Forecasting Rainfall with Time Series Model

M Sidiq
Magister of Information System, Faculty of Postgraduate, Universitas Komputer Indonesia, Indonesia
maulanasisidiq1304@gmail.com

Abstract. The aim of the study was to forecast Rainfall with Time Series Model. Monthly rainfall of 48 data got from Badan Meteorologi dan Geofisika (BMG) Bandung from January 2011 to December 2013 are processed by a computer program to look at the pattern in minitab model ARIMA. This modelling stage begins from the stationarity test data, identify models, parameter estimation, and model verification, to forecasting. Precipitation data is data that is not stationary so that distinguish first. The data obtained from the model results this first distinction is ARIMA (1, 1, 1) with the MAD 82.712 test of the significance of the parameters of the model.

1. Introduction
Indonesia is tropical country. It has high enough rainfall because its islands are surrounded by a vast ocean which has a fairly high daily temperature and humidity [1] [2]. Currently, there are about 40.6 million hectares of agricultural and plantations areas derived from volcanic activity. Agricultural or plantation can be done as long as there is enough water coming from rainfall [3] [4]. In general, the rainfall follows two kinds of seasons, a dry (April to September) and a rainy season (October to March) [1].

Many studies have reported how to identify forecasting rainfall, as shown by Spessa, Field, Pappenberger, Langner, Englhart, Weber and Moore [5], Qian, Robertson, and Moron [6], Adiwijaya, and UN [7]. Although the model they have been referred to by many research reports, they still have limitations, particularly to identify in detail the rainfall in every province in Indonesia.

Time series is basically a measurement data taken in chronological order within a certain time [8]. The ARIMA (Autoregressive Integrated Moving Average) method is used in this study is because the characteristic of each cascading is stationary (has a mean and constant variance also covariance lag that does not depend on where the calculation is done) [9]. The method is also called the Box-Jenkins method as developed by George Box and Gwilym Jenkins in 1976 [8].

This study will analyze and determine the optimum shape models of ARIMA to estimate the monthly rainfall in the past two years to come. As for the benefits of this research is to expand scientific development in particular the application of mathematics in statistics and mathematical models gain rainfall city of Bandung which can serve as a checklist against an existing model. Data processing to complete Modeling and quantitative forecasting was aided by some software, among others, SPSS, Minitab, and Microsoft Excel [10]. Specifically to do with the analysis of forecasting time in this research, used software minitab because computer software has full facilities for problems of ARIMA. The aim of the study was to forecast Rainfall with Time Series Model.
2. Method
Many studies have reported how to identify forecasting rainfall, as shown by Spessa, Field, Pappenberger, Langner, Englhart, Weber and Moore [5], Qian, Robertson, and Moron [6], Adiwijaya, and UN [7], Valipour [11]. Although the model they have been referred to by many research reports, they still have limitations, particularly to identify in detail the rainfall in every province in Indonesia.

3. Results and Discussion
In this study, the data obtained through the collection of several stages of modeling ARIMA:

3.1. Test Data Stationary
Stationary data is data which has the average and variance that is constant over time.

3.2. Identification of the Model
Identification of the model while a distribution done by comparing the coefficient of autocorrelation and literary sources and documentation BMG Bandung. This data is analyzed using the analysis model of ARIMA time, ranging from the identification of the model, stationary data, the estimation models, forecasting models, and verification. Autocorrelation partial coefficients of actual theoretical distribution with.

3.3. Mean Absolute Deviation (MAD).
MAD is the size of the overall forecasting error for a model. The MAD value is calculated by taking the sum of the absolute values of the forecasting error divided by the number of data periods.

\[
MAD = \frac{\sum |actual \ value - forecasting|}{n}
\]

Where:
\( n \) = number of data periods

3.4. Verification of the Model
Testing the feasibility of a model can be done in several ways:
- Overfitting is done when the model is required.
- Testing the residual (error term).

Systematically residual can be calculated by way of reducing the forecast result data with the original data. The selection of models in the method of ARIMA is done by observing the distribution coefficients of autocorrelation and partial autocorrelation coefficients.

a. Autocorrelation Coefficient. The correlation coefficient indicates the direction and the closeness relationship of the two varieties so that describe what happens on one variable if there are changes in other variables. To test the significance of an autocorrelation coefficient.

b. Partial Autocorrelation. Partial autocorrelation coefficients measure the degree of relationship between the closeness \(X_t\) with \(X_{t-k}\), while the influence of time lag of 1, 2, 3, and so on until the \(k-1\) are considered constant (method of forecasting, 2008).

3.5. Using the selected Models for forecasting
Bandung city rainfall data from January 2011 until December 2013 using minitab software so that obtained results are as follows.
Graph trend of original data showed not stationary data due to data growth along the axis of time (fluctuation data is not stationary in the value of its Center). Likewise the graph of the ACF's original data showed not stationary data because some value autocorrelation at lag time-1 quite significantly from zero.

3.5.1. Stationarity Testing Data. To know the not stationary the data views from a plot and ACF trend data. i.e. rk, significant from zero to quite large. Therefore, after this stage, performed on differencing data in this first distinction so that the retrieved data difference one which its graph is as follows (See Figure 2).

**Figure 1.** The plot of Autocorrelation Function and Trend of original Data.

**Figure 2.** Plot trend, ACF, PACF and Data after in Deference’s.
Graph trend data after in deference’s shows stationary on the value of the data center because the graphics look along the horizontal axis of time. ACF Graph towards zero after a lag. Autocorrelation values after a lag 4 does not differ significantly from zero or is within the limits of an autocorrelation values so that shows the stationary data. In this case, limits the significance value of autocorrelation $r_k$, is $0.397324 \leq r_k \leq 0.397324$. This shows a pattern for time series modeling once at deference’s.

3.5.2. Identification. Identification of the Model While the graph of FAKP gap data shows the value of the partial autocorrelation decreases exponentially from lag lag of 1 to 4. This indicates the existence of a pattern of MA (Moving Average) takes one, MA (1) that is not seasonal. The seasonal pattern is still visible on the autocorrelation values data deference’s so strengthen the presence process MA (1) seasonal. There is one value of a very significant partial autocorrelation at lag 1 so assumed the existence of a pattern of AR (Autoregression) takes one or AR (1) is not seasonal. Based on this, the model is whereas ARIMA (1, 0, 1), (0, 1, 1), (1, 1, 1) $D = 1$, and $s = 12$.

3.5.3. Estimation. Parameter estimation model for estimate parameters in the model, the first step is to elaborate on three models of ARIMA (1, 0, 1), (0, 1, 1), (1, 1.1). The above model involves three parameters, namely 1 as AR (1) is not seasonal, W1 as MA (1) seasonal and not 1 as MA (1) seasonal. With the help of the program, the results of the estimation of minitab three parameters is obtained as follows. Final Estimates of Parameters (1,0,1).

| Type | Coef | SE Coef | T   | P   |
|------|------|---------|-----|-----|
| AR   | 1    | 0.8951  | 0.1085 | 8.25  | 0.000 |
| MA   | 1    | 0.2063  | 0.2116 | 0.97  | 0.336 |

Forecasts from period 36

| Period | Forecast | Lower | Upper | Actual |
|--------|----------|-------|-------|--------|
| 37     | 319,302  | 24,034 | 614,569 |
| 38     | 285,822  | -72,725 | 644,369 |
| 39     | 255,853  | -146,276 | 657,982 |
| 40     | 229,027  | -204,877 | 662,930 |
| 41     | 205,013  | -252,762 | 662,787 |
| 42     | 183,517  | -292,523 | 659,556 |
| 43     | 164,275  | -325,910 | 654,459 |
| 44     | 147,050  | -354,180 | 648,280 |
| 45     | 131,632  | -378,277 | 641,540 |
| 46     | 117,830  | -398,927 | 634,587 |
| 47     | 105,475  | -416,705 | 627,655 |
| 48     | 94,416   | -432,069 | 620,901 |

Final Estimates of Parameters (0,1,1)

| Type | Coef | SE Coef | T   | P   |
|------|------|---------|-----|-----|
| MA   | 1    | 0.4503  | 0.1590 | 2.83  | 0.008 |
### Forecasts from period 36

| Period | Forecast | Lower  | Upper  | Actual  |
|--------|----------|--------|--------|---------|
| 37     | 310,275  | 11,208 | 609,342|
| 38     | 310,275  | -30,994| 651,544|
| 39     | 310,275  | -68,522| 689,072|
| 40     | 310,275  | -102,654| 723,204|
| 41     | 310,275  | -134,173| 754,723|
| 42     | 310,275  | -163,599| 784,149|
| 43     | 310,275  | -191,303| 811,852|
| 44     | 310,275  | -217,554| 838,104|
| 45     | 310,275  | -242,560| 863,110|
| 46     | 310,275  | -266,483| 887,033|
| 47     | 310,275  | -289,452| 910,002|
| 48     | 310,275  | -311,574| 932,124|

### Final Estimates of Parameters (1,1,1)

| Type  | Coef  | SE Coef | T    | P     |
|-------|-------|---------|------|-------|
| AR    | 0.3939| 0.1933  | 2.04 | 0.050 |
| MA    | 0.9441| 0.0941  | 10.03| 0.000 |

### Forecasts from period 36

| Period | Forecast | Lower  | Upper  | Actual  |
|--------|----------|--------|--------|---------|
| 37     | 294,082  | 22,105 | 566,059|
| 38     | 245,275  | -52,946| 543,497|
| 39     | 226,052  | -78,832| 530,936|
| 40     | 218,480  | -89,039| 526,000|
| 41     | 215,498  | -93,583| 524,579|
| 42     | 214,323  | -95,972| 524,619|
| 43     | 213,861  | -97,523| 525,245|
| 44     | 213,679  | -98,744| 526,101|
| 45     | 213,607  | -99,833| 527,047|
| 46     | 213,579  | -100,868| 528,025|
| 47     | 213,567  | -101,880| 529,015|
| 48     | 213,563  | -102,882| 530,008|

Seen from the third to the model from the model of ARIMA (1, 0, 1) has the value of AR are significant but for MA was not significant because it’s worth more than 0.050. Then do the test next to the model of ARIMA (0, 1, 1), here we get AR and MA are not significant. Then do a test model with ARIMA (1, 1, 1) bias here we get the AR and MA of significant value is less than 0.050.

3.5.4. Forecasting. Based on the optimum model of ARIMA (1, 1, 1) rainfall forecasts obtained from Bandung city for July 2011 year to December the year 2013, The results showed that the prediction is quite accurate using method of ARIMA in the study area. [12]
Table 1. The results of Comparison Rainfall Forecast are Bandung (mm).

| Month | 2011 | 2012 | 2013 | 2014 | forecast |
|-------|------|------|------|------|----------|
| January | 63   | 82,9 | 216,9 | 309  | 294,1    |
| February | 76,7 | 303,7 | 250  | 88,9  | 245,3    |
| March | 89,4 | 155,5 | 305  | 418,7 | 226,1    |
| April | 381,5 | 290,8 | 286  | 216,6 | 218,5    |
| May | 193,4 | 257,1 | 171  | 176,7 | 215,5    |
| June | 117,6 | 60,5  | 231,5 | 195,5 | 214,3    |
| July | 77,2  | 34,2  | 159  | 180,6 | 213,9    |
| August | 3,1  | 15    | 74   | 119,8 | 213,7    |
| September | 102,8 | 27    | 172  | 106   | 213,6    |
| October | 103,6 | 125   | 234  | 65    | 213,6    |
| November | 321,4 | 537   | 164  | 296,5 | 213,6    |
| December | 259  | 637   | 418  | 316,4 | 213,6    |

4. Conclusions
Based on the research results several conclusions can be drawn as follows.

- On the basis of the monthly precipitation patterns in 2011 - 2013, visual observation of the ACF and PACF plots and calculation of the AIC, the rainfall in the district of Semarang has ARIMA models (1, 1, 1) with the value of the MAD 82.712.
- Based on the Box - Jenkins method of ARIMA, then using monthly rainfall data in 2011 - 2013 in Semarang Regency prediction can be done monthly rainfall for the area in question in 2014.
- Based on the form and function of the ARIMA, MAE and MAPE values are good enough then ARIMA model has fairly good accuracy for prediction of rainfall the following year (2014).
- Forecasting results using ARIMA model will be very useful for planning of agriculture and plantations in Bandung that rely on the water needs of the rainfall in the area concerned.

Acknowledgements
We recognized UNIKOM for this event and help the leading applied research (PTUPT) and applied research (PPT) to support this research.

References
[1] *Dry season and the rainy season in Indonesia* http://www.bmkg.go.id. Retrieved July 11, 2013.
[2] The area of farms and plantations in Indonesia 2013 http://indonesia.go.id/en/potential/natural-resources. Retrieved July 25
[3] Geography, topography, and geology Semarang District Retrieved July 10, 2013.
[4] Site Agricultural Research and Development Department of Agriculture, Republic of Indonesia http://bbsd3p.litbang.deptan.go.id/tamp_komoditas.php. Retrieved July 20, 2013.
[5] Spessa A C, Field R D, Pappenberger F, Langner A, Englhart S, Weber U and Moore J 2015 Seasonal forecasting of fire over Kalimantan, Indonesia *Natural Hazards and Earth System Science* 15(3), 429-442
[6] Qian J H, Robertson A W and Moron V 2013 Diurnal cycle in different weather regimes and rainfall variability over Borneo associated with ENSO Journal of Climate 26(5) 1772-1790

[7] Adiwijaya W and UN N 2014 Study of line search techniques on the modified backpropagation for forecasting of weather data in Indonesia Far East J. Math. Sci. 86(2)

[8] Lutkepohl H 2005 New Introduction to Multiple Time Series Analysis (Berlin: Springer Science + Business Media, Inc.,)

[9] Gujarati D N 2009 Essential of Econometrics (New York: McGraw-Hill Co.)

[10] Prvan T, Reid A and Petocz P 2002 Statistical laboratories using Minitab, SPSS and Excel: A practical comparison Teaching Statistics 24(2) 68-75

[11] Valipour M, Banihabib M E and Behbahani S M R 2013 Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir Journal of hydrology 476 433-441

[12] Nugroho A and Simanjuntak B H 2014 ARMA (Autoregressive Moving Average) Model for Prediction of Rainfall in Regency of Semarang-Central Java-Republic of Indonesia International Journal of Computer Science Issues (IJCSI) 11(3) 27