Comparisons among rainfall prediction of monthly rainfall basis data in Aceh using an autoregressive moving average

I Ramli1*, S Rusdiana2, AYulianur3, A Achmad4

1: Agricultural Engineering Department, Universitas Syiah Kuala, Indonesia
2: Department of Mathematics and Natural Science, Universitas Syiah Kuala, Indonesia
3: Civil Engineering Department, Engineering Faculty, Universitas Syiah Kuala, Indonesia
4: Urban and Regional Planning Department, Engineering Faculty, Universitas Syiah Kuala, Indonesia

*Email: ichwana.ramli@unsyiah.ac.id

Abstract. Climate variability especially rainfall is an important factor in observing climate change. Extreme weather events can disrupt the rice planting calendar system which ultimately causes crop failure. The planting season for each region depends on the rainfall of an area. The purpose of this study is to determine the comparison of the predicted rainfall results in Sabang, Aceh Besar, and Aceh Tengah Municipality. Historical data for 1988-2015 were taken from BMKG. The ARIMA model was divided into three groups, namely autoregressive (AR), moving average (MA), and ARIMA (autoregressive moving average) models. The best prediction model was seen from the model that has the smallest standard error estimate (S) value. The results showed that the three regions have different rainfall than the seasonal plot visualization. The highest peaks of rainfall occurred on October and November in Aceh Besar, on November and December in Aceh Tengah, and only on December in Sabang (Weh Island). The best model for predicting rainfall for January 2016-December 2020 in Sabang was ARIMA (1,0,0)(2,0,0) [12]. The prediction model for the location of Aceh Besar in the future was ARIMA (1,0,0)(2,0,0) [12]. In Aceh Tengah, the best model for predicting January 2016-December 2020 rainfall was ARIMA (1,0,0)(2,0,2) [12]. Forecasting using ARIMA model was good for short-term forecasting, while for long-term forecasting, the accuracy of the forecasting was not good because the trends of rainfall was flat.

1. Introduction
Intensity information and rainfall variability can help resolving several problems related to agriculture, climate change, and natural hazards such as floods and droughts. High and low rainfall that can cause flooding and other times when drought occurs. The occurrence of drought that lasts a long time will result in a decrease in agricultural productivity in addition to causing fires. Time series analysis is important for water resources to estimate hydrological events, and detect trends and shifts in hydrological data [1].

The spread of rainfall to all regions is not the same. Rainfall in highland and coastal areas is relatively higher. To find out the upcoming rainfall a forecasting is carried out [2] has predicted rainfall by using a
vector autoregressive model (VAR). Forecasting results obtained for the coefficient of determination of 66%. A forecasting requires a high level of accuracy, so an analysis is carried out using a method to get the best results, so that the level of accuracy can be maintained. Various attempts have been made to predict rainfall behavior patterns, one of which is using autoregressive integrated moving average model (ARIMA). The ARIMA model includes linear statistical techniques for time series modeling and easily developed rainfall estimates [3].

Research by Rahman et al. used a comparative study of ANFIS and ARIMA model for the weather forecast in Dhaka city and found that ARIMA model performs better than ANFIS [4] [5] Successfully used ARIMA model for predicting rainfall trend of Jordan. Therefore, this paper aims to forecast rainfall in several regions, there are Aceh Besar, Aceh Tengah and Sabang by using Autoregressive Integrated Moving Average (ARIMA).

2. Methodology

The research location is the province of Aceh with the locations of Sabang, Aceh Besar and Aceh Tengah (Figure 1). The rainfall data used in this study is the rainfall data from January to December for 27 years period (1988-2015) from each Climatology and Geophysics Agency of Aceh Besar Climatology station in the research location. Historical data can be used to find relationships, trends, and systematic data patterns [6]. In the data forecasting process, time series data is used (time series). Time series data analysis has a purpose that is to find out changes in data, events or variables and to find patterns of the data so that based on these data pattern, it can be predicted events that will occur. The ARIMA method is processed with R software used to predict rainfall in 2016-2018.

![Figure 1. Location of the study area research](image-url)

The steps taken to predict rainfall with ARIMA are identification, assessment and testing and application of the model. A time series that is not stationary must be converted into stationary data by
performing differencing. If it’s not stationary then differencing must be done again. If the data is not stationary, a logarithmic transformation is carried out. To recognize the existence of seasonal factors, one must look at high autocorrelation.

The classification of the model by using assessment and examination aims to discover if the model is qualified. The Box-Jenkins Model (ARIMA) divided by autoregressive (AR) model, moving average (MA), and ARIMA (autoregressive moving average) combined models that have the specifications of Autoregressive Model (AR) models and Moving Average Models (MA). The usual shape of the autoregressive model with the order \( p \) (AR \((p)\)) or the ARIMA model \((p, 0, 0)\) come from \( \mu \) (constant), \( \epsilon_t \) (error value), \( \phi_p \) (p-autoregressive parameter at t) as shown in (1) as \([7], [8]\)

\[
X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t
\]  

(1)

Where the usual shape of the moving average model of the order \( q \) (MA \((q)\)) or ARIMA \((0,0,q)\): \( \mu \)' = a constant, \( \theta_1 \) to \( \theta_q \) are standard of moving average, \( \epsilon_{t-k} \) (the error value at \( t - k \)) is shown in (2) as

\[
X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}
\]  

(2)

The usual model for pure AR (1) and MA (1) combined processes, eg ARIMA (1,0,1) is declare as:

\[
X_t = \mu + \phi_1 X_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} \text{ or } (1-\phi_1 B)X_t = \mu + (1-\theta_1 B)\epsilon_t
\]  

(3)

If non-stationarity is added to the ARMA process mix, then the general ARIMA \((p, d, q)\) model is fulfilled. The equation for the simple ARIMA case \((1,1,1)\) is as follows:

\[
(1-B)(1-\phi_1 B)X_t = \mu + (1-\theta_1 B)\epsilon_t
\]  

(4)

In general, ARIMA is denoted as ARIMA \((p, d, q)\) where \( p, d, \) and \( q \) sequentially represent the order of auto-regression, differencing, and moving average. Whereas for ARIMA models with seasonal data (SARIMA) expressed by ARIMA notation \((p, d, q)\) \((P, D, Q)\) \( s \), where \( p, d, q = \) non-seasonal parts of the model, \( P, D, Q = \) the seasonal part of the model and \( s \) represents the seasonal period. The thing to note is that most periodic series are non-stationary and that the aspects of AR and MA of the ARIMA model only concern with a periodic series that is stationary.

The prediction model will be going into trial as to see for its accuracy or validation by utilizing Nash-Sutcliffe efficiency (NSE is normalized statistic that decide the relative magnitude of the measured data variance), PBIAS (the deviation of data being evaluated) and RSR (calculated as a ratio of the RMSE and standard deviation of measured data \([9]; [10]\). 

\[
RSR = \frac{\sqrt{\sum_{i=1}^{n}(Y_{obs}-Y_{sim})^2}}{\sqrt{\sum_{i=1}^{n}(Y_{obs}-Y_{mean_{obs}})^2}}
\]  

(5)
PBIAS = \frac{\sum_{i=1}^{n}(Y_{obs} - Y_{sim}) \times 100}{\sqrt{\sum_{i=1}^{n}(Y_{obs})^2}} \tag{6}

NSE = 1 - \frac{\sum_{i=1}^{n}(Y_{obs} - Y_{sim})^2}{\sum_{i=1}^{n}(Y_{obs} - Y_{mean\_sim})^2} \tag{7}

r = \frac{n \sum Y_{obs} \sum Y_{sim} - (\sum Y_{obs})(\sum Y_{sim})}{\sqrt{n(\sum Y_{obs}^2 - (\sum Y_{obs})^2)(n \sum Y_{sim}^2 - (\sum Y_{sim})^2)}} \tag{8}

Where:

- \(Y_{obs}\) = measured rainfall data
- \(Y_{sim}\) = Prediction rainfall data
- \(Y_{mean\_obs}\) = measured rainfall mean data
- \(Y_{mean\_sim}\) = Prediction rainfall mean data

General performance ratings for recommended statistic for accuracy model (Table 1).

| Performance      | RSR          | PBIAS        | NSE           | R      |
|------------------|--------------|--------------|----------------|--------|
| Very good        | 0.0 ≤ RSR ≤ 0.50 | PBIAS < ±10   | 0.75 ≤ NSE ≤ 1.00 | 0.8 ≤ 1 |
| Good             | 0.50 < RSR ≤ 0.60 | ±10 ≤ PBIAS < ±15 | 0.65 < NSE ≤ 0.75 | 0.6 ≤ 0.79 |
| Satisfactory     | 0.60 < RSR ≤ 0.70 | ±15 ≤ PBIAS < ±25 | 0.50 < NSE ≤ 0.65 | 0.49 ≤ 0.59 |
| Unsatisfactory   | RSR > 0.70   | PBIAS ≥ ±25   | NSE ≤ 0.50     | < 0.3  |

3. Result and Discussion

Descriptive analysis and boxplot of rainfall at the location of Aceh Besar each month is unstable and fluctuating. The highest peak of rainfall at the location of Aceh Besar occurred in November. At the location of Aceh Tengah, the peak point (peak season) the highest rainfall occurred in November and December. While rainfall at the location of Sabang (Weh Island) every month is also unstable and fluctuates with the highest peak of rainfall at location Sabang occurring in December figure 2.
Based on the characteristics of the average rainfall at the location of Aceh Besar in 1988-2015 it was 123.7 mm. The lowest rainfall level at the location of Aceh Besar is 0.6 mm, while the highest rainfall is 639.5 mm. The standard deviation of rainfall at the location of Aceh Besar is 87.2 mm. The average rainfall at the location of Aceh Tengah (Takengon) in 1988-2015 was 160.3 mm. The lowest rainfall level at the location of Aceh Tengah (Takengon) is 0.80 mm, while the highest rainfall is 615.60 mm. The standard deviation of rainfall at the location of Aceh Tengah (Takengon) is 116.48 mm. The average rainfall at location C in 1988-2015 was 183.68 mm. The lowest rainfall level at the location of Sabang is 1.0 mm, while the highest rainfall is 723.0 mm. The standard deviation of rainfall at the location of Sabang is 136.92 mm.

Forecasting by using arima begins with testing the data stationarity with an augmented dickey-fuller test. ACF and PACF plots of rainfall data for locations in Aceh Besar, Takengon and Sabang can be taken. Because of the p-value 0.01 < α (0.05), the rainfall data at the location of Aceh Besar, Aceh Tengah and Sabang are stationary. Seasonal data has a tendency to repeat patterns of movement in the seasonal period, the seasonal ARIMA model is an ARIMA model used to complete seasonal time series consisting of two parts, namely non-seasonal (non-seasonal) and seasonal parts.

Forecasting model based on auto.arima function, obtained the best model for predicting January 2016-December 2020 rainfall is ARIMA (0,0,0) (2,0,0)[12]. The AIC value of the ARIMA (0,0,0) (2,0,0) [12] model obtained is 3913.95, AICc is 3914.07 and BIC is 3929.22. In addition, the forecasting model obtained has also been significant. This can be seen from the p-value of each parameter SAR (1) and SAR (2) smaller than α = 0.05. The results of testing the significance of the parameters indicate that the p-value of the two parameters is smaller than α = 0.05. Therefore, the best model for predicting the amount of rainfall at the location of Aceh Besar going forward is ARIMA (0,0,0) (2,0,0) [12].

The best model for predicting rainfall at the location of Takengon in January 2016-December 2020 is ARIMA (1,0,0) (2,0,2) [12]. The AIC value of the ARIMA model (1,0,0) (2,0,2) [12] obtained is 4104.01, AICc is 4104.07 and BIC of 4130.73. In addition, the forecasting model obtained has also been significant. This can be seen from the p-value of each parameter AR (1), SAR (1), SAR (2), SMA (1), and SMA (2) smaller than α = 0.05. The results of testing the significance of the parameters indicate that the p-value of the two parameters is smaller than α = 0.05. Therefore, the best model for predicting the amount of rainfall at the location of Sabang in the future is ARIMA (1,0,0) (2,0,0) [12].

The best model for predicting January 2016-December 2020 rainfall is ARIMA (1,0,0) (2,0,0)[12]. The AIC value of the ARIMA model (1,0,0) (2,0,0)[12] obtained is 4212.27, AICc is 4212.45 and BIC is 4231.36. In addition, the forecasting model obtained has also been significant. This can be seen from the p-value of each parameter AR (1), SAR (1), SAR (2) smaller than α = 0.05. The results of testing the significance of the parameters indicate that the p-value of the two parameters is smaller than α = 0.05. Therefore, the best model for predicting the amount of rainfall in the future Sabang location is ARIMA (1,0,0) (2,0,0) [12].
The Diagnostic Test Model was carried out through the Normality test, the white Noise test. The residual normality assumption was tested using Kolmogorov-Smirnov. The test results show that p-value (1)> α (0.05), therefore H_0 is accept this means that the ARIMA model residuals for the three different locations have normal distribution. A model is said to be white noise if the residuals of the model have met identical assumptions (homogeneous residual variations) and are independent (between residuals not correlated).

Testing the white noise assumption is done by using the Ljung-Box Test. The results of testing in Aceh Besar indicate that p-value (0.09)> α (0.05), so that the residual ARIMA model (0,0,0) (2,0,0) \cite{12} meets the white noise requirement. The white noised test results in Aceh Tengah shows that p-value (0.357)> α (0.05), so that the residual ARIMA model (1,0,0) (2,0,2) \cite{12} meets the white noise requirements. The results of the white noise test at the location of Sabang indicate that p-value (0.2035)> α (0.05), so that the residual ARIMA model (1,0,0) (2,0,0) \cite{12} meets the white noise requirements. The following is a visualization of the residual plot and ACF plot of the rainfall prediction model at the locations of Aceh Besar (a), Aceh Tengah (b) and Sabang (c) as seen in the figure 3.

**Figure 3.** Residual plot and ACF plot of the rainfall prediction model

Visualization in the figure indicates that the residual is normally distributed and the model residual has white noise. This is indicated through a lag line that does not cross the boundary line much on the ACF plot and the normal distribution curve. Therefore the model used to predict rainfall at the locations of Aceh Besar, Aceh Tengah, Sabang for the period January 2016-December 2020 respectively is (0,0,0) (2,0,0) \cite{12}, ARIMA (1,0,0) (2,0,2) \cite{12}, (1,0,0) (2,0,0) \cite{12}. (Figure 4).

**Figure 4.** Rainfall prediction at the locations of Aceh Besar, Aceh Tengah, Sabang
Model Evaluation base on Performance

PBIAS was chosen for recommendation because it has the ability to show poor model performance clearly [3]. PBIAS values for rainfall tend to be more varied but still in the very good category. Based on previous recommendations, a combination of graphical and statistical techniques without dimensions and error indices should be used for model evaluation. NSE, PBIAS, and RSR quantitative statistics are recommended. Performance ratings for recommended statistics are presented in Table 2. In general, simulation models can be rated as "satisfaction" if NSE> 0.65 and RSR <0.30, PBIAS <15% and if r> 60%. Rainfall is very non-linear in nature and very complicated to predict, due to adverse effects of climate change rainfall pattern has also been changing rapidly [10].

Additional considerations, such as a single event simulation, the quality and quantity of the measured data, consideration of procedures, model calculations, time steps of evaluation, and the scope and breadth of the research area, which affect the results of the model. To illustrate the results of the evaluation of the model produced, the predicted rainfall is compared with the measured rainfall.

Table 2. Evaluation Performance rainfall Prediction

| Location     | Year | Performance | RSR | Pbias | NSE  | r    |
|--------------|------|-------------|-----|-------|------|------|
| Aceh Besar   | 2016 | 0.22        | 0.08| 0.78  | 0.71 |
|              | 2017 | 0.29        | 0.21| 0.18  | 0.40 |
|              | 2018 | 0.25        | 0.07| 0.66  | 0.82 |
|              | 2016-2018 | 0.27 | 0.09| 0.43  | 0.56 |
| Aceh Tengah  | 2016 | 0.19        | 0.08| 0.96  | 0.80 |
|              | 2017 | 0.18        | 0.07| 0.97  | 0.81 |
|              | 2018 | 0.25        | 0.10| 0.68  | 0.53 |
|              | 2016-2018 | 0.21 | 0.08| 0.90  | 0.60 |
| Sabang       | 2016 | 0.27        | 0.05| 0.74  | 0.35 |
|              | 2017 | 0.29        | 0.01| 0.04  | 0.07 |
|              | 2018 | 0.29        | 0.13| 0.81  | 0.31 |
|              | 2016-2018 | 0.28 | 0.03| 0.65  | 0.17 |

RSR is a comparison of the root mean square error (RMSE) with the measured standard deviation of the data [11], the RSR is an evaluation of a statistical model that standardizes the RMSE using observational standard deviations, and combines both the error index and additional information and scaling / normalization factors. RMSE value is less than half then the evaluation of the model formed is in good category, RSR varies from the optimal value of 0, which shows zero RMSE or residual variation means perfect model simulation, up to a large positive value. The lower the RSR, the lower the RMSE, and the better the performance of the simulation model. The results of RSR rainfall data prediction with actual rainfall in this study have a performance with a very good category because the value is <0.5.

Nash-Sutcliffe Efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variant noise compared to the measured data variance [10]. NSE shows how well the observed data plot versus the simulated data matches line 1: 1. Performance from the three study sites varies greatly (very good, good and bad categories). The predictions for three years worth very good for central Aceh locations, unsatisfactory NSE scores occurred in 2017 for the locations of Aceh Besar and Sabang.
The size of the correlation abbreviated as r ranges from –1 to +1, including 0. The greater the value of r (close to number 1), the closer the relationship of the two variables in this case the correlation between predicted rainfall and measured rainfall (Figure 5). This r can be positive, but it can also be negative. Following are the interpretations of the sign on the correlation coefficient. Scatter Diagrams actually can also describe the relationship of 2 variables by describing the relationship in graphical form. But the scatter diagram can only estimate the tendency of the relationship whether positive linear, negative linear or does not have linear correlation.

Figure 5. Relation of rainfall measured and rainfall prediction

The weakness of the Scatter Diagram is that it cannot show precisely and also cannot provide a Quantity number about the strength of the relationship between the two variables studied. The results showed predictive rainfall with measured rainfall had a good and moderate correlation for Aceh Tengah and Aceh Besar locations but correlated poorly at the Sabang location. This application is useful for water harvesting systems for flood management, basin management, tourism to overcome extreme and uncertainty rainfall [12].

Conclusion
a. The highest rainfall in three different study areas, namely in the Greater Aceh Region occurred in October and November, the Central Aceh region in December and the Weh Islands (Sabang) region occurred in December.
b. The best prediction model in January 2016-December 2020 in Sabang was ARIMA (1.0.0) (2.0.0)\textsuperscript{(12)}. The prediction model for the location of Aceh Besar in the future was ARIMA (0.0.0) (2.0.0)\textsuperscript{(12)}. In Central Aceh, the best model for predicting January 2016-December 2020 rainfall was ARIMA (1.0.0) (2.0.2)\textsuperscript{(12)}.

c. The NSE, RSR, Pbias, r of ARIMA accuracy test models on predicted rainfall and actual rainfall have varying performance values, but in general the resulting prediction models are satisfactory. Unsatisfactory performance scores occur in 2017.

References

\[1\] Baranowski P., Krzyszczak J., Sławiński C., Hoffmann H., Kozyra J., Nieróbca A., Siwek K., and Gluza A. 2015 Multifractal Analysis of Meteorological Time Series, Assess Climate Impacts. Climate Res., 65, 39-52

\[2\] Ichwana Ramli, siti rusdiana, Hairul Basri, Agus A Munawar, Via Azelia, 2019, Predicted Rainfall and discharge Using Vector Autoregressive Models in Water Resources Management in the High Hill Takengon, The 11th. Aceh International Workshop and Expo on Sustainable Tsunami Distaster Recovery, Banda Aceh, https://iopscience.iop.org/issue/1755-1315/273/1

\[3\] Gupta, H. V., S. Sorooshian, and P. O. Yapo. 1999. Status of auto- matic calibration for hydrologic models: Comparison with mul- tilevel expert calibration. J. Hydrologic Eng. 4(2): 135-143

\[4\] Rahman M, Islam AHMS, Nadvi SYM, Rahman RM 2013 Comparative study of ANFIS and ARIMA model for weather forecasting in Dhaka. In : Informatics, Electronics & Vision (ICIEV), International Conference on; Dhaka. 2013. p. 1–6

\[5\] Momani P.E., Naill M 2009 Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. Am J Environ Sci. 5:599 – 604. ISSN 1553-345X, Science Publications

\[6\] Makridakis, R.J Hyndman, S.C. Wheelwright, 2009 Forecasting: Methods and Applications in Journal of the Operational Research Society 35(1) DOI: 10.2307/2581936 Edition: 3, Publisher: Wiley

\[7\] Box G.E.P., Jenkins G., and Reinsel G., 2008 Time series analysis. Wiley Press, New Jersey, USA

\[8\] Montgomery, D.C. 2008 Introduction To Time Series Analysis and Forecasting

\[9\] Moriasi, D.N, J.G. Arnold, M.W. Van Liew, R.L Bingner, R.D Harmel, dan T.L Veith 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Jurnal American Society of Agricultural and Biological Engineers. ISSN 0001-2351.50 (3): 885-900

\[10\] Nash JE, dan J.V. Suteliffe 1970. River flow forecasting through conceptual Models 1, a Discussion of Principles, J. Hydrol. 10 (1):282-290

\[11\] Murat M, Iwona Malinowska, Magdalena Gos, and Jaromir Krzyszczak, 2018, Forecasting daily meteorological time series using ARIMA and regression models International Agrophysics, 32, 253-264, doi: 10.1515/intag-2017-0007

\[12\] Singh, J., H. V. Knapp, and M. Demissie. 2004. Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004-08. Champaign, Ill.: Illinois State Water Survey. Available at: www.sws.uiuc.edu/pubdoc/CR/ ISWSCR2004-08.pdf

\[13\] Bari S H, M T Rahman, M M Hussain, Sourav Ray, 2015 Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model Civil and Environmental Research www.iiste.org ISSN 2224-5790 (Paper) ISSN 2225-0514 (Online)Vol.7, No.1.