J. Basic. Appl. Sci. Res., 3(6)860-865, 2013
ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Prediction of Precipitation Data at Batu Town Using the GSTAR (1,*p*)-SUR Model

[1,2,*]Atiek Iriany, [3]Suhariningsih, [4]Budi Nurani Ruchjana, [5]Setiawan

[1]Doctoral Program on Mathematics and Natural Sciences, Post Graduate Program on University of Airlangga, Surabaya, East Java of Indonesia
[2]Lecturer on Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, East Java of Indonesia
[3]Lecturer on Department of Physics, Faculty of Mathematics and Natural Sciences, University of Airlangga, Surabaya, East Java of Indonesia
[4]Lecturer on Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Pajajaran, Bandung, West Java of Indonesia
[5]Lecturer on Department of Statistics, Faculty of Mathematics and Natural Sciences, Sepuluh November Institute of Technology Surabaya, East Java of Indonesia

ABSTRACT

In daily life we often get data is not only observed on time series but also together with spatial which we called a space time pheomena, for example a monthly precipitation data at several stations in one region, a daily temperature data at some locations, etc. The space time model such as a Generalized Space Time Autoregressive (GSTAR) model is a stationary model which has a different parameters for each location, and the GSTAR model is applicable for locations which have a heterogeneous characterization. An error assumption of GSTAR model is independent and identically normal distribution, which homogeneous variance. So, we can use an Ordinary Least Squares (OLS) method to estimate the parameters of GSTAR. In real phenomena such as a precipitation data, an error ussualy have a non constant variance, and we need another estimation method to estimate the parameters of GSTAR model. In this paper we propose a Seemingly Unrelated Regression (SUR) method by accommodating various matrixes among locations, which then makes an assumption $\varepsilon_{ij} \sim NID(\mathbf{0},\sum)$ to estimate parameters of GSTAR (1,*p*) model. \We apply the GSTAR(1,*p*)-SUR model for prediction of precipitation data at Batu Town in Station Malang Area including Tlekung, Tinjumoyo, Temas, Ngujung and Ngaglik. The precipitation data at five stations have a correlation between one station to another, so that we use SUR method to overcome a wide range of non-constant variance. The result show that the GSTAR model for precipitation data at Batu Town has a GSTAR ((1),(1,2,12,24))-SUR model and give a determination coefficient 53.84%.

Keywords: GSTAR, OLS, SUR, precipitation

INTRODUCTION

Space-time model is a multivariate time series model which is combine of time and spatial observations simultaneously. The space-time model was first introduced by Pfeifer and Deutsch [1][2], known as *Space-Time Autoregressive* (STAR) model. The STAR model developed by Pfeifer and Deutsch [1][2] had a weaknesses on its parameter flexibility explaining the dependence of time and spatial which are different on a time and location series data. This weaken was already improved by Ruchjana [3] through a model known as a *Generalized Space-Time Autoregressive* (GSTAR) model. Borovkova, *et al*. [4] argued that there was still a chance to make a further analysis related to asymptotic characteristics of an estimated model of generalized space-time and the comparison of accuracy of prediction result compared with models of multivariate time series which was developed previously. Up to now, some research related to GSTAR model has mostly been limited to space time data which is stationary and non-seasonal. This condition is unsatisfied in climate data, particularly in precipitation level data which occasionally contains seasonal and non-stationary data. In addition, most research concerning GSTAR model is limited to an estimation of using a method of *ordinary least squares* (OLS) to obtain an estimator of GSTAR model parameters. [5]. An estimation of OLS model in a model involving a response which contains more than one variable and is correlated each other which will result in an efficient estimator which does not, therefore, qualify a good estimator (BLUE). Nainggolan [6] developed a model of GSTAR-ARCH to overcome heterogeneous or non-constant variance by using a likelihood maximum estimation model despite obscure variance among locations.

One of the estimation methods suited to a model involving a multivariate response and is correlated each other is a *Seemingly Unrelated Regression* (SUR) method [7][8][9][10][11]. By using this SUR method,

***Corresponding Author:** Atiek Iriany, Department of Statistics, Faculty of Mathematics and Natural Sciences, University Brawijaya, Malang, East Java of Indonesia,. Email: atiekiriany@yahoo.com

the relation among response variables can be accomodated by shaping variance-covariance error matrix. The fact indicated that climate data particularly precipitation data which is likely to contain non stationary and seasonal pattern as well as OLS estimation weakness in GSTAR model. So in this research, we performs a further analysis related to developing GSTAR model for non stationary data and seasonal pattern using lag-time and uses SUR estimation which is then termed *Generalized Space-Time Auto regressive–Seemingly Unrelated Regression* (GSTAR-SUR). Estimate parameters for SUR model using *Generalized Least Square* (GLS) method which is developed of OLS method.

Estimate parameters OLS model is $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ using assumption $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2)$. Estimate $\boldsymbol{\beta}$ using SUR approach by accommodation the correlation among location and heterogeneous variance sites, so assumption become $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sum)$. For example using 5 locations we have variance of error as following:

$$\mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{42} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{44} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{55} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{42} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{44} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{55} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\Omega} = (\sum \otimes \mathbf{I})$$

So $\hat{\boldsymbol{\beta}}$ estimate using SUR can be obtained as following:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'\Omega^{-1}X})^{-1}\mathbf{X'\Omega^{-1}y} \qquad (1)$$

The development of GSTAR-SUR model is done to accommodate non-stationary and seasonal data accurately by considering an existing relationship between time and location as well as the location and estimated parameter model, which is systemized which eventually results in a more accurate prediction result.

Up to now there is not yet available software to estimate parameters of GSTAR-SUR model, so in this research the software is also developed in this research for parameter estimation and the prediction of GSTAR-SUR model using R software.

MATERIALS AND METHODS

Data in this research we used a monthly precipitation data in 5 different locations in Batu town, consisting of Tlekung, Tinjumoyo, Temas, Ngujung and Ngaglik for a-16-year period (1996-2011).

Figures 1 Map of Precipitation Stations at Batu Town
Source: Batu Government, East Java (2012)

The procedure of research method using GSTAR(1,*p*)-SUR is started by the initial stage, data exploration used a statistical analysis, which was then followed by an initial identification by making ACF and PACF plots to determine GSTAR model order. Once the data was formed, an estimation using precipitation data was conducted in 5 locations of Batu town.

RESULTS AND DISCUSSION

GSTAR(1,1)–OLS analysed by Ruchjana [5] still has a weaknesses, because of the assumption that there was no error correlation between locations and constant variance for every location or it can be formulated with: $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$.

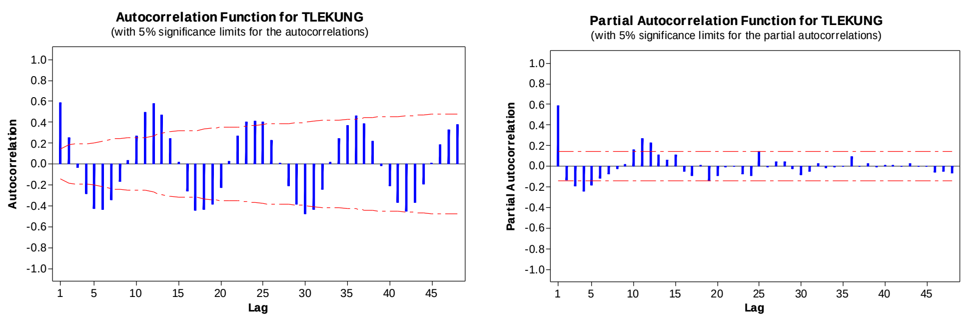
GSTAR (1,1) model can be extended for time lag 1, time lag 2, to time lag *p* or GSTAR (1,*p*) which can be formulated as follows [5][4]:

$$\mathbf{z}(t) = \boldsymbol{\mu} + \boldsymbol{\Phi}_{01}\mathbf{z}(t-1) + \boldsymbol{\Phi}_{11}\mathbf{W}\mathbf{z}(t-1) + \boldsymbol{\Phi}_{02}\mathbf{z}(t-2) + \boldsymbol{\Phi}_{12}\mathbf{W}\mathbf{z}(t-2)$$
$$+\boldsymbol{\Phi}_{03}\mathbf{z}(t-12) + \boldsymbol{\Phi}_{13}\mathbf{W}\mathbf{z}(t-12) + \boldsymbol{\Phi}_{04}\mathbf{z}(t-24) + \boldsymbol{\Phi}_{14}\mathbf{W}\mathbf{z}(t-24) + \boldsymbol{\varepsilon}(t) \qquad (2)$$

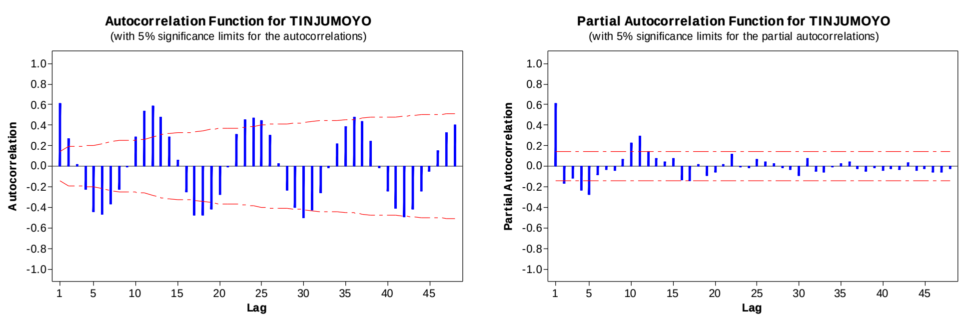
This research is suited to the phenomenon of precipitation level in Batu town which is assumed that a monthly precipitation level has a seasonal pattern that can be seen from ACF and PACF plots so that GSTAR model can be expanded by including a seasonal element through time lag.

On the other side, OLS estimation method can be developed to a *Seemingly Unrelated Regression* method with an assumption that there is an error correlation between location and non-constant error variance. To analyze GSTAR-SUR model, the number of locations was chosen which is N=5, the amount of time t = 1,2,.........,T and **W** is a uniform weight matrix.

A model of precipitation level in 5 locations of station, began with the identification of univariate time series model using ACF and PACF in each location. We have an ACF and PACF as following figure.

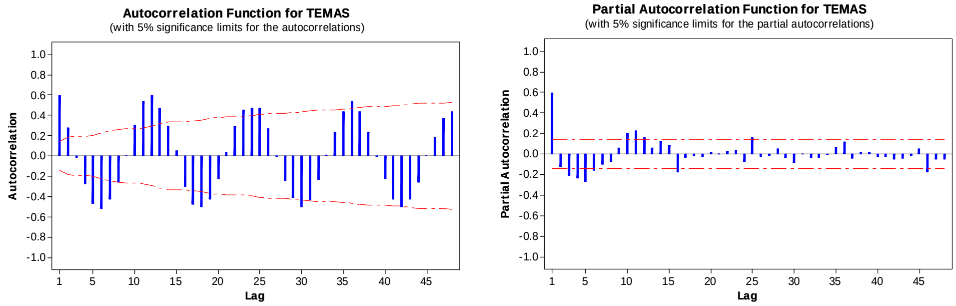


(a)

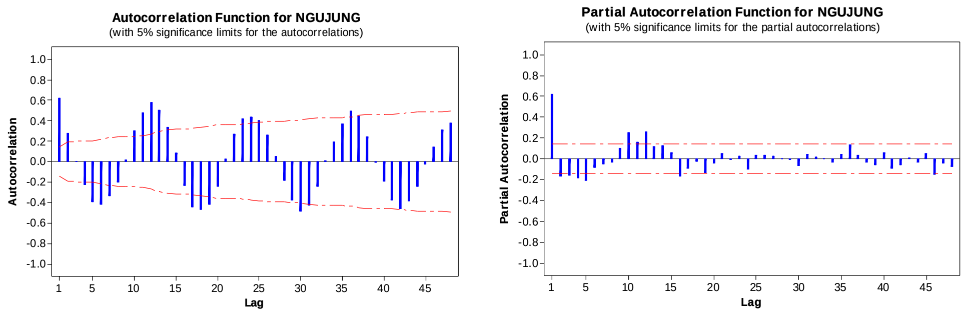


(b)

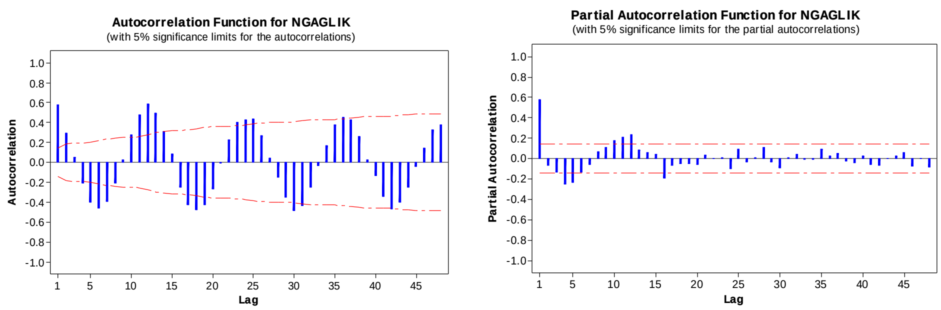
Figures 2 ACF and PACF Plots in 5 Locations of Precipitation Data



(c)



(d)



(e)

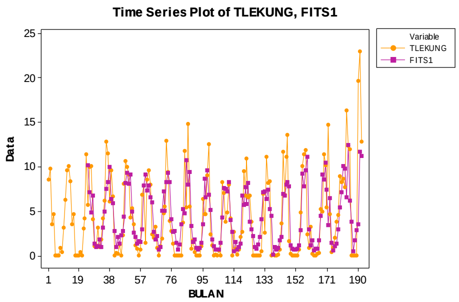
Figures 3 ACF and PACF Plots in 5 Locations of Precipitation Data

ACF and PACF plots illustrated above depict that data contains a seasonal element, that can be seen from ACF pattern which experiences an increase and a decline every month. We identified order of AR model based on ACF and PACF plot. The PACF is cut off at time lag first, second, and then from ACF we have a seasonal at time lag 12th and 24th, so the space time model of the GSTAR(1,*p*)-SUR can be written as GSTAR((1),(1,2,12,24)-SUR.

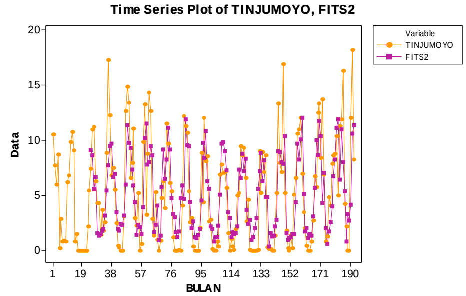
The approach of $\boldsymbol{\beta}$ estimation using SUR method used matrix $\boldsymbol{\Omega}$ obtained from residual covariance result of GSTAR(1,*p*)-OLS model.

$$\boldsymbol{\Omega} = \begin{bmatrix} 10.7902 & 7.9277 & 7.7036 & 7.8304 & 8.4692 \\ 7.9277 & 10.2432 & 7.4426 & 8.4664 & 8.5837 \\ 7.7036 & 7.4426 & 8.9410 & 7.5976 & 8.1768 \\ 7.8304 & 8.4664 & 7.5976 & 9.7263 & 8.6803 \\ 8.4692 & 8.5837 & 8.1768 & 8.6803 & 10.3331 \end{bmatrix}$$

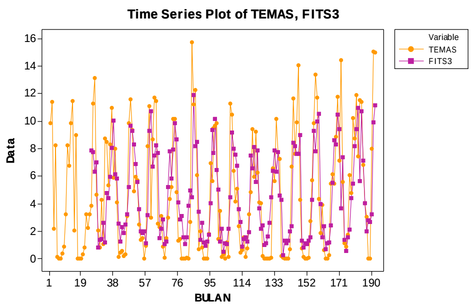
The result of predicted precipitation in Tlekung, Tinjumoyo, Temas, Ngujung and Ngaglik is depicted in time series plot below:



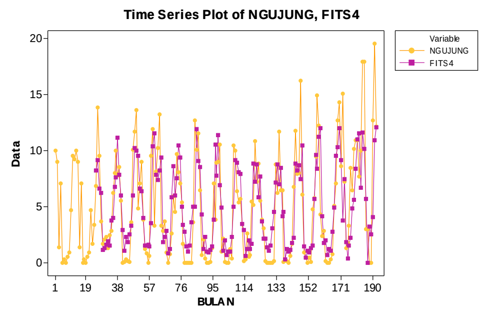
(a)



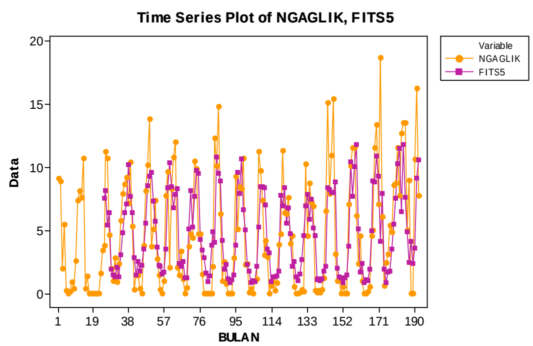
(b)



(c)



(d)



(e)

Figure 5 The GSTAR ((1),(1,2,12,24))-SUR prediction results at 5 locations

Table 1 RMSE and R^2 *prediction* of GSTAR ((1),(1,2,12,24))-SUR Model

Location	RMSE	R^2 *Prediction*
Tlekung	3.35402	0.488798
Tinjumoyo	3.21698	0.525305
Temas	3.01042	0.523936
Ngujung	3.18033	0.551680
Ngaglik	3.25027	0.484559

R^2 *prediction* for 5 locations is 53, 84%. The greater value of R^2 *prediction* the obtained models can explain the distribution of precipitation. Greatest R^2 *prediction* at location Ngujung is 55,19% . It means that the precipitation leevel at Ngujung can be explained by influenced of precipitiuation at Ngujung itself and four other stations which have the same pattern of ACF and PACF using GSTAR((1),(1,2,12,24))-SUR model.

CONCLUSION

Based on the result and discussion, it can be drawn that a monthly precipitation in Batu can be modelled with a model of GSTAR ((1), (1, 2, 12, 24)), which has R^2 Prediction is 53.84%, which means GSTAR(1,*p*)-SUR is an alternative model to predict a precipitation levcel with high value enough of R^2 prediction.

REFERENCES

1. Pfeifer, P.E and Deutsch, S.J. 1980. Stationarity and Invertibility Regions for Low Order STARMA Models. *Communications in Statistics-Simulation and Computation* 9 (5). P. 551-562.

2. Pfeifer, P.E and Deutsch, S.J. 1980. Identification and Interpretation of First Order Space-Time ARMA *Models techno metrics*, 22 (1), 397-408.

3. Ruchjana, B.N. 2001. Study on the Weight Matrix in the Space-Time Autoregressive Model. *Proceeding of the Tenth International Symposium on Applied Stochastic Models and Data Analysis (ASMDA).* Gerard Govaert, et al. Editor. Universite de Technologie Compiegne. Prancis. Vol. 2/2. p. 789-794.

4. Borovkova, S.A., Lopuha, H.P. and Ruchjana, B.N. 2002. Generalized S-TAR with Random Weights. *Proceeding of the 17th International Workshop on Statistical Modeling.* Chania-Greece.

5. Ruchjana, B.N. 202. A Generalized Space Time Autoregressive Model and its Application to Oil Production Data. *Dissertation* Unpublished. Bandung : ITB

6. Nainggolan, Nelson. 2012. Development of GSTAR Model with Assumption ARCH and its Application to Inflation. *Dissertation* Unpublished. Bandung: ITB.

7. Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association* 57: 348–368.

8. Srivastava, V.K. and Giles, D.E.A. (1987). *Seemingly unrelated regression equations models: estimation and inference*. New York: Marcel Dekker. ISBN 978-0-82-477610-7

9. Setiawan. 1992. Study on Seemingly Unrelated Regression (SUR) and its application to Almost Ideal Demand System (AIDS). *Thesis* Unpublished. Bogor : IPB.

10. Davidson, Russell and MacKinnon, James G. 1993. *Estimation and Inference in Economics*. Oxford University Press. UK.

11. Greene, W.H. 2002. *Econometric Analysis* (5th ed). Prentice Hall