Application Cluster Analysis on Time Series Modelling with Spatial Correlations for Rainfall Data in Jember Regency

Ira Yudistira, Alfian Futuhul Hadi, Dian Anggraeni, Budi Lestari
Mathematics Depart. University of Jember Indonesia
e-mail: Irayudistira91@gmail.com

Abstract—Forecasting is a statistical analysis to obtain an overview the development of event in the future. Forecasting performed on time series data, is a series of data observation data that affected by previous data. In addition, time series data is also affected by the location of research, it is called spatial correlations. This correlation can be analyzed by cluster analysis method. Cluster analysis aims to group objects based on similar characteristics. Variability of rainfall in Jember Regency depends on time and space so that there is a spatial correlation. Cluster analysis is expected to form groups that optimal in the data so that the forecasting results more optimal. Selection of the best forecasting models in this study is determined by the smallest RMSE value.

Keywords—Forecasting, Time Series, Cluster Analysis, Spatial Correlations

INTRODUCTION

Cluster analysis is a multivariate technique that has the main purpose to classify objects based on the characteristics possessed. Cluster analysis to classify objects so that any objects that are closest parallels are in the same cluster. The resulting clusters have internal homogeneity and external heterogeneity high. Another goal of cluster analysis is to find a subset of the variables that have the highest correlation, so that these variables can describe subsets without losing important information [1]. Cluster analysis can be used to group data on a time series forecasting.

Time series data is data obtained from the points of time, where there is a correlation between the structure of its values[2]. Time series data from several adjacent sites often have a mutually dependent relationship[3]. This relationship is called spatial correlation. Forecasting time series data that is linked to the time and location modeled with Generalized Space Time Autoregressive (GSTAR). This model is the development of *Space Time Autoregressive* (STAR) models, which tend to be inflexible in the face of the locations that have the characteristics of heterogeneous [1]. Unlike the STAR model, GSTAR not require that the values of the parameters the same for all locations. Therefore, GSTAR model more realistic, because in fact more prevalent models with different parameter model for different locations [4]. The difference between locations are shown in the form of weighting matrix.

Rainfall data is one of the time series data that have spatial correlation. Rainfall variability is heavily dependent on space and time. According to the scale of the space, precipitation variability was much influenced geography, altitude, latitude location, topography and general wind direction [5]. So each location tend to have different rainfall. Based on this, researchers are interested in deploying a cluster analysis on the modeling of time series forecasting with spatial relationships for rainfall data in Jember district. Cluster analysis is used to classify locations homogeneous rainfall station in the same cluster and forecasting carried out on the data that has been clustered. Cluster analysis used in this study is the K-Means cluster analysis and forecasting models that will be used is GSTAR.

RESEARCH METHODS

a. Research Data

The data used in this study is the rainfall data of 77 rainfall stations in the Jember Regency period January 2005 to December 2015. This data is derived from research data [6]. Data is divided into two parts, ie the data in-sample and out-sample of data. In-sample data used to establish a model, while the out-sample data is used to check the power of prediction models created from the data in-sample.

1. Data in-sample: in January 2005 to December 2014 as many as 120 data.
2. Data out-sample: in January 2015 to December 2015 as many as 12 data.

List of the 77 rainfall station can be seen in **Table 1**.

Table 1 Data rainfall station in the Jember regency

No	Rainfall Station	No	Rainfall Station
1	Watuurip	40	Sabrang DM. 4
2	Tanggul	41	Sabrang SB. 1
3	Darungan	42	Sumberejo
4	Sukorejo	43	Tempurejo
5	Dam Langkap	44	Sanenrejo
6	Dam Kijingan	45	Kr. Kedawung
7	Dam Tugusari	46	Renes (Ajung)
8	Wringin Agung	47	Dam Makam
9	Pladingan	48	Dam Pecoro
10	Pondokwaluh	49	Rambipuji
11	Kencong	50	Rawatamtu
12	Kencong	51	Curahmalang
13	Wonorejo	52	Dam Klatakan
14	Bagorejo	53	Dam Karanganom
15	Gumukmas(BT)	54	Dam Pono
16	Bedodo	55	Dam Manggis
17	Gumukmas(KT)	56	Dam Semangir
18	Menampu	57	Bintoro
19	Pondokjoyo	58	Dam Sembah
20	Pondokjoyo	59	Kopang
21	Semboro	60	Wirolegi
22	Paleran	61	Jember
23	Puger	62	Dam Tegal Batu
24	Grenden	63	Sukowono
25	Jambearum	64	Sukorejo
26	Balung	65	Sbr. Jambe
27	Karangduren	66	Cumedak
28	Gumelar Timur	67	Sbr. Kalong
29	Tamansari	68	Ajung
30	Glundengan	69	Ledokombo
31	Lojejer	70	Suren
32	Ampel	71	Sumberjati
33	Tanjungrejo	72	Silo
34	Kesilir	73	Seputih
35	Dam Talang	74	Jatian
36	Jenggawah	75	Kottok
37	Kemuningsari	76	Pakusari
38	Jatisari	77	Dam Arjasa
39	Karang Anyar		

b. Grouping Data with K-Means Cluster Analysis

In this study 77th rainfall station are grouped into 4 clusters. Forecasting is performed on each grouping. The grouping data in this study using the K-Means cluster analysis. K-Means is one of the non-hierarchical cluster method used to classify data that have a large size [7]. The process of grouping with K-means is as follows:

1. Determine the number of clusters (k)
2. Determine the center of the cluster. Cluster centers is the average of the whole object in the group.

3. Determine the shortest distance (Euclidean) each object with the center of the cluster.
4. Grouped object to a cluster based on the flats nearby.
5. Recount central cluster if the object is moved from the starting position.
6. Repeat step 2 until no further transfer of objects between clusters.

c. Identification of Stationary Data

Time series data is said to be stationary if the mean and variance constant or no systematic change. Formally to identify stationary data is done by using Augmented Dickey Fuller (ADF) or by looking at the cross correlation matrix scheme MACF and MPACF. If the plot MACF and MPACF down slowly, the data is not stationary to mean so needs differencing. Similarly if the value up and down value on lambda is less than zero then the data has not been stationary on the variant, so it is necessary to transform Box Cox [8].

ADF test is a test to determine whether a stationary time series data contains unit root. The model used is

$$\Delta Y_t = \beta_1 + \delta Y_{(t-1)} + e_1 \quad (1)$$

where $\delta = \rho - 1$ and the hypothesis is

$H_0: \delta = 1$ (Y_t is not stationary)

$H_1: \delta < 1$ (Y_t is stationary)

The significance test for this hypothesis using test τ (tau) with a test statistic is $\tau = \frac{\rho}{(SE(\rho))}$. Hypothesis H_0 rejected if τ is smaller than the value of τ in ADF with certain real level or *p-value* < 5% [9].

d. Identification GSTAR Model

GSTAR is a model that has a relationship between time and location where the locations studied had heterogeneous characteristics. In GSTAR, parameter values at the same spatial lag allowed different. In matrix notation, GSTAR models with p autoregressive degrees and degrees of spatial $\lambda_1, \lambda_2, \cdots, \lambda_p$ is modeled by

$$Z(t) = \sum_{k=1}^{p} [\emptyset_{k0} + \emptyset_{k1} W] Z(t-1) + e(t) \quad (2)$$

where is

$\emptyset_{k0} = di \quad (\phi_{k0}^1, \cdots, \phi_{k0}^n)$

$\emptyset_{k1} = dia \quad (\phi_{k1}^1, \cdots, \phi_{k1}^n)$, and

W = Weight with $W_{ii} = 0$ and $\sum_{i\neq j} W_{ii} = 0$

$e(t)$ = noise vector size ($n \times 1$) multivariate normal distribution with mean 0 and variance-covariance matrix $\frac{\sigma_2}{N}$ [10].

e. Selection of Order GSTAR Models

The selection of spatial order GSTAR generally restricted to the order of 1, because the higher order will be difficult to interpret[11]. Meanwhile, on the order of time (autoregressive) can be determined by AIC (Akaike Information Criterian). Selection of the best order models can be determined by looking at the smallest AIC value.

The calculation of the value AIC formulated with

$$AIC(p) = \ln \left| \sum_{u}^{\sim} (p) \right| + \frac{2p}{T} K^2$$

Where $\sum_{\tilde{u}}(p) = T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ is a matrix of the estimated residual vector autoregressive models (p) and T and K respectively are the number of residual and number of variable [4].

f. Weight Calculation Model Location GSTAR

One of the locations weight on GSTAR is the inverse distance. This weight value is obtained by calculating the distance between the actual locations. The closer two locations will have greater weight value. Weighting by the inverse distance refers to the distance between sites, for example the distance between the four locations are defined:

r_1 = distance between location 1 and 2
r_2 = distance between locations 1 and 3
r_3 = distance between locations 1 and 4
r_4 = distance between locations 2 and 3
r_5 = distance between locations 2 and 4
r_6 = distance between locations 3 and 4

Written in matrix form:

$$\boldsymbol{W} = \begin{bmatrix} 0 & W_{12} & W_{13} & W_{14} \\ W_{21} & 0 & W_{23} & W_{24} \\ W_{31} & W_{32} & 0 & W_{34} \\ W_{41} & W_{42} & W_{43} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \frac{r_2+r_3}{r_1+r_2+r_3} & \frac{r_1+r_3}{r_1+r_2+r_3} & \frac{r_1+r_2}{r_1+r_2+r_3} \\ \frac{r_4+r_5}{r_1+r_4+r_5} & 0 & \frac{r_1+r_5}{r_1+r_4+r_5} & \frac{r_1+r_4}{r_1+r_4+r_5} \\ \frac{r_4+r_6}{r_2+r_4+r_6} & \frac{r_2+r_6}{r_2+r_4+r_6} & 0 & \frac{r_2+r_4}{r_2+r_4+r_6} \\ \frac{r_5+r_6}{r_3+r_5+r_6} & \frac{r_3+r_6}{r_3+r_5+r_6} & \frac{r_3+r_5}{r_3+r_5+r_6} & 0 \end{bmatrix}$$

The above matrix is standardized in the form Wij^* to get $\sum_{i\neq j} Wij^{(l)}$ $i \neq j = 1$ [12].

g. Parameter Estimation GSTAR Model

Parameter estimation GSTAR be done using Least Squares Method. GSTAR with order $p = 1$ and order spatial 1 will result model

$$Z_i(t) = \emptyset_{10}{}^{(i)} Z_i(t-1) + \emptyset_{11}{}^{(i)} \sum_{j=1}^{N} W_{ij} Z_j(t-1) + e_i(t)$$

With $Z_i(t)$ expressed observations on $t = 0,1,\cdots,T$ for $i = 1,2,\cdots,N$ location.

This applies to linear form $Y_i = X_i \beta_i + e_i$

$$\boldsymbol{Y_i} = \begin{bmatrix} Z_i(1) \\ Z_i(2) \\ \vdots \\ Z_i(t) \end{bmatrix}, \quad \boldsymbol{X_i} = \begin{bmatrix} Z_i(0) & V_i(0) \\ Z_i(1) & V_i(1) \\ \vdots & \vdots \\ Z_i(t-1) & V_i(t-1) \end{bmatrix},$$

$$\boldsymbol{e_i} = \begin{bmatrix} e_i(1) \\ e_i(2) \\ \vdots \\ e_i(t) \end{bmatrix}$$

$$\boldsymbol{\beta} = \left(\emptyset_{10}{}^{(1)}, \emptyset_{10}{}^{(2)}, \cdots, \emptyset_{10}{}^{(N)}; \emptyset_{11}{}^{(1)}, \emptyset_{11}{}^{(2)}, \cdots, \emptyset_{11}{}^{(N)}\right)$$

with

$$V_i(t) = \sum_{j=1}^{N} W_{ij} Z_j(t)$$

Match the model for all the linear model is $Y = \beta X +$ with $Y = (Y_1', Y_2', \cdots, Y_N')'$

$X = (X_1, X_2, \cdots, X_N)$

$\beta = (\beta_1', \beta_2', \cdots, \beta_N')'$ and

$e = (e_1', e_2', \cdots, e_N')'$

so the form of the least squares estimation of $\hat{\beta}_T$ is $\hat{\beta}_T = (X'X)^{-1} X'y$ [13].

h. Feasibility GSTAR Model

GSTAR be feasible if it meets the constant variance (white noise). Testing assumptions need to be conducted to determine whether the residual meet the white noise. if the layout contained in the AIC values lag AR (0) and MA (0) then the residual meet the assumption of white noise[4]. Fulfillment Assuming white noise can be done using the portmanteau test with significance level of 5%.

RESULTS AND DISCUSSION

a. Identification of Data

77th rainfall station are grouped into 4 and 6 clusters using the K-Means. The results are summarized in table 2.

Table 2. The Result of Grouped Data

No.	4 Cluster	Number of Object	6 Cluster	Number of Object
1.	Cluster 1 (Z_{41})	15	Cluster 1 (Z_{61})	14
2.	Cluster 2 (Z_{42})	24	Cluster 2 (Z_{62})	22
3.	Cluster 3 (Z_{43})	29	Cluster 3 (Z_{63})	25
4.	Cluster 4 (Z_{44})	9	Cluster 4 (Z_{64})	3
5.	-	-	Cluster 5 (Z_{65})	5
6.	-	-	Cluster 6 (Z_{66})	8
	Total	77	Total	77

Forecasting is performed on each grouping. The data used is the average value of rainfall from each station in these groups. Data used in forecasting must be stationary on the mean and variance. Stationary on the variance can see from plot of data.

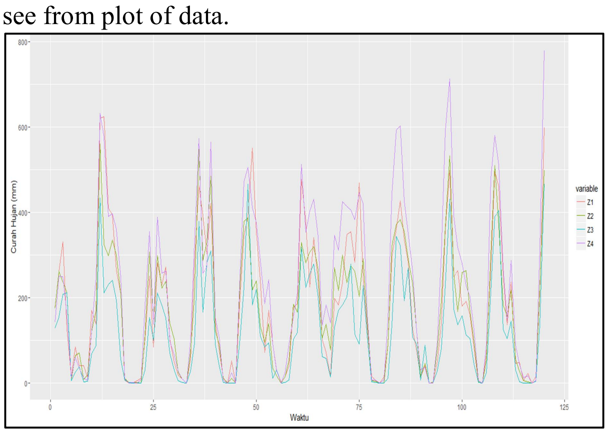

Fig 1. Plot of Data with 4 grouping

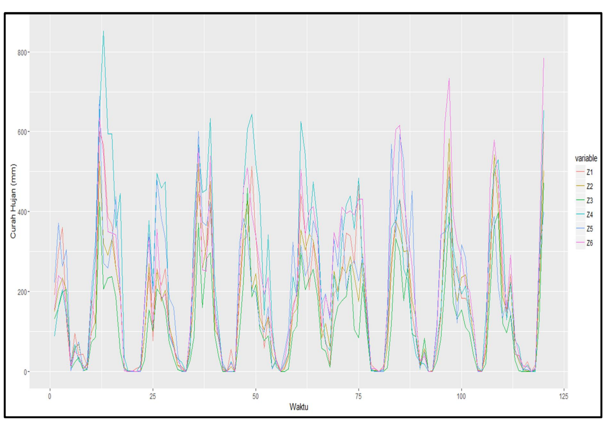

Fig 2. Plot of Data with 6 grouping

Plot data on Figure 1. and Figure 2. have not significant changes so can be said that the data has been stationary for variant. Stationary on the mean can see from ADF test.

Table 3. The Result of ADF Test

Clustering	Location	ADF-test	p-value	Description
4 Cluster	Z_{41}	-7.8237	0.01	stationary
	Z_{42}	-7.6768	0.01	stationary
	Z_{43}	-7.7669	0.01	stationary
	Z_{44}	-7.7301	0.01	stationary
6 Cluster	Z_{61}	-7.7472	0.01	stationary
	Z_{62}	-7.5623	0.01	stationary
	Z_{63}	-7.7967	0.01	stationary
	Z_{64}	-8.3247	0.01	stationary
	Z_{65}	-6.9225	0.01	stationary
	Z_{66}	-7.5013	0.01	stationary

From table we can see that ADF test generate p-value $< 5\%$, so we can say that the data is stationary on the means.

GSTAR Modeling

GSTAR models depend on spatial and autoregressive order. Spatial order of GSTAR model restricted to only one. While the autoregressive order model can be determined from the value of the smallest MIC. The smallest MIC value lies in AR(1) and MA(0) amounting to 32.533335 for 4 grouping and 49.144803 for 6 Grouping. so that got p value is 1 and the models formed by

$$Z_i(t) = [\emptyset_{10} + \emptyset_{11}W]Z(t-1) + e(t)$$
$$= \emptyset_{10}Z(t-1) + \emptyset_{11}WZ(t-1) + e(t)$$

b. Weights Locations

Location weight in this study using inverse distance weighting. The distance between location calculated based on the shortest distance between stations on each cluster. Inverse distance weighting presented in the following matrix.

$$W_{ij} = \begin{bmatrix} 0 & 0.763 & 0.575 & 0.663 \\ 0.614 & 0 & 0.728 & 0.658 \\ 0.635 & 0.857 & 0 & 0.508 \\ 0.699 & 0.813 & 0.488 & 0 \end{bmatrix} \text{ for 4 grouping}$$

and

$$W_{ij} = \begin{bmatrix} 0 & 0.924 & 0.888 & 0.932 & 0.364 & 0.892 \\ 0.753 & 0 & 0.826 & 0.652 & 0.913 & 0.856 \\ 0.843 & 0.925 & 0 & 0.722 & 0.768 & 0.741 \\ 0.931 & 0.892 & 0.800 & 0 & 0.438 & 0.939 \\ 0.618 & 0.984 & 0.901 & 0.666 & 0 & 0.831 \\ 0.840 & 0.935 & 0.728 & 0.911 & 0.586 & 0 \end{bmatrix}$$

for 6 grouping.

c. Parameter estimation

Parameter estimation in this study using OLS (Ordinary Least Square). Parameter estimation results are summarized in the following table.

Table 4. The Result of Parameter estimation results for 4 grouping

Parameter	Estimasi	Std. Error	t-value	p-value	Sig
$\emptyset_{10}$	0.19894	0.18510	1.075	0.28303	
$\emptyset_{20}$	0.33726	0.27760	1.215	0.22501	
$\emptyset_{30}$	-0.48306	0.25659	-1.883	0.06038	.
$\emptyset_{40}$	1.29705	0.21466	6.042	3.1e-09	***
$\emptyset_{11}$	0.29153	0.10029	2.907	0.00382	**
$\emptyset_{21}$	0.15833	0.13271	1.193	0.23346	
$\emptyset_{31}$	0.42742	0.09525	4.487	9.1e-06	***
$\emptyset_{41}$	-0.40856	0.14727	-2.774	0.00575	**

Table 5. The Result of Parameter estimation results for 6 grouping

Parameter	Estimasi	Std. Error	t-hitung	p-value	Sig
$\emptyset_{10}$	-0.02353	0.24149	-0.097	0.922413	
$\emptyset_{20}$	0.18773	0.06082	3.087	0.002105	**
$\emptyset_{30}$	-0.46668	0.30212	-1.545	0.122870	
$\emptyset_{40}$	0.26518	0.06998	3.789	0.000164	***
$\emptyset_{50}$	-0.44221	0.26471	-1.671	0.095260	.
$\emptyset_{60}$	0.19598	0.04650	4.215	2.83e-05	***
$\emptyset_{11}$	0.50106	0.12064	4.153	3.68e-05	***
$\emptyset_{21}$	0.09230	0.04254	2.170	0.030365	*
$\emptyset_{31}$	0.52564	0.12834	4.096	4.70e-05	***
$\emptyset_{41}$	0.02366	0.03451	0.685	0.493279	
$\emptyset_{51}$	1.09123	0.18616	5.862	7.04e-09	***
$\emptyset_{61}$	-0.12523	0.05918	-2.116	0.034674	*

From the table shows that there are some p-value greater than 0.05 is the parameter $\emptyset_{10}$, $\emptyset_{20}$, $\emptyset_{21}$ for 4 grouping and for $\emptyset_{10}$, $\emptyset_{20}$, $\emptyset_{41}$ 6 grouping. This indicates that parameter is not significant but the parameter is still used in modeling in forecasting GSTAR because there is no guarantee that the models created from the significant parameters will generate small error values. So the models for 4 grouping obtained are as follows

$$\hat{Z}_{41}(t) = 0.199Z_{41}(t-1) + 0.222Z_{42}(t-1) + 0.168Z_{43}(t-1) + 0.193Z_{44}(t-1) + e_1(t)$$
$$\hat{Z}_{42}(t) = 0.337Z_{42}(t-1) + 0.097Z_{41}(t-1) + 0.115Z_{43}(t-1) + 0.104Z_{44}(t-1) + e_2(t)$$
$$\hat{Z}_{43}(t) = -0.483Z_{43}(t-1) + 0.271Z_{41}(t-1) + 0.366Z_{42}(t-1) + 0.217Z_{44}(t-1) + e_3(t)$$
$$\hat{Z}_{44}(t) = 1.297Z_{44}(t-1) \pm 0.286Z_{41}(t-1) + 0.332Z_{42}(t-1) + 0.199Z_{43}(t-1) + e_4(t)$$

with RMSE 109.7019. And the models for 6 grouping obtained are as follows

$$\hat{Z}_{61}(t) = -0.024Z_{61}(t-1) + 0.924Z_{62}(t-1) + 0.888Z_{63}(t-1) + 0.932Z_{64}(t-1) + 0.364Z_{65}(t-1) + 0.892Z_{66}(t-1) + e_1(t)$$
$$\hat{Z}_{62}(t) = -0.467Z_{62}(t-1) + 0.753Z_{61}(t-1) + 0.826Z_{63}(t-1) + 0.652Z_{64}(t-1) + 0.913Z_{65}(t-1) + 0.856Z_{66}(t-1) + e_2(t)$$
$$\hat{Z}_{63}(t) = -0.442Z_{63}(t-1) + 0.843Z_{61}(t-1) + 0.925Z_{62}(t-1) + 0.722Z_{64}(t-1) + 0.768Z_{65}(t-1) + 0.741Z_{66}(t-1) + e_3(t)$$
$$\hat{Z}_{64}(t) = 0.501Z_{64}(t-1) + 0.931Z_{61}(t-1) + 0.892Z_{62}(t-1) + 0.800Z_{63}(t-1) + 0.438Z_{65}(t-1) + 0.939Z_{66}(t-1) + e_4(t)$$
$$\hat{Z}_{65}(t) = 0.526Z_{65}(t-1) + 0.618Z_{61}(t-1) + 0.984Z_{62}(t-1) + 0.901Z_{63}(t-1) + 0.666Z_{64}(t-1) + 0.831Z_{66}(t-1) + e_5(t)$$
$$\hat{Z}_{66}(t) = 1.091Z_{66}(t-1) + -0.105Z_{61}(t-1) + -0.117Z_{62}(t-1) + -0.091Z_{63}(t-1) + -0.114Z_{64}(t-1) + -0.073Z_{65}(t-1) + e_6(t)$$

with RMSE 105.2595.

d. Feasibility of Model

Feasibility of Model using Portmanteau test. Portmanteau test generate p-value is greater than 0.05. This suggests that the residuals still meet the assumptions of "white noise". Therefore we can conclude the model fit for use. Portmanteau test results are summarized in the following table 6.

Table 6. The Result of Portmanteau test

lag	4 Cluster		6 Cluster	
	df	p-value	df	p-value
2	16	0.0832	36	0.0823
3	32	0.5175	72	0.5066
4	46	0.2063	108	0.7007
5	64	0.1342	144	0.7003
6	80	0.2094	180	0.6696
7	96	0.3957	216	0.7095
8	112	0.2789	252	0.8578
9	128	0.0724	288	0.9381
10	144	0.0824	324	0.9269

CONCLUSION

Based on data analysis, RMSE values generated at grouping by 6 groups is smaller than grouping by 4 groups. So we can conclude that grouping by 6 groups making better forecast model.

REFERENCES

[1] N. Hajarisman, "Analisis Statistika Multivariat Terapan", Bandung: Universitas Islam Bandung, 2003.

[2] T. W. Liao, "Clustering of Time Series Data-Survey", Pattern Recognition, vol. 38, pp 1857–1874, 2005.

[3] S. A. Borovkova, H. P. Lopuhaa and B. N. Ruchjana, "Generalized STAR Model with Eksperimental Weights". In M. Stasionopoulus and G. Toulomi (Eds.), Proceeding of the 17th International Workshop on Statistical Modeling, Chania, pp. 139-147, 2002.

[4] R. S. Tsay, "Analysis of Financial Time Series". New Jersey: John Wiley & Sons, 2005.

[4] E. Aldrian and R. D. Susanto, "Identification Of Three Dominant Rainfall Regions Within Indonesia And Their Relationship To Sea Surface Temperature", International Journal Of Climatology, vol. 23, pp. 1435-1452, 2003.

[5] F. Nurcahyani, "Pengelompokkan Stasiun Hujan untuk Model *Generalized Space Time Autoregressive* (GSTAR) pada Peramalan Curah Hujan Kabupaten Jember dengan Tiga Pembobotan", Thesis, Jember: Jurusan Matematika FMIPA Universitas Jember, 2016.

[6] S. A. P. Rani, H. Kusdarwati and E. Sumarminingsih, "Pemodelan Generalized Space Time Autoregressive (GSTAR(p1)): Penerapan pada Data Kesakitan Penyakit ISPA di Kota Malang", Malang: Universitas Brawijaya, 2013.

[7] S. Makridakis, S. C. Wheelwriht and V. E. McGee, "Metode dan Aplikasi Peramalan", Jakarta: Erlangga, 1992.

[8] R. Harris and S. Robert, "Applied Time Series Modelling and Forecasting", England: John Wiley & Sons, Inc, 2003.

[9] R. Sitepu, "Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan", Jurnal Penelitian Sains, vol. 14,issue 14303, pp. 11-15, 2011.

[10] D. W. Wutsqa, Suhartono and B. Sujito, "Generalized Space-Time Autoregressive Modeling", Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications (ICMSA2010), Universitas Tunku Abdul Rahman, Kuala Lumpur, Malaysia, 2010.

[11] D. Anggraeni, A. Prahutama and S. Andari, "Aplikasi Generalized Space Time Autoregressive (GSTAR) pada Pemodelan Volume Kendaraan Masuk Tol Semarang", Media Statistika, vol. 6, issue 2, pp. 71-80, 2013.

[12] S. Borovkova, H. P. Lopuhaa and B. N. Ruchjana, "Consistency and Asymptotic Normality of Least Squares Estimators in Generalized STAR Models", Journal, 2008.