CAUCHY –Jurnal Matematika Murni dan Aplikasi
Volume 5(4) (2019), Pages 203-210
p-ISSN: 2086-0382; e-ISSN: 2477-3344

GSTAR-X-SUR Model with Neural Network Approach on Residuals

Diana Rosyida[1], Atiek Iriany[2], Nurjannah[3]

[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Science, University of Brawijaya, Malang, Indonesia

Email: dianarosyida94@gmail.com, atiekiriany@yahoo.com, nj_anna@ub.ac.id

ABSTRACT

One of the models that combine time and inter-location elements is Generalized Space Time Autoregressive (GSTAR) model. GSTAR model involving exogenous variables is GSTAR-X model. The exogenous variables which are used in GSTAR-X model can be both metrical and non-metrical data. The case study of this research is the forecasting of precipitation, which is the exogenous variable is non-metrical data of precipitation intensity of a certain location. The approach of parameter estimation method employed by Seemingly Unrelated Regression (SUR) model, which can solve the correlation between residual models. Now, the phenomenon of precipitation possesses patterns and characteristics is difficult to identify, and can be interpreted as a non-linear model. The non-linear model is much developed by Neural Network (NN). This research employed GSTARX-SUR modelling with neural network approach on residuals. The data used in this research were the records of 10-day precipitations in four regions in West Java, namely Cisondari, Lembang, Cianjur, and Gunung Mas, from 2005 to 2015. The GSTARX-SUR NN modelling resulted in precipitation deviation average of the forecast and the actual data at 4.1385 mm. This means that this model can be used as an alternative in forecasting precipitation.

Keywords: GSTAR-X, SUR, neural network, precipitation

INTRODUCTION

Multivariate time series analysis has been implemented in forecasting since there are many problems that cannot be solved with univariate forecasting [1]. One of them is Space Time Autoregressive (STAR) model that was first introduced by Pfeifer and Deutsch [2]. That model considers the element of time and location. The STAR model was developed by Ruchjana [3] to be a more general model for all locations, which is known as the Generalized Space Time Autoregressive (GSTAR).

In time series model, there is a model that contains exogenous variables, namely ARIMAX and VARIMAX. Similarly, space time model has GSTARX model. Exogenous variable which is used in GSTAR model can be in a form of metrical data and non-metrical data. Some researches on GSTAR using exogenous variables with metrical data are Kurnia, et al [4] and Astuti, et al [5], while that of exogenous variable with non-metrical data are Ditago, et al [6] and Suhartono, et al [7].

One of the methods in estimating the parameter in multivariate analysis is Seemingly Unrelated Regression (SUR) [8] [9]. Parameter estimation of GSTAR model with Ordinary Least Square method conducted by Ruchjana [3] found residuals which are mostly correlated and thus the estimation value acquired is not efficient [10].

One of the non-linear phenomena is precipitation. Currently, precipitation possesses patterns which are difficult to be identified and predicted. Neural network

Submitted: November 10, 2018 Reviewed: Pebruary 11, 2019 Accepted: May 31, 2019

model is a non-linear model which has been commonly developed. Some researches which implement neural network in space time model are Suhartono [11] and Sulistyono, et al [12]. These researches employed GSTARX-SUR-NN modelling aiming at acquiring accurate forecast.

METHODS

The data used in this research include 10-day precipitation data in four locations in West Java, namely Cisondari, Lembang, Cianjur, and Gunung Mas regions. The period of precipitation data used is from 2005 to 2015, resulting in 360 observations from each observed region.

The exogenous variable used is dummy variable. There are four dummy variables employed, namely precipitation of more than 20 mm in Cisondari, precipitation of more than 20 mm in Lembang, precipitation of more than 25 mm in Cianjur, and precipitation of more than 30 mm in Gunung Mas. The code of dummy variables is 1 if the data fill the criteria of four exogenous variable and 0 when the data do not. This ways is analog with the reasearch of Ditago [6] uses Ramadhan month criteria, in which the data in Ramadhan month are given 1 and 0 for the data in other months.

Location value used in the GSRARX modelling is the normalization value of cross correlation initially introduced by Suhartono and Atok [13]. The normalization value of cross correlation taking place in the corresponding lag [14] $W_{ij} = \frac{r_{ij}(s)}{\sum_{k \neq s} |r_{ik}(s)|}$ where $i \neq j$ and this value meets $\sum_{i \neq j} W_{ij} = 1$.

SUR model possesses the assumption $E[e|\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_m}] = 0$ and $E[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}|\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_m}] = \boldsymbol{\Omega}$ where $\boldsymbol{\Omega}$ is the variance covariance matrix [15].

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11}\boldsymbol{I_{mT}} & \sigma_{12}\boldsymbol{I_{mT}} & \cdots & \sigma_{1m}\boldsymbol{I_{mT}} \\ \sigma_{21}\boldsymbol{I_{mT}} & \sigma_{22}\boldsymbol{I_{mT}} & \cdots & \sigma_{2m}\boldsymbol{I_{mT}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}\boldsymbol{I_{mT}} & \sigma_{m2}\boldsymbol{I_{mT}} & \cdots & \sigma_{mm}\boldsymbol{I_{mT}} \end{bmatrix} \tag{1}$$

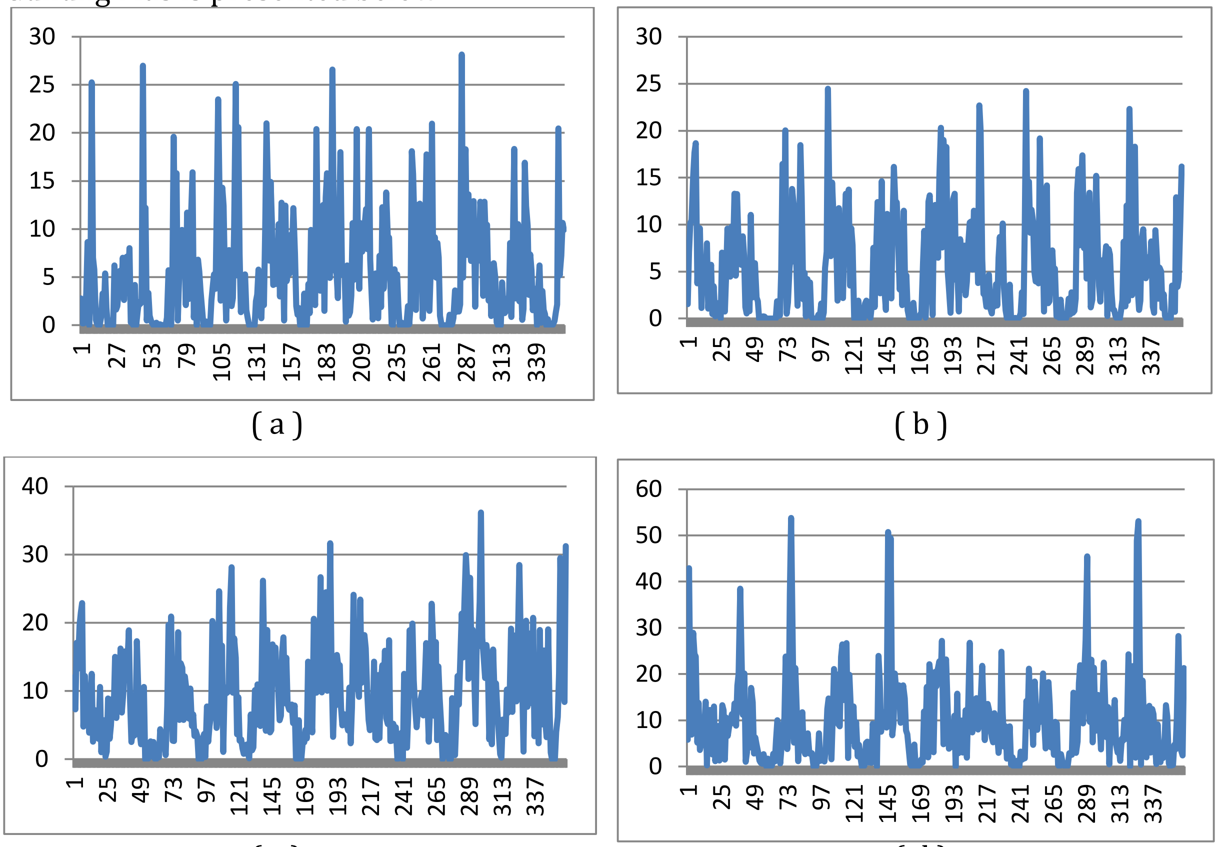
where $\boldsymbol{I_{mT}}$ as the identity matrix with $(mT \ x \ mT)$ size and σ_{ij} is the error variance from each equation for $i = j$ and covariance error between equations $i \neq j$. Therefore, estimation of parameter using SUR is shown in the equation below (2)

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \tag{2}$$

Essential components in the neural network model are the layer input, hidden layer, and output layer. Input layer used in this case is the residual of GSTAR-SUR model. This current research merely employs one hidden layer; however, the number of neurons used in the hidden layer is based on the lowest RMSE value. The output layer employed involves four variables. Resilient Algorithm backpropagation is used in estimating the neural network model. This algorithm has been employed in researches by Apriliyah, et al [16] and Fadil, et al [17] in estimating the sales of electricity load.

RESULTS AND DISCUSSION

The illustration of precipitation data plot in Cisondari, Lembang, Cianjur, and Gunung Mas is presented below.



(a) (b)

(c) (d)

Figure 1. Precipitation Data Plot

Figure 1.a is precipitation data plot in Cisondari, Figure 1.b for Lembang, Figure 1.c for Cianjur, and Figure 1.d for Gunung Mas. In Figure 1, it can be seen that the precipitation data patterns in these four different locations are randomly horizontal. Gunung Mas region has shown higher level of precipitation compared to those in the other three regions, with the precipitation of 50 mm being the highest. The highest precipitation in Cisondari is 28.15 mm; while precipitations in Lembang and Cianjur respectively are at 24.45 mm and 36.18 mm. The increase of precipitation in these four regions has happened in different time period and therefore it is hard to predict.

Precipitation data plot illustrated in Figure 1 shows that the data possess big variance. Therefore, Box-Cox transformation must be conducted in the precipitation data in the four regions. If the lambda in Box-Cox transformation equals to 1, the data can be classified as stationary against the variance.

Table 1. Box-Cox Transformation

Location	Transformation
Cisondari	$(Z_{1t}+1)^{-0.16}$
Lembang	$[\ln(Z_{4t}+1)+1]^{0.25}$
Cianjur	$(Z_{3t}+1)^{0.21}$
Gunung Mas	$(Z_{4t}+1)^{0.12}$

Precipitation data from the transformation based on Table 1 have shown to be variance stationary. It is then tested to find if the data are stationary against the average value. Dickey Fuller test result has shown that the stationary variance data are stationary against the average value. This means that differencing is not necessary.

GSTAR model order is determined by the MPACF and MACF schemes; while spatial order used is order 1.

Schematic Representation of Partial Cross Correlations															
Variable/Lag	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
cisondari	+.-.	+...	....	..+.	....	....	.+..	....	....	....	.+-.	....	..+.	....	....
lembang	.+++	....	....	....	...+	....	....	+...	..-.	....	....	....	....	....	....
cianjur	-+++	.+..	....	....	....	....	+...	....	....	....	....	....	....	.-.-	....
gunung_mas	.+.+	....	...+	....	....	....	....	....	.-..	....	.-..	...-	...-	....	..-.
+ is > 2*std error, - is < -2*std error, . is between															

(a)

Schematic Representation of Cross Correlations																
Variable/Lag	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
cisondari	+---	+---	+---	+---	+---	+---	+---	....	....	....	....	.+.+	-+++	-+++	-+++	-+++
lembang	-+++	-+++	-+++	-+++	-+++	-+++	-+++	....	....	....	+...	+---	+---	+---	+---	+---
cianjur	-+++	-+++	-+++	-+++	-+++	-+++	-+++	.+.+	....	....	+...	...-	+---	+---	+---	+---
gunung_mas	-+++	-+++	-+++	-+++	-+++	-+++	-+++	-+++	....	....	....	.---	+---	+---	+---	+---
+ is > 2*std error, - is < -2*std error, . is between																

(b)

Figure 2. MPACF and MACF schemes

Figure 2.a is of a MPACF scheme and Figure 2.b is of a MACF scheme. Figure 2 shows that MPACF scheme cuts off at lag 1 and MACF scheme possesses sinus wave pattern, so the model developed is autoregressive with lag 1 [18]. Therefore, in this research, the order of GSTARX-SUR model used is GSTARX-SUR (1_1).

Neural network modelling is to be performed on the residuals from GSTARX model (1_1). There are four inputs of input layers used, namely $a_{1,t-1}, a_{2,t-1}, a_{3,t-1}, and\ a_{4,t-1}$. The inputs are layers used by Suhartono and Endharta [19] as one of the inputs in neural network modelling. There are four neurons in the output layer of some variables used in GSTARX-SUR (1_1). Hidden layer is limited to 1 layer, where neuron used is limited to 1 to 10 neurons. The selection of the number of neurons used in hidden layer is drawn from the lowest RMSE value.

Table 2. The RMSE Value in Numbers of Neurons in Hidden Layer

The number of Neurons in Hidden Layer	RMSE Value	The number of neurons in Hidden Layer	RMSE Value
1	0.1435	6	0.1346
2	0.1430	7	0.1352
3	0.1421	8	0.1403
4	0.1411	9	0.1327
5	0.1405	10	0.1303

Based on Table 2, 10 neurons possess the lowest RMSE values. Therefore, the neural network model developed is NN (4, 10, 4). The best architectural design for neural network based on the residual GSTARX-SUR model (1_1) is shown in Figure 3.

The mathematical equations developed from the GSTARX-SUR (1_1) – NN (4,10,4) from each location are as follows.

1. Cisondari

$$\begin{aligned}\hat{Z}_{1t} &= Z_{1t,GSTARX} + Z_{1t,NN}\\ \hat{Z}_{1t} &= 0.7717\, Z_{1,t-1} + 0.0516\, Z_{2,t-1} + 0.0468\, Z_{3,t-1} + 0.0404\, Z_{4,t-1} - 0.1656\, D_1\\ &\quad - 0.0786\, D_2\\ Z_{1t,NN} &= -0.615 - 0.0473\, f(h_1) + 0.6355\, f(h_2) - 0.2415\, f(h_3) + 0.8345\, f(h_4)\\ &\quad + 0.2273\, f(h_5) - 0.5837\, f(h_6) - 0.4523\, f(h_7) - 0.2578\, f(h_8)\\ &\quad - 0.6448\, f(h_9) + 1.1062\, f(h_{10})\end{aligned} \tag{3}$$

2. Lembang

$$\begin{aligned}\hat{Z}_{2t} &= Z_{2t,GSTARX} + Z_{2t,NN}\\ \hat{Z}_{2t} &= 0.4027\, Z_{2,t-1} + 0.1928\, Z_{1,t-1} + 0.2077\, Z_{3,t-1} + 0.1976\, Z_{4,t-1} + 0.0994\, D_1\\ &\quad + 0.0796\, D_2\\ Z_{2t,NN} &= 1.1521 + 0.0785\, f(h_1) - 0.7041\, f(h_2) - 0.1386\, f(h_3) - 1.1537\, f(h_4)\\ &\quad - 0.4983\, f(h_5) + 0.9914\, f(h_6) + 0.0946\, f(h_7) + 0.5594\, f(h_8)\\ &\quad - 0.4155\, f(h_9) - 0.5196\, f(h_{10})\end{aligned} \tag{4}$$

3. Cianjur

$$\begin{aligned}\hat{Z}_{3t} &= Z_{3t,GSTARX} + Z_{3t,NN}\\ \hat{Z}_{3t} &= 0.4711\, Z_{3,t-1} + 0.2305\, Z_{1,t-1} + 0.2713\, Z_{2,t-1} + 0.2191\, Z_{4,t-1} + 0.1923\, D_1\\ &\quad + 0.1197\, D_2 + 0.2225\, D_3\\ Z_{3t,NN} &= -0.9406 + 0.2818\, f(h_1) + 0.2951\, f(h_2) - 1.1130\, f(h_3) - 2.5219\, f(h_4)\\ &\quad - 1.2281\, f(h_5) + 1.7891\, f(h_6) + 1.0893\, f(h_7) + 1.424\, f(h_8)\\ &\quad + 0.8216\, f(h_9) + 0.7601\, f(h_{10})\end{aligned} \tag{5}$$

4. Gunung Mas

$$\begin{aligned}\hat{Z}_{4t} &= Z_{4t,GSTARX} + Z_{4t,NN}\\ \hat{Z}_{4t} &= 0.2083 Z_{4,t-1} + 0.2482\, Z_{1,t-1} + 0.3161\, Z_{2,t-1} + 0.268\, Z_{3,t-1} + 0.1239\, D_1\\ &\quad + 0.2187\, D_4\\ Z_{4t,NN} &= 0.1742 + 0.1249\, f(h_1) - 1.1336\, f(h_2) - 0.1868\, f(h_3) - 0.84\, f(h_4)\\ &\quad - 0.777\, f(h_5) + 0.7616\, f(h_6) + 2.6908\, f(h_7) + 0.8405 f(h_8)\\ &\quad - 0.2835\, f(h_9) - 0.4847\, f(h_{10})\end{aligned} \tag{6}$$

$f(h_i)$ is the activation function of logistic sigmoid in the hidden unit which is defined as follow

$$f(h_i) = \frac{1}{1 + e^{-(h_i)}} \;, \quad i = 1,2,\ldots,10 \tag{7}$$

where,

$$\begin{aligned}
h_1 &= -54.5045 + 179.7641\, a_{1,t-1} + 491.5302\, a_{2,t-1} - 35.5407\, a_{3,t-1} - 316.5188\, a_{4,t-1}\\
h_2 &= 0.3103 + 7.4173\, a_{1,t-1} - 1.4297\, a_{2,t-1} + 1.7059\, a_{3,t-1} + 0.4799\, a_{4,t-1}\\
h_3 &= -0.7417 - 21.0108\, a_{1,t-1} - 4.3276\, a_{2,t-1} - 2.0921\, a_{3,t-1} + 4.6632\, a_{4,t-1}\\
h_4 &= 0.3549 + 14.6467\, a_{1,t-1} - 3.4917\, a_{2,t-1} - 0.5961\, a_{3,t-1} + 0.5499\, a_{4,t-1}\\
h_5 &= 7.1522 - 29.3058\, a_{1,t-1} + 128.514\, a_{2,t-1} - 183.3922\, a_{3,t-1} + 3.3083\, a_{4,t-1}\\
h_6 &= 0.1041 + 11.7434\, a_{1,t-1} - 8.4182\, a_{2,t-1} - 0.516\, a_{3,t-1} + 2.8594\, a_{4,t-1}\\
h_7 &= -0.4036 + 0.481\, a_{1,t-1} - 0.5068\, a_{2,t-1} + 0.1383\, a_{3,t-1} + 0.4666\, a_{4,t-1}\\
h_8 &= 3.1722 - 10.3635\, a_{1,t-1} + 64.1415\, a_{2,t-1} - 84.2384\, a_{3,t-1} - 2.4503\, a_{4,t-1}\\
h_9 &= 9.2704 - 5.0167\, a_{1,t-1} + 6.2685\, a_{2,t-1} - 18.4776\, a_{3,t-1} + 15.4932\, a_{4,t-1}
\end{aligned}$$

$$h_{10} = -0.3992 - 13.9493\, a_{1,t-1} - 0.9574 a_{2,t-1} - 2.3135\, a_{3,t-1} + 1.5081\, a_{4,t-1}$$

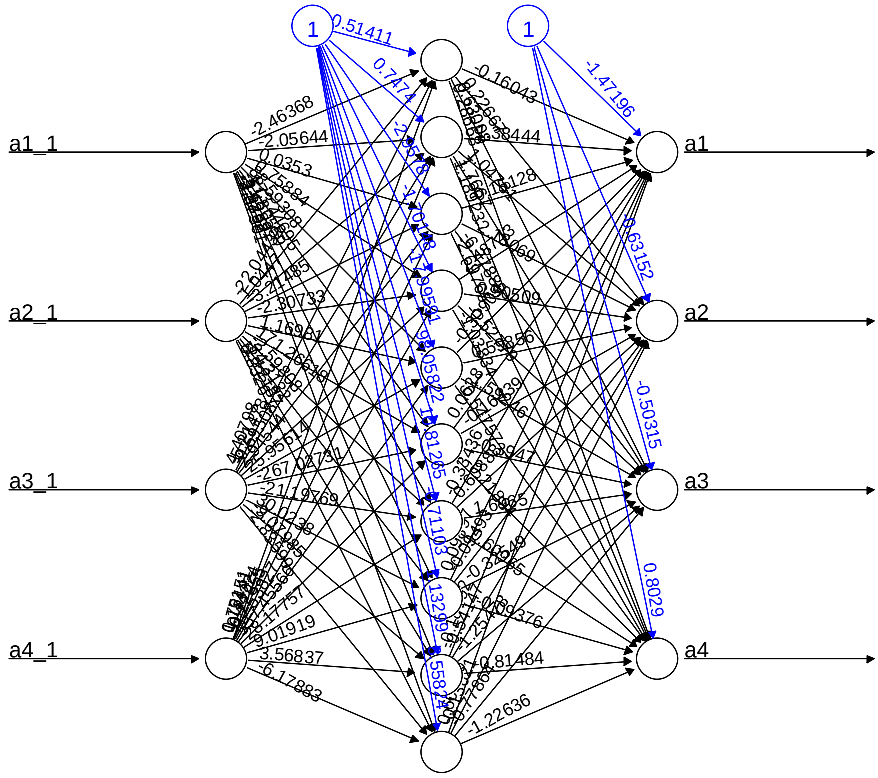
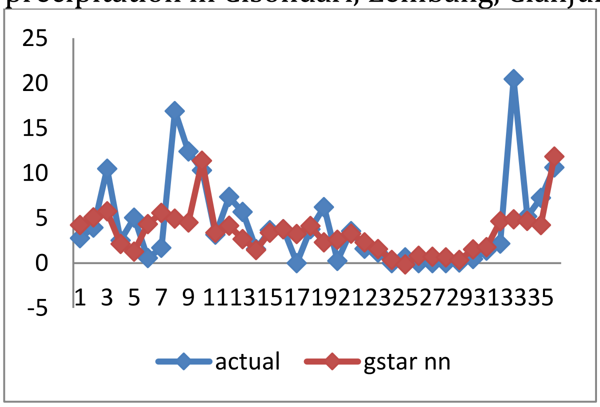
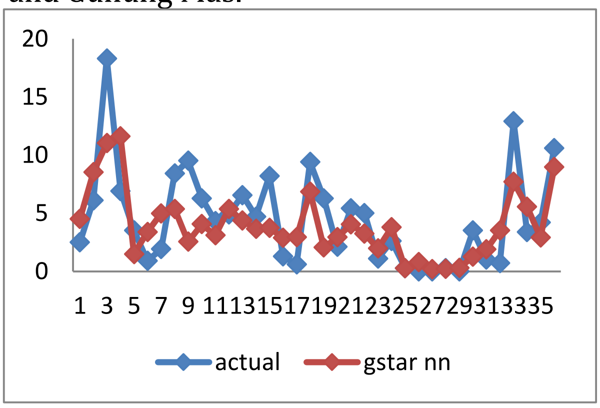


Figure 3. The Best Architectural Design for GSTARX-SUR Residual Model Neural Network (1_1)

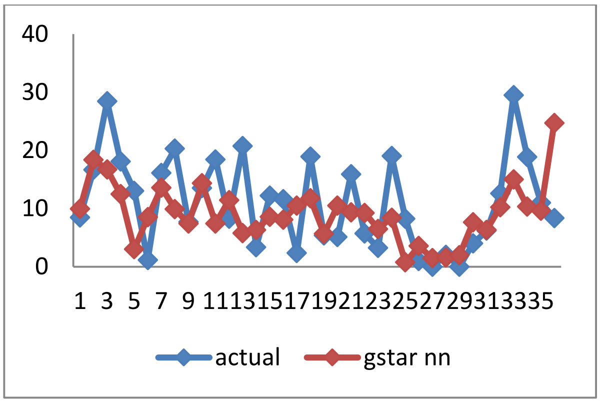
The already developed model is to be forecasted and compared to the actual data in order to find out whether or not the models are able to illustrate the real condition of precipitation in Cisondari, Lembang, Cianjur, and Gunung Mas.

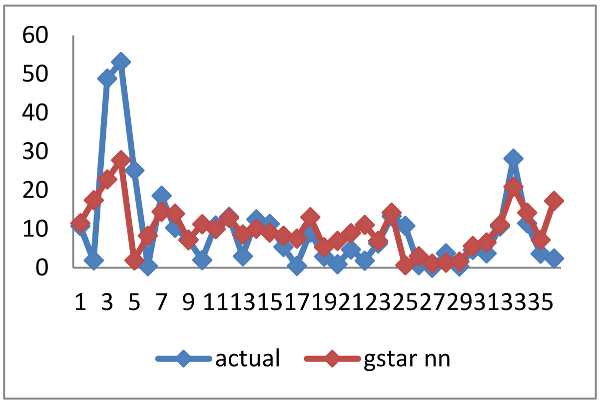


(a)



(b)





(c) (d)

Figure 4. Precipitation Forecast Plot in 2015

Figure 4.a is precipitation forecast result plot of Cisondari, Figure 4.b, 4.c, and 4.d respectively are precipitation forecast result plots of Lembang, Cianjur, and Gunung Mas. The precipitation forecast result using GSTARX-SUR (1_1) – NN (4,10,4) possesses similar pattern to the actual data. However, the performance of GSTARX-SUR (1_1) – NN (4,10,4) model cannot be seen in detailed from the graph. The performance of GSTARX-SUR (1_1) – NN (4,10,4) model can be seen from the RMSE and MAD values.

Table 3. GSTARX-SUR (1_1) – NN (4,10,4) Performance Model

Regions	RMSE Value	MAD Value
Cisondari	5.4012	3.5767
Lembang	4.9883	3.3882
Cianjur	5.7111	4.2282
Gunung Mas	7.8792	5.3608

The result of precipitation forecast in Lembang has shown the lowest RMSE and MAD values. The average deviation of precipitation from the forecast result and the actual data of Lembang is only at 3.3882 mm. On the other hand, Gunung Mas has the highest RMSE value and precipitation average deviation from forecast result and actual data at 5.3882 mm.

CONCLUSION

The Model of precipitation forecasting in Cisondari, Lembang, Cianjur, and Gunung Mas that are formed with GSTARX-SUR (1_1) - NN (4,10,4). Based on the value of forecasting in each location and compared by the actual data, the result of forecasting is good because the average of MAD value is 4.1385 mm, so GSTAX-SUR model with a neural network approach on the side can be used as a good alternative to predict precipitation.

ACKNOWLEDGMENTS

Acknowledgments to BMKG, the head of Cisondari, Chincona, Cianjur and Gunung Mas villages so that this research can be complete and help people in the villages.

REFERENCES

[1] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *Int. J. Forecast.*, vol. 22, pp. 443–473, 2006.

[2] P. E. Pfeifer and S. J. Deutsch, "Stationarity and Invertibility Regions for Low Order STARMA Models," *Commun. Stat. Comput.*, vol. 9 no. 5, pp. 551–562, 1980.

[3] B. N. Ruchjana, "A Generalized Space Time Autoregressive Model and its Application to Oil Production Data," ITB, 2002.

[4] J. Dwi and S. Puteri, "The simulation studies for Generalized Space Time Autoregressive-X (GSTARX) model," pp. 30–37, 2015.

[5] B. N. S. Astuti, Dewi; Ruchjana, "Generalized space time autoregressive with exogenous variable model and its application," 2017.

[6] A. P. Ditago, "Simulation Study of Parameter Estimation Two - Level GSTARX - GLS Model," pp. 167–168, 2015.

[7] S. Suhartono, S. R. Wahyuningrum, S. Setiawan, and M. S. Akbar, *GSTARX-GLS Model for Spatio-Temporal Data Forecasting*, vol. 10. 2016.

[8] A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias," *J. Am. Stat. Assoc.*, vol. 57, pp. 348–368, 1962.

[9] Setiawan, "Study on Seemingly Unrelated Regression (SUR) and its application to Almost Ideal Demand System (AIDS)," IPB, 1992.

[10] A. Iriany and B. N. Ruchjana, "Prediction of Precipitation Data at Batu Town Using the GSTAR (1 , p) -SUR Model," *J. Basic Appl. Sci. Res.*, vol. 3, no. 6, pp. 860–865, 2013.

[11] Suhartono, "Feedforward Neural Networks Untuk Pemodelan Runtun Waktu," Universitas Gadjah Mada, 2007.

[12] A. D. Sulistyono, W. H. Nugroho, R. Fitriani, and A. Iriani, "Hybrid Model GSTAR-SUR-NN For Precipitation Data," *Cauchy*, vol. 4, no. 2, p. 74, 2016.

[13] Suhartono and R. M. Atok, "Pemilihan Bobot Lokasi yang Optimal pada Model GSTAR," in *Prosiding Koerensi Nasional Matematika XIII, Semarang*, 2006.

[14] Suhartono and Subanar, "The Optimal Determination of Space Weight in Gstar Model by Using Cross-Correlation Inference," *Quant Methods*, vol. 2, no. 2, pp. 45–53, 2006.

[15] W. H. Greene, *Econometrics Analysis*, 7th ed. Boston: Prentice Hall, 2012.

[16] Apriliyah, W. F. Mahmudi, and A. W. Widodo, "Perkiraan Penjualan Beban Listrik Menggunakan Jaringan Saraf Tiruan Resilent Backpropagation," *Kursor*, vol. 4, no.2, pp. 41–47, 2008.

[17] J. Fadil, O. Penangsang, and A. Soeprijanto, "Load Forecasting for the Distribution Network of South and Middle Kalimantan Using Artificial Neural Networks Resilient Propagation," in *Proceedings of National Seminar on Apllied Technology, Science, and Arts (1st APTECS)*, 2009.

[18] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate*, 2nd ed. USA: Pearson Education Inc, 2006.

[19] Suhartono and A. J. Endharta, "Double Seasonal Recurrent Neural Networks for Forecasting Short Term Electricity Load Demand in Indonesia," *Recurr. Neural Networks Temporal Data Process.*, 2011.