The forecasting model of discharge at Brantas sub-basin using autoregressive integrated moving average (ARIMA) and decomposition methods

G Idfi¹, A Yulistyorini¹, T Rahayuningsih¹, V A K Dewi¹ and E Setyawan¹

¹Department of Civil Engineering, Faculty of Engineering, Universitas Negeri Malang, Malang, Jawa Timur, Indonesia.

gilang.idfi.ft@um.ac.id

Abstract. A watershed is a combination of several rivers and tributaries with certain boundaries that function to drain rainwater into a lake or sea. One of the hydrological data contained in the watershed is discharge data. If there is incomplete discharge data, it must be extended based on historical data. ARIMA and decomposition are methods that can predict time series data. The purposes of this research are to determine the historical discharge patterns of Brantas Sub-basin, to know the discharge forecasting model of Brantas Sub-basin, to know the results of forecasting Brantas Sub-basin discharge, and to compare the accuracy between ARIMA and decomposition methods. The accuracy is obtained by calculating MSE and RMSE values. The best method is a method that has the smallest MSE and RMSE values. The results of the research showed that Brantas Sub-basin discharge data in 2007-2017 has a seasonal pattern. The best ARIMA model is ARIMA (0,0,3)(1,0,1)¹2 model, while the best decomposition model is the additive decomposition model. The Decomposition method has better accuracy than the ARIMA method in predicting discharge of Brantas Sub-basin.

1. Introduction

Some rivers and tributaries will form a watershed which has certain boundaries on land and sea that are used as rainwater container and then flow into lake or sea [1]. The watershed can consist of one or more sub-basin that function to receive and drain rainwater through tributaries to the main river. One of the hydrological data contained in watersheds is discharge data [2]. According to Limantara, the discharge is a volume of water that passes through a river cross-section which is calculated per unit of time [3].

Some rivers and tributaries will form a watershed which has certain boundaries on land and sea that are used as rainwater container and then flow into lake or sea [4]. The watershed can consist of one or more sub-basin that function to receive and drain rainwater through tributaries to the main river [5]. One of the hydrological data contained in watersheds is discharge data. According to Zhang, the discharge is a volume of water that passes through a river cross-section which is calculated per unit of time forecasting can be done according to right method [6]. These patterns include seasonal, cyclic, trend, and irregular patterns.

Auto-Regressive Integrated Moving Average (ARIMA) is a method that can be used to extend data using time series data [7]. According to Juwono in Pramujo, time-series data used in the ARIMA model must be stationary [8]. This makes the ARIMA model forecasting results better than the ARMA model. There is also a decomposition method that can be used to forecast data. According to Meliano state that
components of the decomposition method include trend factor, seasonal factor, cycle factor, and a random factor [9]. The advantage of this method is that patterns or components can be calculated separately, so that more accurate forecasting results are obtained.

Discharge data is an example of time series data in the field of hydrology, so it can be extended using stochastic methods. Some previous research states that the ARIMA method can be used to predict discharge data and provide good results [10]. Based on this, the research tries to compare the accuracy of the ARIMA method with the decomposition method. ARIMA and decomposition methods are used to forecast discharge data with the help of Minitab 18 software.

The research was conducted at Brantas Sub-Basin to predict the availability of water resources in the future, this is because the communities around the site need large amounts of water that need to be used in industry, food crop agriculture, plantation, and tourism. This is in accordance with the Decree of Minister of Settlement and Regional Infrastructure Number: 360/KPTS/M/2004 in 2004 which states that river discharge data forecasting is needed to estimate water availability needed. The results of forecasting are used when planning, designing, building, operating and maintaining water resources. Data commonly used are annual, seasonal, monthly, semi-monthly, and ten-daily discharge data. Discharge forecasting can also be used to predict flooding.

The purpose of this research is to conduct a comparative analysis of ARIMA and decomposition methods for the extension of discharge data. The results of this research are expected to obtain information on the reliability level of both methods in predicting the extension of discharge data.

2. Research method

Discharge forecasting of Sumber Brantas Sub Watershed uses ARIMA and Decomposition Methods which is modeled with Minitab 18 software. The historical data used is discharge data of Brantas Sub-basin with a discharge gauge at Sengkaling Dam. The data began in 2007 until 2018 which is divided into two parts, namely in-sample data (data processed for model determination) and out sample data (data used to calibrate forecast results for the next period). Discharge data in 2007-2017 are used as in sample data, while discharge data in 2018 are used as out sample data.

2.1. Deletion of outlier data and pattern analysis

Outlier data is data that has value too far from other data. The data needs to be deleted so that forecast results have good accuracy [11]. Analysis of historical data pattern needs to be done before making time series forecasting. This is important because data pattern will affect type of forecasting method that will be used.

2.2. Forecasting with arima method

The ARIMA method consists of several processes with different models. The models is :

2.2.1. Auto regressive (AR) process

The equation for auto regressive (AR) is

ARIMA (p,0,0)  
\[ X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \]  
(1)

Where:

\( \mu \) = constanta  
\( \phi_p \) = autoregresif parameters  
\( \epsilon_t \) = margin errors

2.2.2. Moving average (MA) process

The equation for moving average (MA) is

ARIMA (0,0,q) or MA (q)  
\[ X_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} \]  
(2)

Where:
\[ \theta_1 \text{ until } \theta_q = \text{ moving average parameters} \]
\[ e_{t-q} = \text{ margin errors} \]
\[ \mu = \text{ constant} \]

2.2.3. Steps forecasting
The stages of forecasting with ARIMA methods are:
- Input in sample data in Minitab 18 software.
- Test stationarity of data on variant and average.
- Determine ARIMA models based on ACF and PACF patterns.
- Test parameters of model, white noise assumption and residual normality.
- Determine the best ARIMA model.
- Forecast based on the best ARIMA model.

2.3. Forecasting with decomposition method
Decomposition is a method for analyzing time series data by calculating each component separately [12]. Forecasting results from the method are obtained by combining the estimates of each component. This method depend on 4 component:
- Trend factor: the trend factor shows that data has increased, decreased or been constant over a long period of time. Trends can be represented by straight lines, exponential S curves and other long-term patterns.
- Cycle factor: the cyclical factor shows that data has a varying pattern where the data moves from large values to small values and returns to large values in a longer period of time than seasonal factors.
- Seasonal factor: seasonal factors show that data has a steady increase and decrease over a long period of time, such as monthly, quarterly and yearly.
- Random factor: the random factor is assumed to be the difference between the combined effect of the three decomposition factors and the actual data

2.3.1. Decomposition equation model
The general model of the decomposition method can be seen in equation below
\[ X_t = f(I_t, T_t, C_t, E_t) \] (3)

where
- \( X_t \) = time series value in \( t \) period,
- \( I_t \) = season indeks in \( t \) period
- \( T_t \) = trend component in \( t \) period
- \( C_t \) = cycle component in \( t \) period
- \( E_t \) = random component in \( t \) period

2.3.2. Decomposition aditif model
The aditif model of the decomposition method can be seen in equation below
\[ X_t = (I_t + T_t + C_t) + E_t \] (4)

2.3.3. Decomposition multiplikatif model
The multiplikatif model of the decomposition method can be seen in equation below
\[ X_t = (I_t \times T_t \times C_t) \times E_t \] (5)

2.3.4. Steps forecasting
The stages of forecasting with decomposition methods are:
- Input in sample data in Minitab 18 software.
- Analyze data using additive and multiplicative decomposition models.
- Determine the best decomposition model.
• Forecast based on the best decomposition model.

**Figure 1** Flow chart of research
2.4. Fitting test
Data accuracy testing must be done in every forecasting process. This can be done by minimizing the error rate, which is the difference between the actual value and the forecast value [13]. There are several ways that can be used to test the accuracy of forecasting results, including MSE and RMSE [14].

Mean Square Error (MSE) is the result of dividing the error value squared by the amount of data [15]. The MSE calculation formula can be seen in equation below:

\[
MSE = \frac{\sum (X_t - F_t)^2}{n}
\]

where:
- \(X_t\) = actual value in t period
- \(F_t\) = forecast value in t period
- \(n\) = forecast period

Root Mean Square Error (RMSE) is the root of the MSE value[16]. The RMSE calculation formula can be seen in equation below:

\[
RMSE = \sqrt{\frac{\sum (X_t - F_t)^2}{n}}
\]

where:
- \(X_t\) = actual value in t period
- \(F_t\) = forecast value in t period
- \(n\) = forecast period

To make the implementation of the research does not deviate from the formulation of the problem and the purpose, a flow diagram is used as in Figure 1. The diagram is compiled from the data which were obtained from the relevant institutions or survey research locations.

3. Results and discussion
Pattern of Brantas Sub-basin discharge data in 2007-2017 can be seen in Figure 2 below.
transformation graph [11]. Data stationarity test on average is done by looking at trend pattern of data
Result of data stationary test on variant and test for average can be seen in Figure 3 and 4 below

Figure 3 Result of data stationary test on variant

Figure 4 Result of data stationary test for average

The best comparison of the graph of the actual discharge value and the ARIMA fit value
$(0,0,3)(1,1,3)^2$ can be seen in Figure 5 below
The tests are also carried out on other ARIMA models and then calculate MSE and RMSE values of each model [13]. Recapitulation of significance tests, MSE and RMSE calculation results can be seen in Table 1 below.

### Table 1 Recapitulation of Significance Tests, MSE and RMSE Calculation Results

| No | Models                          | Parameter Estimation | Diagnostic Test                      | MSE  | RMSE  |
|----|---------------------------------|----------------------|--------------------------------------|------|-------|
|    |                                 |                      | White Noise Test                      |      |       |
|    |                                 |                      | Normality Test                        |      |       |
| 1  | ARIMA (0,0,3)(0,1,1) [12]       | Fulfilled            | Fulfilled                             | 7,095| 2,664 |
| 2  | ARIMA (0,0,3)(1,0,0) [12]       | Fulfilled            | Fulfilled                             | 3,093| 1,759 |
| 3  | ARIMA (0,0,3)(1,0,1) [12]       | Fulfilled            | Fulfilled                             | 2,865| 1,693 |
| 4  | ARIMA (0,0,3)(1,1,3) [12]       | Fulfilled            | Fulfilled                             | 6,933| 2,633 |
| 5  | ARIMA (0,1,1)(0,1,3) [12]       | Fulfilled            | Fulfilled                             | 7,680| 2,771 |
| 6  | ARIMA (0,1,2)(1,0,1) [12]       | Fulfilled            | Fulfilled                             | 3,410| 1,847 |
| 7  | ARIMA (1,0,0)(0,1,1) [12]       | Fulfilled            | Fulfilled                             | 6,987| 2,643 |
| 8  | ARIMA (1,0,0)(0,1,3) [12]       | Fulfilled            | Fulfilled                             | 6,650| 2,579 |
| 9  | ARIMA (1,0,0)(1,0,1) [12]       | Fulfilled            | Fulfilled                             | 2,912| 1,707 |
| 10 | ARIMA (1,1,1)(0,1,1) [12]       | Fulfilled            | Fulfilled                             | 7,843| 2,800 |

Results of additive model analysis include time series, component analysis graph and seasonal analysis graph. Time series plot of additive decomposition model can be seen in Figure 6 below.
Figure 6. Comparison of observed discharge compared to the results of decomposition model.

The analysis is also carried out on multiplicative model and then calculate MSE and RMSE values of each model. Recapitulation of MSE and RMSE calculation results can be seen in Table 2 below.

| No | Models     | MSE   | RMSE  |
|----|------------|-------|-------|
| 1  | Additive   | 4,506 | 2,123 |
| 2  | Multiplicative | 4,621 | 2,150 |

Comparison of accuracy between ARIMA and decomposition methods can be seen in Table 3 below.

| No | Methods     | In Sample | Out Sample |
|----|-------------|-----------|------------|
|    |             | MSE       | RMSE       |
| 1  | ARIMA       | 2,865     | 1,693      |
| 2  | Decomposition | 4,506 | 2,123 |

|    |             | MSE       | RMSE       |
| 1  |             | 11,167    | 3,342      |
| 2  |             | 7,970     | 2,823      |

4. Conclusions

Discharge data of Brantas Sub-basin in 2007-2017 has seasonal pattern. It can be seen from existence of monthly seasonal factor that affect data. Therefore, forecasting process can be carried out using ARIMA and decomposition methods. Discharge forecasting model of Brantas Sub-basin are made using ARIMA (0,0,3)(1,0,1) and additive decomposition model. Both models are chosen as forecasting models because they have the best accuracy among other ARIMA and decomposition models.

The results of forecasting Brantas Sub-basin discharge data in 2018 using decomposition method have MSE and RMSE values that are smaller than ARIMA method. It shows that the results forecasting using decomposition method has smaller error values and data variance that is closer to actual data variance. Decomposition method has better accuracy than ARIMA method in predicting discharge data of Brantas Sub-basin. That is because MSE and RMSE values of decomposition method is smaller than ARIMA method for out sample criteria.
References

[1] Harto S 1993 *Analisis Hidrologi (Hydrological Analysis)* (Jakarta: Gramedia Pustaka)

[2] Limantara L M 2010 *Hidrologi Praktis (Practical Hydrology)* (Bandung: Lubuk Agung)

[3] Viessman W, Knapp J W, Lewis G L and Harbaugh T E 1977 *Introduction to Hydrology* (New York: Harper and Row)

[4] Nathan R J and McMahon T A 1990 *Water Resources Research* 26 1465–1473

[5] Stewart M K 2015 *Hydrol. Earth Syst. Sci.* 19 2587–2603

[6] Zhang R, Li Q, Chow T L, Li S and Danielescu S 2013 *Hydrol. Process* 27 2659–2665

[7] Xu C 2002 *Hydrologic Models* (Uppsala: Uppsala University Department of Earth Sciences Hydrology)

[8] Pramujo B, Juwono P T and Soetopo W 2014 *Discharge Modeling Using ARIMA Method to Determine Selorejo Reservoir Operation Pattern (Journal of Water Resources Engineering)* (Malang: Brawijaya University)

[9] Meilano A 2016 *Forecasting Number of Regional Library Visitors at Central Java Provincial Using Decomposition Method* Undergraduate thesis published (Semarang: Faculty of Mathematics and Natural Sciences, State University of Semarang)

[10] Danish Hydraulic Institute 1982 *NAM Model Users Guide* (Denmark: Danish Hydraulic Institute)

[11] Tsutsumi A, Jinno K and Berndtsson R 2004 *Hydrological Sciences* 49(2) 205-226

[12] Shakti D W, Hadiani R and Setiono 2014 Discharge Analysis Based on 15 Daily Cumulative Rain by Backpropagation Artificial Neural Network Method at Tirtomoyo Watershed for Sulingi Point *Journal of Civil Engineering Matrix* 768-775

[13] Utami T N, Hoyyi A and Rusgiyono A 2017 Penggunaan Metode Kombinasi Trend Deterministik dan Stokastik pada Data Jumlah Penumpang Kereta Api (Studi Kasus: KA Argo Muria) (The Use of Forecasting Methods Combination of Deterministic and Stochastic Trends on the Number of Train Passengers Data (Case Study: Argo Muria Train)) *Jurnal Gaussian* 6(1) 131-140

[14] Kuok K K, Harun S, and Chiu PC 2011 *Int. J. Advance. Soft Comput. Appl.* 3(3) 1-17

[15] Jain S K 1993 *Sciences Journal-des Sciences Hydrologiques* 38(5) 431-441

[16] Yuni S, Talakua M W, Lesnussa Y A 2015. Forecasting Number of Pattimura University Library Visitors at Ambon Using Decomposition Method *Journal of Mathematical Sciences and Applied* 9(1) 41-50