Monthly Rainfall Components in Ambon City: Evidence from the Serious Time Analysis

Lexy J. Sinay¹, Esther Kembauw²*

¹Department of Mathematics, Pattimura University, Indonesia
²Departement of Agriculture, Pattimura University, Indonesia
Email: lexyjz@gmail.com¹; ekembauw@yahoo.co.id²

*ekembauw@yahoo.co.id

Abstract. The purpose of this study is to prove whether rainfall in Ambon City contains seasonal components based on the characteristics of monthly rainfall history data in Ambon City. The data used is monthly rainfall data in Ambon City 2005.01 - 2017.12. The data is time series data from the BMKG Meteorology Station observation in Maluku Province. The time series analysis method used is ARIMA / SARIMA and Holt-Winter Exponential Smoothing. The results of this study have a good level of accuracy, namely using the results of analysis based on the values of information criteria, Ljung-Box Test, and RMSE. The results of the analysis prove that rainfall in Ambon City has a seasonal pattern.

1. Preliminary

Ambon City is the capital of Maluku Province and is located in the eastern part of Indonesia. Based on the astronomical location of Ambon City has a tropical climate with two seasons namely the rainy season and the dry season. In addition, the climate conditions in Ambon City are also influenced by the geographical conditions of Ambon Island. This condition resulted in Ambon City always experiencing rain every month.

Based on habits, Ambon City people know the rainy season in Ambon City occurs from June to August. In these months rainfall in Ambon City is usually quite high and is often referred to as the east season. This is usually marked by a thunderous roar. Does statistical rainfall in Ambon City have a seasonal pattern? This will be proven using monthly rainfall historical rainfall data in Ambon City. Generally, rainfall data available in Ambon City is the result of observations made by the Geophysical Station and the Meteorological Station, BMKG (2016) [3] Maluku Province in two different locations. The results of the study using rainfall data in Ambon City have been done before. In 2015, Sinay & Aulele [12] modeled the causal relationship between rainfall and the number of rainy days in Ambon City using the Vector Autoregression Model (VAR). The data used in this study is the observation of BMKG Geophysics Station in Maluku Province. In 2017, Kafara et. al [8] use the same data to forecast rainfall using the Seasonal Autoregressive Integrated Moving Average (SARIMA). On the other hand, using data from the BMKG Meteorological Station in Maluku Province, Sinay et. al (2016) [13] modeled and predicted monthly rainfall in Ambon City using the Box-Jenkins method. Then, Sinay et. al (2017) [14] forecast rainfall using the Holt-Wintes Exponential Smoothing Method. Akpanta (2015) [2] In several studies, the seasonal ARIMA (SARIMA) is a suitable model for modeling rainfall data that has seasonal patterns such as in Nigeria and in Sudan Etuk et al (2014) [5], Hassan et al (2015) [6]. These studies focus more on causal relationships, models and predictions of rainfall in Ambon City.
Thus, this study was made different from previous studies. This study aims to prove whether rainfall in Ambon City contains seasonal components based on the characteristics of monthly rainfall history data in Ambon City. Because the data used is time series data, the analysis method applied is the time series method. The time series method used is ARIMA / SARIMA and Holt-Winter Exponential Smoothing. In addition, the data used is the latest data with a longer range than previous studies.

2. Data
The data used is the Ambon City monthly rainfall data for the period January 2005 - December 2017 (2005.01 - 2017.12). This data is the result of observations of the Meteorology, Climatology and Geophysics Agency (BMKG, 2016) [3] of Meteorological Station Class II Pattimura Ambon at coordinates 03°42’25”LS 128°05’23” BT, with an elevation of 15.4 m.

3. Method
This study uses a method that is able to detect seasonal patterns in rainfall. The analytical method used in this study is the ARIMA/SARIMA method and the Holt-Winter Exponential Smoothing method. Both of these methods have the ability to model non-seasonal data and seasonal data with good accuracy.

The analysis procedure carried out in this study is divided into two parts, namely modeling using the ARIMA/SARIMA method and modeling using the Holt-Winter Exponential Smoothing method. The analysis procedure using the ARIMA/SARIMA method is 1) Detecting seasonal patterns using a line plot; 2) Detect data stationary, patterns and seasonal periods using a plot of the Autocorrelation Function / Partial Autocorrelation Function (ACF / PACF); 3) Preparation of the ARIMA / SARIMA model; and 4) Identification of the best model based on information criteria values such as Akaike (1974) [1] Information Criteria \[ AIC = n\log_e(\sigma^2) + 2(p + q + 1) \], Schwarz Bayesian Information Criteria \[ SIC = n\log_e(\sigma^2) + (p + q + 1)\log_e(n) \] and mean of square error [MSE]. In addition, serial correlation checks are performed using Ljung-Box (1978) [9] test statistics \[ Q = n(n + 2)\sum_j \rho(j)^2 / (n - j) \]. The procedure of analysis using the Holt-Winter Exponential Smoothing method is 1) Detect seasonal patterns using line plots; 2) Preparation of the Holt-Winter Exponential Smoothing model based on smoothing parameter values; and 3) Identification of the best model based on the RMSE value. The following are given the basic theories that support this research.

3.1 ARIMA/SARIMA Model
ARIMA model, first introduced by Box-Jenkins in 1970 [4]. The ARIMA model is a combination of the result model of differentiation with \( ARMA(p, q) \) denoted by \( ARIMA(p, d, q) \) where \( p \) is the autoregressive component, \( d \) is the result order of differentiation, and \( q \) is the moving average component. The general form \( ARIMA(p, d, q) \) is written as follows

\[
\Phi_p(B)(1 - B)^dX_t = \theta_q(B)Z_t
\]

\( ARIMA(p, d, q) \) is a stationary model.

The generalized ARIMA model for dealing with seasonal problems is the SARIMA model. Generally this model is denoted by \( ARIMA(p, d, q)(P, D, Q)_s \) or \( SARIMA(p, d, q)(P, D, Q)_s \). The general form of the model is

\[
\Phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t
\]

Where \( B \) is the backward shift operator, \( \Phi_p, \Phi_P, \theta_q, \Theta_Q \) is the polynomial-polynomial with order \( p, P, q, Q \) in sequence, \( Z_t \) is a random process and

\[
W_t = \nabla^d\nabla^D^sX_t
\]
3.2 Model Holt-Winters Seasonal Smoothing

Exponential smoothing data was introduced by Holt (1957) [7] and Winter (1960) [15]. There are two types of models, namely the multiplicative model and the additive model.

1. Multiplicative Model

Assume a time series model
\[ X_t = (\beta_1 + \beta_2 t)S_t + \epsilon_t \]

Where \( \beta_1 \) is a constant, \( \beta_2 \) is a linear trend component, \( S_t \) is a seasonal component with index, and \( \epsilon_t \) is an irregular/random error component. The smoothing equation for the model with three parameters (\( \alpha \), and \( \gamma \)) is given as follows

Level
\[ S_t = \alpha \frac{X_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + T_{t-1}) \]

Trend
\[ T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \]

Sessional
\[ I_t = \gamma \frac{X_t}{S_t} + (1 - \gamma)I_{t-L} \]

Where \( T \) is the trend component, \( I \) is the seasonal adjustment factor, and \( L \) is the length of the period of relaxation. For forecasting \( m \) period is given by equation.

\[ F_{t+m} = (S_t + T_t m)I_{t-L+m} \]

Rosadi. D (2011) [10], Rosadi. D (2012) [11]

2. Additive Model

Assume a time series model
\[ X_t = \beta_1 + \beta_2 t + S_t + \epsilon_t \]

Smoothing equation for the model with three parameters (\( \alpha \), , and \( \gamma \)) that is

Level
\[ S_t = \alpha(X_t - I_{t-L}) + (1 - \alpha)(S_{t-1} + T_{t-1}) \]

Trend
\[ T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \]

Sessional
\[ I_t = \gamma(X_t - S_t) + (1 - \gamma)I_{t-L} \]

The equation for forecasting \( m \) periods with initial values is

\[ F_{t+m} = S_t - mT_t + I_{t+m-L}, m = 1, 2, ..., L \]
\[ F_{t+m} = S_t - mT_t + I_{t+m-2L}, m = L + 1, L + 2, ..., 2L \]

etc. [13], [14]

4. Results And Discussion

The rainfall data used is the result of observations for 13 years with the number of observations being 156 months. Statistically, the data is described in Table 1. Average rainfall indicates that rainfall in Ambon City is classified as moderate. Based on the Jarque-Bera test statistic, the data is not normally distributed at the 95% confidence level.

Table 1. Description of Statistics Of Monthly Rainfall Data In Ambon City Period 2005.01 – 2017.12
|                          |        |
|--------------------------|--------|
| Maximum rainfall         | 1923   |
| Minimum rainfall         | 3      |
| Rate of rainfall         | 312.1603 |
| Standard Deviation       | 329.0676 |
| Normality test:          |        |
| Statistic test Jarque-Bera (JB) | 322.7822 (0.0000) |

Visually, the line plot of monthly rainfall data in Ambon City in the period 2005.01 - 2017.12 is shown in Figure 1. The figure shows that rainfall data in Ambon City does not contain trends and is random. This indicates that rainfall data in Ambon City is stationary in the mean. The highest rainfall occurred in 2013, while the lowest rainfall occurred in 2005. Rainfall fluctuated every year with a high tendency of rainfall to occur from May to August, while in other months rainfall tends to be lower.

**Figure 1.** Actual Monthly Rainfall Data Plot Line in Ambon City Period 2005.01 - 2017.12

In addition to Figure 1, the ACF and PACF plots shown in Figure 2 can be used to identify trends and seasonality. The ACF plot for 36 lags does not slow down to 0, but forms a sinusoidal pattern. Whereas PACF plots do not form a certain pattern. This means that there is no trend, but indicates a
seasonal effect in rainfall data in Ambon City. The results of the ACF plot observations show that 12, 24, 36 lags are positive and are above the limit. Thus, the seasonal period formed is 12 months.

4.1 ARIMA vs SARIMA
It is known that the ARIMA / SARIMA model is based on stationary data. The results of data stationarity analysis state that the results of the transformation of natural logarithms are data that is stationary (Figure 3). Thus, the data is good for modeling ARIMA and SARIMA.

Table 2. 20 Top Arima Models Vs 20 Top Sarima Models

| Model       | AIC   | SIC   | Model       | AIC   | SIC   |
|-------------|-------|-------|-------------|-------|-------|
| (2,4)(0,0)  | 2.4715* | 2.6279 | (3,1)(1,1)  | 2.3450* | 2.5014 |
| (3,4)(0,0)  | 2.4840  | 2.6600 | (2,0)(1,1)  | 2.3545  | 2.4718* |
| (2,3)(0,0)  | 2.4904  | 2.6272* | (2,2)(1,2)  | 2.3604  | 2.5364 |
| (4,4)(0,0)  | 2.4960  | 2.6915 | (1,1)(1,1)  | 2.3624  | 2.4979 |
| (3,3)(0,0)  | 2.5602  | 2.7166 | (2,0)(1,2)  | 2.3628  | 2.4997 |
| (2,2)(0,0)  | 2.5654  | 2.6827 | (4,2)(1,1)  | 2.3632  | 2.5587 |
| (2,1)(0,0)  | 2.5664  | 2.6641 | (2,0)(2,1)  | 2.3637  | 2.5006 |
| (3,1)(0,0)  | 2.5679  | 2.6852 | (2,1)(1,1)  | 2.3671  | 2.5040 |
| (3,0)(0,0)  | 2.5701  | 2.6678 | (3,0)(1,1)  | 2.3672  | 2.5040 |
| (0,2)(0,0)  | 2.5710  | 2.6492 | (1,2)(1,1)  | 2.3686  | 2.5055 |
Table 2 shows 20 ARIMA models and 20 best SARIMA models based on the information criteria values of AIC and SIC. Based on the table, the smallest AIC and SIC values are obtained by the seasonal models ARIMA(3, 0, 1)(1, 1, 1) and ARIMA(2, 0, 0)(1, 1, 1). Based on these results, the SARIMA model is better than the non-seasonal ARIMA model.

Then the serial correlation is examined using Ljung-Box et al (1978) [9] test statistics shown in Table 3. The table compares the Q test statistic for 4 models, namely Model 1 = ARIMA(2,0,4), Model 2 = ARIMA(2,0,3), Model 3 = ARIMA(3, 0, 1)(1, 1, 1) and Model 4 = ARIMA(2, 0, 0)(1, 1, 1). These results state that for lags 21 and so on Model 3 and Model 4 do not contain serial correlation at the 95% confidence level. While Model 1 and Model 2 contain serial correlation at the 95% confidence level. This means that the SARIMA model is better than the non-seasonal ARIMA model.

Table 3. Ljung-Box test statistics

| Lag | Model 1 Q-Stat | p value | Model 1 p value | Model 2 Q-Stat | p value | Model 2 p value | Model 3 Q-Stat | p value | Model 3 p value | Model 4 Q-Stat | p value |
|-----|---------------|---------|----------------|---------------|---------|----------------|---------------|---------|----------------|---------------|---------|
| 21  | 20.9380       | 0.1390  |                |               |         |                |               |         |                |               |         |
| 22  | 33.6110       | 0.0060  | 39.6760        | 0.0340        | 18.8150 | 0.2780         | 20.2410       | 0.3190  |                |               |         |
| 23  | 35.0030       | 0.0060  | 40.8360        | 0.020         | 22.2230 | 0.1760         | 23.0530       | 0.2350  |                |               |         |
| 24  | 35.7220       | 0.0080  | 41.3400        | 0.020         | 23.1060 | 0.1870         | 24.0630       | 0.2400  |                |               |         |
| 25  | 36.0730       | 0.0100  | 41.3610        | 0.0030        | 23.8240 | 0.2030         | 24.7770       | 0.2570  |                |               |         |
| 26  | 38.2220       | 0.0080  | 44.5020        | 0.0020        | 23.8860 | 0.2470         | 24.8730       | 0.3030  |                |               |         |
| 27  | 39.0590       | 0.0100  | 45.7500        | 0.0020        | 24.3150 | 0.2780         | 25.2480       | 0.3380  |                |               |         |
| 28  | 40.8590       | 0.0090  | 48.3650        | 0.0020        | 26.1990 | 0.2430         | 27.0240       | 0.3030  |                |               |         |
| 29  | 43.2040       | 0.0070  | 50.4930        | 0.0010        | 26.8100 | 0.2640         | 27.7000       | 0.3220  |                |               |         |
| 30  | 43.3570       | 0.0090  | 50.6360        | 0.0020        | 27.1060 | 0.3000         | 27.9140       | 0.3630  |                |               |         |
| 31  | 44.2380       | 0.0100  | 52.0930        | 0.0020        | 27.1230 | 0.3500         | 27.9140       | 0.4150  |                |               |         |
| 32  | 44.4780       | 0.0130  | 52.2230        | 0.0020        | 27.2530 | 0.3960         | 28.1620       | 0.4560  |                |               |         |
| 33  | 44.4880       | 0.0180  | 52.2230        | 0.0040        | 29.3740 | 0.3430         | 30.6330       | 0.3830  |                |               |         |
| 34  | 44.5850       | 0.0240  | 52.2430        | 0.0050        | 29.7970 | 0.3730         | 31.2870       | 0.4010  |                |               |         |
| 35  | 44.5930       | 0.0320  | 52.3220        | 0.0070        | 30.0790 | 0.4100         | 31.3970       | 0.4460  |                |               |         |
| 36  | 49.5370       | 0.0140  | 56.6940        | 0.0030        | 30.0790 | 0.4620         | 31.4310       | 0.4950  |                |               |         |
4.2 Holt-Winter (HW)

The results of the analysis of rainfall data in Ambon City using the Holt (1957) [7]-Winter (1960) [15] Exponential Smoothing method are summarized in Table 4. The table shows three types of Holt-Winter models, namely non-seasonal models, seasonal multiplicative models, and seasonal additive models. The three models are based on optimal parameter values.

| Parameter | HW Non Seasonal | HW Multiplicative | HW Additive |
|-----------|-----------------|-------------------|-------------|
| $\alpha$  | 0.6600          | 0.2400            | 0.1300      |
| $\beta$   | 0.0000          | 0.0000            | 0.0000      |
| $\gamma$  | 0.0000          | 0.0000            | 0.0000      |

| Coefficient | Linear | Seasonal |
|-------------|--------|----------|
| $\beta_1$ (Mean) | 197.7427 | 459.3269 | 421.3710 |
| $\beta_2$ (Trend)  | 6.9359   | 240.5519 | 1.4925   |
| RMSE       | 325.8882 | 240.5519 | 243.7276 |

Table 4 shows the RMSE values of the three HW models. Based on the RMSE values, the HW seasonal multiplicative model is better than the other two HW models. This is because the RMSE value of the seasonal multiplicative model is smaller than the two models.

5. Conclusion

Statistically, it was found that the average rainfall in Ambon City was classified as moderate and the data was not normally distributed. Based on the results of the study using time series analysis, it is found that the model that best describes rainfall in Ambon City is a seasonal model. This is based on the results of comparative seasonal models with non-seasonal models, both in ARIMA / SARIMA and Holt-Winter Exponential Smoothing. Thus, it can be concluded that rainfall in Ambon City has a seasonal pattern.
References

[1] Akaike H. A New Look at the Statistical Model Identification. Annals of Statistics: IEEE Trans. Automat. Contr., 1974. Vols. Ac 19. pp. 716-723.

[2] Akpanta, A. C., Okorie, I. E. & Okoye, N. N., SARIMA Modelling of the Frequency of Monthly Rainfall in Umuahia, Abia State of Nigeria, American Journal of Mathematics and Statistics, 2015, vol. 5(2): pp. 82-87, DOI: 10.5923/j.ajms.20150502.05

[3] BMKG, Stasiun Klimatologi Kairatu, Buletin Desember 2016 Provinsi Maluku.

[4] Box, G. E. P. & Jenkins, G. M., 1970. Time Series Analysis: Forecasting and Control. San Fransisco: Holden-Day.

[5] Etuk, E.H., Mohamed T.M., Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by SARIMA Methods, International Journal of Scientific Research in Knowledge, 2014, vol. 2(7), pp. 320-327.

[6] Hassan, H. M. & Mohamed, T. M., 2014, Rainfall Drought Simulating Using Stochastic SARIMA Models for Gadaref, Sudan, University of Khartoum, Khartoum Technical College, MPRA Paper No. 61153, posted 8 January 2015 13:38 UTC, Online at https://mpra.ub.uni-muenchen.de/61153/.

[7] Holt, C.C. (1957). Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. ONR Memorandum, Vol. 52, Carnegie Institute of Technology, Pittsburgh. Available from the Engineering Library, University of Texas, Austin.

[8] Kafara, Z., Rumlawang, F. Y. & Sinay, L. J., Peramalan Curah Hujan Dengan Pendekatan Seasonal Autoregressive Integrated Moving Average (SARIMA), Barekeng: Jurnal Ilmu Matematika dan Terapan, 2017, vol. 11 (1), pp. 63-74.

[9] Ljung, G. M. & Box G. E. P., On a Measure of Lack of Fit in Time Series Models, Biometrika, 1978, vol. 65, pp. 297.

[10] Rosadi, D., 2011. Analisis Ekonometrika & Runtun Waktu Terapan dengan R (Aplikasi untuk bidang ekonomi, bisnis, dan keuangan). Yogyakarta: Andi Offset.

[11] Rosadi, D., 2012. Ekonometrika & Analisis Runtun Waktu Terapan dengan Eviews (Aplikasi untuk bidang ekonomi, bisnis, dan keuangan). Yogyakarta: Andi Offset.

[12] Sinay, L. J. & Aulele, S. N., Rainfall and Number of Rainy Days Prediction in Ambon Island using Vector Autoregression Model, Proceeding 1st International Seminar of Basic Science 2015, pp. 93-101, Faculty of Mathematics and Science, Pattimura University, Ambon

[13] Sinay, L. J., Patty, H. W. M. & Leleury, Z. A., Analisis Model Curah Hujan Di Kota Ambon Menggunakan Metode Box-Jenkins, Prosiding Seminar Nasional Pendidikan Matematika 2016, pp. 176-197, FKIP UNPATTI, Ambon.

[14] Sinay, L. J., Pentury, T. & Anakotta, D., Peramalan Curah Hujan Di Kota Ambon Menggunakan Metode Holt-Winters Exponential Smoothing, Barekeng: Jurnal Ilmu Matematika dan Terapan, 2017, vol. 11 (2), pp. 101-108.

[15] Winter, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science 6 (3), 324-342.