Rainfall Forecast of Merauke Using Autoregressive Integrated Moving Average Model

Yenni P. Pasaribu[1], *Hariani* Fitrianti[2], and *Dessy* Rizki Suryani[3]

[1] Department of Chemistry Education, Faculty of Teacher Training and Education, Musamus University, Merauke - Indonesia
[2] Department of Mathematics Education, Faculty of Teacher Training and Education, Musamus University, Merauke - Indonesia
[3] Department of Mathematics Education, Faculty of Teacher Training and Education, Musamus University, Merauke - Indonesia

Abstract. Climate is an important element for human life, one of them is to agriculture sector. Global climate change leads to increased frequency and extreme climatic intensity such as storms, floods, and droughts. Rainfall is climate factor that causes the failure of harvest in Merauke. Therefore, rainfall forecast information is very useful in anticipating the occurrence of extreme events that can lead to crop failure. The purpose of this research is to model rainfall using autoregressive integrated moving average (ARIMA) model. The ARIMA model can be used to predict future events using a set of past data, including predicting rainfall. This research was conducted by collecting secondary data from Agency of Meteorology, Climatology, and Geophysics (BMKG) from 2005 until 2017, then the data was analyzed using R.3.4.2. software. The analysis result showed that ARIMA model (2.0,2) as the right model to predict rainfall in Merauke. The result of forecasting based on ARIMA model (2.0,2) for one period ahead is 179 mm of average rainfall, 46 mm of minimum rainfall, and 295 mm of maximum rainfall. Thus it can be concluded that the intensity of rainfall in Merauke has decreased and there was a seasonal shift from the previous period.

Keywords: Climate; crop failure; rainfall; Merauke; ARIMA model.

1 Introduction

Climate is an important element for human life in some fields, namely fishery, agriculture, etc. Climate change is one of the natural phenomena, in which there are value changes of climate elements, either in natural process or accelerated process due to human activities. Nowadays, climate change becomes common problem between communities, interagency, interstate, and even global scope to gain serious solution because it affects many aspects of life. One of the climate change phenomena is the increased frequency and intensity of extreme climate such as storms, floods, and droughts [1]. Carbon emissions will hence lead adverse climate change on short and long time scale.

Global warming has resulted in climate change, which is very influential to agriculture because this sector very depends on climate condition [2]. Climate change adversely affects water availability and tends to decrease the harvest quality [3].

Agriculture sector is susceptible to climate change due to its effect on cropping pattern, planting time, production, and yield quality [4]. Additionally, agriculture is one of the important sectors in the provision of food. One of the main commodities of food crops in Indonesia is rice. Papua province is one of the areas with high diversity of biological resources, especially in Merauke City, which is the largest rice producer of rice crops in Papua and potential to become one of the national rice production centers [5]. One of the determinants of plant cultivation is rainfall [6].

Rainfall is the amount of rainwater that falls on an area in a certain time. The reduced rain intensity is one of the biggest reasons for the decrease in farmer yields [7]. The high rainfall can cause flood, while the low rainfall can cause drought; both of them can lead to crops failure. This is shown on Fitriani's research [5] stating that crops failure in Merauke is due to rainfall. Therefore, rainfall forecast information is really useful to anticipate extreme events that may lead to crops failure.

The ARIMA model is one of the time series analysis models that can be used to predict future events using a set of past data, including predicting rainfall in a period of time [8][9].

The advantage of ARIMA model is it can be used to analyze random, seasonal, even cyclic situation in the time series analyzed. Therefore, ARIMA model is used to predict rainfall in one future period. A research on rainfall forecast using ARIMA model is conducted by

---

* Corresponding author: pasaribu@unmus.ac.id

Weesakul and Lowanichchai [10]. It is about the rainfall forecast in Thailand for rice cultivation water supply.

2 Method

2.1. Data Plot

Before identifying the model, the first step that must be conducted is visual form of data plot, so it is convincing whether the data is stationer or not. However, it is more convincing to plot the autocorrelation values down to zero quickly after certain lag, so the data can be said as stationer. Whereas, if the autocorrelation is down into zero slowly or is different significantly from zero, it means that the data is not stationary. If the data is not stationary, it is necessary to conduct modification to obtain stationary data. One of the commonly-used ways is differencing method mean and transformation on variance.

2.2. Stationary Test

2.2.1. Correlogram

Correlogram is a stationary identification technique of time series using autocorrelation function (ACF) that is obtained through plotting between ρk and k (lag). Plotting between ρk and k is called population correlogram. Practically, we can only count the sample autocorrelation function. For stationary data, correlogram decreases rapidly as the increase of k. Meanwhile, for non-stationary data, correlogram tends not to zero (slow down).

2.2.2. Unit Root Test

Unit root test used is Augmented Dickey-Fuller (ADF test). Hypothesis used is as follow:

$H_0 : \rho k = 0$
$H_1 : \rho k \neq 0$

Reject H_0 if *p-value* < 0.05 then it can be said that the data is stationary.

2.3. Model Identification

If the data is stationary in mean and variance, it can proceed to see plot ACF and PCAF. Based on ACF and PCAF, it can be identified some possible suitable models. Model identification is conducted to know the existence of autocorrelation and data stationary, so it can be decided whether transformation or differencing process should be performed.

2.4. Parameter Estimation

After determining models based on identification of plot ACF and PCAF, model parameter estimation is conducted. One of the method used to predict model parameter ARIMA is ordinary least square (OLS) [11]. Parameter coefficient significance test is conducted to

know the proper model for predicting. The test used is *t-*test [10].

2.5. Residual Assumption Test (*diagnostic checking*)

Diagnostic checking is conducted to identify whether the model estimated is suitable enough for the data. Diagnostic checking is based on the residual analysis. The basic assumption of ARIMA model is that residual is an independent random variable distributed normally with zero mean constant variance.

2.5.1. Independence Test

It is to identify whether independent residual can use *Ljung-Box* test [11]. This test uses sample autocorrelation from residual to examine the initial residual. The hypothesis is as follow:

$H_0 : \rho k = 0$
$H_1 : \rho k \neq 0$

Reject H_0 if *p-value* < 0,05 it can be said that residual has no autocorrelation

2.5.2. Normality Test

To test residual assumption, Kolmogorov–Smirnov is used [12]. The formulation hypothesis is as follow:

H_0 : error data are normally distributed
H_1 : error data are not normally distributed

Criteria for rejected hypothesis H_0 test statistics > critical value.

1.6. Selection of the Best Model

The selection of the best model uses Akaike's Information Criterion (AIC). This model is based on maximum likelihood estimation (MLE). The selection of the best model with minimum AIC value [13].

1.7. Forecasting

The most important purpose of *times series* analysis is to predict future value [14]. If the best model has been selected, the model is ready to use for forecasting. Forecasting method is expected to increase the level of trust towards the next data.

3 Result and Discussion

3.1. Data Collection

Data used is monthly rainfall data from 2005 to 2017 that is obtained from Meteorology, Climatology, and Geophysics Agency (BMKG) in Merauke. However, the data used in modeling is rainfall data from 2005 to 2016, while data of 2017 is compared to the forecasting result. The data is obtained by asking the related officer about the information by bringing letter of permission to conduct a research. The data obtained is monthly rainfall data from 2005 to 2016.

3.2. Plot of rainfall data

The first step of modeling time series is data plot to see whether the data is stationary in mean and variance. If the data has not been stationary to the mean, it is necessary to conduct differencing, and if it is not stationary to the variance, it needs to perform transformation. Plot of rainfall from 2005 to 2016 is shown in Fig 1.
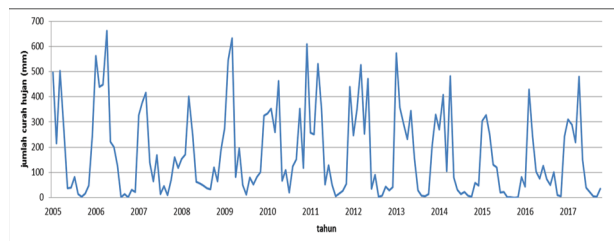

Fig 1. Plot of rainfall from 2005 to 2016

Based on the picture above, it can be seen that there is a shift in rainfall, and the intensity of rainfall is decreased. The rainfall peaks its maximum in April 2006, which is 663 mm, and the minimum point happens in October 2015, there is no rain at all (0 mm). The average of rainfall from 2005 to 2016 is 172 mm with deviation standard is 171 mm. A visual observation conducted towards Picture 1 shows that the rainfall data has been stationary towards it mean and variance. To make it sure, there is stationary data, which uses Augmented Dickey-Fuller (ADF).

3.3. Rainfall Data Stationary Test

Visually, it can be seen that the rainfall data has been stationary towards its mean and variance. However, to make it more accurate, it is conducted a stationary test using Augmented Dickey-Fuller (ADF) test [15]. The formulation of hypothesis is:

H_0 : non-stationary rainfall data
H_1 : stationary rainfall data

With test statistics is:
Criteria for rejected hypothesis H_0 if $p - value < \alpha$

Based on the assistance of software R 4.2.1, ADF obtain
obtained $p - value = 0,01 < \alpha = 0,05$ then H_0 is rejected. So, it can be concluded that the rainfall data is stationary.

3.4. Identification of ACF and PACF

When the data has been stationary, data is ready to be identified to obtain the prediction of ARIMA model (p,d,q). Identification is obtained through plot ACF and PACF of the rainfall data. Plot ACF and PACF of the rainfall data in Merauke is seen in Fig 2.
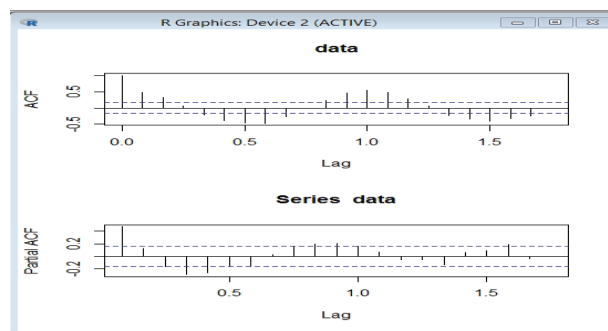

Fig 2. ACF and PACF of the rainfall data in Merauke

Based on plot ACF and PACF of rainfall data, it can be seen that significance value on the lag is small, so it is proposed an order for Moving Average (MA), which is order 1,2,3 and for Autoregressive is 1,2, 3.

3.5. Prediction of ARIMA Model

Model determining used is based on identification of plot ACF and PACF, so there is obtained 9 ARIMA models combined from the possible models, which is order-p (1,2,3) and order-q (1,2,3). Combination of ARIMA model is as seen in Table 1.

Table 1. Prediction of ARIMA model (p,dq)

Model	MA(1)	MA(2)	MA(3)
AR(1)	ARIMA (1,0,1)	ARIMA (1,0,2)	ARIMA (1,0,3)
AR(2)	ARIMA (2,0,1)	ARIMA (2,0,2)	ARIMA (2,0,3)
AR(3)	ARIMA (3,0,1)	ARIMA (3,0,2)	ARIMA (3,0,3)

3.6. Parameter Estimation of ARIMA Model

After being identified, there are 9 models combined based on Table 1. The next step is to find out parameter estimation of the model. Results of estimation using software R. 3.4.2 are listed in Table 2.

Result of Table 2 shows that the best ARIMA model is ARIMA model (2,0,2) compared to other models based on AIC criteria. It shows that ARIMA model (2,0,2) has the smallest AIC value. However, it needs to know whether the parameter is significant or not. Therefore, it is necessary to conduct a parameter significance test. The test for parameter of ARIMA model (2,0,2) is t-test. Results of significant test are listed in Table 3.

Table 2. Result of parameter estimation of ARIMA model (p,d,q) using *software R. 3.4.2*

Model	Intercept	α_1	α_2	α_3	β_1	β_2	β_3	AIC
ARIMA (1,0,1)	174,3	0,58	0	0	-0,13	0	0	1857
ARIMA (2,0,1)	174,3	0,3	1,28	0	0,29	0	0	1858
ARIMA (3,0,1)	172	1,06	-0,01	-0,36	-0,70	0	0	1839
ARIMA (1,0,2)	174,02	0,48	0	0	-0,08	0,23	0	1854
ARIMA (2,0,2)	170,6	1,73	-0,99	0	-1,69	1	0	1796
ARIMA (3,0,2)	170,78	1,72	-0,98	-0,01	-1,69	1	0	1798
ARIMA (1,0,3)	174,18	0,34	0	0	0,07	0,26	0,16	1854
ARIMA (2,0,3)	170,59	1,73	-0,998	0	-1,71	1,02	-0,01	1798
ARIMA (3,0,3)	170,6	0,73	0,72	-0,99	-0,69	-0.69	0,99	1799

**Table 3.** The result of parameter significance test for ARIMA model (2,0,2)

Parameter	Koefisien parameter	Standard error	t_{test}	t_{table}
α_1	1,73	0,004	433	1,98
α_2	0,998	0,003	333	1,98
β_1	1,69	0,03	56	1,98
β_2	1	0,4	3	1,98
Intercept	170,6	10,9	16	1,98

The initial hypothesis of ARIMA model (2,0,2) is insignificance parameter. However, based on the Table, $t_{test} > t_{table}$ with $\alpha = 0,05$ and n = 144, so the hypothesis is rejected. Therefore, it can be concluded that the parameter of ARIMA model (2,0,2) is significant.

3.7. Selection of the Best Model

3.7.1. Residual Assumption Test (Diagnosis Checking) of ARIMA model

After conducting parameter estimation and significance test, white noise test is carried out for ARIMA model (2,0,2). The aim of this test is to diagnose whether residual from the model obtained is no autocorrelation. The result of diagnostic test of ARIMA model using software R. 3.4.2 is seen in Fig 3.
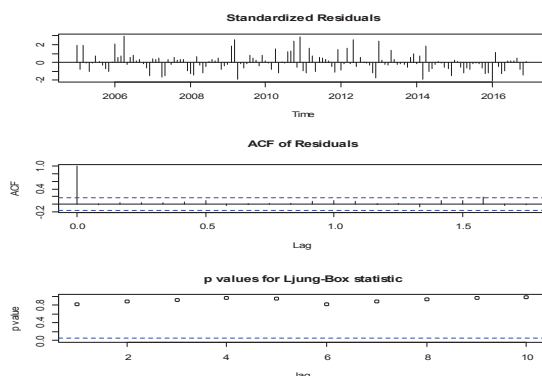
Fig 3. Plot of *white noise* test

The result on Picture 3 shows that Ljung-Box test obtains $p-value > \alpha = 0,05$. Therefore, it can be concluded that the model is *white noise,* meaning that

ARIMA model (2,0,2) meet the requirement and can be used for forecasting.

3.7.2. Residual Normality Test

A good model is a model without autocorrelation and model residual from population with normal distribution. Therefore, to obtain a good model, it is necessary to test whether residual of ARIMA model (2,0,2) is from population with normal distribution. With equation (4), it is obtained test statistics = 0,64 and critical values with $\alpha = 0,05$ it is obtained critical values $\frac{1,36}{\sqrt{n}} = 0,11$. Based on the result, it is obtained test statistics > critical value, so initial hypothesis (H0) is rejected. It can be concluded that residual data of ARIMA model comes from normal distributed population.
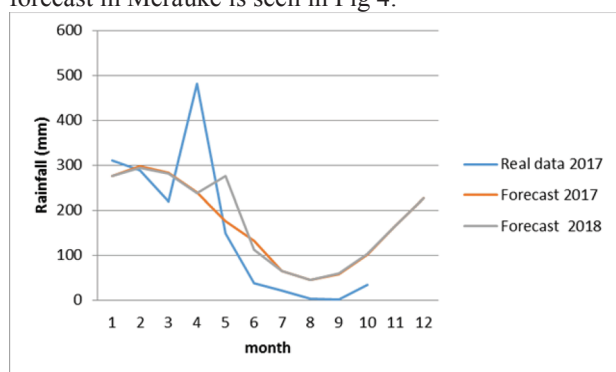
3.8. Forecasting

Because all steps of ARIMA modeling have been conducted, and the requirements of ARIMA model have been met, the next step is to forecast rainfall using the model for one period ahead. Table 4 shows is the result of rainfall forecast using ARIMA (2,0,2):

Table 4. Result of Rainfall Prediction in 2018

Month	Original Data 2017	Forecast in 2017	Forecast in 2018
January	310,1	277	277
February	288,7	298	295
March	219,1	284	282
April	480,8	445,0	238
May	149	177	276
June	38,6	133,0	113
July	22,9	65,0	66
August	4,6	45,0	46
September	2,4	58,0	60
October	34,7	102,0	104
November		165,0	165
December		228,0	228

Prediction of rainfall is divided into three categories determined by BMKG, that is, 'low' category (0 mm – 100 mm), 'medium' category (100 mm – 300 mm), and 'high' category (more than 300 mm). based on Table 4, there is obtained a rainfall forecast with 'low' category' on July, August, and September; 'medium' category on January-June 2018 and October-December 2018, and there is no 'high' category. Moreover, based on the result of rainfall forecast, the maximum intensity of rainfall decreases, and there is season shift, that is, the maximum rainfall in 2017 is on April, and the rainfall forecast in 2018 is on February. Meanwhile, based on visual observation, it can be seen that the plot of rainfall of original data and forecast data have the same pattern. The plot time series of rainfall forecast in Merauke is seen in Fig 4.

Fig 4. Rainfall forecast in 2018

The result shows that rainfall in Merauke in 2018 using ARIMA model (2,0,2) peaks its maximum point on February (295 mm) and minimum point on August (46), while the average of rainfall is 179 mm. This rainfall forecast can be used by farmers to determine the best rice planting time in the medium category. In this time, farmers do not need much additional water from the river, thus reducing the use of water pumping machines. This gives the advantage of reducing the use of fuel oil so as to reduce carbon emissions in Merauke.

4. Conclusion

Based on the research result of the best model for rainfall forecasting in Merauke based on AIC is ARIMA model (2,0,2), where the maximum frequency of rainfall in 2018 decreases and there is season shift from the previous period. The average of rainfall and the minimum point of rainfall increases.

This model can be used to determine the best planting time so as to reduce the use of fuel oil for water pump which has a positive impact on reducing carbon emissions in Merauke.

References

1. Hukom , E, L M Limantara *at al*. "Pengaruh Perubahan Iklim terhadap Optimasi Ketersediaan Air di Irigasi Way Mital Provinsi Maluku." *Jurnal Teknik Pengairan* 3 No 1 (2012): 24-32.
2. Rasmikayati, E *at al*. "Dampak Perubahan iklim terhadap Perilaku dan Pendapatan Petani." *J. Manusia dan Lingkungan* 22 No 3 (November 2015): 372-379.
3. Hidayati, I N, and Suryanto. "Pengaruh Perubahan Iklim terhadap Produksi Pertanian dan strategi Adaptasi pada Lahan Rawan Kekeringan." *Jurnal Ekonomi dan Studi Pembangunan* 16 No 1 (2015): 42-52.
4. Nurdin. *Antisipasi Perubahan Iklim untuk Keberlanjutan Ketahanan Pangan.* Sulawesi Utara: Universitas Negeri Gorontalo, 2011.
5. Fitrianti, H. "Menentukan Premi Murni menggunakan Generalized Linear Models dan Model Copula." *Magistra* 3 No 1 (2016): 69-81.
6. Purba Novelina *at al*. "Pemodelan Data Curah Hujan Di Kabupaten Banyuwangi Dengan Metode Arima Dan Radial Basis Function Neural Network". jurnal sains dan seni ITS Vol.5No.2 (2016):339-342.
7. Angles, Chinnadurai, and Sundar. "Awareness on Impact of Climate Change on dryland Agriculture and Coping Mechanisms of Dryland Farmers." *Indian journal of Agricultural Economics* 66 (2011): 365-372
8. Hasria. "Perkiraan Curah Hujan Bulanan Kota Kendari Dengan Model ARIMA". *Jurnal Aplikasi Fisika* Volume 8 No 1 (2012): 25-30
9. Elvani, S P, A R Utari, and R Yudaruddin. "Peramalan Jumlah Produksi Tanaman Kelapa Sawit dengan Menggunakan ARIMA." *Jurnal Manajemen* 8(1) (2016): 95-112.
10. Weesakul U, Lowanichchai S, 2005. Rainfall forecast for agricultural water allocation planning in Thailand. Thammasat International Journal of Science and Technology, 10(3): 18–27.
11. Gujarati. "Basic Econometrics". Fourth Edition.The McGraw-Hill Company Inc. 2004.
12. Klugman, S.A., Harry, H.P. & Gardon, E.W. (2004). "Loss Models: *From Data to Decisions*". Second Edition. Jhon Wiley & Sons: Hoboken.
13. Cryer, D J. *Time Series Analysis With Applications in R.* New York USA: Springer, 2008.
14. Wei, W. *Time Series Analysis Univariate and Multivariate Methods.* Amerika: Addison Wesley Publishing Company, Inc. (2006).
15. Nugroho Adi, Hartati Sri, Subanar, dan Mustofa Khabib. " Vector Autoregression (Var) Model for Rainfall Forecast and Isohyet Mapping in Semarang – Central Java – Indonesia". International Journal of Advanced Computer Science and Applications, Vol. 5, No. 11, 2014.