Applied Mathematics and Computational Intelligence
Volume 10, No.1, Dec 2021 [101-126]

Prediction of Rainfall Using ARIMA Mixed Models

Miftahuddin[1], Norizan Mohamed[2*], Maharani.A. Bakar[2], Nur Shima[3], Fadhli[4], and Ichsan Setiawan[5]

[1]Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Aceh, Indonesia
[2] Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu
[3]Department of Meteorology, Indonesian School for Meteorology Climatology and Geophysics
[4]Department of Physics, Faculty of Mathematics and Natural Sciences, Syiah Kuala University
[5]Department of Marine Sciences, Faculty of Marine and Fishery, Syiah Kuala University

[*] Corresponding author: norizan@umt.edu.my

Received: 23 December 2020; Accepted: 3 May 2021; Available online: 24 June 2021

ABSTRACT

The average rainfall in Aceh Barat every year is different pattern and it is influenced by several factors. In this paper we used rainfall dataset, which is changing time to time. The change is caused by an element of fluctuate and volatility in the data. The purpose of this study was to find the best ARIMA mixed models as combination with ARCH and GARCH models. The data used in this study are rainfall data and the number of rainy days in Aceh Barat district from the period January 2008 to December 2017. The results showed that stationary rainfall in the transformation results of Z_t 0.27 and the first differencing (d=1) and test results Lagrange multiplier-ARCH for rainfall data and the number of rainy days shows significant lag 4. The best model for predicting rainfall uses the ARIMA(2,1,0)-ARCH(3) model and for the number of rainy day using the ARIMA(2,0.2) model. The calculation results obtained prediction accuracy value for rainfall using ARIMA(2,1,0)-GARCH(1,3) model with MAD, RMSE, MAE, and MASE values of 1,175, 1.163, 0.941 and 0.720 respectively and for the number of rainy days using ARIMA(2,0,2) model were accuracy value respectively 4.448, 3.849, 3.189 and 0.737.

Keywords: rainfall, ARIMA model, GARCH model, ARIMA mixed model

1 INTRODUCTION

Aceh region is western of Indonesia Islands. A coast western of Aceh is influenced by phenomena in the Indian Ocean. To obtain display local climate the western Aceh, we investigate rainfall phenomena in Aceh Barat district that have direct border in Indian Ocean. In addition, due to geographical separation by Barisan mountain along the Sumatra island, rainfall pattern including Aceh, is separated by this mountain into two distinct annual mean rainfall patterns, west and east mountain [1]. To represent these regions, we select Meulaboh station for the western part of the mountain. Aceh Barat district is located on 04°06'-04°47'NL (north latitude) and 95°52'- 96°30'ET (east longitude) with total area 2.927,95 km^2. Aceh Barat district in north side has border with Aceh Jaya and Pidie districts, east side has border with Aceh Tengah and Nagan Raya districts, south side restricted by Indian Ocean and Nagan Raya district and west side restricted Indian Ocean [2]. Region of Aceh Barat district has potential area for the cultivation of various agricultural commodities because they are supported by adequate weather and climate. One of the weather factors that is intended is rainfall.

Aceh Barat district is one [3] based on data from the Central Bureau of Statistics, during last 13 years, the highest rainfall was recorded in August 2011, while the lowest rainfall occurred in April

2007. Impact of high rainfall, on December 2, 2018 floods hit Aceh Barat district. Woyla Barat sub district is one of the sub districts with the highest rainfall. Due to the high level of rainfall causing the river water to overflow. As a result, villages in Woyla Barat sub district adjacent to the river were flooded that reached 50 -80 cm.

According to the Meteorology, Climatology and Geophysics Agency (MCGA), the weather conditions in an area can change. These changes it is caused by many factors that occur in the past, present, and future condition. In addition, weather variable changes also occur due to volatility indicator. Volatility is a condition where fluctuations are relatively large and usually followed by low or high return fluctuations (mean and inconstant variance). One model that assumes residual variants are inconstant in time series was developed by Engle [4] called heteroscedasticity autoregressive conditional (ARCH) models and refined by Bollerslev [5] known as generalized autoregressive conditional heteroscedasticity (GARCH) models. Heteroscedasticity can occur because in time series data shows an element of volatility. According to Engle [6] the ARCH/GARCH models can be used to show volatility of time series data, such as rainfall dataset. Therefore, to construct ARIMA mixed model of the rainfall dataset that has volatility effects, the time-series approach that can be used to extend of ARIMA (Autoregressive Integrated Moving Average) models, such as ARIMA-ARCH and ARIMA-GARCH models.

2 MATERIAL AND METHODS

Rainfall is the amount of water that falls to the surface of the earth in a certain time. The average rainfall in Indonesia every year is different and the rainfall in the territory of Indonesia is influenced by several factors including the shape of the terrain, topography, geography, the direction of the slope of the field, wind direction parallel to the direction of the coast and air pressure [7]. The types of rain based on the amount of rainfall according to the MCGA are divided into four, as presented in Table 1:

Table 1. Types of rain based on the amount of rainfall

Type of rain	Amount of rainfall per day (mm)
Weak	under 20
Medium rain	20 - 50
Heavy rain	50 - 100
Very heavy rain	above 100

2.1 Model and Data

Rainfall intensity is a measure of the amount of rain per unit of time during the rain. Rainfall is generally divided into 5 levels according to its intensity as presented in Table 2 below

Table 2. The level of rainfall is based on its intensity [30]

Level	Intensity (mm/minute)
Very weak	< 0.02
Weak	0.02 – 0.05
Medium	0.05 – 0.25
Heavy	0.25 – 1
Very Heavy	> 1

A time series is fulfilling a stationary requirement where stationary data is divided into two, namely stationary data in mean and variance. If the data is non stationary against the variance, then the data transformation is performed. While the data is non stationary to the mean, it is carried out differencing [8]. According to Box and Cox [9] a data is said to have been stationary to the variance, if it has a value 1. The value of the parameter λ can be predicted through a likelihood function with the following equation:

$$L(\lambda) = -\frac{1}{2}(n \log \{S(\lambda; z)\}/n) \tag{1}$$

If we used data stationarity with the mean, then testing was used through the Dickey Fuller (ADF) approach with hypothesis as follows:

$H_0: \phi = 1$ (there is a root unit/not stationary)

$H_1: \phi \neq 1$ (no root/stationary unit)

The significant level used is α (5%). Test statistics:

$$ADF = t_{ratio} \frac{\hat{\varphi}}{se(\hat{\varphi})} \ . \tag{2}$$

Decision criteria: reject H_0 if $t_{ratio} >$ critical value of ADF or p-value < α. The models used in the time series analysis are as follows:

In general, the autoregressive model with the order-p AR(p) or in the ARIMA(p,0,0) model is written as follows:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + \alpha_t \ , \tag{3}$$

where Z_t is stationary data, t is time unit ($t = 1,2,3,\cdots,k$), $\phi_1, \phi_2, \cdots, \phi_p$ are autoregressive parameter *p-th* with $\phi_p \neq 0$, α_t is error in t. While in moving average model with order-q MA(q) or equal in the ARIMA(0,0, q), model is written as follows:

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \tag{4}$$

where $\theta_1, \theta_2, \ldots, \theta_q$ are moving average parameter [9].

Generally, the ARMA equation is stated as follows:

$$Z_t = \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \tag{5}$$

where $\phi_p \neq 0$, $\theta_q \neq 0$ [10]. If the time series data shows that it is not stationary or shows a trend, then the data can be differenced. Differencing data that is processed by using the extend of ARMA model or is named ARIMA process with parameters (p,d,q) with order-p as the operator of AR and order-q as the operator of the MA. This model is used for time series data that has been differencing or are already stationary in the mean, where d is the order of differencing. The form of the formulation of ARIMA(p,d,q) model is as follows:

$$\phi(B)(1-B)^d \breve{z}_t = \theta(B) a_t \tag{6}$$

where $\phi(B)$ is AR parameter and $\theta(B)$ is MA parameter, both parameters of differencing with d is differencing order [11].

2.1.1 *ARIMA Model Identification*

i) Autocorrelation Function (ACF)

The autocorrelation function is used to explain a stochastic process regarding the correlation between adjacent data. A z_t process that is stationary that has $E[z_t] = \mu$ and $Var(z_t) = E[z_t - \mu^2] = \sigma^2$ constant covariance $Cov(z_t, z_{t-k})$ can be written as follows:

$$\gamma_k = Cov(z_t, z_{t-k}) = E[z_t - \mu][z_{t-k} - \mu] \tag{7}$$

and correlation between z_t and z_{t-k} is

$$\rho_k = \frac{Cov(z_t, z_{t-k})}{\sqrt{Var(z_t)}\sqrt{Var(z_{t-k})}} = \frac{\gamma_k}{\gamma_0} \tag{8}$$

where $Var(z_t) = Var(z_{t-k})$, for z_t is a variable at time t, and z_{t-k} is a variable at time $t-k$, γ_k is the auto covariance function at k and ρ_k is ACF at k [12].

ii) Partial Autocorrelation Function (PACF)

The PACF is used to measure magnitude of the association variables between z_t and z_{t+k} which occurs when the lag time is omitted by the equation below:

$$\hat{\varphi}_{kk} = \frac{\widehat{\rho_k} \sum_{j=1}^{k-1} \hat{\varphi}_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\varphi}_{k-1,j}\rho_{k-j}} \tag{9}$$

iii) Parameter Estimation and Parameter Testing of Model

There are two ways to estimate parameter, namely by trial and error; and by iterative improvements [13]. The method that used to estimate parameters in this study is the likelihood function. The likelihood function for the parameters if it is known that observation data are [11]:

$$L(\phi,\theta,\sigma_a^2 \mid Z) = (2\pi\sigma_a^2)^{-1/2}\left(\exp(-\frac{1}{2\sigma_a^2}S(\phi,\theta)\right) \quad (10)$$

If θ is a parameter and $\hat{\theta}$ is the estimated value of that parameter, and $se(\hat{\theta})$ is the standard error of the estimated value $\hat{\theta}$, then the parameter significance test as follows:

a) Hypothesis, H_0: = 0 (significant parameter)

H_1: ≠ 0 (parameter not significant)

b) Statistic Test

$$t_{count} = \frac{\hat{\theta}}{se(\hat{\theta})} \quad (11)$$

c) Critical area: if p-value < α or | t_{count} | > $t_{(1-\alpha/2)\ db = n-p}$, then reject H_0.

iv) Diagnostic Model

a) Residual with white noise property

This residual test is carried out applying the Ljung-Box test

$$Q = n(n+2)\sum_{k-1}^{K}(n-k)^{-1}\hat{\rho}_k^2 \quad . \quad (12)$$

If $Q > \chi^2$ or $p\text{-value} < \alpha$, then H_0 is rejected.

b) Residual normality test

This residual normality test was carried out using the Kolmogorov-Smirnov test

$$D = sup\ |S(z) - F_0(z)| \quad (13)$$

If $D_{count} > D_{(1-\alpha,n)}$ or $p\text{-value} < \alpha$, then H_0 is rejected.

3 AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTIC MODEL (ARCH(Q)) AND GENERALIZED-ARCH

Engle [6] introduced the ARCH model for the first time and solved the issue of the weights to be used for the variance term and counted these weights as a parameter which is to be estimated. This model is used to overcome residual variances that are inconstant for time series. The residual variance in the ARCH(q) model is strongly influenced by residuals in the previous period ε_{t-1}^2 [12]. The equation for the ARCH(q) model is as follows:

$$\sigma_t^2 = \alpha_0 + \alpha_0\varepsilon_{t-2}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2\ , \quad (14)$$

where σ_t^2 is variance residual at time t, α_0 is constant component, α_q is parameter ARCH with order q, ε_{t-q}^2 is residual square at time $t - q$.

The Lagrange Multiplier ARCH model (LM-ARCH) is used if the residuals in time series indicate the presence of heteroscedasticity [14].

a) Hypothesis, H_0 : $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$ (no effect ARCH until *lag-k*)

H_1 : (there are at least one $\alpha_j \neq 0$, j=1, 2,...,k (there is effect ARCH)

b) Statistics Test by LM-ARCH

$$\tau = nR^2 \tag{15}$$

c) Decision Criteria: If $\chi^2_{calculated} > \chi^2_{(\alpha,df)}$ or p-value < α then H_0 is rejected.

Thus, the data are allowed by the model to determine the best weights to use in forecasting the variance. ARCH models are specific for low order [15], so that if there is a significant lag that causes the model to be inefficient, then generalized ARCH model, namely GARCH, is needed.

GARCH model is an extended of the ARCH model [5]. A very important generalization of this model was introduced by [5] called GARCH parameterization. The GARCH model is also a weighted average of the past squared residuals, but it has declining weights which ever reaches completely to zero. The property provided useful models which can handle and estimate easily and has confirmed effectiveness in forecasting the conditional variances. These models are broadly used for best forecasting of variance in the coming period as a weighted average of the long run variance.

In this investigation, a GARCH(p,q) model was tested because to obtain comparison of forecasting accuracy using ARCH and GARCH methods. The formulation of the GARCH(p,q) model is:

$$\sigma_t^2 = a_0 + a_1\varepsilon_{t-1}^2 + \ldots + a_p\,\varepsilon_{t-p}^2 + \beta_1\sigma_{t-1}^2 + \ldots + \beta_q\sigma_{t-q}^2, \tag{16}$$

where a_0 is constant component, a_p is parameter of GARCH with order p, β_q is parameter of GARCH with order q, ε_{t-p}^2 is quadratic residual at time t-p, σ_{t-q}^2 is variance residual at time t-q.

3.2 Model Selection

To select the best model fitting of time series data, there are several tools such as the Akaike's Information Criterion (*AIC*), Generalized *AIC* (*GAIC*), Bayesian Information Criterion (*BIC*), Schwarz Information Criterion (*SIC*) and AIC correction (*AICc*) values. The third method is based on Maximum Likelihood Estimation, i.e. *AIC*, *BIC*, and *AICc*. The *AIC* formulation is as follows:

$$AIC = n + n\log 2\pi + n\log(RSS/n) + 2(p+1). \tag{17}$$

The mathematics equation can also be written as

$$AIC = \ln \sigma_p^2 + \frac{m+2p}{m} , \quad (18)$$

where σ_p^2 is variance residual with m is number of observation, and p is number of parameter of the model [11]. *BIC* formulation is as follows:

$$BIC = n + n\log 2\pi + n\log(RSS / n) + (\log n)(p+1) . \quad (19)$$

In Cavanaugh, *AICc* equation is:

$$AICc = n \ln(RSS / n) + \frac{n(n+p)}{n-p-2} . \quad (20)$$

The best model criteria are the model that has a value of *AIC*, *BIC*, and *AICc* minimum [13]. To assess forecasting model, we used mean absolute deviation (*MAD*) and root mean square error (*RMSE*), mean absolute error (*MAE*) and mean absolute scaled error, as follows:

$$MAD = \frac{1}{n}\left|\sum_{t=1}^{n} Z_t - \hat{Z}_t\right| \quad (21)$$

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{1}{n} (\hat{Z}_t - Z_t)^2} \quad (22)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (23)$$

$$\text{MASE} = \text{mean}(|q_t|), \text{ where } q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n} |z_t - z_{t-1}|} \quad (24)$$

In [19], the authors compared different types of forecasting models, including the random walk, historical mean, moving average, exponential smoothing, linear regression models, autoregressive models, and various GARCH models to forecast petroleum prices. The researchers used WTI daily futures prices of crude oil, heating oil, and unleaded gasoline covering the period from February 5, 1988 to January 31, 2003. The findings indicate that for heating oil and natural gas, the TGARCH model fits the best, whereas for crude oil and unleaded gasoline, the GARCH model fits the best. Therefore, GARCH type models outperform the other techniques.

Similarly, [20] employed several GARCH types of models for forecasting the daily WTI crude oil prices volatility and pin pointed some indistinct models, though, the results obtained from this study were incompatible and their respective performance exposed by some diverse measures and statistical tests. In this research, the sample periods was from December 31, 1991 to May 02, 2005. Additionally, [21] also incorporated several GARCH types of models for forecasting the daily crude oil prices for future volatility. They worked on NYMEX Exchange from January, 1995 to November, 2005. The authors then concluded that no model performs well on regular basis, some supported research findings with several statistical tests were provided. They used different performance measures tests such as MSE (with adjusted heteroscedasticity), MAE, Diebold Mariano, and success ratio.

Furthermore, [22] developed a new method by inculcating nonparametric system in model for forecasting the return volatility of crude oil prices. The outcomes determined that the nonparametric GARCH model outperform the parametric GARCH models regarding the out sample forecasting volatility using WTI data from January 6, 1992 to October 23, 2009. Likewise, [23] also forecasted the WTI and Brent daily crude oil spot prices by implementing the GARCH type of models. They used time series data of WTI excluding the public holidays spanning January 01, 1986 to September 30, 2006 and Brent from May 20, 1987 to September 30, 2006. The main focus of the study was the demonstration of the advantages and disadvantages of the linear and nonlinear models and fitted the different GARCH type of models namely GARCH(G, N, T) for WTI and Brent crude oil daily spot prices. The output of all these three models were different because not a single model performed well for both data sets, the GARCH-G model best fitted the WTI crude oil spot prices whereas for Brent crude oil spot prices the best candid model was the GARCH-N.

In [24] , the authors pointed out the short term of the results which are reliable non-switching models, while Markov switching GARCH model performed well and produced a higher accuracy in terms of forecasting the long-term volatility in crude oil. The researchers used the daily data of WTI from July 01, 2003 to April 02, 2014. Furthermore, [25] examined the return volatility of Brent crude oil returns through GARCH, E-GARCH, Glosten-Jagannathan-Runkle (GJR), GJR-GARCH and Markov Regime Switching (MRS), MRS-GARCH models. All of them were modeled under normal, generalized error distribution and Student's t distributions. The best model was chosen based on AIC and BIC values and the model MRS-GARCH outclasses all other alternate models. The study used the time series from December 01, 1998 to January 30, 2015.

The data source used is secondary data in regular intervals of time that obtained from the Aceh Provincial Statistics Agency. The data is time series (in monthly unit) with data periods from January 2008 to December 2017.

4 RESULTS AND DISCUSSION

We use performance on monthly-time rainfall and the amount of rainy days' data in our study to obtain realistic phenomena and data updating. Besides it is given several statistical characteristics, also its plotting in time series type.

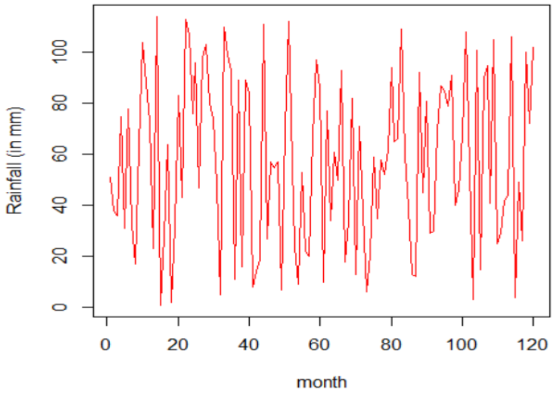

Figure 1a. Plot of rainfall on 2008 -2017 in Aceh Barat district

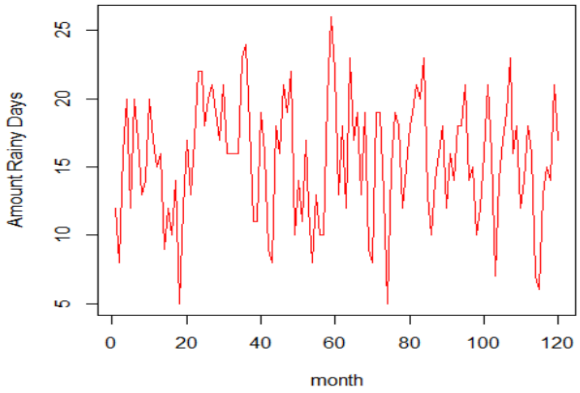

Figure 1b. Plot of the amount of rainy days on 2008-2017 in Aceh Barat district

Based on Figures 1a and 1b, the plots show the irregular patterns with original data, furthermore, with the descriptive analysis for rainfall and the amount of rainy days in Aceh Barat district as follows:

Table 3. Measurement of location for rainfall & the amount of rainy days in Aceh Barat

Variable	N	min	max	mean	Q_1	Q_2 (Median)	Q_3	trimmed	mad
Rainfall	120	63.1	774.3	322.395	209.775	300.3	396.825	310.96	138.7
The amount of rainy days	120	5	26	15.575	12	16	19	15.71	4.45

According to the rainfall classification by MCGA, if it is known that the mean is 322.395 mm divided 30 days (assumption) then the rainfall amount is 10.7465 mm/day. While if it is known the maximum is 774.3 mm, divided 30 days, which equals to 25.81 mm/day (medium).

Table 4. Dispersion Measurement of rainfall & the amount of rainy days in Aceh Barat

Variable	range	mean deviation	σ^2 (variance)	σ (std dev)	skew	kurtosis	se
Rainfall	711.2	122.4848	23582.3935	153.5656	0.66	-0.11	14.02
The amount of rainy days	21	3.6413	19.8934	4.4602	-0.23	-0.51	0.41

Tables 3 and 4 showed that for one month, the average number of rainy days is 16 days, minimum occurs 5 rainy days and reaches a maximum of 26 days, with mean deviation of 122.4848 mm and 3.6413 days respectively.

4.1 Stationarity Data

4.1.1 Stationarity Test Against Variance

By using equation (1), we examine stationary in variance of the rainfall and amount rainy days' data obtained as in Table 5.

Table 5. Result of Stationary Data Test Against Variance

Variable	λ (lambda)
Rainfall	0.34
Number of rainy days	1.20

Table 5 showed that the value of λ is 0.34 for the rainfall variable. So, it can be concluded that the rainfall variable is not stationary to the variance. General transformation can be used as Box-Cox transformation with λ value [12], as seen Table 6

Table 6. The Box-Cox Transformation

λ value	Transformation
-1.0	$1/Z_t$
-0.5	$1/\sqrt{Z_t}$
0.0	$\ln Z_t$
0.5	$\sqrt{Z_t}$
1.0	Z_t (no transformation)

Therefore, transformation must be done with $(Z_t)^x$, where the x values that have been tested are 0.3, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, and 0.37. After the data transformation is done at $Z_t^{0,34}$, the value of $\lambda = 1$ is obtained for the rainfall variable. So it can be concluded that the rainfall variable is stationary against the variance. While for the number of rainy day variable, the value of λ is 1.20. So, it can be concluded that the number of rainy day variable is stationary for variance because the value of $\lambda \approx 1$.

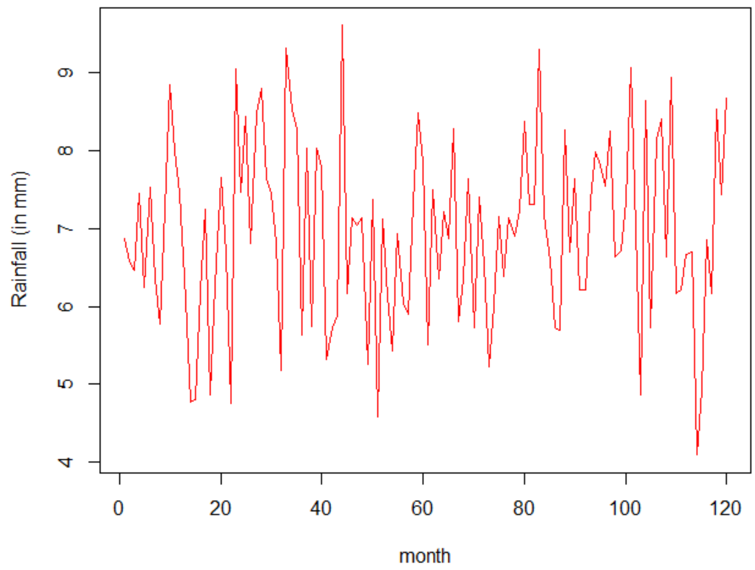

Figure 2. Rainfall post-transformation

4.1.2 Stationary of Data Test Against Mean

We used Augmented Dickey-Fuller (ADF) test to examine the stationary data of rainfall and number of rainy days in mean context.

Table 7. Results of Augmented Dickey-Fuller Test Pre-Differencing

Variable	**ADF**	***p*-value**	***Lag* order**	***A***	**Decision**
Rainfall	- 4.499	0.01	4	0.05	H_0 rejected
Number of Rainy Days	- 4.924	0.01	4	0.05	H_0 rejected

Based on Table 7, *p*-value of 0.01 of rainfall and number of rainy days of variables are obtained. By using *A* =0.05 can be seen that the rainfall and number of rainy days' variables are stationary with the mean in the fourth lag. However, for rainfall variable performed differencing, due to forecasting of time series model by using differencing data is more appropriate than forecasting using model without differencing. The following results are stationary to the mean post-differencing (d = 1), we can see in Table 8.

Table 8. Results of Augmented Dickey-Fuller Test Post-Differencing

Variable	**ADF**	***p*-value**	***Lag* order**	***A***	**Decision**
Rainfall	- 7.577	0.01	4	0.05	H_0 rejected
Number of Rainy Days	- 7.653	0.01	4	0.05	H_0 rejected

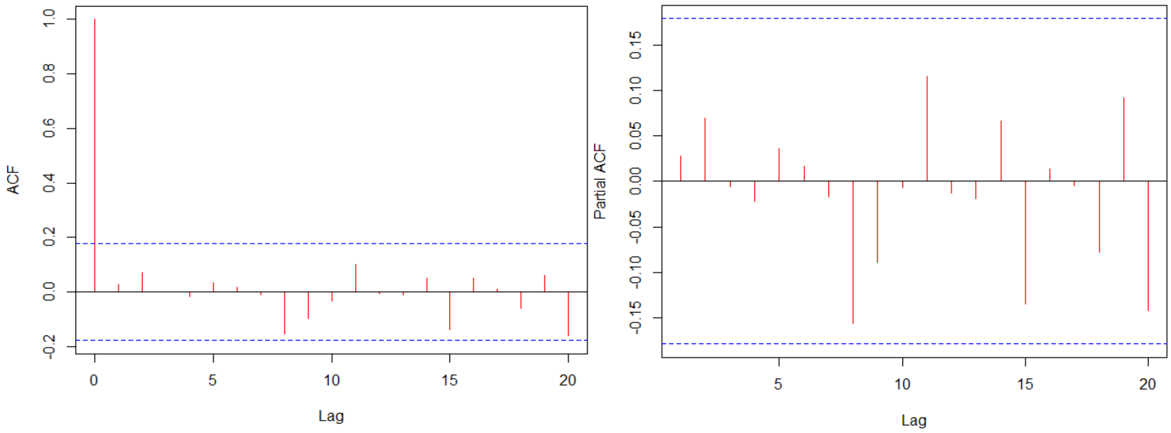

Figure 3: ACF and PACF of rainfall pre-transformation and differencing with *d* =1

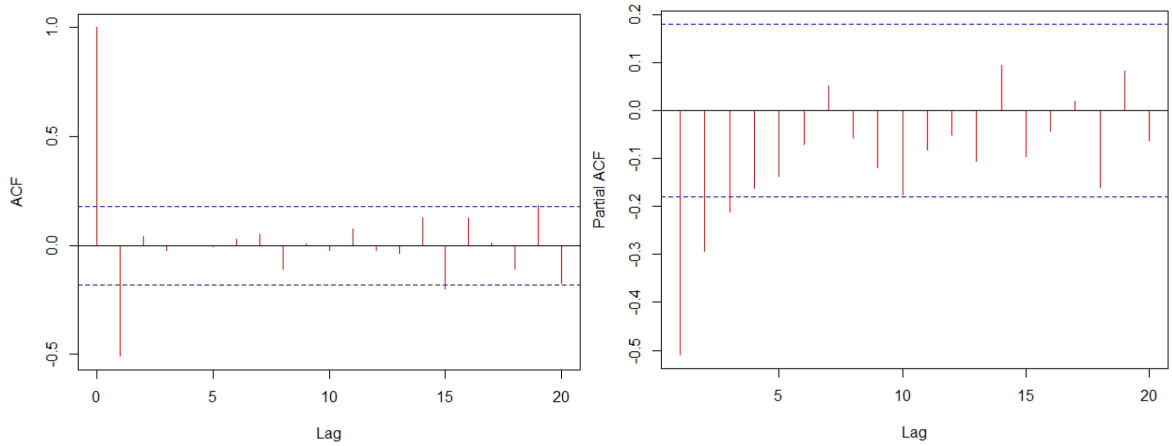

Figure 4. ACF and PACF of rainfall post-transformation and differencing with *d* =1

4.2 Model Identification

In this section, we can see ACF and PACF plots of rainfall and number of rainy days in Aceh Barat district. Model identification is an important procedure to construct of the model of rainfall and number of rainy days in time series modelling. By ACF and PACF plots, we have obtained patterns of the data in lag perspective.

As seen on Figure 5, the ACF plot decreased drastically (cut off) and PACF plot decreased exponentially for rainfall. While Figure 6 shows that the ACF plot decreased exponentially and PACF decreased drastically (cut off). So that, it can be estimated several tentative ARIMA models that appropriate to forecast rainfall and number of rainy days variables.

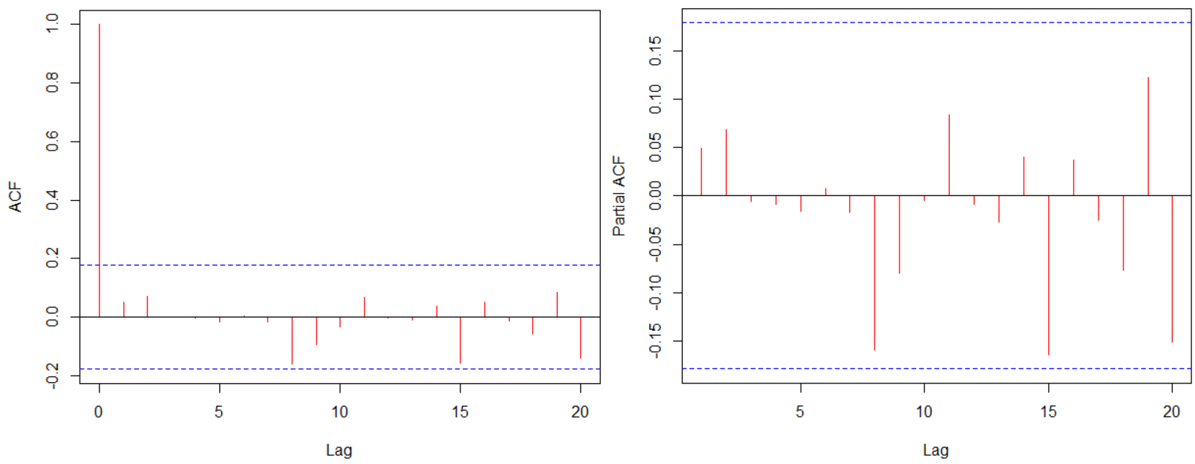

Figure 5. ACF & PACF plots of rainfall post-transformation

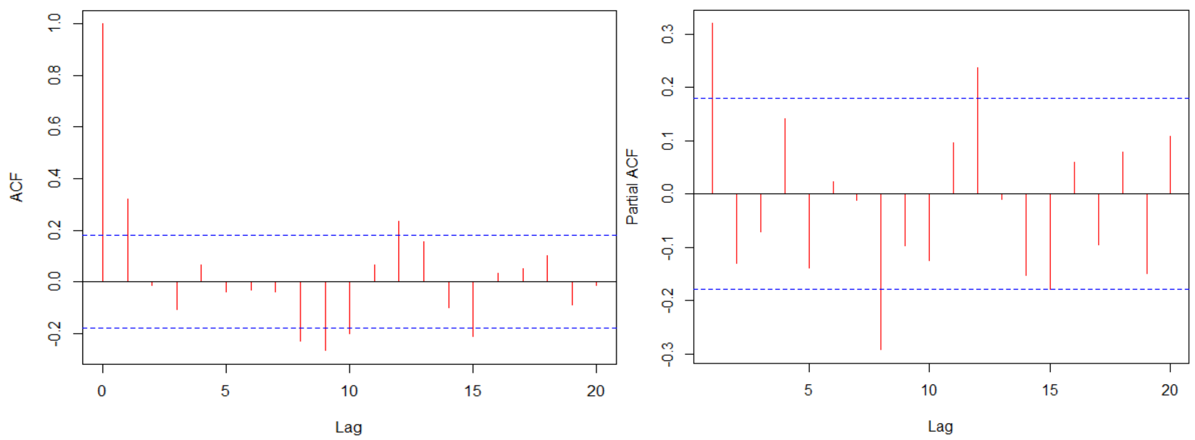

Figure 6. ACF & PACF plots of number of rainy days pre-differencing

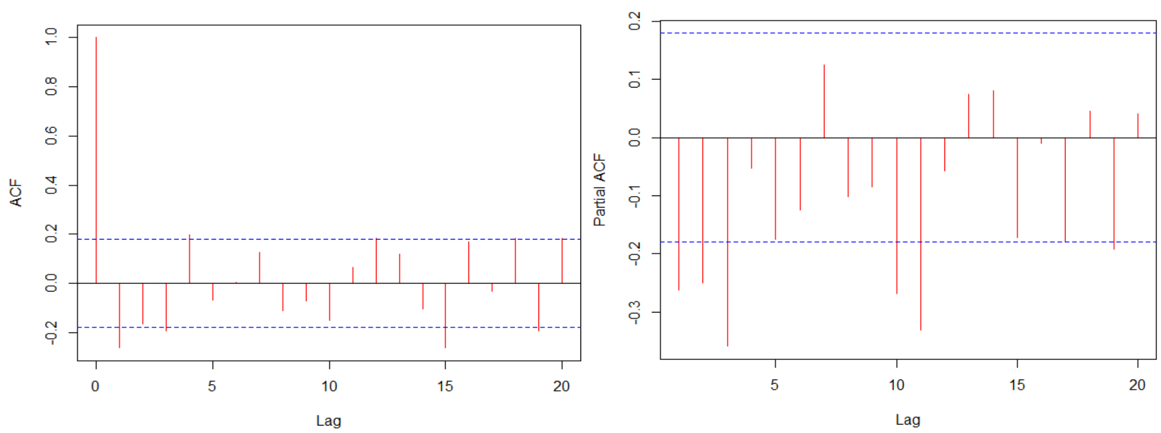

Figure 7. ACF and PACF of number of rainy days with differencing *d* =1

We have two ways to get model appropriate for data fitting. Firstly, through the tentative ARIMA models that estimated based on trial and error for rainfall are ARIMA(1,1,0), (1,1,1), (1,1,2), (1,1,3), (2,1,0), (2,1,1), and (2,1,2) models. While for number of rainy days with tentative ARIMA models that estimated based on the same way are ARIMA(1,0,0), (1,0,1), (1,0,2), (1,0,3), (2,0,0), (2,0,1), (2,0,2), and (2,0,3) models.

4.3 Estimation and Parameter Testing of ARIMA Models

We obtained several model identifications for rainfall and number rainy days as in Tables 9 and 10,

Table 9. Tentative of ARIMA Models of Rainfall post-transformation

Model	Coefficient	Std Error	*p*-value	*AIC*	Decision	Conclusion
ARIMA (1,1,0)	AR (1) = -0.508	0.079	9.734×10^{-11}	419.067	H_0 rejected	Significant (***)
ARIMA (1,1,1)	AR (1) = 0.058 MA(1) = -0.999	0.092 0.027	0.528 $< 2 \times 10^{-16}$	385.787	H_0 accepted H_0 rejected	No Significant Significant
ARIMA (1,1,2)	AR (1) = -9.606×10^{-1} MA(1) = -6.515×10^{-7} MA(2) = -9.999×10^{-1}	3.193×10^{-2} 3.979×10^{-2} 3.979×10^{-2}	$< 2 \times 10^{-16}$ 1 $< 2 \times 10^{-16}$	386.881	H_0 rejected H_0 rejected H_0 rejected	Significant No Significant Significant
ARIMA (1,1,3)	AR(1) = 0.165 MA(1) = -1.110 MA(2) = 0.182 MA(3) = -0.072	1.115 1.110 1.045 0.116	0.882 0.317 0.862 0.534	389.067	H_0 rejected H_0 rejected H_0 rejected H_0 rejected	No Significant No Significant No Significant No Significant
ARIMA (2,1,0)	AR(1) = -0.656 AR(2) = -0.290	0.087 0.087	6.062×10^{-14} 0.000887	**410.533**	H_0 rejected H_0 rejected	**Significant** **Significant**
ARIMA (2,1,1)	AR(1) = 0.056 AR(2) = 0.078 MA(1) = -0.999	0.092 0.092 0.026	0.547 0.396 $< 2 \times 10^{-16}$	387.069	H_0 rejected H_0 rejected H_0 rejected	No Significant No Significant Significant
ARIMA (2,1,2)	AR(1) = 0.097 AR (2) = 0.076 MA(1) = -1.042 MA(2) = 0.042	1.157 0.116 1.158 1.158	0.933 0.513 0.368 0.971	389.067	H_0 accepted H_0 accepted H_0 accepted H_0 accepted	No Significant No Significant No Significant No Significant

As displayed on Table 9, we can see that model that has a significant parameter value is ARIMA(1,1,0) and ARIMA(2,1,0) models. In addition, based on the significant model parameter values, the best model is selected based on the smallest AIC value. So that it can be concluded that the ARIMA model for the rainfall variable that was chosen as the best model in prediction rainfall in Aceh Barat was the ARIMA(2,1,0) model. Furthermore, the following shows the ARIMA tentative model for the number of rainy days variable in Aceh Barat in Table 10, with non-included constant, since constant is not significant.

Table 10. Tentative of ARIMA Models for Number of Rainy Days

Model	Coefficient	Std Error	*p*-value	*AIC*	Decision	Conclusion
ARIMA (1,0,0)	AR(1) = 0.3053		0.0016	556.217	H_0 rejected	Significant
ARIMA (1,0,1)	AR(1) = 0.1592		0.4988	557.757	H_0 accepted	No Significant
	MA(1) = 0.1653		0.4609		H_0 accepted	No Significant
ARIMA (1,0,2)	AR (1) = -0.6701		2.998×10^{-7}	553.848	H_0 rejected	Significant
	MA(1) = 1.1470		4.296×10^{-16}		H_0 rejected	Significant
	MA(2) = 0.5304		1.375×10^{-5}		H_0 rejected	Significant
ARIMA (1,0,3)	AR(1) = -0.4418		0.0909	552.341	H_0 accepted	No Significant
	MA(1) = 0.8580		0.0006		H_0 rejected	Significant
	MA(2) = 0.2468		0.1457		H_0 accepted	No Significant
	MA(3) = -0.2293		0.0617		H_0 rejected	Significant
ARIMA (2,0,0)	AR(1) = 0.3322		0.001	557.670	H_0 rejected	Significant
	AR(2) = -0.0895		0.3864		H_0 accepted	No Significant
ARIMA (2,0,1)	AR(1) = 0.4669		0.4316	559.412	H_0 accepted	No Significant
	AR(2) = -0.1328		0.5057		H_0 accepted	No Significant
	MA(1) = -0.1351		0.8195		H_0 accepted	No Significant
ARIMA (2,0,2)	AR(1) = -0.8455		5.364×10^{-12}	**549.833**	H_0 rejected	Significant
	AR (2) = -0.5651		2.990×10^{-7}		H_0 rejected	Significant
	MA(1) = 1.2459		$< 2.2 \times 10^{-16}$		H_0 rejected	Significant
	MA(2) = 0.9531		$< 2.2 \times 10^{-16}$		H_0 rejected	Significant
ARIMA (2,0,3)	AR (1) = -0.9501		7.927×10^{-11}	551.520	H_0 accepted	No Significant
	AR (2) = -0.6278		1.266×10^{-9}		H_0 accepted	No Significant
	MA(1) = 1.3880		3.286×10^{-15}		H_0 accepted	No Significant
	MA(2) = 1.1378		4.220×10^{-7}		H_0 accepted	No Significant
	MA(3) = 0.1129		0.5211		H_0 accepted	No Significant

Based on Table 9 can be concluded that ARIMA model for amount of rainy days variable has significant parameter and smallest *AIC* value is ARIMA(2,0,2) model. The following ARIMA(2,1,0) models for rainfall variables are shown in equation (25) and the ARIMA(2,0,2) model for number of rainy day variable which refers to equation (26).

$$Z_t = -0.656Z_{t-1} - 0.290Z_{t-2} \quad (25)$$

$$Z_t = -0.8455Z_{t-1} - 0.5651Z_{t-2} - 1.2459a_{t-1} - 0.9531a_{t-2} \quad (26)$$

4.4 Diagnostic Model

4.4.1 White Noise Test

White noise testing is done using Ljung-Box test which refers to equation 11.

Table 11. Result of white noise testing

Variable	ARIMA Model	χ^2	*p*-value
Rainfall	ARIMA (2,1,0)	0.744	0.388
Number of Rainy Days	ARIMA (2,0,2)	0.094	0.757

As seen on Table 11, the obtained *p*-value for rainfall variable in the amount of 0.388, whereas *p*-value for number of rainy days' variable in the amount of 0.757. By using α = 0.05 can be concluded that residual of ARIMA(2,1,0) model of rainfall and residual of ARIMA(2,0,2) model of number of rainy days variables fulfill white noise assumption

4.4.2 Residual Normality Test

The residual normality test is done on ARIMA(2,1,0) and ARIMA(2,0,2) models using *Kolmogorov-Smirnov* test which refers to equation (12).

Table 12. Results of residual normality test

Variable	ARIMA Model	*D*	*p*-value
Rainfall	ARIMA (2,1,0)	0.124	0.095
Number of Rainy Days	ARIMA (2,0,2)	0.339	2.28 x 10^{-10}

As displayed on Table 12, the obtained p-value for rainfall variable is 0,095 and the number of rainy days' variable is 2.28 x 10^{-10}. By using α = 0.05 it is known that the residuals of the ARIMA(2,1,0) model are normally distributed. Whereas the residual of the ARIMA(2.0,2) model is not normally distributed. Testing for residual normal assumptions in the number of rainy days is not fulfilled allegedly because of the non-constant variance. So it should be suspected that there is a problem with residual heterocedasticity in the ARIMA(2.0,2) model.

4.5 Testing for Effect of ARCH Model

Testing the ARCH effect for the rainfall variable was tested on the residual of ARIMA(2,1,0) model and for the number of rainy days variable tested in the ARIMA(2,0,2) model. This test uses the ARCH-Lagrange Multiplier (ARCH-LM) test which refers to equation (14). The following shows the results of the ARCH-LM test on each variable.

Table 13. Result of ARCH-LM Model Testing

Variable	*Order*	*LM*	*p*-value
Rainfall	4	17.479	0.0005
Number of Rainy Days	4	7.866	0.0489

Based on Table 13, we can see that on lag 4 for rainfall variable is obtained *p*-value 0.0005 and for number of rainy days' variable is 0.0489. These show that with α =0.05 decision that can be recommended is H_0 rejected. So that, we can conclude that there is an ARCH effect or a heteroscedasticity problem in the data. The presence of ARCH and GARCH elements can also be seen based on the ACF and PACF plots of the residual squared model. The following shows the plot of ACF and PACF from the residual squares of the ARIMA(2,1,0) and ARIMA (2,0,2) models.

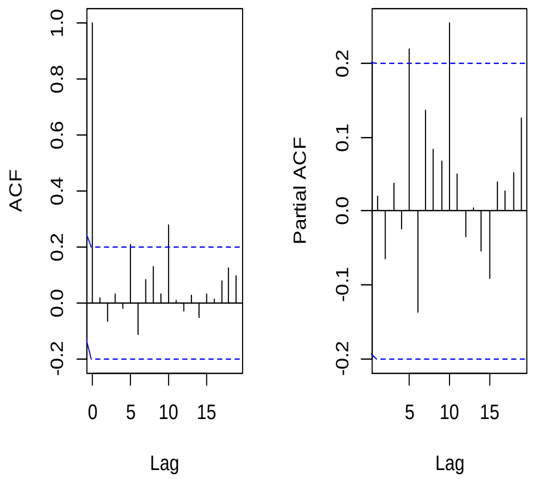

Figure 8. ACF and PACF plots of residual square of rainfall

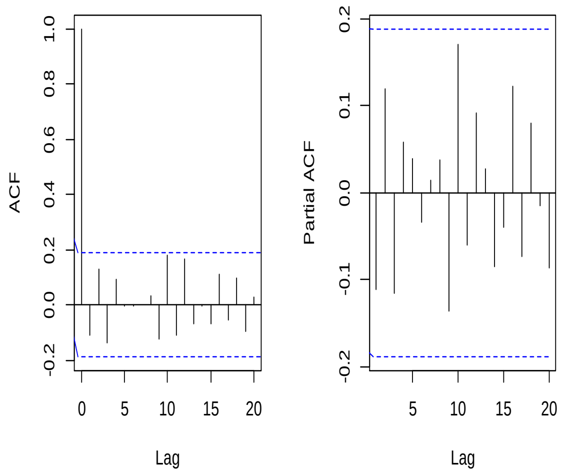

Figure 9. ACF and PACF plots of residual square of number of rainy days

There are lags that cross the Barlett line and Figure 9 shows that there is no lag across the Barlett line. So that, we can be concluded that heteroscedasticity problem only find in the ARIMA(2,1,0) model for rainfall variable. Whereas ARIMA(2,0,2) model for number of rainy day variable is no heteroscedasticity problem.

i) Identification of ARCH Model

Identification of the ARCH-LM model is done by conducting a residual check. Based on the results of the ARCH-LM test in Table 12 it can be concluded that there is an ARCH effect on lag 4. Therefore, the ARCH model is suitable to predict the rainfall and the number of rainy days in Aceh Barat district.

Table 14. ARCH (q) Model

Variable	Order ARCH	Parameter Model					
		α_0	α_1	α_2	α_3	α_4	*AIC*
Rainfall	c(0,1)	$4.669 \times 10^{-1*}$	5.951×10^{-15}	-	-	-	203.172
	c(0,2)	$4.621 \times 10^{-1*}$	2.577×10^{-15}	1.967×10^{-2}	-	-	204.203
	c(0,3)	$3.991 \times 10^{-1*}$	2.768×10^{-15}	$2.254 \times 10^{-}$	1.574×10^{-1}	-	**201.846**
	c(0,4)	$3.941 \times 10^{-1*}$	3.291×10^{-2}	3.747×10^{-2}	9.885×10^{-2}	2.024×10^{-15}	203.659
Number of Rainy Days	c(0,1)	$1.4283 \times 10^*$	4.768×10^{-13}	-	-	-	534.027
	c(0,2)	$1.405 \times 10^*$	1.966×10^{-13}	7.511×10^{-2}	-	-	527.948
	c(0,3)	$1.327 \times 10^*$	1.880×10^{-2}	7.882×10^{-2}	6.614×10^{-14}	-	525.426
	c(0,4)	$1.249 \times 10^*$	1.610×10^{-2}	7.684×10^{-2}	9.151×10^{-15}	5.499×10^{-2}	**522.267**

As seen in Table 12 and based on the smallest AIC value obtained in Table 14 it can be concluded that the best ARCH model for the rainfall variable is the ARCH(3) model. While the best ARCH model for the number of rainy day variable is the ARCH(4) model. So next process, a diagnostic check of the models on ARCH(3) and ARCH(4) is carried out.

4.6 Diagnostic Model ARCH

4.6.1 White Noise Test

White noise test is done on ARCH(3) and ARCH(4) models using Ljung-Box test which refers to equation 11.

Table 15. White Noise Test of ARCH Model

Variable	ARCH	χ^2	*p*-value
Rainfall	c(0,1)	0.617	0.432
Number of Rainy Days	c(0,4)	0.119	0.729

p-value for rainfall variable is 0.432 and for number of rainy day variable, the obtained *p*-value is 0.729. Therefore, by using α=0.05, it can concluded that residual of ARCH model for rainfall and number of rainy days has fulfill white noise assumption.

4.6.2 Residual Normality Test

Residual normality test is done on ARCH(3) and ARCH(4) models by using Kolmogorov-Smirnov test which refers to equation (12).

Table 16. Residual Normality Test of ARCH Model

Variable	ARCH	*D*	*p*-value
Rainfall	c(0,1)	0.064	0.808
Number of Rainy Days	c(0,4)	0.069	0.744

As seen in Table 16, the p-value for the rainfall is 0.808 and the p-value for the ARCH days are 0.7445. So, by using α = 0.05, it can be concluded that the residual of ARCH model and the ARCH model of rainy days have normal distribution. Based on the diagnostic testing of the model carried out by the ARCH(3) and the ARCH(4) models it is suitable to be used to model the rainfall and number of rainy days in Aceh Barat, respectively.

The prediction using ARCH model is done with compound ARIMA and ARCH best model previously obtained, i.e. ARIMA(2,1,0)-ARCH(3) models for rainfall variable which refers to equation (27) and ARIMA(2,0,2)-ARCH(4) models for number of rainy days variable which refers to equation (28) become

$$Z_t = -0.6284Z_{t-1} + (-0.2675)Z_{t-2} + 0.03991 + 2.768 \times 10^{-15}(\varepsilon_{t-1}^2) + 2.254 \times 10^{-3}(\varepsilon_{t-2}^2) + 0.1574(\varepsilon_{t-3}^2) \tag{27}$$

$$Z_t = -0.8455Z_{t-2} - 0.5655Z_{t-2} - 1.245a_{t-1} - 0.953a_{t-2} + 1.249 \times 10 + 1.610 \times 10^{-2}\varepsilon_{t-1}^2 + 7.648 \times 10^{-2}\varepsilon_{t-2}^2 + 9.151 \times 10^{-14}\varepsilon_{t-3}^2 + 5.499 \times 10^{-2}\varepsilon_{t-4}^2 \tag{28}$$

4.7 Identification of GARCH Model

In [18] used GARCH type of models for the ability of volatility forecasting. The research uses daily spot prices of crude oil of Brent and WTI spanning January 4, 1993 to December 31, 2008. The out-of-sample forecasting accuracy is estimated for 5, 20, 60, and 100 day horizons. The results indicate that in the case of Brent crude oil prices, the standard short memory GARCH normal and student-t models outperform for the 5-days and 20-days horizon forecasts and GARCH models that account to asymmetric reaction of oil volatility to price change perform better at longer horizons. Thus, a single model is not uniformly superior to predicting changes in oil price volatility. The rainfall and number of rainy days can be constructed as a GARCH(p,q) model as in Table 17.

Table 17. The GARCH(p,q) Model

Variable	**Order GARCH**	**Parameter Model**							***AIC***
		$\boldsymbol{\alpha_0}$	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\alpha_3}$	$\boldsymbol{b_1}$	$\boldsymbol{b_2}$	$\boldsymbol{b_3}$	
Rainfall	c(1,1)	4.442×10^{-1}	1.463×10^{-15}	-	-	5.715×10^{-2}	-	-	205.169
	c(1,2)	4.293×10^{-1}	7.300×10^{-15}	2.869×10^{-2}	-	6.960×10^{-2}	-	-	206.194
	c(1,3)	3.782×10^{-1}	7.782×10^{-16}	1.083×10^{-2}	1.505×10^{-2}	4.127×10^{-2}	-	-	**204.149**
	c(2,1)	4.200×10^{-1}	6.729×10^{-15}	-	-	5.718×10^{-2}	6.418×10^{-2}	-	206.024
	c(2,2)	4.042×10^{-1}	7.151×10^{-15}	1.548×10^{-2}	-	7.025×10^{-2}	$6,419 \times 10^{-2}$	-	208.138
	c(2,3)	3.524×10^{-1}	3.479×10^{-3}	3.374×10^{-15}	1.555×10^{-1}	3.843×10^{-2}	5.389×10^{-2}	-	529.493

	c(3,1)	3.968×10^{-1}	1.586×10^{-16}	-	-	5.817×10^{-2}	6.404×10^{-2}	5.591×10^{-2}	206.869
	c(3,2)	3.784×10^{-1}	1.954×10^{-15}	2.305×10^{-2}	-	6.802×10^{-2}	6.223×10^{-2}	6.153×10^{-2}	209.033
	c(3,3)	3.287×10^{-1}	2.279×10^{-16}	8.220×10^{-3}	1.513×10^{-2}	4.099×10^{-2}	5.495×10^{-2}	4.764×10^{-2}	207.992
Number of Rainy Days	c(1,1)	1.405×10	9.374×10^{-14}	-	-	6.015×10^{-2}	-	-	535.994
	c(1,2)	1.327×10	2.832×10^{-13}	7.455×10^{-2}	-	4.379×10^{-2}	-	-	530.011
	c(1,3)	1.249×10	2.000×10^{-2}	7.669×10^{-2}	1.510×10^{-13}	5.399×10^{-2}	-	-	**527.533**
	c(2,1)	1.327×10	1.412×10^{-13}	-	-	5.575×10^{-2}	5.193×10^{-2}	-	530.927
	c(2,2)	1.249×10	1.734×10^{-14}	7.458×10^{-2}	-	4.522×10^{-2}	4.721×10^{-2}	-	532.003
	c(2,3)	1.171×10	1.689×10^{-2}	7.745×10^{-2}	2.707×10^{-13}	5.415×10^{-2}	5.491×10^{-2}	-	529.492
	c(3,1)	1.249×10	7.901×10^{-14}	-	-	5.815×10^{-2}	5.365×10^{-2}	5.664×10^{-2}	528.118
	c(3,2)	1.171×10	7.529×10^{-15}	7.804×10^{-2}	-	4.612×10^{-2}	4.820×10^{-2}	5.055×10^{-2}	529.193
	c(3,3)	1.093×10	1.670×10^{-2}	7.766×10^{-2}	3.913×10^{-14}	$5.377\ 10^{-2}$	5.466×10^{-2}	5.668×10^{-2}	531.449

As displayed in Table 16, based on the smallest AIC value it can be concluded that the best GARCH model for the rainfall variable and for the number of rainy day variable is the GARCH(1,3) model. So, a diagnostic check of the model at GARCH(1,3) is carried out.

4.8 Diagnostic Model GARCH

4.8.1 4.7.1. White Noise Test

The white noise test has done in GARCH(1,3) model to rainfall and amount of rainy days variables by using Ljung-Box test.

Table 18. White Noise Test of GARCH Model

Variable	GARCH	χ^2	p-value
Rainfall	c(1,3)	0.7372	0.391
Number of Rainy Days	c(1,3)	0.0508	0.821

Based on *p*-value of rainfall variable is 0.391 and for rainy days variable, *p*-value is 0.821. It can be known that the residual model of GARCH rainfall and amount of rainy day variables have fulfill the white noise assumption.

4.8.2 Residual Normality Test

Residual Normality Test of GARCH(1,3) model for rainfall and amount of rainy days variables as follows.

Table 19. Residual Normality Test of GARCH Model

Variable	GARCH	D	*p-value*
Rainfall	c(0,1)	0.0486	0.970
Number of Rainy Days	c(1,3)	0.0722	0.689

It obtained p-value for GARCH model of rainfall is 0.9701 and p-value for GARCH model of rainy day is 0.6894. By using $\alpha = 0.05$, it be known that residual of GARCH model for rainfall and GARCH model of number of rainy days have normal distribution. Whereas prediction using GARCH is done by combining the best ARIMA and GARCH models obtained previously, namely the ARIMA (2,1,0) -GARCH (1,3) models for rainfall variables which can be written as refers in equation 11 and ARIMA(2,0,2)-GARCH(1,3) model for number of rainy days variable which refers to equation 12.

$$Z_t = -0.6284Z_{t-1} - 0.2675Z_{t-2} + \alpha_t 0.3782 + 7.782 \times 10^{-16}(\varepsilon_{t-1}^2) + 1.083 \times 10^{-2}(\varepsilon_{t-2}^2) + 1.505 \times 10^{-16}(\varepsilon_{t-3}^2) + 4.127 \times 10^{-2}\sigma_{t-1}^2 \quad (29)$$

$$Z_t = -0.8455Z_{t-2} - 0.5655Z_{t-2} - 1.245a_{t-1} - 0.953a_{t-2} + 1.249 \times 10 + 2.000 \times 10^{-2}\varepsilon_{t-1}^2 + 7.669 \times 10^{-2}\varepsilon_{t-2}^2 + 1.510 \times 10^{-13}\varepsilon_{t-3}^2 + 5.399 \times 10^{-2}\varepsilon_{t-1}^2 \quad (30)$$

4.9 Prediction of ARIMA, ARIMA-ARCH, and ARIMA-GARCH Models

Prediction is done using the best model of selected rainfall and number of rainy day variables. Testing prediction uses rainfall and number of rainy day data from January 2016 - December 2017. The results of forecasting testing for the rainfall and rainy days are shown in Table 20.

Table 20. Prediction Test

Rainfall				Number of Rainy Days			
Actual Data	ARIMA Prediction	ARIMA ARCH Prediction	ARIMA GARCH Prediction	Actual Data	ARIMA Prediction	ARIMA ARCH Prediction	ARIMA GARCH Prediction
495	421.794	425.121	426.435	15	17	20	21
261	381.249	384.615	385.908	10	14	18	17
271	494.801	498.156	499.450	12	12	16	15
354	261.045	264.400	265.689	16	10	14	14
653	271.118	274.474	275.760	21	13	16	17
303	354.334	357.718	358.974	16	17	21	21
105	652.685	656.039	657.324	7	18	21	22
566	302.837	306.222	307.476	14	11	15	15
170	105.496	108.872	110.135	17	9	13	13

477	565.639	569.004	570.278	19	14	18	18
523	170.342	173.707	174.981	23	17	21	21
262	477.081	480.449	481.720	16	20	24	24
625.9	522.774	526.129	527.413	18	19	22	23
209.7	262.399	265.780	267.038	12	16	20	20
214.6	625.519	628.878	630.158	14	15	18	19
264.8	209.74	213.097	214.379	18	12	17	16
268.2	214.685	218.040	219.324	16	15	18	19
63.1	264.838	268.193	269.477	7	16	21	20
109	268.03	271.444	272.669	6	11	15	15
287.7	63.181	66.561	67.820	13	6	9	10
210	109.214	112.568	113.853	15	8	12	13
545.5	287.657	291.016	292.296	14	4	17	17
364	210.371	213.743	215.010	21	14	18	17
573.9	545.354	548.709	549.993	17	16	20	20

As seen in Table 19 prediction testing for rainfall and number of rainy days' variables, visually can be seen in compound prediction models ARIMA, ARIMA-ARCH, dan ARIMA-GARCH as in Figure 8 and 9, respectively.

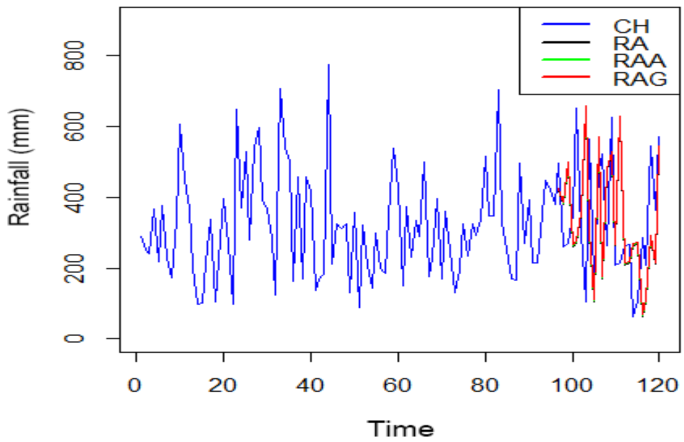

Figure 10. Prediction of rainfall models
(CH=rainfall, RA=ARIMA, RAA=ARIMA-ARCH, RAG=ARIMA-GARCH)

As displayed in Figure 10, it can be seen that for rainfall prediction by using ARIMA, ARIMA-ARCH, and ARIMA-GARCH models have values that are close to the actual data. The following shows a combined plot of prediction number of rainy days using the ARIMA, ARIMA-ARCH, and ARIMA-GARCH models.

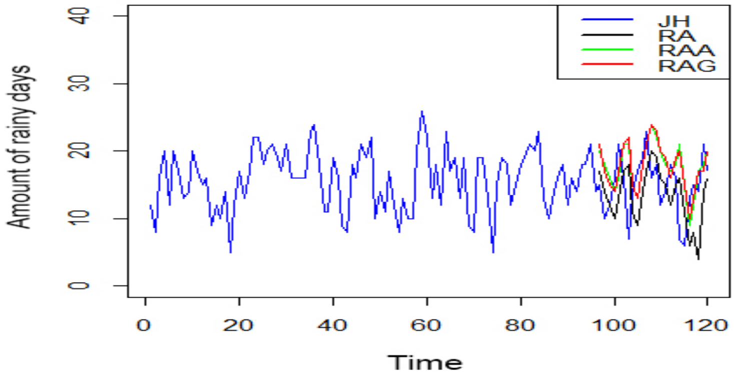

Figure 11. Compound plots of number of rainy days prediction (JH=amount of rainfall)

Based on Figure 11, it can be seen that for prediction of the number of rainy days using the ARIMA, ARIMA-ARCH, and ARIMA-GARCH models, there are several values that are close and there are also values that are far from the actual data. The following is a test of prediction accuracy by calculating the values of MAD, RMSE, MAE and MASE. This value is obtained from rainfall data and number of rainy days from January 2016 - December 2017. The results of the calculation of these values are shown in Table 21.

Table 21. Forecasting Accuracy

Model	Variable	Measurement			
		MAD	RMSE	MAE	MASE
ARIMA	Rainfall	1.778	1.316	1.073	0.821
	Number of Rainy Days	**4.448**	**3.849**	**3.189**	**0.737**
ARIMA-ARCH	Rainfall	1.510	1.165	0.948	0.726
	Number of Rainy Days	4.992	3.859	3.199	0.739
ARIMA-GARCH	Rainfall	**1.175**	**1.163**	**0.941**	**0.720**
	Number of Rainy Days	4.623	4.072	3.315	0.766

Table 21 shows that to predict rainfall variable using ARIMA(2,1,0) model has MAD value of 1.778, RMSE value of 1.316, MAE value of 1.073 and MASE 0,821. Interestingly, using ARIMA(2,1,0)-ARCH(3), model is obtained MAD value of 1.510, RMSE of 1.165, MAE value of 0.948 and MASE 0.726. Whereas using ARIMA(2,1,0)-GARCH(1,3) model, we obtained MAD value of 1.175, RMSE of 1,163, MAE of 0.941 and MASE 0,720. In other words, the smallest value of MAD, RMSE, MAE and MASE for rainfall variable is using ARIMA-GARCH model.

For rainy days variable in ARIMA(2,0,2) model has MAD of 4.448, RMSE value of 3.849, MAE value of 3.189 and MASE 0.737, using ARIMA(2,0,2)-ARCH(4) MAD value of 4.992, RMSE value of 3.859, MAE value of 3.199 and MASE 0.739, and using ARIMA(2,0,2)-GARCH(1,3) model is obtained MAD value of 4.623, RMSE of 4.072, MAE of 3.315 and MASE 0.766. MASE shows under 1, implies that actual forecast performance better than a naïve method [29].

The best prediction can be obtained through the best selection model with the accuracy of the smallest value of MAD, RMSE, MAE and MASE. Based on the results of the prediction accuracy test on Table 21, it can be concluded that the best rainfall uses ARIMA(2,1,0)-GARCH(1,3) model and the best prediction number of rainy day using ARIMA(2,0,2) model.

5 CONCLUSION

Model selected in rainfall prediction in the Aceh Barat district is ARIMA(2,1,0), ARIMA(2,1,0)-ARCH(3) and ARIMA(2,1,0)-GARCH(1,3) models and for prediction of number of rainy days by using ARIMA(2,0,2), ARIMA(2,0,2)-ARCH(4), and ARIMA(2,0,2)-GARCH(1,3) models. Rainfall prediction is more appropriate using ARIMA(2,1,0)-GARCH(1,3) model, whereas for number of rainy days more appropriate using ARIMA(2,0,2) model.

Climate change is a natural phenomenon that can impact the earth's life either directly or indirectly. Future climate change impacts will occur, such as increased rainfall, tropical storm intensity, prolonged forest fires, and droughts in some regions. Due to these impacts specifically rainfall to the climate change, here we proposed ARIMA mixed models to predict rainfall and number of rainy days in Aceh Barat district, Indonesia.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Statistics, Department of Physics, Faculty of Mathematics & Sciences, Faculty of Fisheries & Marine, Institute for Research & Community Service USK, Department of Meteorology, Indonesian School for Meteorology Climatology, Faculty of Ocean Engineering, Technology and Informatics, Universiti Malaysia Terengganu & Geophysics & Directorate of Research and Community Service Ristekdikti Jakarta. 2019.

REFERENCES

[1] A. Aryo, "The overview of rainfall anomaly in northern Sumatra," Master Thesis, Bandung Institute of Technology, Bandung, Indonesia, 2011.

[2] Badan Pusat Statistik Kabupaten Aceh Barat, "Aceh Barat Regency in Figures". Provinsi Aceh: BPS of Aceh Barat Agency, Banda Aceh, 2016.

[3] Badan Meteorologi Klimatologi dan Geofisika. "Prakiraan Curah Hujan", Indonesia : BMKG, Jakarta, 2016.

[4] R. Engle, "Autoregressive conditional heteroscedasticity with estimate of the variance of United Kingdom inflation", *Econometrica: Journal of the Econometric Society*, vol. 15, no. 4, pp. 987–1007, 1982.

[5] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity", *Econometrica: Journal of the Econometric Society*, vol. 31, pp. 307-327, 1986.

[6] R. Engle, "The use of ARCH/GARCH models in applied econometrics", *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 157-168, 2001.

[7] Sunardi, Anton Yudhana, and Ghufron Zaida Mufih, “Sistem prediksi curah hujan bulanan menggunakan jaringan saraf tiruan backpropagation”*, Jurnal Sistem Informatika Bisnis,* vol. 02, pp. 155-162, 2020.

[8] O. D. Anderson, *Time Series Analysis and Forecasting the Box-Jenkins Approach*. London: Butterworths, 1976.

[9] Box, E.P. George, and R. Cox. David "An analysis of transformations", *Journal of the Royal Statistical Society*, vol. 26, no. 2, pp. 211-243, 1964.

[10] Box, G. E. P. and Jenkins, G.M. *Time series analysis, Forecasting, and Control,* Second edition, California: Holden-Day, Oakland, 2005

[11] R. H. Shummay and D. S. Stoffer, *Time Series Analysis and Its Application with R Example, Second edition*, New York: Springer-Verlag, 2006.

[12] Wiliam W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods.* Boston: Pearson Addison Wesley, 2006.

[13] S. Makridakis, S.C. Wheelwright and C.E. McGee, *Metode dan Aplikasi Peramalan, Edisi Kedua*, Jakarta: Erlangga, 1999.

[14] C. Brooks, *Introductory Econometrics for Finance*, 2nd ed., New York: Cambridge University Press, 2008.

[15] D.N. Gujarati,. *Ekonometrika Dasar. Alih Bahasa Sumarno Zain*, Jakarta: Erlangga, 1997

[16] R. F. Eagle, “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation”, *Econometrical Journal of the Econometric Society*, pp 987-1007, 1982.

[17] J.J.M. Moreno, A.P. Pol, A.S. Abad and B.C. Blasco, “Using the R-MAPE index as a resistant measure of forecast accuracy”, *Psicothema,* vol. 25, no. 4, pp. 500-506, 2013.

[18] C. W. Cheong, “Modeling and forecasting crude oil markets using ARCH-type models”, *Energy policy*, vol. 37, no. 6, pp. 2346-2355, 2009.

[19] P. Sadorsky, “Modeling and forecasting petroleum futures volatility”, *Energy Economics*, vol. 28, no. 4, pp. 467-488, 2006.

[20] P. Agnolucci, “Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models”, *Energy Economics,* vol. 31, no. 2, pp. 316-321, 2009.

[21] M. Marzo and P. Zagaglia, “Volatility forecasting for crude oil futures”, *Applied Economics Letters*, vol. 17, no. 16, pp. 1587-1599, 2010.

[22] A. Hou and S. Suardi, “A nonparametric GARCH model of crude oil price return volatility”, *Energy Economics*, vol. 34, no. 2, pp. 618-626, 2012.

[23] R. A. Ahmed and A. B. Shabri, “Fitting GARCH models to crude oil sport price data”. *Life Science Journal*, vol. 10, no. 4, 2013.

[24] Herrera, Ana Maria, Hu, Liang, and Pastor, Daniel, "Forecasting crude oil price volatility", *International Journal of Forecasting*, vol. 34, no. 4, 2014.

[25] S. Gunay, "Markov regime switching GARCH model and volatility modeling for oil returns", *International Journal of Energy Economics and Policy*, vol. 5, no. 4, 2015.

[26] L. C. Hsu and C. H. Wang, "Forecasting integrated circuit output using multivariate grey model and grey relational analysis", *Expert Systems with Applications*, vol. 36, no. 2, pp. 1403-1409, 2009.

[27] O. B. Shukur and M. H. Lee, "Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA", *Renewable Energy*, vol. 76, pp. 637-674, 2015, doi:10.1016/j.renene.2014.11.084.

[28] W. C Wang, K. W. Chau, C. T. Cheng and L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series", *Journal of Hydrology*, vol. 374, no. 3-4, pp. 294-306, 2009.

[29] R. B. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy", *International Journal of Forecasting,* vol. 22, no. 4, pp. 679-688, 2006.

[30] Syamsudin Noor, "BMKG Stasiun Meteorologi Kelas II", *Buletin Meteorologi*, vol. 5, no. 10, 2017.