# R Software for Parameter Estimation of Spatio Temporal Model

**Conference Paper** · December 2016



Budi Nurani Ruchjana
Universitas Padjadjaran


Atje Setiawan Abdullah
Universitas Padjadjaran


Mindra Jaya
Universitas Padjadjaran

# R Software for Parameter Estimation of Spatio Temporal Model

Budi Nurani Ruchjana, Atje Setiawan Abdullah, I Gede Nyoman Mindra Jaya, Eddy Hermawan

***Abstract*****—**In this paper, we present the application package to estimate parameters of spatio temporal model based on the multivariate time series analysis using the R open-source software. We build packages mainly to estimate the parameters of the Generalized Space Time Autoregressive (GSTAR) model. GSTAR is a combination of time series and spatial models that have parameters vary per location. We use the method of Ordinary Least Squares (OLS) and use the Mean Average Percentage Error (MAPE) to fit the model to spatio temporal real phenomenon. For a case study, we use oil production data from a volcanic layer at Jatibarang Indonesia or climate data such as rainfall in Indonesia. Software R is very user-friendly and it makes calculation easier, processing the data is accurate and faster. R script for the estimation of parameters GSTAR is still limited to a stationary time series model. Therefore, the R program under windows can be developed either for theoretical studies and application.

***Keywords*****—**GSTAR model, MAPE, OLS method, oil production, R software.

## I. Introduction

IN daily life, we always have simultaneous observations from space and time called spatio temporal phenomenon. In this paper, we study one of the spatio temporal models of an alternative based on the analysis of time series of [3]. Famous model for the stationary time series univariate autoregressive model is called Model AR. Time series analysis can be considered as a stochastic process, which is indexed by the observation time. In another way, we have a stochastic process indexed by space or location, we called spatial phenomena [5]. If we have the high correlation between observations, we can extend the univariate time series model to be bivariate or multivariate time series [6]. Furthermore, if we add the information of characteristic of locations in multivariate time series data, we can have a Space Time Autoregressive (STAR) model as proposed by [4] and [7]. The STAR model from [7] is only capable for homogeneous locations. So, [8] expand STAR into Generalized STAR or GSTAR which has the assumption that the model parameters were varied per location, so it can be used for heterogeneous locations.

One of the problems in the estimation of spatio temporal parameters of both STAR and GSTAR models yet provides a user-friendly package. R software is an open source software provided many applications for univariate and multivariate time series [11], but not directly available to the spatio temporal. So, we extend the command from R software under windows to be a script file for parameter estimation of spatio temporal model, especially for GSTAR model. So, this paper addresses two issues, viz:

To propose R software for parameters estimation of GSTAR model using OLS method

To apply the R package of GSTAR to real phenomena such as oil production data or rainfall data at several locations in Indonesia.

Budi Nurani Ruchjana is with the Department of Mathematics Universitas Padjadjaran, Jl. Raya Bandung Sumedang Km 21 Jatinangor, Sumedang 45363, Indonesia (phone/fax: 62-22-7794696; e-mail: budi.nurani@unpad.ac.id).

Atje Setiawan Abdullah is a with the Department of Computer Science Universitas Padjadjaran, Jl. Raya Bandung Sumedang Km 21 Jatinangor, Sumedang 45363, Indonesia (e-mail: atje.setiawan@unpad.ac.id).

I Gede Nyoman Mindra Jaya is with the Department of Statistics Universitas Padjadjaran, Jl. Raya Bandung Sumedang Km 21 Jatinangor, Sumedang 45363, Indonesia (e-mail: jay.komang@gmail.com).

Eddy Hermawan is with the Atmospheric Modeling Division of Atmospheric Science and Technology Center of National Institute of Aeronautics and Space (LAPAN), Jl. Dr. Djundjunan 133 Bandung, Indonesia (e-mail: eddy_lapan@yahoo.com).

## II. Literature Review

### 2.1 The Weight Matrix on Spatio Temporal Model

One characteristic of the spatio temporal STAR and GSTAR models is the weight matrix W. The weight in the spatio temporal model can be determined fixed and subjective by the researchers. It may be chosen to reflect physical properties of the observed system. It can describe the spatial correlation. For regular systems, the weight can be determined through the coding method [1]. This method gives the uniform weights. It means that the weights are only determined by the total number of neighbors of a certain location at a certain distance. Therefore, the spatial lag 1 is the first group of locations where their positions are the same distance or a certain radius.

To determine the weightings within the regular system, we can use the grid system. The lack of this method, we can have a lot of zero weighting for certain spatial lag. But, this is an easy way to gain weight matrix W. Another weight matrix is a binary weight. If the binary weight is standardized, it will be a uniform weight. The position of the location in 2D can be displayed as Fig. 1. So, the center of location 1 has 4 neighbors (location 2, 3, 4, and 5) in lag spatial 1. The uniform weight at spatial lag one can be determined by (1):

$$w_{ij} = \begin{cases} \dfrac{1}{n_i^{(1)}}, i \, and \, j \, are \, 1^{th} \, order \, neighbors \\ 0 \; , otherwise \end{cases} \quad (1)$$

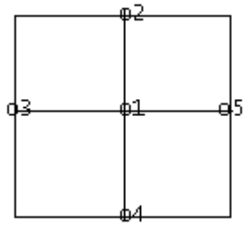


Fig. 1 Spatial Lag One at 2D

where $n_i^{(1)}$ is is the total number of locations in a spatial lag 1 for a certain location i. In a regular system of Fig. 1, using equation (1) we can determine a uniform weight matrix W [7]. The weight has properties:

$$w_{ij} > 0, \sum_{i=1}^{N} w_{ij} = 1, \forall i \text{ and } \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} = N \qquad (2)$$

Set of weights in a spatial lag one is defined as the weight matrix $\mathbf{W}^{(1)}$. It is a square matrix ($N$x$N$), where $N$ is the total locations at an observed system. We can use another weight matrix, such as binary weight, inverse distance weight and also spatial weight using semivariogram [8].

### 2.2 The Generalized STAR Model

The GSTAR model is a natural generalization of STAR models, allowing the autoregressive parameters to vary per location, so the GSTAR model is applicable for the heterogeneous characteristic of sample locations. The GSTAR($\lambda_k$,p) model is written as [8]:

$$\mathbf{z}_{(N\times 1)}(t) = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k} \mathbf{\Phi}_{kl(N\times N)} \mathbf{W}^{(l)}_{(N\times N)} \mathbf{z}_{(N\times 1)}(t-k) + \mathbf{e}_{(N\times 1)}(t) \qquad (3)$$

where:

$\mathbf{z}(t)$ : ($N$x$1$) of observation vector at time $t$

$\lambda_k$: spatial order of the $k^{th}$ autoregressive term

$\Phi_{kl}$ : the diagonal matrices with the diagonal elements as autoregressive and the space time for each location ($\phi_{kl}^{(1)},\ldots,\phi_{kl}^{(N)}$)

$\mathbf{e}(t)$ : the white noise with mean vector $\mathbf{0}$ and variance-covariance matrix $\sigma^2\mathbf{I}$

The GSTAR(1,1) model which both of parameters $\phi_0^{(i)}$ and $\phi_1^{(i)}$ are vary per location is presented by the equation:

$$\mathbf{z}(t) = \mathbf{\Phi}_{10}\mathbf{W}^{(0)}\mathbf{z}(t-1) + \mathbf{\Phi}_{11}\mathbf{W}^{(1)}\mathbf{z}(t-1) + \mathbf{e}(t) \qquad (4)$$

where the $\Phi_{10}$ is the diagonal matrices with the diagonal elements as the autoregressive parameters of lag time 1 for each location ($\phi_0^{(1)}, \ldots, \phi_0^{(N)}$). The $\Phi_{11}$ is the diagonal matrices with the diagonal elements as the space time parameters in the lag spatial 1 and lag time 1 for each location ($\phi_1^{(1)},\ldots,\phi_1^{(N)}$).

The GSTAR(1,1) can be written as the vector autoregressive order one model or VAR(1) [7]:

$$\mathbf{z}(t) = \mathbf{\Phi}\ \mathbf{z}(t-1) + \mathbf{e}(t) \qquad (5)$$

where $\mathbf{\Phi} = \mathbf{\Phi}_{10} + \mathbf{\Phi}_{11}\mathbf{W}$

### 2.3 Least Squares Estimator of GSTAR(1;1)

Consider the GSTAR(1,1) model define through (4). Suppose we have a linear model for GSTAR:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i \qquad (6)$$

where,

$$\boldsymbol{Y}_i = \begin{bmatrix} z_i(2) \\ z_i(3) \\ \vdots \\ z_i(T) \end{bmatrix}, \quad \boldsymbol{X}_i = \begin{bmatrix} z_i(1) & V_i(1) \\ z_i(2) & V_i(2) \\ \vdots & \vdots \\ z_i(T-1) & V_i(T-1) \end{bmatrix}, \quad \boldsymbol{e}_i = \begin{bmatrix} e_i(2) \\ e_i(3) \\ \vdots \\ e_i(T) \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \phi_{10}^{(1)} \\ \vdots \\ \phi_{10}^{(N)} \\ \phi_{11}^{(1)} \\ \vdots \\ \phi_{11}^{(1)} \end{bmatrix} \text{ and}$$

$$V_i(t) = \sum_{j=1}^{N} w_{ij} z_j, \ \forall\ i \neq j \qquad (7)$$

Consequently, the model equations for all locations simultaneously has the linear model structure. Note that for each site $i$ = 1, 2,…, $N$, we have a separate linear model in equation (6), which means that for each site the least squares estimator for $\beta_i$ can be computed separately. However, the value of the estimator does depend on the values of $Z_i(t)$ at other sites , because of the form of $V_i(t)$. So, we have the OLS estimator of GSTAR(1,1) model as written in equation (8):

$$\hat{\mathbf{\Phi}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad (8)$$

where

$$X = diag(X_1, \dots, X_N),$$

The properties of OLS estimator of GSTAR model is consistent and has an asymptotic normality [9].

## III. Research Method

We built a script file of OLS method for estimation of parameters of GSTAR(1,*p*) model for the equation of GSTAR model in equation (7) via R command under windows for time series analysis [10]. The algorithm of OLS method for GSTAR model using R software [11] can be ordered:

1). Determine an order of Autoregressive (AR) model for each location 1, 2, …, p, it can be done also using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)
2). We choose the locations which have the same order of AR to be neighbors
3). Determine weight matrix using the map or position of locations, we can choose a uniform binary or inverse distance weight
4). Select the X matrix for GSTAR(1,1), GSTAR(1,2) or GSTAR(1,p)
5). Use the OLS method to estimate the parameters of GSTAR (1,p)
6). We get the estimator of OLS, R-square and significance test for each parameter at each location of GSTAR(1,p) model.

In this study, we build a package GSTAR using the R software for Windows developed for estimating parameters Generalized Space Time Autoregressive model. GSTAR package has two functions, GSTAR and Model. GSTAR is used to estimate the parameter model and Model is used to select the best order p. For using GSTAR package we have to install this package at R library. The several steps to use GSTAR can be explained:

1). Get the GSTAR package and put this package in specific folder in your computer
2). Install GSTAR package using Install package(s) from a local zip file.
3). Active GSTAR package with type library (GSTAR)
4). Set your working directory seated ("F:/GSTAR")
5). Prepare data set and put in R windows. For example, data set is datax.txt, so we can retrieve data
6). R> Data<-read.table("datax.txt, header=TRUE)
7). Define your Weight matrix
W<-matrix(c(0,0.5,0.5,0.5,0,0.5,0.5,0.5,0),3,3)
8). Estimate your model using GSTAR

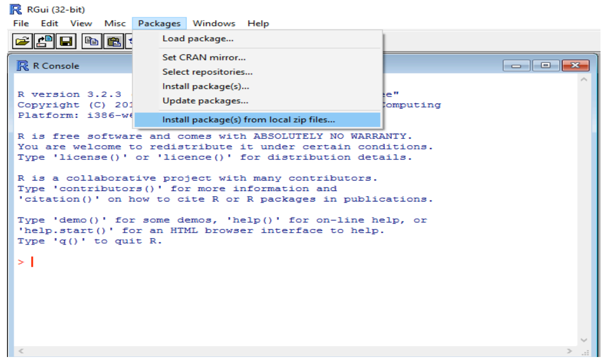


Fig 2. Activation of GSTAR Package using R under Windows

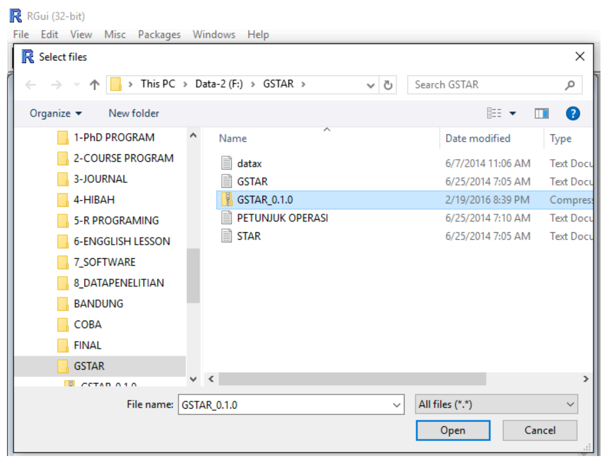


Fig 3. Choosing a GSTAR Package

## IV. Case Study

To give an illustration of using script GSTAR via R software under windows, we use the data of oil production from several wells at a volcanic layer in Jatibarang-West Java-Indonesia. It is a heterogeneous oil field, so the GSTAR model is suitable for this dataset. We built script file using R software to estimate the parameters of GSTAR(1;p) [9].

For running the script file, we use the command:
1). Read data
2). Determine order p of AR model
3). Determine weight matrix W
4). Running program R for GSTAR-OLS

>data
p<-1

We have the plot of 3 oil wells at the volcanic layer in Jatibarang Field:

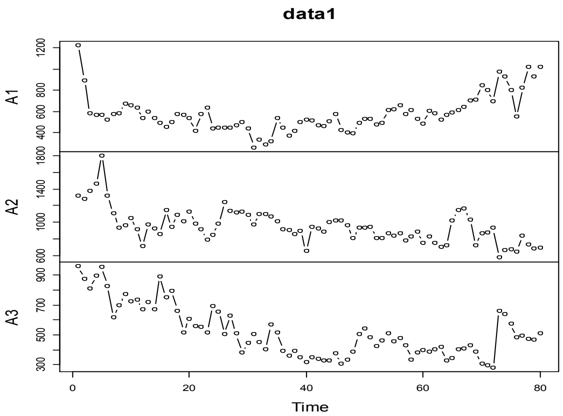


Fig 4. Time Series Plot of Oil Production at 3 Oil Wells

```
> W
# weight matrix

     [,1]  [,2]  [,3]
[1,] 0.000 0.605 0.395
[2,] 0.657 0.000 0.343
[3,] 0.566 0.444 0.000

>GSTAR(data,W,p)
```

and then we get the least squares estimator of GSTAR(1;1) as Fig 5.

```
$GOF
      MSE      MAE      MAPE      MADP      AIC   Rsquare AdjRSquare
1 12546.6 79.58704 0.1201356 0.1173383 2917.117 0.9767942  0.9761915

$Fit

Call:
lm(formula = ZT ~ . - 1, data = GSTAR)

Residuals:
    Min      1Q  Median      3Q     Max
-391.41  -58.24    1.33   55.77  385.34

Coefficients:
     Estimate Std. Error t value Pr(>|t|)
X1    0.90704    0.06219  14.585  < 2e-16 ***
X2    0.85373    0.04749  17.978  < 2e-16 ***
X3    0.79661    0.08731   9.124  < 2e-16 ***
X1.1  0.05843    0.04922   1.187  0.23645
X2.1  0.22830    0.08217   2.778  0.00592 **
X3.1  0.12970    0.06196   2.093  0.03740 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 112 on 231 degrees of freedom
Multiple R-squared:  0.9768,    Adjusted R-squared:  0.9762
F-statistic:  1621 on 6 and 231 DF,  p-value: < 2.2e-16
```

Fig 5. Output for Least Squares Estimator of GSTAR(1;1)

For $p$ =2, we have the complete output 6 parameters of time series and 6 parameters of spatio temporal as Fig 6.

```
$GOF
       MSE      MAE      MAPE      MADP      AIC   Rsquare AdjRSquare
1 12385.63 77.41093 0.1183601 0.1148661 2883.029 0.9774032  0.9761818

$Fit

Call:
lm(formula = ZT ~ . - 1, data = GSTAR)

Residuals:
    Min      1Q  Median      3Q     Max
-345.54  -55.50   -2.07   52.57  401.22

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
X1     1.00285    0.12540   7.997 6.96e-14 ***
X2     0.72989    0.09067   8.050 5.00e-14 ***
X3     0.74092    0.15109   4.904 1.82e-06 ***
X4    -0.07784    0.12734  -0.611    0.542
X5     0.12350    0.09105   1.356    0.176
X6     0.05757    0.14795   0.389    0.698
X1.1 -0.05974    0.15630  -0.382    0.703
X2.1  0.07109    0.17530   0.406    0.685
X3.1  0.06161    0.14200   0.434    0.665
X4.1  0.11251    0.15566   0.723    0.471
X5.1  0.15914    0.17174   0.927    0.355
X6.1  0.06766    0.14401   0.470    0.639
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 111.3 on 222 degrees of freedom
Multiple R-squared:  0.9774,    Adjusted R-squared:  0.9762
F-statistic: 800.2 on 12 and 222 DF,  p-value: < 2.2e-16
```

Fig 6. Output Least Squares Estimator of GSTAR(1;2)

Finally, we can explain that the least squares estimator of the GSTAR(1,1) and GSTAR(1,2) using uniform weight have a fit the oil production model, which the GSTAR(1,1) model has a minimum AIC. For both of GSTAR models, there are significant differences in parameters values at different locations. For all pairs of locations, we tested the parameters. The estimator of GSTAR are significantly different to zero and it has properties of consistent and asymptotic normality [2, 9]

## V. Conclusion

We explain the estimation parameter of spatio temporal, GSTAR(1,$p$) model using OLS method via R software. The package still has a restriction, it is only applicable for stationary data, so we can extend the script for nonstationary data to be GSTARI model or another estimation method, such as Seemingly Unrelated Regression (SUR) for non-constant variance of error.

## APPENDIX

### Script R for GSTAR(1,$p$) using OLS Method

Script R for estimation parameters of GSTAR model is written by Budi Nurani Ruchjana and I Gede Nyoman Mindra Jaya, which has a Copyright from Ministry of Justice and Security Republic of Indonesia, Number 075641 the year 2015.

```
#' Generalized Spatio Temporal Autoregressive
#' Estimation using Ordinary Least Square
#' @param Data, W, p
#' @return Estimate Parameters GSTAR
#' @export

GSTAR<-function(Data,W,p){
  Series<-ts(Data)
```

```
k<-ncol(Data)                          # Number Location
n<-nrow(Data)                          # Number Time
zt<-(stack(as.data.frame(t(Data)))[,1])          # Zt with Lag

# Function Create Lag Data
##############################################
MD<-function(zt){
 M<-matrix(0,(n*k),k)
 z=0
 for( i in 1:(n)){
  for(j in 1:k){
   z<-z+1
   M[z,j]<-zt[z]
  }
 }
 M
}
##############################################
M1<-MD(zt)
MA<-matrix(0,(n*k-k*p),k*p)
MAW<-matrix(0,(n*k-k*p),k*p)
W1<-kronecker(diag(n), W)
ztw<-W1%*%zt
M2<-MD(ztw)
zt<-as.matrix(zt)
ZT<-zt[-(1:(k*p)),]
for (i in 1:p){
 MA[(1:(n*k-k*p)),(k*(i-1)+1):(k*((i-1)+1))]<-M1[(k*(p-
i)+1):(n*k-(k*i)),]
 MAW[(1:(n*k-k*p)),(k*(i-1)+1):(k*((i-1)+1))]<-M2[(k*(p-
i)+1):(n*k-(k*i)),]
}
XT<-MA
WXT<-MAW
GSTAR<-data.frame(ZT,XT,WXT)
GSTARfit<-lm(ZT~.-1,data=GSTAR)
fit<-summary(GSTARfit)
MSE<-(fit$sigma)^2
MAE<-sum(abs(GSTARfit$residuals))/(n*k-k*p)
MAPE<-sum(abs(GSTARfit$residuals)/ZT)/(n*k-k*p)
MADP<-sum(abs(GSTARfit$residuals))/sum(ZT)
AIC<-AIC(GSTARfit)
R2<-fit$r.squared
R2a<-fit$adj.r.squared
Coef<-GSTARfit$coefficients
Error<-data.frame(MSE=MSE, MAE=MAE, MAPE=MAPE,
MADP=MADP, AIC=AIC, Rsquare=R2, AdjRSquare=R2a)
Coefficient<-data.frame(Coef)
Zt<-as.data.frame(matrix(ZT,n-p,k,byrow=T))
Zhat1<-as.data.frame(matrix(GSTARfit$fitted.values,(n-
p),k,byrow = T))
Residual<-as.data.frame(matrix(GSTARfit$residuals,(n-
p),k,byrow = T))
Result<-data.frame(Zt,Zhat1)
Result<-ts(Result)
windows()
plot(Series,plot.type="single", lty=1:3, col =
4:2,xlab="Time",main="Data Series Plot")
windows()
plot(Result,plot.type="multiple", lty=1:3, col =
4:2,main="Predictive vs Observed")
windows()
plot(Result,plot.type="single", lty=1:3, col =
4:2,main="Predictive vs Observed")
Residual<-ts(Residual)
windows()
plot(Residual,plot.type="single", lty=1:3, col =
4:2,main="Residual")
Summary<-list(GOF=Error,Fit=fit)
return(Summary)
}

#' Select the Best p
#' @param p
#' @return p
#' @export

Model<-function(p){
 Error<-matrix(0,p,7)
 for (i in 1:p){
  Error[i,]<-as.matrix(GSTAR(Data,W,i)[[1]])
 }
 colnames(Error)<-c("MSE", "MAE", "MAPE", "MADP",
"AIC", "Rsquare", "AdjRSquare")
 Result<-list(Error=Error)
 return(Result)
}


#Example Data of Oil Production (m3/month) at 3 Oil Wells
>data
```

| A1 | A2 | A3 |
|---|---|---|
| 1223 | 1322 | 961 |
| 895 | 1285 | 872 |
| 581 | 1372 | 809 |
| 565 | 1463 | 895 |
| 568 | 1801 | 955 |
| 525 | 1320 | 828 |
| 574 | 1105 | 618 |
| 584 | 939 | 698 |
| 676 | 966 | 770 |
| 657 | 1054 | 726 |


## ACKNOWLEDGMENT

The authors thank Rector Universitas Padjadjaran who gave funding for this research through Academic Leadership Grant (ALG) Universitas Padjadjaran Indonesia 2016, under the research project entitled on " the Development of Space-time Model and Its Application to Industry, Organization and Environment to Support the West Java Common Goals for the Year of 2015-2019". The authors thank tHigher Education Ministry of Research, Technology and Higher Education for funding research through scheme International Collaboration, and also thank PT Pertamina Jatibarang and LAPAN Bandung for discussion and data for case studies.

## REFERENCES

[1] J. S. Besag, "Spatial interaction and the statistical analysis of lattice system," *J. Roy. Statist. Soc. B*, **Vol. 36**, pp 197-242. 1974.

[2] S. A. Borovkova, H.P. Lopuhaa, and B. N. Ruchjana, "Consistency and Asymptotic Normality of Least Squares Estimators of Generalized STAR models," *Statistica Neerlandica*, 62, nr 4, pp 482-508. 2008.

[3] G. E. P. Box and G. M. Jenkins, "*Time series analysis: forecasting and control*," Holden- Day Inc. San Francisco, 1976.

[4] A. D. Cliff and J. Ord, 1976. "Model building and the analysis of spatial pattern in human geography," *Journal of the Royal Statistical Society, Series B* **37**, 297–348, 1976.

[5] N. Cressie, "Statistics for Spatial Data," John Wiley & Sons., Inc., New York. 1993.

[6] E. J. Hannan, "*Multiple time series*," John Wiley & Sons, New York, 1970

[7] P. E. Pfeifer, and Deutsch, S.J. 1980. "A three stage iterative procedure for space-time m odeling," *Technometrics* **22**, 35–47.

[8] B. N. Ruchjana, "The Generalized Space Time Autoregressive Order One Model and Its Application to Oil production Data," Unpublished Dissertation, Department Mathematics, Institut Teknologi Bandung, Bandung, 2002.

[9] 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 1450, ICREM 5*: The 5th International Conference on Research and Education Mathematics, Bandung, 2005.

[10] R. S. Tsay, " Multivariate Time Series Analysis with R and Financial Applications.," John Wiley, 2014.

[11] A. Coghlan, "A Little Book of R For Time Series *Release 0.2* ", Booklet R, Trust Sanger Institute, Cambridge, 2016.