International Journal of Engineering & Technology, 8 (1.9) (2019) 161-165

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Rainfall Forecasting Using Gstar-Sur-Nn Approach in West Java Province

Atiek Iriany[1*], Wayan Firdaus Mahmudi[2], Waego Hadi Nugroho[3], Agus Dwi Sulistyono[4]

[1]*Faculty of Mathematics and Natural Sciences, Brawijaya University, Indonesia*
[2]*Faculty of Computer Science, Brawijaya University, Indonesia*
[3]*Faculty of Mathematics and Natural Sciences, Brawijaya University, Indonesia*
[4]*Faculty of Mathematics and Natural Sciences, Brawijaya University, Indonesia*
**Corresponding author E-mail: agus.dwi.sulistyono@gmail.com*

Abstract

Potato is one of food commodities which is expected to serve as a diversification option of carbohydrate source. One of the negative factors to influence the productivity of potato in the last few years is uncertain climate condition. This problem can be overcome with the development of season forecasting method that produces reliable season forecasting model and has a precise forecasting accuracy, especially in extreme climate conditions. The second year stage of study has reached 70% in completion which includes the implementation of Year I research result, surveying potential locations for potato plants by digging information from farmers and agriculture experts in West Java, the identification of potato crop, identification of factors that influence potato crop growth, identification of rainfall patterns in West Java and exploration of rainfall data of each research rain post. From data exploration and data identification, GSTAR model((1,2,3,4,13,33)(1)–Sur is obtained.

Keywords: *potato, GSTAR-SUR*

1. Introduction

In line with the government program to increase potato production as an alternative source of carbohydrate to support food security, it is necessary to optimize the production of potato. However, potato productivity in Indonesia is still low at 16.51 t ha-1 (Central Board of Statistics or BPS, 2010). This is because crop production, soil condition, and farmer's socioeconomic condition depend on the climatic conditions, one of which is the level of rainfall. Information on the level of rainfall in an area will be influential in determining the correct timing of potato planting, so it will have a significant effect on the increase of potato production.

One approach that can be performed to solve this problem is through tactical approach. Tactical approach is an anticipatory effort through the development of a more reliable season forecasting method and technique, and through the application

of various models and types of data (Peragi and Perhimpi, 1994).

Currently Indonesian Agency for Meteorology, Climatology and Geophysics (Badan Meteorologi, Klimatologi, dan Geofisika or BMKG) uses three methods to forecast climate, i.e. ARIMA, wavelet transformation, and Adaptive Neuro-Fuzzy Inference Systems or ANFIS (Indragustari, 2005a; 2005b; Nuryadi, 2005).

Some of the problems that arise are the validation results of those three methods show different performance in each location. It means each method has good performance only in a specific location. To overcome this, the method that will be studied and developed is two-level hybrid model of ARIMAX-NN combination, the combination of Fuzzy-NN or ANFIS, and an ensemble hybrid model of both approaches. ARIMA is a univariate forecasting model that does not involve predictor variables. Transfer function or ARIMAX is a modeling of time series data involving predictor variables. NN and ANFIS method are modeling data with or without predictor variables. Ensemble method is a method that combines the results of several models or forecasting methods to produce a forecast.

1.1. Arima Model

Seasonal ARIMA model is part of a group of flexible time series model that can be used to model some seasonal types as in non-seasonal time series. Seasonal ARIMA model is expressed as follows (Wei, 1990; Box et al., 1994):

1.2. Generalized Space Time Autoregressive (GSTAR) Model

Generalized Space Time Autoregressive (GSTAR) is a generalization of Space Time Autoregresssive (STAR) model. GSTAR model is built for heterogeneous locations, such as the position of oil wells in the layers of volcanic earth (Rucjana, 2002). In general GSTAR(1.1) model can be expressed as follows (Ruchjana, 2002):

$$\mathbf{z}(t) = \boldsymbol{\mu} + \left[\boldsymbol{\Phi}_{01}^{(i)} + \boldsymbol{\Phi}_{11}^{(i)}\mathbf{W}\right]\mathbf{z}(t-1) + \boldsymbol{\varepsilon}(t)$$

i-location can be defined as follows:

$$\mathbf{z}_i(t) = \boldsymbol{\mu}_i + \phi_{01}^{(i)}\mathbf{z}_i(t-1) + \phi_{11}^{(i)}\sum_{j=1}^{N}\mathbf{w}_{ij}\mathbf{z}_j(t-1) + \boldsymbol{\varepsilon}_i(t)$$

Model equation for all locations follows the structure of linier model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Therefore, it can be concluded that the estimation of the smallest square $\widehat{\beta}_T$ for $\beta = (\phi_{01}, \phi_{11}, ..., \phi_{0N}, \phi_{1N})'$ complies the equation $\hat{\boldsymbol{\beta}}_T = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

1.3. GSTAR(1,p)–Sur Model

GSTAR(1.1)-OLS studied by Ruchjana (2002) still has a weakness because the model assumes that there is no correlation between error among locations and constant variance for each location, or it can be expressed by $\sim$. Besides, this model does not pay attention to seasonal factors.

The pattern of rainfall data, either daily, 10-day, or monthly rainfall always has an element of seasonality. In such conditions, GSTAR(1.1) is not appropriate. In this study, the model is built to fix the weakness of GSTAR(1.1) by accommodating seasonal patterns contained in the rainfall data by entering time lag in the model. In general, time lag 1, time lag 2, up to time lag p or abbreviated as GSTAR(1,p) at M location can be written as follows (Iriany, 2013)

$$\mathbf{z}(t) = \boldsymbol{\mu} + \boldsymbol{\Phi}_{01}\mathbf{z}(t-1) + \boldsymbol{\Phi}_{11}\mathbf{W}\mathbf{z}(t-1) + ... + \boldsymbol{\Phi}_{0p}\mathbf{z}(t-p) + \boldsymbol{\Phi}_{1p}\mathbf{W}\mathbf{z}(t-p) + \boldsymbol{\varepsilon}(t)$$

1.4. Selection Criteria for The Best Model

Selection criteria for the best model can be seen through Mean Absolute Percentage Error (MAPE) value, Root Mean Squared Error (RMSE) value, and R2 Prediction.

1.4.1. Mean Absolute Percentage Error (MAPE)

MAPE is often used to evaluate cross-sectional estimates. MAPE which is also known as Mean Absolute Percentage Deviation (MAPD) is a measure of the accuracy of a method for establishing the values of time series, especially in trend estimate. Here is the calculation of MAPE (Ruchjana, 2002):

$$\text{MAPE}=\sum_{t=2}^{T}\left[\left|\frac{\hat{\varepsilon}(t)}{z(t)}\right|\right] \text{x} \frac{1}{(T-1)} \quad \text{or} \quad \sum_{i=1}^{N}\sum_{t=2}^{T}\left[\left|\frac{\hat{\varepsilon}_i(t)}{z_i(t)}\right|\right] \text{x} \frac{1}{N(T-1)}$$

1.4.2. Root Mean Squared Error (RMSE)

Estimation using the true value of observation is called Root Mean Squared Error (RMSE). The function of RMSE is to obtain an overview of the standard deviation that occurs when there are differences among the models. RMSE value is obtained from the following formula:

$$\text{RMSE}=\sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{l=1}^{n}(Z_{n+1} - \hat{Z}_n(l))^2}$$

1.4.3. R^2 Prediction

R^2 is a quantity that is used to determine model feasibility. The larger the $R^{2,}$ the better the model obtained because it can explain the distribution of data.

$$R^2 = 1 - \frac{SSE}{SST}$$

The purpose of this research is to make new breakthroughs in the field of science, especially in the field of climatology statistics and computation, namely models that are suitable for climate forecasting from various central areas of potato production and software such as climate forecasting system in the central areas of potato production.

2. Research Design and Methods

2.1. Data Source

This research is carried out in stages (within three years) to obtain appropriate android-based forecasting software for the climate in a region of Indonesia. At this stage, the research will be carried out under a group of computational statistics field, focusing on creating a spatial climate forecasting system and developing android-based climate forecasting software for models developed in open source software to be easily accessible on smart phones. The research sites are:

a. Potato-producing areas in East Java, Sumatra, Central Java and Sulawesi
b. Climatology Laboratory of Faculty of Agriculture, University of Brawijaya
c. Statistics Laboratory, University of Brawijaya
d. Information system Laboratory of Information and Technology and Computer Science Program or PTIIK, University of Brawijaya

2.2. Research Stages

Research stages are completed within 3 years as follows:

a. First year
The result of the study in the first year will be further refined in the second year, which includes forecasting models being studied, improving software prototype to make it more user-friendly, and other outcomes.
b. Second year, Improving Climate Forecasting System
There are two main activities conducted in the second year of research, namely:
i. Implementation of system created in the first year
ii. The collection of facts and data from various central areas of potato production in East Java, Central Java, Yogyakarta and West Java. After all facts and data are collected, data exploration and statistical modeling are carried out.
iii. Completion of Forecasting System based on the results of activities i) and ii
c. Third year
In the third year there are two activities, namely
i. Evaluating forecasting system performance generated in the second year.
ii. Making package or library R and submit it to the CRAN r-project.

3. Data and Findings

Implementation of research is conducted in Year I with several activities, namely installing mini weather station for climate forecasting, manufacturing desktop-based rainfall forecasting software, and making a website for the publication of research result.

In year II of the research, the data used in the study are monthly rainfall data at six rain post locations in West Java region, during the period of 2005-2015 (11 years). The average highest rainfall in Cisondari rain post occurred in November and December, while in Chincona, Cibuni and Lembang rain post it occured in December. In Cianjur rain post the average highest rainfall was in November, and in Gunung Mas rain post it took place in January. The average highest rainfall for the six rainfall posts located in Gunung Mas in January amounted to 20.28 mm and the lowest was in Lembang in August (only 0.89 mm). Based on the description of standard deviation value it can be seen that the degree of variability in the data of average monthly rainfall in six rain posts tends to be high, with the highest degree of variability generated by the average rainfall data in Gunung Mas as much as 9.29. After that, a time series plot is made and the result can be seen in Figure 1.

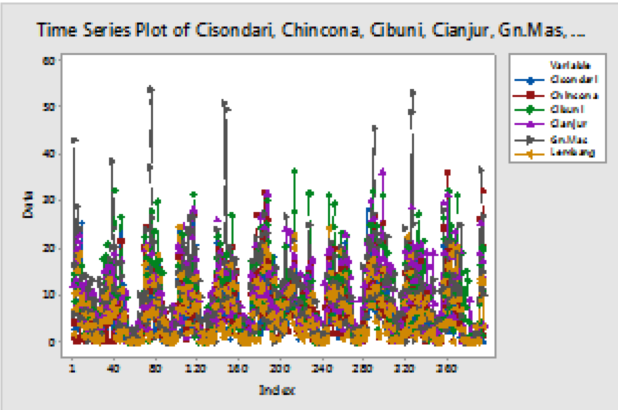

Fig 1: Time series Plot of rainfall data

From figure 1. it is known that the movement patterns of rainfall data in six rain posts tend to be similar. The correlation value is the basis to determine the relevance of rainfall data in each location at the same time. The correlation value on rainfall data in all locations for corresponding time generates quite high correlation value, which means the relevance is relatively high. Therefore, multivariate modeling can be applied to this data. Meanwhile, to understand the pattern of rainfall data in each of these locations data pattern is tested using Terasvirta test. The initial hypothesis of this test is linear data pattern, and the alternative hypothesis is non-linear data pattern. The next step is to determine the stationarity of the data.

The examination result of the stationarity against the mean, namely p-value of Augmented Dickey Fuller test of rainfall data from six rain posts, is less than the value of $\alpha = 0.05$. This means that rainfall data at the six rain posts is stationary against the mean.

Model identification is used to find the order of autoregresive GSTAR model. Autoregresive Order is obtained from the identification of a real MPACF lag, and then the best one is selected from several real lags by means of AIC. Lag with the smallest AIC value will be used as the order of autoregresive GSTAR model. Based on the result of MPACF scheme it is shown that there is a real MPACF lag at lag 1 to lag 6. From several real lags, one is then selected by looking at the smallest AIC value, where the result of AIC for a real lag is shown in Table 1.

Table 1: AIC Value for The Selection of Model Order

Model	AIC
GSTAR (1)	19.18639
GSTAR (2)	19.08502
GSTAR (3)	19.08238
GSTAR (4)	19.08008
GSTAR (5)	19.11327
GSTAR (6)	19.15656
GSTAR (7)	19.16115

Table 1 shows the value of AIC on GSTAR (4) has the smallest value so that the chosen model is GSTAR order 4. In addition to viewing the MPACF plot, the determination of GSTAR order is also viewed from univariate time series plot, i.e. ACF and PACF plot to detect seasonal pattern. Rainfall modeling at each rain post location is begun with the identification of ACF and PACF in each location.

ACF and PACF patterns show that the patterns increase and decrease every month, which means the data contain seasonal elements. ACF and PACF patterns indicate that rainfall data of West Java region are seasonal at time lag 13 and 33 because in lag 13 and 33 the pattern of ACF and PACF is out of limit. This condition is consistent with the fact that the pattern of rainfall, for example dry season in this year will be recurrent in the next year, so it can be said that the pattern of rainfall is an annual pattern. According to Iriany et al. (2013) GSTAR model with seasonal data can be developed by incorporating seasonal elements which are out of limit so that GSTAR model formed for rainfall in West Java region is GSTAR (1,2,3,4,13,33)(1)-Sur.

Parameter estimator significance testing is conducted by two tests, simultaneously by using F test and partially by using t test. The estimation parameter of GSTAR model((1,2,3,4,13,33)(1)-Sur in this study results in 96 parameters. Based on the test result, it is found that there are several insignificant parameters at $\alpha = 5\%$ (p-value > 0.05). The insignificant parameters should be withdrawn and then reestimation is done. Konstenko (2008) states that the insignificant variables can still be used for forecasting. By considering the influence of weight of each location, the insignificant parameters are still incorporated into the model.

Based on the model, the result of prediction for each location (rain post) can be obtained in Figure 2.

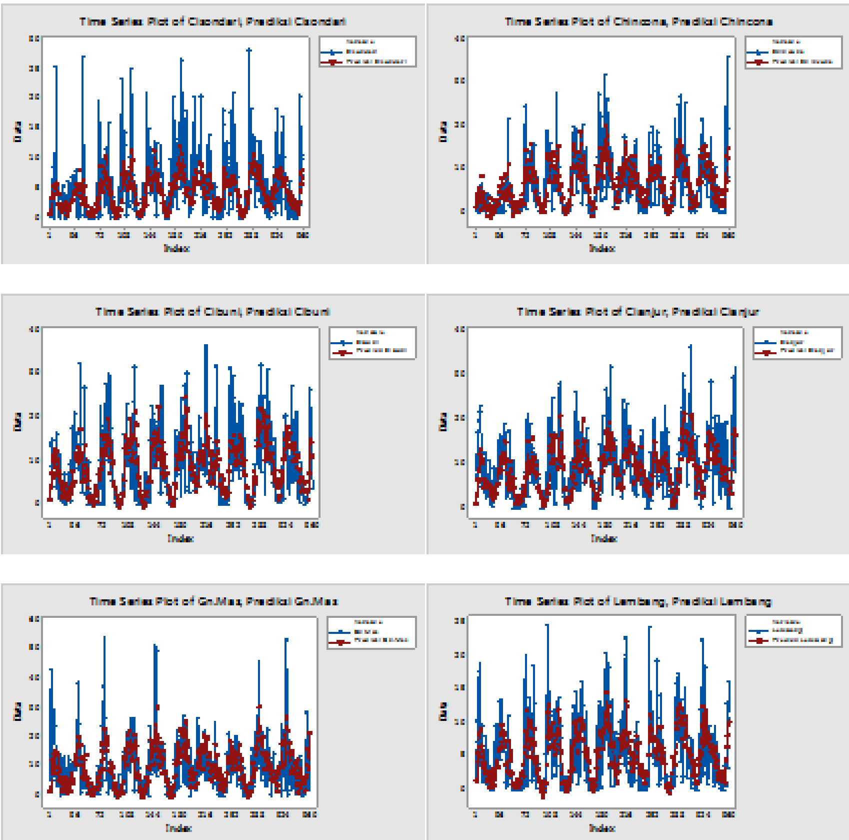

Fig 2: The prediction result of GSTAR((1,2,3,4,13,33)(1)-Sur at each location

Judging from the actual plot and the forecast in Figure 2, it is shown that the prediction result indicates a pattern that approaches the actual data. It can be seen that the plot of actual data and data of prediction result have the same pattern.

Validation of GSTAR((1,2,3,4,13,33)(1)-Sur shows that RMSE has a relatively small value. This shows that GSTAR((1,2,3,4,13,33)(1)-Sur model formed can be used to predict rainfall in the future period. R2 prediction for all locations is still not big enough. The greater the value of R2 prediction, the bigger the model that can explain rainfall distribution. The biggest R2 prediction is obtained for Cisondari location, that is 0.759. It means approximately 75.9% rainfall distribution in Cisondari can be explained by GSTAR((1,2,3,4,13,33)(1)-Sur.

In Year II the first stage has been running for 7 months. The result of the first stage has reached 70% which is still at the stage of implementation of Year I research result (2015) and the creation of rainfall forecasting model for West Java region. For the next stage, it is necessary to create rainfall forecasting model for Central Java region, enhance the software for rainfall forecasting system, and improve mini station that has been installed in Tengger region (first year implementation).

4. Discussion and Conclusion

70% of the research has been reached in Year II stage 1, and GSTAR((1,2,3,4,13,33)(1)-Sur modeling result for rainfall in West Java region is obtained. In addition, the implementation of Year I research result, i.e. the installation of mini stations in Tengger region has been implemented. The next stage of research is the improvement of mini station and model refinement in West Java and Central Java province.

References

[1] 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.
[2] Davidson, Russell and MacKinnon, James G. 1993. Estimation and Inference in Economics. Oxford University Press. UK.
[3] Greene, W.H. 2002. Econometric Analysis (5thed). Prentice Hall.
[4] Iriany, A, (2013), Climate Change Impact to Potato Farming in The Java of Indonesia in The Mountain Range of Batu East Java Indonesia, Journal of Applied Environmental and Biological Science, 3(6): 48-55.
[5] Iriany, A., Suhariningsih, Ruchjana, B.N., and Setiawan. 2013. Prediction of Precipitation Data at Batu Town using The GSTAR(1,p)-SUR. Journal of Basic and Applied Scientific Research.3(6): vol 3 no 6.
[6] Nainggolan, Nelson. 2012. Development of GSTAR Model with Assumption ARCH and its Application to Inflation. Dissertation Unpublihed. Bandung: ITB.
[7] Pfeifer, P.E and Deutsch, S.J. 1980a. Stationarity and Invertibility Regions for Low Order STARMA Models. Communications in Statistics-Simulation and Computation 9 (5). P. 551-562.
[8] 1980b. Identification and Interpretation of First Orde Space-Time ARMA Models.Technometrics, 22 (1), 397-408.
[9] 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.
[10] .2002.A Generalized Space Time Autoregressive Model and its Application to Oil Production Data. Dissertation Unpublihed. Bandung : ITB.
[11] Suhartono dan Atok, R.M. 2006. Pemilihan Bobot Lokasi yang Optimal pada Model GSTAR. Prosiding Konferensi Nasional Matematika XIII, Universitas Negeri Semarang.
[12] Suhartono and Subanar. 2006. The Optimal Determination of Space Weight in GSTAR Model by Using Cross-sorrelation Inference Journal of Quantitative Methods: Journal Devoted the Mathematical and Statistical Application in Various Field.
[13] Setiawan. 1992. Study on Seemingly Unrelated Regression (SUR) and its application to Almost Ideal Demand System (AIDS). Thesis Unpublished. Bogor : IPB.
[14] 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.
[15] Wutsqa, D.U dan Suhartono. 2010. Peramalan Deret Waktu Multivariat Seasonal pada Data Pariwisata dengan VAR-GSTAR. Jurnal Ilmu Dasar Vol. II no. 1.
[16] 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.