Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/120316

Implementation of Generalized Space Time Autoregressive (GSTAR)-Kriging Model for Predicting Rainfall Data at Unobserved Locations in West Java

Atje Setiawan Abdullah[1,∗], *Setiawan Matoha*[2], *Deltha Airuzsh Lubis*[3], *Annisa Nur Falah*[4], *I. G. N. Mindra Jaya*[5], *Eddy Hermawan*[6] *and Budi Nurani Ruchjana*[7]

[1] Department of Computer Science, Universitas Padjadjaran, Indonesia.
[2] Department of Statistics, Institut Teknologi Sepuluh Nopember Surabaya, Indonesia.
[3] Statistics of South Tapanuli Regency, Indonesia.
[4] Department of Mathematics, Universitas Padjadjaran, Indonesia.
[5] Department of Statistics, Universitas Padjadjaran, Indonesia.
[6] Atmospheric Science and Technology Center National Institute of Aeronautics and Space (LAPAN), Bandung, Indonesia.
[7] Department of Mathematics and Indonesian Ethnoscience Studies Center, Universitas Padjadjaran, Indonesia.

Received: 13 Mar. 2018, Revised: 28 Mar. 2018, Accepted: 6 Apr. 2018
Published online: 1 May 2018

Abstract: A Generalized Space Time Autoregressive or GSTAR is a special model of Vector Autoregressive (VAR) model which is a combination of time series and spatial models which has the assumption of autoregressive parameter and space time parameter having different value for each location of observation. In addition, it assumes stationary time series data at the mean and variance levels and applies to locations with heterogeneous characteristics. One disadvantage of the GSTAR model is that it can not be used to predict at unobserved locations. In this paper we combine the GSTAR model with the Ordinary Kriging (OK) technique, named GSTAR-Kriging model so that the GSTAR model can be used to predict in unobserved locations. GSTAR parameters are estimated using the Ordinary Least Squares (OLS) method and these are used as inputs for the Kriging technique. Furthermore, by using linear semivariogram we can obtain simulations to predict the GSTAR parameters. For the case study we applied the model to annual rainfall data in wet season (Desember, January and February) from several locations in West Java, Indonesia, such as Majalengka, Kuningan and Ciamis Regencies. The GSTAR (1;1) model in observed location have Mean Average Percentage Error (MAPE) value overall less than 15 percent and residual of model have identically independent distributed normal. The results of GSTAR-Kriging model show that the GSTAR-Kriging parameter at unobserved locations are almost similar to GSTAR parameter at observed locations.

Keywords: GSTAR-Kriging, OLS, Semivariogram, Ordinary Kriging, MAPE, Rainfall

1 Introduction

A combination time series model with spatial data is called spatiotemporal model. Spatiotemporal model is used mostly in various sector as geology, climatology, economy and others. Generalized Space Time Autoregressive (GSTAR) model is a general form of Vector Autoregressive (VAR) model which used the spatial weight on its model and estimation [1], [2]. GSTAR model requires the characteristic in every location are heterogenous. The GSTAR model can forecast the data for some periods ahead at observed locations that is trained in the model. But it can not be used to predict the data at unobserved locations. In Geostatistics, we can predict the data at unobserved locations based on the random variable at observed locations around [3]. The dependency between locations are usually assessed by semivariogram. Furthermore, semivariogram value gives us information how the neighbourhood show the similarity value of random variable. To obtain the fitted semivariogram, it requires a lot of pairs observed locations at different distances or spatial lag. The related work [4] have used 164 observed locations that divided by 15 pairs observed locations based on their distance. But in this paper we have the experimental semivariogram value at less observed

∗ Corresponding author e-mail: atje.setiawan@unpad.ac.id

locations. If we obtained the semivariogram value, we could predict the value of random variable at unobserved locations with interpolation technique. One of the interpolation technique is the ordinary kriging [5]. The spatial data mining using Spatial Autoregressive-Kriging (SAR-Kriging) has been used to predict a quality of education in Indonesia, especially for unobserved sample with province as a unit sample in Indonesia [11]. Otherwise, one of disadvantage of kriging technique is only used to predict the data at one period, so we can not get the result of prediction in long period. The GSTAR model and kriging technique have advantage where by combining both of them we can predict the value of random variable at unobserved location in some periods in the future. Combination of GSTAR model and kriging technique is called GSTAR-Kriging model. The GSTAR model parameters in observed locations are used as input data on kriging technique to predict the parameter of GSTAR model at unobserved locations. Furthermore, the application of Clustering GSTAR-Kriging is extended to predict oil production at volcanic layer in Jatibarang-Indonesia which is clustered by porosity [6]. In this paper, we implemented the GSTAR-Kriging model to rainfall data in some regencies in West Java at wet session DJF (December-January-February). We proposed a combination of the observed and unobserved locations from these regencies so we can compare the results of GSTAR-Kriging model using MAPE criteria.

2 Generalized Space Time Autoregressive (GSTAR) Model

GSTAR model was introduced by [1] and [2] as:

$$\mathbf{Z}(t) = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} [\Phi_{kl} W^{(l)} \mathbf{Z}(t-k) + \mathbf{e}(t)] \tag{1}$$

where:

λ_k	:	spatial order from autoregressive form of order k
$\mathbf{Z}(t)$	:	vector random with size $(n \times l)$ in time t
$\mathbf{Z}(t-k)$	:	vector random with size $(n \times l)$ in time $(t-k)$
Φ_{kl}	:	diag. $((\Phi_k)^{(1)}), \ldots, ((\Phi_{kl})^{(n)})$, is diagonal matrix of autoregressive parameters in time lag k and spatial lag l size $(n \times n)$
$W^{(l)}$	:	weight matrix size $(n \times n)$ in spatial lag l (where $l = 0, 1, \ldots$) and the weight choosen for $w_{ii} = 0$ and $\sum_{i \neq j} w_{ij}^{(1)} = 1$
$\mathbf{e}(t)$	:	error vector with size $(n \times l)$ in time t, with assumption that $\mathbf{e}(t) \sim$ iid $N(\mathbf{0}, \sigma^2 \mathbf{I})$

The GSTAR model in (1) with spatial lag $l = 1$, time lag $k = 1$ and the number of location $N = 2$ can be rearranged as matrix equation below:

$$\begin{bmatrix} Z_1(t) \\ Z_2(t) \end{bmatrix} = \begin{bmatrix} \phi_{10} & 0 \\ 0 & \phi_{20} \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_2(t-1) \end{bmatrix} + \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{21} \end{bmatrix} \begin{bmatrix} 0 & w_{12} \\ w_{21} & 0 \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_2(t-1) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} \tag{2}$$

if $V_i = \sum_{j=1}^{N} w_{ij} Z_j$, equation (2) can be written as:

$$\begin{bmatrix} Z_1(t) \\ Z_2(t) \end{bmatrix} = \begin{bmatrix} Z_1(t-1) & 0 & V_1(t-1) & 0 \\ 0 & Z_1(t-1) & 0 & V_1(t-1)\ 0 \end{bmatrix} \begin{bmatrix} \phi_{10} \\ \phi_{20} \\ \phi_{11} \\ \phi_{21} \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} \tag{3}$$

Equation (3) can be arranged as linear form to estimate GSTAR model. The linear form is showed below:

$$Y = X\Phi + e \tag{4}$$

where $e \sim$ iid $N(0, \sigma^2)$. The parameter linear form in (4) can be estimated by ordinary least square method used the formula [1]:

$$\hat{\Phi} = (X'X)^{-1}(X'Y) \tag{5}$$

where X is the first matrix in the right part in equation (3).

The GSTAR model can be extended with addition of exogenous variable , it is called GSTAR-X model [12]. We also can choose a calendar variation for exogenous variable [14]. If the error has a correlation between location, so we can use Seemingly Unrelated Regression (SUR) to estimate the GSTAR model, it is called GSTAR-SUR [15].

3 Semivariogram

Semivariogram is a diagram of half variance of observation value difference at two locations with distance h. The experimental semivariogram model can be showed below [3], [5]:

$$\hat{\gamma} = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \tag{6}$$

The assumptions of semivariogram are:

a. $E[Z(x_i + h) - Z(x_i)] = 0$, the mean value of $Z(x_i)$ at all points x_i is the same.

b. $Var[Z(x_i + h) - Z(x_i)] = 2\gamma(h)$, there are variance of every $[Z(x_i + h) - Z(x_i)]$ and depend on interval h and independent in location x_i.

Semivariogram has anisotropic and isotropic properties. Anisotropic if affected by distance and direction while isotropic if affected only by distance. In this paper, we only use isotropic. Semivariogram model commonly used is the spherical model in which tends to form a straight line for some values around the origin. The equation of theoretical semivariogram for the spherical model is [3], [5]:

$$\gamma(h) = \begin{cases} c\left[\frac{3}{2}\left(\frac{h}{a}\right) - \left(\frac{1}{2}\right)\left(\frac{h}{a}\right)^3\right], & \text{h< a} \\ c, & \text{h}\geq \text{a} \end{cases} \tag{7}$$

where:
c : sill
a : range
h : distance

The spherical model is displayed in Figure 1 below.

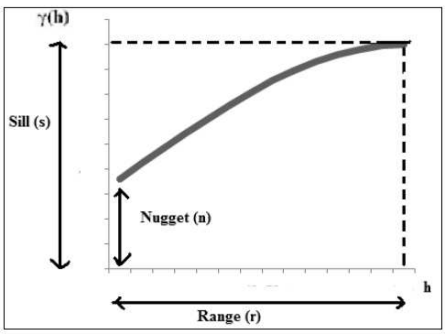


Fig. 1: Spherical Model of Semivariogram

If we just have two locations as sample, then the equation (7) can be written as [6], [7]:

$$\gamma(h) = \begin{cases} c\left[\frac{\hat{\gamma}(r)}{r}h\right], & \text{h< r} \\ \hat{\gamma}(r), & \text{h}\geq \text{r} \end{cases} \tag{8}$$

and the graphic at Figure 1 can be displayed as the linear model, it is showed in Figure 2.

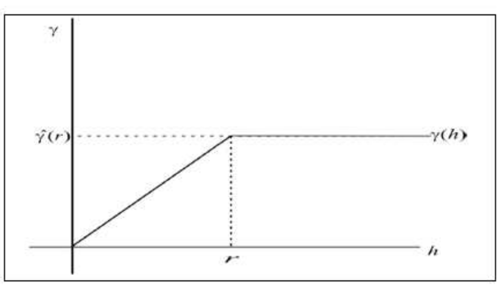


Fig. 2: Linear Model of Semivariogram at Two Locations

4 Ordinary Kriging

Kriging is a technique for predicting data in unobserved locations located around the observed location that used a weighted average of its point value. Kriging is based on spatial data. Ordinary Kriging (OK) is one of the simplest form of Kriging that assumed stationary (mean and variance do not vary significantly in space) [6] , [13] or $E[Z(x)] = m = E[Z(x_i)]$. OK is also assumed normally distributed and its estimator is best linear unbiased estimator (BLUE). OK is linear because its estimator depends on linear combination of data; unbiased because the error mean is assumed to be a constant and it is expected to zero and the variance of error is expected has a minimum value. The Ordinary Kriging equation is formulated [3]:

$$Z^* = \sum \lambda_i Z_i \tag{9}$$

where:
Z^* : unobserved variable in unobserved location
λ_i : Kriging weight i ; and $\sum_{i=1}^{n} \lambda_i = 1$
Z_i : observation value in sampled location i

The mean of estimation error is:

$$E[\sum_{i=1}^{n} \lambda_i [Z(x_i) - \hat{Z}(x_i)]] = m = \sum_{i=1}^{n} [\lambda_i - 1] \tag{10}$$

From (10), the Ordinary Kriging in two locations can be arranged in an equation below [3]:

$$\begin{bmatrix} 0 & \gamma_{12} & 1 \\ \gamma_{21} & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_m \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ 1 \end{bmatrix} \tag{11}$$

From (11) can be derived the equations below [6], [7]:

$$\lambda_1 = \frac{1}{2} + \frac{\gamma_{2V} + \gamma_{1V}}{2\gamma_{12}} \; ; \; \lambda_2 = \frac{1}{2} + \frac{\gamma_{1V} - \gamma_{2V}}{2\gamma_{12}} \tag{12}$$

where:
γ_{12} : semivariogram of two observed locations
$\gamma_{1V};\gamma_{2V}$: semivariogram between observed and unobserved location that obtained on sperical model in (8)

5 GSTAR-Kriging Model

In this paper, we proposed the order of GSTAR model at observed locations as GSTAR (1;1) model. In GSTAR (1;1) at two observed locations, we can predict the value of parameter $\hat{\phi}_{1j}$ at unobserved locations by using the GSTAR-Kriging model below:

$$\hat{\phi}_{1j}(s_0) = \sum_{i=1} 2\lambda_i \phi_{1j}(s_i) \; ; \; \text{where} \sum_{i=1} 2\lambda_i = 1 \text{ and } j = 0,1 \tag{13}$$

6 Data Set and Procedure

The GSTAR-Kriging model is applied to monthly rainfall data at Majalengka, Kuningan and Ciamis Regencies. These locations also are called East Priangan Region. The data is obtained from Climate Hazard Infra Red Precipitation with Station (CHRIPS) website (http://chg.geog.ucsb.edu/) in [10],[16]. The rainfall data is annual data in wet session or DJF season (December, January and February) from year 1981 to 2016. The spatial data consists of latitude and longitude data in these regencies. The rainfall data is divided into two part of data, training data and testing data. Training are 30 years to estimate the parameter of GSTAR(1;1) model and testing data are 3 years to validate the GSTAR(1;1) model. The locations of research is showed at Figure 3.

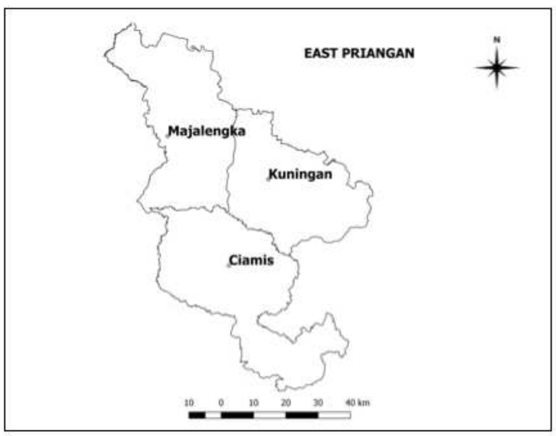


Fig. 3: Map of Research Locations

Procedures of GSTAR-Kriging model can be summarized as [6], [7]:
1). Compute the correlation rainfall data in three locations and check the stationary of time series data
2). Specify the observed and unobserved locations at some combinations
3). Estimate the parameter GSTAR model with Ordinary Least Square (OLS) method at observed locations
4). Check the residual of GSTAR (1;1) assumption as [8], [9]:
a. Homoscesdasticity test, to check that the variance of residual is constant or not. The null hypothesis is the constant variance of error model (homoscedasticity).
b. Multivariate Normal Test. The null hypothesis is the error of model distributed multivariate normal. We can use the Royston test.
c. White Noise Test, to check the independency of error model. The null hypothesis is the error model have multivariate white noise, we can use the Pormanteau test.
5). Validation model GSTAR (1;1) can be showed by MAPE value. If the MAPE values is less than 10 percent, it means that the model has the high forecasting accurately [9].
6). Find the experimental and theoretical semivariogram
7). Find the estimate of Kriging weight
8). Furthermore we predict the parameters of GSTAR-Kriging model at unobserved locations

7 Result and Discussion

7.1 Descriptive Statistics

Data processing on this paper used the R-software and Microsoft Excel. In Figure 4, we plot the rainfall data in wet season DJF fluctuate from 1981 to 2016 at each location, for example at Majalengka Regency. Despite fluctuating, the rainfall in the wet season does not have the high variance. We use the three stages of time series analysis from Box-Jenkins: identification, estimation parameter and checking diagnostic before we forecast the rainfall for the future time [17].

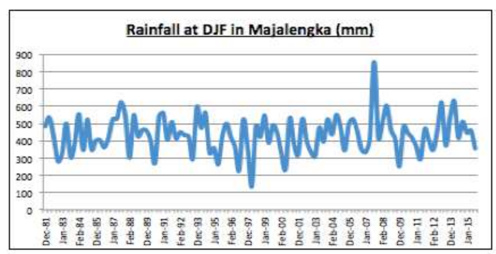


Fig. 4: Plot of Rainfall Data DJF at Majalengka Regency in West Java

Table 1 shows the summary rainfall data in three regencies. Majalengka Regency has the high mean value of rainfall data and Ciamis Regency has the highest variation of rainfall data. The highest maximum value of rainfall data is in Ciamis Regency and the smallest minimum values in in Kuningan Regency.

Table 1: Summary of Rainfall Data (mm)

Regency	Sample	Mean	Std	Max	Min
Majalengka	36	1303.2	188.002	1782.6	951.7
Kuningan	36	1251.5	179.84	1712.6	925.7
Ciamis	36	1290.4	214.87	1840.5	947.1

The correlation of rainfall data at wet season in three regencies are high that showed at Table 2.

Table 2: Correlation of Rainfall Data at Several Regencies in West Java

Regency	Majalengka	Kuningan	Ciamis
Majalengka	1.00	0.94	0.93
Kuningan	0.94	1.00	0.91
Ciamis	0.93	0.91	1.00

Furthermore, we can use the rainfall DJF data at three regencies for implementing the GSTAR-Kriging model. Following the procedure of the GSTAR-Kriging in sub section 6 above, so we have the result of stationary test, estimation parameters GSTAR, diagnostic checking, model validation, experimental semivariogram, estimation of Kriging weight and GSTAR-Kriging model for rainfall data.

7.2 Stationary Test

GSTAR is the spatiotemporal model that requires the stationary data in mean and variance. Table 3 shows that the result of Phillips Perron test and Box-Cox Lambda. The *p*-value in Phillips Perron test is less than 0.05. It means the data have stationary in mean. Overall, the Box-cox Lambda values are higher than 1 that can be concluded the rainfall data are stationary in variance [8].

Table 3: Stationary Test for Rainfall Data at Several Regencies in West Java

Regency	Phillips Perron Test (*p*-value)	Box-Cox Lambda
Majalengka	0.01	1.2015
Kuningan	0.01	1.4929
Ciamis	0.01	-0.9999

7.3 Estimation of GSTAR (1;1) Model

GSTAR (1;1) combine three sets of GSTAR model based on observed and unobserved locations. The combinations have been showed in Table 4.

Table 4: Combination Location of Research

Combination	Observed Location	Unobserved Location
Model I	Majalengka and Kuningan Regencies	Ciamis Regency
Model II	Majalengka and Ciamis Regencies	Kuningan Regency
Model III	Kuningan and Ciamis Regencies	Majalengka Regency

In this paper we use the binary weight matrix where the contiguous locations have the value 1 and non contiguous locations have the value 0 [1]. The locations in GSTAR (1;1) model in this paper are restricted by two regencies so that we can write the binary weight matrix as:

$$W = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

We use the OLS method to estimate the GSTAR(1;1) model and get the result as displayed in Table 5.

Table 5: GSTAR (1;1) Estimation Parameter Model using OLS

Combination	Observed Location	$\hat{\phi}_0$	$\hat{\phi}_1$
I	Majalengka	-0.7388	0.7468
	Kuningan	0.4473	-0.5361
II	Majalengka	-0.3537	0.2603
	Ciamis	0.2678	-0.3927
III	Kuningan	-0.1091	0.0206
	Ciamis	-0.059	-0.0047

By using R software, we can estimate the parameter of GSTAR (1;1) model with OLS method. The OLS estimator of GSTAR parameters model have properties of consistency and asymptotic normality [18]. The result of estimation is displayed in Table 5. The estimation of parameters can be written into estimation model as written in equation (2):

GSTAR (1;1) model at combination I

$$\begin{bmatrix} \hat{Z}_1(t) \\ \hat{Z}_2(t) \end{bmatrix} = \begin{bmatrix} -0.7388 & 0 \\ 0 & 0.4473 \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_2(t-1) \end{bmatrix} + \begin{bmatrix} 0.7468 & 0 \\ 0 & -0.5361 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_2(t-1) \end{bmatrix}$$

$$\begin{bmatrix} \hat{Z}_1(t) \\ \hat{Z}_2(t) \end{bmatrix} = \begin{bmatrix} -0.7388 & 0.7468 \\ -0.5361 & 0.4473 \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_2(t-1) \end{bmatrix}$$

GSTAR (1;1) model at combination II

$$\begin{bmatrix} \hat{Z}_1(t) \\ \hat{Z}_3(t) \end{bmatrix} = \begin{bmatrix} -0.3537 & 0 \\ 0 & 0.2678 \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_3(t-1) \end{bmatrix} + \begin{bmatrix} 0.2603 & 0 \\ 0 & -0.3927 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_3(t-1) \end{bmatrix}$$

$$\begin{bmatrix} \hat{Z}_1(t) \\ \hat{Z}_3(t) \end{bmatrix} = \begin{bmatrix} -0.3537 & 0.2603 \\ -0.3927 & 0.2678 \end{bmatrix} \begin{bmatrix} Z_1(t-1) \\ Z_3(t-1) \end{bmatrix}$$

GSTAR (1;1) model at combination III

$$\begin{bmatrix}\hat{Z}_2(t)\\ \hat{Z}_3(t)\end{bmatrix} = \begin{bmatrix}-0.1091 & 0\\ 0 & 0.059\end{bmatrix}\begin{bmatrix}Z_2(t-1)\\ Z_3(t-1)\end{bmatrix} + \begin{bmatrix}0.0206 & 0\\ 0 & -0.0047\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}Z_2(t-1)\\ Z_3(t-1)\end{bmatrix}$$

$$\begin{bmatrix}\hat{Z}_2(t)\\ \hat{Z}_3(t)\end{bmatrix} = \begin{bmatrix}-0.1091 & 0.0206\\ -0.0047 & 0.059\end{bmatrix}\begin{bmatrix}Z_2(t-1)\\ Z_3(t-1)\end{bmatrix}$$

7.4 Diagnostic Checking

The GSTAR model assumes that the residual have independent identically distributed normal so that we have to check the distribution of the residual.

Table 6: The Result of Homoscedasticity Residual Test

Combination	p-value	Conclusions
Model I	0.999	The Variance of Residual Constant
Model II	0.998	The Variance of Residual Constant
Model III	0.999	The Variance of Residual Constant

By using the R software, we have the p-value of test is bigger than 0.05 that showed in Table 6. It means the model have constant variance or there is no the Autroregressive Conditional Heteroscedasticity (ARCH) error effect.

Table 7: The Result of Multivariate White Noise Residual Test

Lag	Model I		Model II		Model III	
	Q-Stat	p-val	Q-Stat	p-val	Q-Stat	p-val
1	0.146	0.998	0.079	0.999	0.031	0.999
2	1.853	0.985	4.704	0.789	2.941	0.938
...	...	...	...	...	...	...
29	53.381	0.999	56.592	0.999	54.485	0.999
30	53.460	1.000	56.674	0.999	54.5424	1.000

Table 8: The Result of Multivariate Normal Residual Test

Combination	p-value	Conclusion
Model I	0.2096	Residual Multivariate Normal
Model II	0.2665	Residual Multivariate Normal
Model III	0.6305	Residual Multivariate Normal

Table 8 shows the result of Royston test to check the multivariate normal residual. The p-value at all combination model are bigger than 0.05. It means that the residual of model is distributed multivariate normal.

7.5 Model Validation

MAPE is used to measure the accuracy of model that obtained by dividing the percentage of error absolute to actual data [8]. In Table 9 we show that the MAPE value at combination model I and III are less than 10 %. It means the accuracy of model to forecast is high [9].

Table 9: Mean Average Percentage Error (MAPE)

Combination	Location	MAPE by Location	Overall MAPE
Model I	Majalengka	4.67%	8.97%
	Kuningan	13.28%	
Model II	Majalengka	10.95%	12.51%
	Ciamis	14.08%	
Model III	Kuningan	9.05%	7.72%
	Ciamis	6.39%	

7.6 Experimental Semivariogram

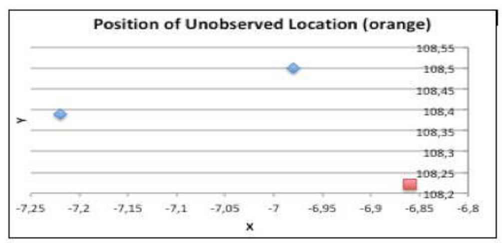


Fig. 5: Position of Unobserved Location of Rainfall Data using GSTAR-Kriging

Figure 5 shows the plot of location points. The coordinate points for every regency are Majalengka (-6.86 ; 108.22), Kuningan (-6.98 ; 108.50) and Ciamis (-7.22 ; 108.39). From the coordinate points we can find the distance of locations by Euclidean distance formula.

Table 10: Experimental and Theoretical Semivariogram

Combination	Experimental (γ_0)			Theoretical	
	ϕ_0	ϕ_1		ϕ_0	ϕ_1
I	0.0205	0.0448	γ_{1V}	0.0205	0.045
			γ_{2V}	0.0184	0.040
II	0.0008	0.00005	γ_{1V}	0.0006	0.0006
			γ_{2V}	0.0005	0.0005
III	0.0054	0.0032	γ_{1V}	0.0054	0.0032
			γ_{2V}	0.0054	0.0032

Table 10 shows the experimental and theoretical semivariogram. The experimental semivariogram can be found by using (6). It shows the semivariogram between observed locations. The theoretical semivariogram can be found by using (8) and it shows the semivariogram between observed locations and unobserved location. If the semivariogram value is small, it means that the value of random variable between locations is similar. For combination II, we have that the values of experimental and theoretical semivariogram are less than others.

7.7 Estimation of Kriging Weight

Ordinary Kriging in equation (9) shows that Kriging weight can be used for predicting the observation at unobserved location. We can find the Kriging weight value using (12). Table 11 shows the Kriging weight values in every location of simulation.

Table 11: Estimation of Kriging Weight

Model	Observed Location	Unobserved Location	Kriging Weight (λ)	Φ_0	Φ_1
I	Majalengka	Ciamis	λ_1	0,4473	0,4473
	Kuningan		λ_2	0,5527	0,5527
II	Majalengka	Kuningan	λ_1	0,4608	0,4608
	Ciamis		λ_2	0,5392	0,5392
III	Kuningan	Majalengka	λ_1	0,5000	0,5000
	Ciamis		λ_2	0,5000	0,5000

If theoretical semivariogram value is small, it shows the similarity of random variable between locations so that these locations will have the bigger Kriging weight value. At combination I, we have the result that the theoretical semivariogram between Kuningan-Ciamis is less than Majalengka-Ciamis so the Kriging weight value Kuningan-Ciamis is bigger. It means to predict the GSTAR parameter in Ciamis Regency is most influenced by Kuningan Regency.

7.8 GSTAR-Kriging Model

By using the Kriging weight values and parameters estimator of GSTAR(1;1) model in observed location, we can predict the GSTAR(1;1) model parameter at unobserved locations. We have the result in Table 12.

If we compare Table 5 and Table 12 at combination I, we have the result of prediction parameter model in unobserved location (Ciamis Regency) similar to the value of parameter in Majalengka Regency. At combination II, we have the result of prediction parameter model in unobserved location (Kuningan Regency) similar to the value of parameter in Majalengka Regency. At combination III, we have the result of prediction parameter model in unobserved location (Majalengka Regency) similar to the value of parameter in Ciamis Regency. It happened because the result of Kriging prediction depends on distance of locations. If the unobserved locations and observed locations is neighbouring, their GSTAR parameter value is almost similar. So that if we predict the rainfall data in unobserved locations, it will be obtained the rainfall data with a pattern similar to observed locations as its neighbour. For example, if we use combination I to predict the rainfall data in Ciamis Regency by generating its GSTAR model based on parameter that obtained from GSTAR-Kriging model with the estimator of GSTAR as an input. $\hat{\phi}_0^* = -0.63$ and $\hat{\phi}_1^* = 0.58$, we obtained the results is similar to rainfall data in Majalengka. We can compare these results with the original rainfall data in Ciamis to validate the GSTAR-Kriging Model. Furthermore, we can apply the GSTAR-Kriging model to other locations that it has not the rainfall data. So, using the GSTAR-Kriging model we can predict the observation of rainfall data at unobserved locations in the future time based on the rainfall data at surrounding locations.

Table 12: The Result of GSTAR-Kriging Model Prediction

Combination	Unobserved Location	Prediction of GSTAR-Kriging ϕ_0^*	ϕ_1^*
I	Ciamis	-0.63	0.58
II	Kuningan	-0.37	0.26
III	Majalengka	-0.06	-0.02

8 Perspective

The annual rainfall data at wet season have high correlation in Majalengka, Kuningan and Ciamis Regencies, and the data have stationary in mean and variance so that GSTAR (1;1) model can be applied. MAPE of GSTAR model in all combination locations show the high enough and good accuracy. The result of GSTAR-Kriging model shows that the prediction parameter at unobserved location similar to observed location as its neighbour at surrounding location. In this paper, we developed idea to generate the GSTAR model in unobserved location by using the prediction of parameters previously in GSTAR model, so that we can forecast the rainfall data in unobserved location in the future time and influenced by another locations as a neighbours in a certain region. Furthermore, the development of GSTAR-Kriging model can be extended to be the GSTAR-X-Kriging or GSTARI-X-Kriging using SUR method to fulfil the assumption of correlated error and exogenous variable in real phenomena.

Acknowledgement

This paper is funded by ALG Unpad contract: 872/UN6.3.2/LT/2017, RKDU contract: 855/UN6.3.1/PL/2017 and PKLN contract: 718/UN6.3.1/PL/2017. The authors thank Rector Universitas Padjadjaran and Ministry of Research, Technology and Higher Education Republic of Indonesia. Also thanks to the anonymous referee for valuable checking to improve this paper.

References

[1] B. N. Ruchjana, The Generalized Space Time Autoregressive Model and Its Application to Oil Production, Unpublished, Dissertation, Institut Teknologi Bandung (2002).

[2] S. A. Borovkova, H. P. Lopuhaa, and B. N. Ruchjana, Generalized STAR with Random Weights, Proceeding of the 17th International workshop on Statistical Modelling, Chania-Greece, pp. 139-147 (2002).

[3] R. A. Olea, Geostatistics for Engineer and Earth Scientists, Kansas Geological Survey, The University of Kansas, Springer Science + Business Media, LLC (1999).

[4] A. N. Falah, A. S. Abdullah, K. Parmikanti, and B. N. Ruchjana, Prediction of Cadmium Pollutant with Ordinary Point Kriging Method Using GStat-R, AIP Conference Proceedings International Conference on Statistics and Its Applications, Vol. 1827, pp. 1-11 (2016).

[5] M. Armstrong, Basic Linear Geostatistics, New York: Springer (1998).

[6] B. N. Ruchjana, A. S. Abdullah, I G. N. M. Jaya, and T. Toharudin, Clustering spatial on the GSTAR model for replacement new oil well, AIP Conference Proceedings Padjadjaran International Physics Symposium, Vol. 1554, pp. 205-209, https://doi.org/10.1063/1.4820321 (2013).

[7] B. N. Ruchjana, Oil Reservoir Characterization using Spatio Temporal Approach, Research Report Universitas Padjadjaran (2005).

[8] W. W. S. Wei, Time Series Analysis Univariate and Multivariate Method, Department of Statistic Temple University: Addison Wesley Publishing Company (2006).

[9] H. Bonar, B. N. Ruchjana and G. Darmawan, Development of Generalized Space Time Autoregressive Integrated with ARCH Error (GSTARI-ARCH) Model based on Consumer Price Index Phenomenon at Several Cities in North Sumatera Province, AIP Conference Proceedings International Conference on Statistics and Its Applications, Vol. 1827, pp. 1-8 (2016).

[10] E. Hermawan, B. N. Ruchjana, A. S. Abdullah, I G. N. M. Jaya, S. B. Sipayung, and S. Rustiana, Development of the statistical ARIMA model: an application for predicting the upcoming of MJO index, Journal of Physics.: Conf. Series., Vol. 893, pp. 1-8, doi:10.1088/1742-6596/893/1/012019 (2017).

[11] A. S. Abdullah, Implementation of Spatial Data Mining using Spatial Autoregressive-Kriging (SAR-Kriging) Model to Predict a Quality of Education in Indonesia. Unpublished. Dissertation. Universitas Gadjah Mada (2009).

[12] D. Astuti, Soemartini, and B. N. Ruchjana, Generalized space time autoregressive with exogenous variable model and its application, Journal of Physics: Conf. Series, Vol. 893, pp. 1-9, doi:10.1088/1742-6596/893/1/012038 (2017).

[13] A. N. Falah, B. Subartini. and B. N. Ruchjana, Application of universal kriging for prediction pollutant using GStat R, Journal of Physics: Conf. Series, Vol. 893, pp. 1-7, doi:10.1088/1742-6596/893/1/012022 (2017).

[14] Setiawan, Suhartono, I. S. Ahmad, and N. I. Rahmawati, Configuring calendar variation based on time series regression method for forecasting a monthly currency inflow and outflow in Central Java, AIP Conference Proceedings Innovation and Analytics Conference and Exhibition, Vol. 1691, pp. 1-8, doi: 10.1063/1.4937106 (2015).

[15] Setiawan, Suhartono, and M. Prastuti, S-GSTAR-SUR Model for Seasonal Spatio Temporal Data Forecasting, Malaysian Journal of Mathematical Sciences, Vol. 10(S) March, pp. 53- 65 (2016).

[16] S. Rustiana, B. N. Ruchjana, A. S. Abdullah, E. Hermawan, S. B. Sipayung, I G. N. M. Jaya, and Krismianto. Rainfall prediction of Cimanuk watershed regions with canonical correlation analysis (CCA), Journal of Physics: Conf. Series, Vol. 893, pp. 1-7 , doi:10.1088/1742-6596/893/1/012021 (2017).

[17] G. E. P. Box and W. L. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day Inc., San Fransisco (1976).

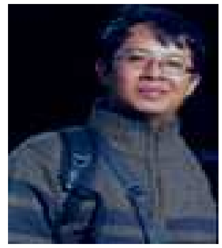
[18] B. N. Ruchjana, S. A. Borovkova, and H. P. Lopuhaa, Least squares estimation of Generalized Space Time Autoregressive (GSTAR) model and its properties. AIP Conference Proceedings International Conference on Research and Education in Mathematics , Vol. 1450, pp. 61-64; https://doi.org/10.1063/1.4724118 (2012).

Atje Setiawan Abdullah is Associate Professor at Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia. He received the PhD degree in Computer Science at Universitas Gadjah Mada, Indonesia (2009). His research interests are in the areas of data mining including theoretical and application specially in ethno-informatics. He is a member of Consortium Spatio Temporal Data Mining and Optimization Modeling in Indonesia and a Member of Indonesian Ethnoscience Studies Center Universitas Padjadjaran.

Setiawan Matoha is Associate Professor at Department of Statistics, Faculty of Mathematics, Computation and Data Science, Institut Teknologi Sepuluh Nopember Surabaya, Indonesia. He received the PhD degree in Statistics at Institut Pertanian Bogor , Indonesia (2007). His research interests are in the areas of econometrics, economics and business statistics and spatial econometrics. He is a member of Consortium Spatio Temporal Data Mining and Optimization Modeling in Indonesia and a Head of Research Center for Regional Development and Community Empowerment (PDPM) LPPM ITS.

Deltha Airuzsh Lubis is a Statistician and Staff of Division of Production Statistics in Statistics of South Tapanuli Regency, Indonesia. He received the Magister Degree in Applied Statistics at Universitas Padjadjaran, Indonesia (2018). His research interests are in the areas of Multivariate Time Series and application.

Annisa Nur Falah is an assistant lecturer at Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia. She received the Magister degree in Mathematics at Universitas Padjadjaran, Indonesia (2018). Her research interests are in the areas of Geostatistics , Spatial Analysis and application using R Software.



I Gede Nyoman Mindra Jaya is Assistant Professor at Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia. He received the Magister degree in Statistics at Institut Pertanian Bogor, Indonesia (2009). His research interests are in the areas of spatial econometrics specially in Bayesian method for dengue disease mapping.

Eddy Hermawan is Professor at Atmospheric Modeling at Division of Atmospheric Science and Technology Center of National Institute of Aeronautics and Space (LAPAN) Bandung. He received the PhD Degree of Atmospheric Science of Geophysics and Meteorology Department of Kyoto University of Japan (1999). He is a Senior Research Scientist and his research interests are in the areas of atmospheric modeling.

Budi Nurani Ruchjana is Professor of Statistics at Department of Mathematics, Faculty of mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia. She received the PhD degree in Mathematics and Natural Sciences at Institut Teknologi Bandung, Indonesia (2002). Her research interest is Spatio Temporal Modeling, Stochastics Processes, Geostatistics and its applications. She is a Head of Consortium Spatio Temporal Data Mining and Optimization Modeling in Indonesia and also she is a Head of Indonesian Ethnoscience Studies Center Universitas Padjadjaran. She was a President of the Indonesian Mathematical Society and the Dean of Faculty of Mathematics and Natural Sciences Universitas Padjadjaran for period 2012-2016. Since 2015 she is a Member of Commission for Developing Countries International Mathematical Union (CDC IMU).