)} + 0{,}003a_{t-7}^{(1)} - 0{,}005a_{t-7}^{(2)} + 0{,}003a_{t-7}^{(3)}$
3	GSTARMA-SUR($[4]_1$,$[3,7]_1$)	Inverse distance	$Y_t^{(3)} = -0{,}118Y_{t-4}^{(1)} - 0{,}060Y_{t-4}^{(2)} - 0{,}072Y_{t-4}^{(3)} - 0{,}214a_{t-3}^{(1)} - 0{,}109a_{t-3}^{(2)} + 0{,}536a_{t-3}^{(3)} - 0{,}116a_{t-7}^{(1)} - 0{,}059a_{t-7}^{(2)} + 0{,}040a_{t-7}^{(3)}$

The GSTARMA model equation in Table 2 shows that PM_{10} data on AMS1, AMS2, and AMS3 are affected by events 3, 4, and 7 days before at the same location and different locations.

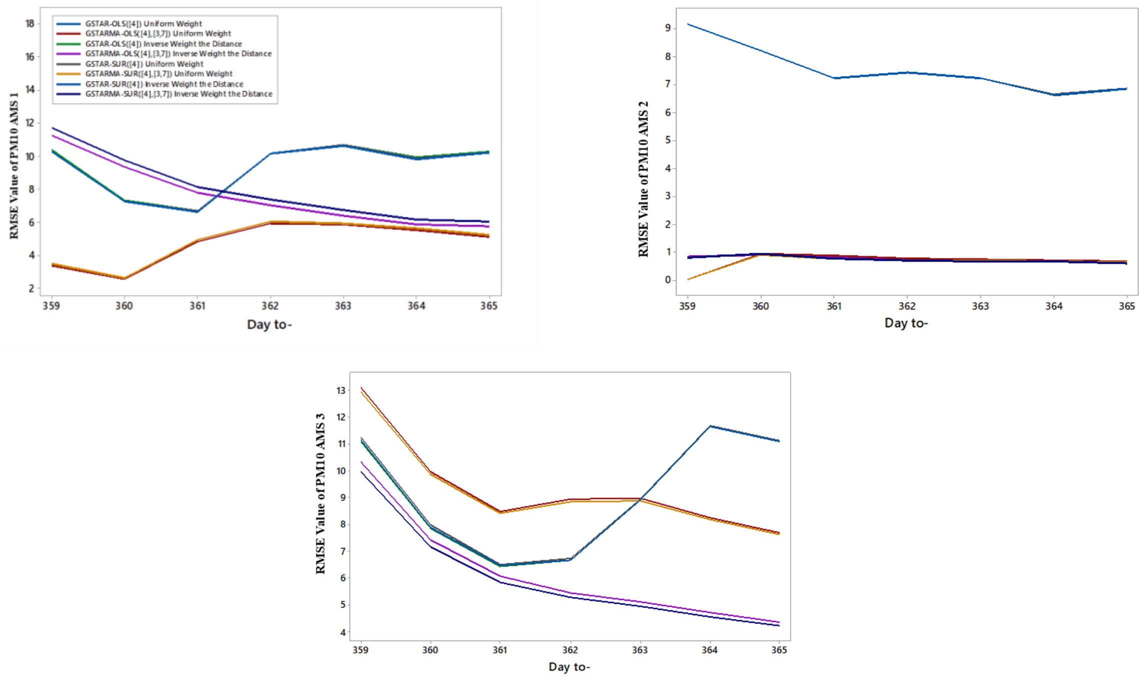

Figure 5. RMSE Value from PM_{10} Testing Data at AMS1, AMS2, and AMS3 using GSTAR and GSTARMA Model

In general, the RMSE value in Figure 5 shows that the best model for each AMS can predict well for the next 2 days. Based on the smallest RMSE value in each AMS, it can be concluded that the GSTARMA model tends to be better than the GSTAR model to model PM_{10} data in Surabaya. It is shown that the development of more sophisticated forecasting models, in this research that the development of GSTAR into GSTARMA, can provide better accuracy values for the model. These findings are in line with the results of the M4-Competition, where the hybrid (complex) forecasting models tend to be better than individual forecasting models [19].

4. CONCLUSION

This work model the PM_{10} data in Surabaya from three stations namely AMS1, AMS2, and AMS3. The models used are GSTAR and GSTARMA with estimation parameters method using OLS and SUR. The weights matrix used in this research are uniform and inverse distances weights. Based on the smallest RMSE value it can be concluded that the best model for PM_{10} data measured in AMS1 is GSTARMA-OLS($[4]_1$,$[3,7]_1$) using uniform weights, the best model in AMS2 is GSTARMA-SUR($[4]_1$,$[3,7]_1$) with uniform weights, and the best model in AMS3 is GSTARMA-SUR($[4]_1$,$[3,7]_1$) with inverse distance weights. Thus, it can be concluded that the GSTARMA model outperform the GSTAR model to model the PM_{10} data in Surabaya.

Acknowledgments. We would like to thank the Environment Agency of Surabaya city for the data provided. We are also grateful to Institut Teknologi Sepuluh Nopember for supporting this research.

RECEIVED: JANUARY, 2020.
REVISED: JUNE, 2020

REFERENCES

[1] CLIFF A. and ORD J. (1975): Space-Time Modeling with an Application to Regional Forecasting. **Transactions of the Institute of British Geographers**, 64, 119-128.

[2] PFEIFER P. and Deutsch S. (1980): A Three Stage Iterative Procedure for Space Time Modeling. **Technometrics**, 22(1), 35-47.

[3] RUCHJANA, B. N. (2002): "**A Generalized Space-Time Autoregressive Model and Its Application to Oil Production**, Unpublished Ph.D. Thesis," ITB, Bandung.

[4] SUHARTONO and SUBANAR (2006): The Optimal Determination of Space Weight in GSTAR Model by Using Cross-Correlation Inference. **Quantitative Methods**, 2, 45-53.

[5] RUCHJANA B., BOROVKOVA S. and LOPUHAA H. (2012): Least Squares Estimation of Generalized Space Time Autoregressive (GSTAR) Model and Its Properties. **AIP Conference Proceedings**, **1450,** 61-64.

[6] SUHARTONO, WAHYUNINGRUM, S. R., SETIAWAN, and AKBAR, M. S. (2016): GSTARX-GLS Model for Spatio-Temporal Data Forecasting. **Malaysian Journal of Mathematical Sciences**,10(S)**,** 91-103.

[7] AKBAR, M. S., SETIAWAN, SUHARTONO, RUCHJANA, B. N., and RIYADI, M. A. A. (2018): GSTAR-SUR Modeling With Calendar Variations And Intervention To Forecast Outflow Of Currencies In Java Indonesia. **Journal of Physics: Conference Series**, 974, 012060.

[8] MIN, X., HU, J., and ZHANG, Z. (2010): Urban Traffic Network Modeling and Short-Term Traffic Flow Forecasting Based on GSTARIMA Model. **IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC.** 1535–1540

[9] NISAK S. (2016): Seemingly Unrelated Regression Approach for GSTARIMA Model to Forecast Rain Fall Data in Malang Southern Region Districts. **CAUCHY**, 4(2), 57-64.

[10] ANDAYANI N., SUMERTAJAYA, I. M., RUCHJANA, B. N., and AIDI M. (2018): Comparison of GSTARIMA and GSTARIMA-X Model by using Transfer Function Model Approach to Rice Price Data. **IOP Conference Series: Earth and Environmental Science.** 187.

[11] QOMARIYAH, L., TOHARUDIN, T. and SOEMARTINI (2016): Generalized Space Time Autoregressive Integrated Moving Average (GSTARIMA) Model to Forecast Cocoa Export Volume. **Proceeding of The 2nd International Conference on Applied Statistics 2016**. ISSN : 2579-4361

[12] CHRISDAYANTI B. and SUHARSONO A. (2015): Forecasting the content of Particular Matter (PM_{10}) in Ambient Air Surabaya City Using Double Seasonal ARIMA (DSARIMA). **Jurnal Sains dan Seni ITS,** 4(2).

[13] Ministerial Decree of Environment and Forestry No. 45 Year 1997. The Air Pollution Standard Index. Oktober 1997, 13rd. **Ministry of Environment and Forestry**. Jakarta.

[14] GUNAWAN G. and KUSMININGRUM N. (2008): Polusi Udara Akibat Aktivitas Kendaraan Bermotor di Jalan Perkotaan Pulau Jawa dan Bali. **Jurnal Jalan-Jembatan 25** (3): 314-326.

[15] RAHIM F. and YEREMIAH, C. R. (2018): Kondisi Kualitas Udara (SO2, NO2, PM10 dan PM2,5) di dalam Rumah di Sekitar Cilegon dan Gangguan Pernapasan yang Diakibatkannya. **Journal of Biology**, 82-90.

[16] SUHARTONO, PRABOWO H., PRASTYO, D. D. and LEE, M. H. (2019): New Hybrid Statistical Method and Machine Learning for PM_{10} Prediction. **International Conference on Soft Computing in Data Science**, 142-155.

[17] WEI, W. W. S. (2006): **Time Series Analysis: Univariate and Multivariate Methods**, Pearson Education. Inc.

[18] AKBAR, M. S., SETIAWAN, SUHARTONO, RUCHJANA, B. N., PRASTYO, D. D., MUHAIMIN A. and Setyowati E. (2019): A Generalized Space-Time Autoregressive Moving Average (GSTARMA) Model for Forecasting Air Pollutant in Surabaya: **International Conference on Mathematics: Pure, Applied and Computation (ICoMPAC 2019), Surabaya**

[19] MAKRIDAKIS S., SPILIOTIS E. and ASSIMAKOPOULOS V. (2018): The M4 Competition: Results, Findings, Conclusion and Way Forward. **International Journal of Forecasting**, 34(4): 802-808.