# Flood Analysis in Karkheh River Basin using Stochastic Model

Karim Hamidi Machekposhti [a], Hossein Sedghi [a*], Abdolrasoul Telvari [b], Hossein Babazadeh [a]

[a] *Department of Water Sciences and Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.*

[b] *Department of Civil Engineering, Islamic Azad University, Ahvaz, Iran.*

## Abstract

This study analyzed the annual streamflow of Karkheh River in Karkheh river basin in the west of Iran for flood forecasting using stochastic models. For this purpose, we collected annual stremflow (peak and maximum discharge) during the period from 1958 to 2015 in Jelogir Majin hydrometric station (upstream of Karkheh dam reservoir). A time series model (stochastic model or ARIMA) has three stages consists of: model identification, parameter estimation and diagnostic check. Model identification was done by visual inspection on the Autocorrelation and Partial Autocorrelation Function. Three types of ARIMA(p,d,q) models (0,1,1), (1,1,1) and (4,1,1) suggested for the studied series. The suggested model parameters were computed using the Maximum Likelihood (ML), Conditional Least Square (CLS) and Unconditional Least Square (ULS) methods. In model verification, the chosen criterion for model parsimony was the Akaike Information Criteria (AIC) and the diagnostic checks include independence of residuals. The best ARIMA model for this series was (4,1,1), with their AIC values of 88.9 and 77.8 for annual peak and maximum streamflow respectively. Forecast series up to a lead time of ten years future (2006 to 2015) were generated using the accepted ARIMA models. Model accuracy was checked by comparing the predicted and observation series by coefficient of determination (R2). Results show that the ARIMA model was adequate for the flood analysis in Karkheh River and the forecast of the series in short time at future.

*Keywords:* Stochastic Model; Flood Analysis; Maximum Likelihood; Karkheh River Basin.

## 1. Introduction

Flood analysis is a form of extreme value analysis in nature. The main interest in analyzing extreme hydrological events is not in what has occurred but possibilities that further extreme events such as flood will occur in the future. Flood analysis in particular, allows hydrologists and statisticians to estimate future flood occurrence probabilities as well as the peak magnitude of streamflow. Another reason flood analyses are important is that the design and operation of hydraulic structures such as dams and reservoirs are determined based on them. Flood modelling depends on available data to generate efficient estimations. There are several approaches for hydrological modelling such as deterministic, probabilistic and stochastic. The stochastic models are related to the probability models in the sense that both types of models have random variables. Time series analysis and regression techniques are applied in order to build a stochastic model in flood analysis. The chosen method of study falls under the category of time series modelling. Time series is commonly used in the field of hydrology and water resource management. The beauty of time series modelling is that future values of a variable can be estimated using its historical values. A time series often exhibits trends, sometimes shifts (jumps), seasonality and periodicity. These attributes are referred to as components in Equation 1. The components this equation are trend ($T_t$), seasonal component ($S_t$), cyclical component ($C_t$) and irregular component ($e_t$).

$$X_t = T_t + S_t + C_t + e_t \tag{1}$$

Hamidi machekposhti et al. (2017) studied the stochastic model to inflow of Karkheh dam at Iran and suggested ARIMA (4,1,1) is the best stochastic model for annual mean streamflow [4]. Time series modelling of annual maximum flow (AMF) of River Indus at Sukkur India to be examined by Shakeel et al. (1993) and they found that the suitable stochastic model was ARIMA (2,1,1) in this river [15]. The applied of ACF and PACF for annual flow of Australian streams were studied by Srikanthan et al. (1983). They determined the appropriate form of stochastic models (Box-Jenkins time series models) for annual flow [17]. O'Connel (1977) suggested a type of stochastic models (ARMA (1,1)) to generate synthetic flow series and predict streamflow in future. [13]. Stojković et al. (2015) suggested that the stochastic flows (annual discharges) simulated by the stochastic model can be used for hydrological phenomenon simulations in river basins of large European rivers [18]. In a study, Vijaya kumar and Vennila (2016) found that an ARMA (2,4) model is the suitable model for generated and predicted annual inflow of Krishnagiri reservoir in the state of Tamilnadu at India [20]. Musa (2013) studied ARMA model for flow discharge from the Shiroro River (about 22 years (1990-2011)) and analyzed with 3 different models namely; AR, ARMA and ARIMA models. Based on the model analysis and evaluations, appropriate predictions were made for the effective usage of the flow from the river for farming activities and generation of power for both industrial and domestic [10]. Huang et al. (2016) analyzed the annual maximum stage readings of three rivers in Langat River basin in Malaysia for flood predicting using stochastic model (ARIMA model). They found that ARIMA(1,1,0), (1,1,0) and (1,1,1) were appropriate models for the Dengkil, Kg. Lui and Kg. Rinching series respectively, with their AIC values of 133.736, 55.348 and 42.292 [5].

Tian et al. (2011) studied extreme value analysis of stream flow time series in Poyang Lake Basin, China [19]. Ghanbarpour et al. (2010) studied stochastic modeling of surface streamflow at different time scales for Sangsoorakh Karst basin at Iran Their results indicate that autoregressive integrated moving average (ARIMA) models perform better than deseasonalized autoregressive moving average (DARMA) models for weekly, monthly and bimonthly flow forecasting applications in the study area [3]. Shering et al. (2009) used stochastic time series modeling for prediction of rainfall and runoff in Lidder catchment of Lidder river in south Kashmir and concluded that AR(1) model can be effectively used for prediction this series [16]. The further researchers like Mohamed and Etuk (2017) [6], Muhammad et al. (2017) [8], Otache et al. (2011) [14], Mujumdar and Nagesh Kumar (2009) [9], Nguyen et al. (2007) [11], Adeli et al. (2015) [1], Chakraborty et al. (2010) [2] and Nigam et al. (2014) [12] studied the application of stochastic model in hydrologye and, water reseource managment in the all world.

This study develop stochastic models (ARIMA models) for prediction of flood (annual extreme streamflow such as peak and maximum discharge) using Box-Jenkins methodology in Jelogir Majin hydrometric station (upstream of Karkheh dam reservoir) in Karkheh river of Karkheh river basin in Khuzestan state at Iran.

The purpose of this study is:

(1) To generate or develop stochastic time series model (ARIMA model) for prediction of flood in Karkheh river basin.

(2) To estimate parameters of ARIMA model for annual streamflow (annual peak and maximum discharge) and.

(3) To test the validity of the annual predicted streamflow with measured and evaluated the performance of the best selected model.

## 2. Materials and Methods

### 2.1. Study Area and Data Collection

The study area is the Karkheh river basin in west of the Iran, located in the central and southern regions of the Zagros mountain range and its area is more than 50000 km$^2$. In terms of the geographical coordination, this region has been extended between 46° 57′- 49° 10′ E longitudes and 31° 48′- 34° 56′ N latitudes. The Karkheh River is the third largest river in Iran with 900 km long. This river is directly connected to the Karkheh dam, the largest surface reservoir in the region, which has an important role in supplying water to the region. The annual streamflow (annual peak and maximum discharge) selected for modelling from 1958 to 2015 (58 years) in Jelogir Majin hydrometric station (station number 9 in Figure 1). This station is located at the upper reaches of the reservoir of Karkheh dam and is the supplier of the most water entering the dam reservoir and has the greatest impact on reservoir water dam. The plotted of data are shown in Figures 3 and 4. This data were taken from Iran Water Resources Management Organization (IWRMO). The goal of this study is to decrease the flood problems in Karkheh river basin through developing stochastic models (ARIMA model) for the study Karkheh river using Box-Jenkins approach and then, forecast future annual peak and maximum streamflow (discharge) values in this river by the best ARIMA model.
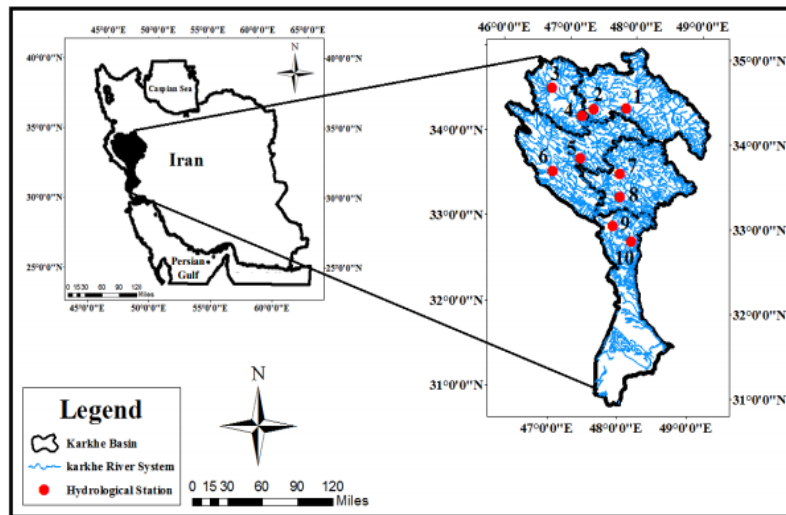
**Figure 1. The study area location**

## 2.2. ARIMA Model

The ARIMA modelling is actually an approach that has the flexibility to fit a model which is adapted from the data structure itself. The time series' stochastic nature can be modelled by the help of the computed Auto-correlation and Partial Auto-correlation Function (ACF and PACF) and fundamental information such as trend, periodic components, random components and serial correlation can be obtained. The Box-Jenkins approach to ARIMA modelling is an iterative model building process where the best models have to be determined through trial and error. However, with the advent of computers and statistical software, this iterative process can be simplified. Commonly SAS, SPSS, MINITAB and STATISTICA software use for this purpose. The basic methodology of ARIMA development is shown in Figure 2. The ARIMA model has three main components of an ARIMA model are AR, I and MA. The AR component represents the auto-correlation between past and current observations, the MA component describes the autocorrelation structure of error and I component represents the level of differencing required to transform a non- stationarity series into a stationary series. A non-seasonal ARIMA model is usually denoted by (p,d,q). The order of the AR, I and MA components are denoted by p, d and q respectively. The general ARIMA (p,d,q) model is:

$$U_t = \phi_1 U_{t-1} + \phi_2 U_{t-2} + \cdots + \phi_p U_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{2}$$

$$U_t = X_t - X_{t-d} \tag{3}$$

Which $\Phi_p$ = auto-regressive parameter, $\varepsilon_t$ = residual, $\theta_q$ = moving-average parameter, U=$d^{th}$ difference of the dependent variable and X= dependent variable.
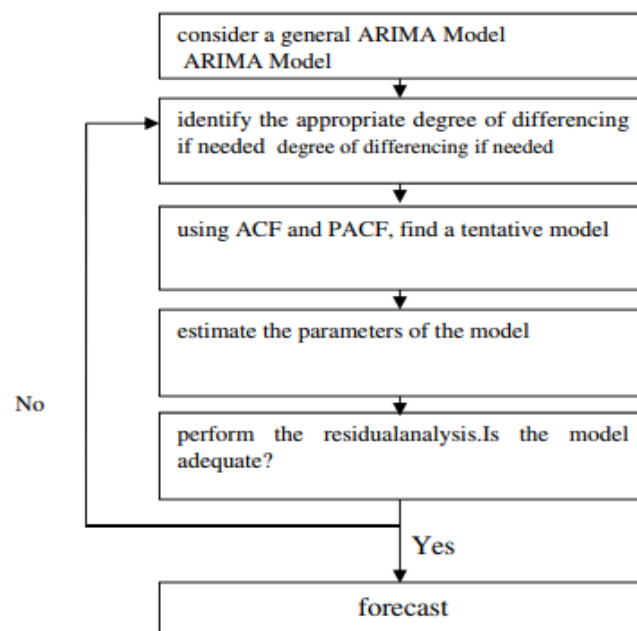


**Figure 2. ARIMA model development**

### 2.2.1. Stationary

Before the Box-Jenkins approach can be carried out, we need to recognize if time series is stationarity or non-stationarity. For this purpose, there are many ways to determine stationary such as Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and trend tests (Mann-Kendall trend test).

### 2.2.2. Independence

The basic assumption in Box-Jenkins approach is that the residuals of an ARIMA model are white noise. A white noise series have uncorrelated random shock with zero mean and constant variance. If the residuals are independent, it means that there is no more information that could be extracted from the series. One of the ways to determine the independence is to visually inspect the correlogram of the residuals. If the correlogram shows values that are close to zero, the residuals are uncorrelated and independent.

### 2.2.3. Transformation

Many statistical analyses are done based on the assumption that the population being investigated is normally distributed with a common variance. One of the popular transformation methods is the Box-Cox transformation. Also another method is Ln transformation of natural series.

### 2.3. Stationarity Tests

Unit root tests such as the Augmented-Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test were carried out to test the presence of a unit root while the Mann Kendall trend test was performed to check for the presence of a trend. The presence of a unit root or a trend should indicate non-stationarity of the series. The significance level used was 5%. If the series is non-stationary, differencing is required to transform it into a stationary series. On the other hand, if the series is stationary, the series is modelled as an ARMA process instead, which requires no differencing.

### 2.4. Differencing

The series was initially differenced once (d=1) and the ACF and PACF of the differenced series were plotted and analyzed. If the ACF and PACF decay rapidly then it indicates stationarity is achieved. Another indicator is the standard deviation of the differenced series. The optimum differenced series should have the lowest standard deviation. If the standard deviation of the current series is lower than that of the previous series, then the current series has the optimum order of differencing.

### 2.5. Plotting the Series and Its ACF and PACF

The main tools used for identification of model parameter were the visual displays of the series, which included the plot of natural data against time, which will show up important aspect of a time series such as trend, seasonality, outliers and etc., ACF and PACF. By using the annual streamflow as the input time series, the auto covariance function ($c_k$), the autocorrelation coefficients ($r_k$) and the partial autocorrelation coefficients ($\phi_k(k)$) were calculated and the series with its ACF and PACF were plotted using SAS and SPSS software. The number of lags k should fall between N/4 and N, therefore the chosen number of lags in this study was sufficient.

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}), 0 \leq k \leq N \tag{4}$$

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{5}$$

$$\widehat{\Phi}_{k+1}(k+1) = \left[r_{k+1} - \sum_{j=1}^{k}\widehat{\Phi}_k(j)r_{k+1-j}\right]\bigg/\left[1 - \sum_{j=1}^{k}\widehat{\Phi}_k(j)r_j\right] \tag{6a}$$

$$\widehat{\Phi}_{k+1}(j) = \widehat{\Phi}_k(j) - \widehat{\Phi}_{k+1}(k+1)\widehat{\Phi}_k(k-j+1) \tag{6b}$$

If all the ACF and PACF values are insignificant and fall within the confidence band, it indicates that the observations are independent. In such a case the time series is a white noise process and no modelling could be performed. A stationary time series has a rapidly decaying ACF. If the ACF is slow decaying, it indicates that the series may be non-stationary and requires differencing.

### 2.6. Identifying p and q

Having identified the optimum order of differencing (d), the next step was to identify the order of the autoregressive (p) and moving average (q) parameters using ACF and PACF of series.

## 2.7. Choosing the Best ARIMA Model

The previous step gave an indication of the order of p and q that should be fitted in the model. However, it was recommended to try a few different values of p and q to get the best model while preserving the parsimony of the parameters. To test for the parsimony of parameters, the Akaike Information Criteria (AIC) was used. The model with the minimum AIC was selected as the best model. The SPSS and SAS software can find the best model based on the AIC values calculated for a range of p and q. In this study, the maximum p and q were selected four.

## 2.8. Diagnostic Checks

After estimating the model parameters, the diagnostic checking is applied to see if the model is adequate or not. Therefore the following statistical tests are used:

### A. Port Manteau Lack of fit Test

Port manteau lack of fit test is used for this purpose. It is a test of the residual independency and uses the Q-statistic defined as:

$$Q = (N - d - DS) \sum_{k=1}^{M} r_k^2 a_t \tag{7}$$

Where $r_k$ ($a_t$) is the autocorrelation coefficient of the residual ($a_t$) at lag k, and M is the maximum lag considered (about N/4), ARIMA model is considered adequate if p>chi square is greater than the level of significant 0.05.

### B. Residual autocorrelation Function Test

The second test is the independency of the resulting ($a_t$) series, the correlogram of this series are computed for lag (M=N/5). The figures of RACF and RPACF (residual autocorrelation and partial autocorrelation function) show that the most of computed lags inside the tolerance interval ($\pm 2/\sqrt{N}$, at 95% confidence limits).

## 2.9. Series Comparison and Forecasting

Forecasting can be categorized into long term and short term forecasting. Short term forecasting can predict values that are a few time periods (a few years) into the future while long term forecasting can predict values for time periods that extend far beyond that. In terms of applications, long term forecasts are used for strategic planning while short term forecasts are used for project developments as well as operation management. Statistical methods are good for short term forecasting because the historical data normally exhibit inertia and do not show drastic changes [7]. Short term forecasting is based on identifying, modelling and extrapolating the patterns found in the data. The best model that passed the diagnostic checking, will predict the data series in future. We determine the degree of similarity between the predicted and observation data series by coefficient of determination ($R^2$). If the pattern of the predicted series appears similar to the pattern of the original series, then the fitted model is a good model. The final step was to generate a forecast of future values. The ARIMA model can predict future values as well as its confidence interval using the calculated model parameters. In this study, the chosen number of predicted values was ten, which means that the values were predicted for the next ten years after the last observation.

# 3. Results and Discussion

Before to start of modelling, we must test the two series for normal distribution. In this study, the natural series have not normality therefore we gave Ln transformation of natural series. Plotting the observations of Ln natural data against time (Figures 5 and 6) show that there is increase trend for studied series. In this study, the Ln of natural series is not stationary and then has first differencing of Ln natural data to achieving stationary series (d=1). This plot shows in Figures 7 and 8. It is stationary and there is not trend.

## 3.1. Model Identification

The ACF and PACF plot of Ln series for the Karkheh River show in Figures 9 and 10. Since some the ACF and PACF values are significant and do not fall within the confidence band, it indicates that the time series are not white noise processes and modelling could be performed. The ACF and PACF of the once differenced (d=1) series decayed rapidly compared to the ACF and PACF of the Ln series. Therefore, the optimum level of differencing for the series was one and the d value used in the ARIMA model would be one (d=1). The ACF and PACF of this series show in Figures 11 and 12. Tjelogir majin in Figures 9-12 and 13-14 mean Ln transformation of data in Jelogir majin station. The original data in this station was not normal distribution therefore the Ln transformation is necessary. The suggested model for p and q are 1, 2, 3, 4 of ACF and PACF this series and we start to estimation parameters some of models.

## 3.2. Estimation of the Model Parameters

After the identification of model using the parameter estimation methods such as Maximum Likelihood (ML), Conditional Least Square (CLS) and Unconditional Least Square (ULS) are done. The values of parameters estimation are shown in Tables 1 and 2. The value of this tables showed that all selected model are suitable for entrance to next

stage because it have two conditional stationary and inevitability. The model that gives the minimum AIC is selected as best fit model. Obviously, model ARIMA (4,1,1) has the smallest values of AIC, then one would temporarily have a model ARIMA (4,1,1). The goodness of fit statistic and Akaike Information Criterion (AIC) values for the different ARIMA models are shown in Table 5 (AIC=88.87 in CLS estimation method for annual peak streamflow and AIC=77.75 in ML estimation method for annual maximum streamflow). In this case, the initial suggested model structure has the minimum AIC value and has been chosen as best model structure for annual streamflow time series. The results of Port manteau lack of fit test indicate in Tables 3 and 4. This tables shows that all three models are adequate for forecasting of studied series data. In conclusion, the ARIMA (4,1,1) model is the best model for annual peak and maximum streamflow (discharge) in Karkheh River at Karkheh river basin. Also the RACF and RPACF residual autocorrelation Function test are shown in Figures 13 and 14. This figures show that the suggested model can be considered as appropriate model.

### 3.3. Forecasting

In this step, we forecasted annual data for ten years ahead of original data for the period from 2006 to 2015 by applying the best model (ARIMA (4,1,1)) in CLS and ML estimation parameter method for annual peak and maximum discharge respectively. After obtaining the forecasted series ($Z_t$ for t=1, 2, 3…10), then the final series ($X_t$) is determined by reversing (Ln) transformation. The forecasting actuals were brought in Table 6. In this table we can see forecasting values in ten futures. Figures 13 and 14 show predicted of series for these data. The figures show us that forecasting is appropriate. The corresponding observed values are also shown in the Figures 15 and 16 and since agreement between observed and forecasted values of annual peak discharge ($R^2=0.84$) and annual maximum discharge ($R^2=0.87$) are very good, it is confirmed that the ARIMA (4,1,1) model is adequate for forecasting of annual peak and maximum discharge. The forecasted values in ten years future uses for calibration and verification of the best selected model.
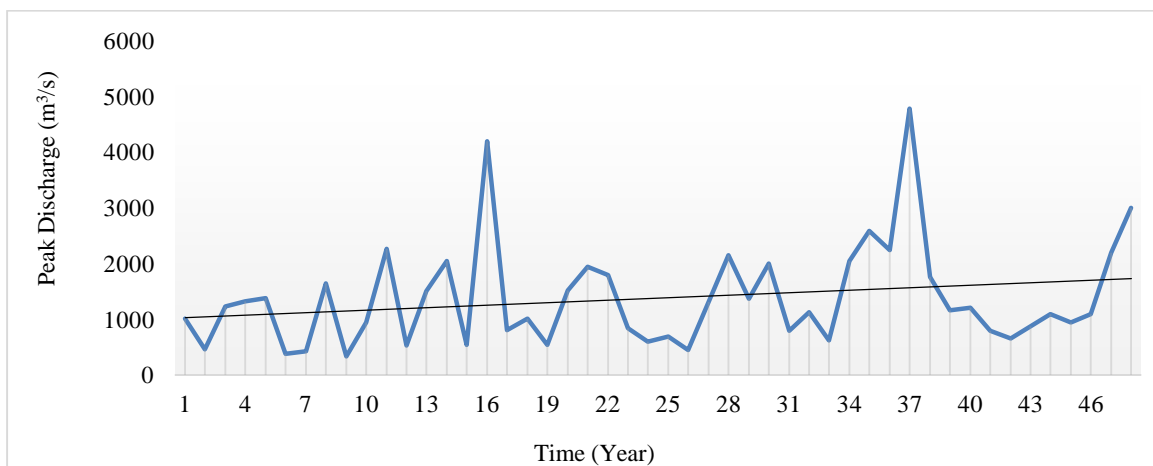


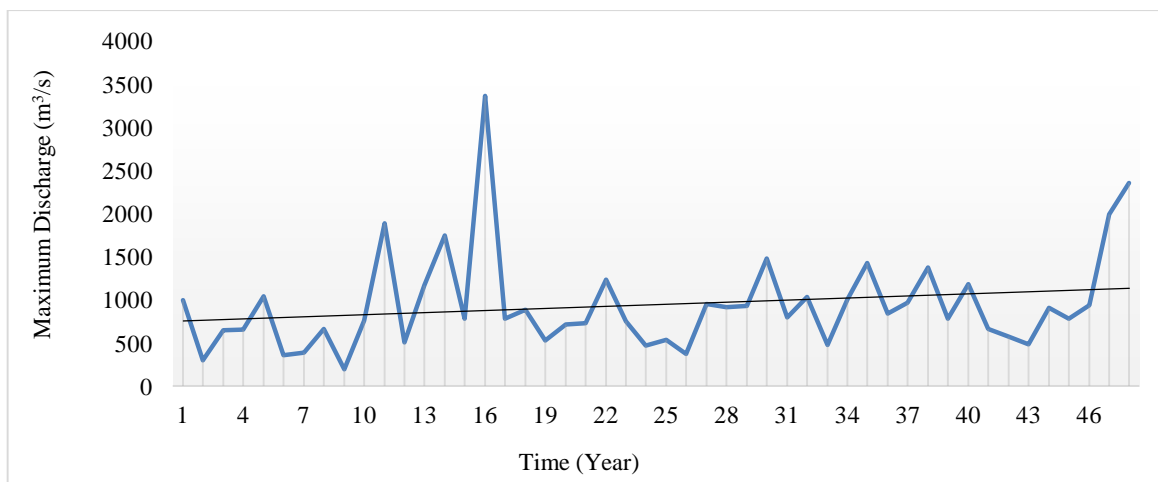**Figure 3. Original annual peak streamflow (discharge) time series**



**Figure 4. Original annual maximum streamflow (discharge) time series**
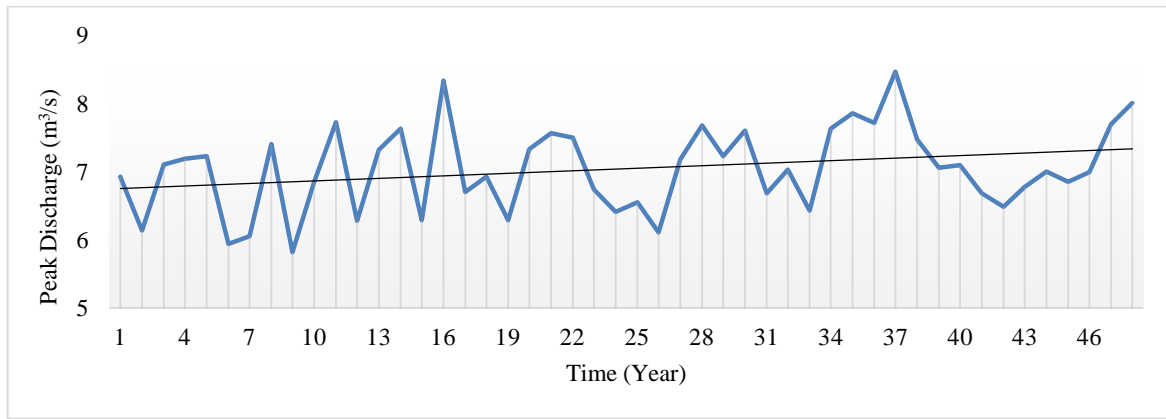
**Figure 5. Ln transformation of annual peak streamflow (discharge) time series**
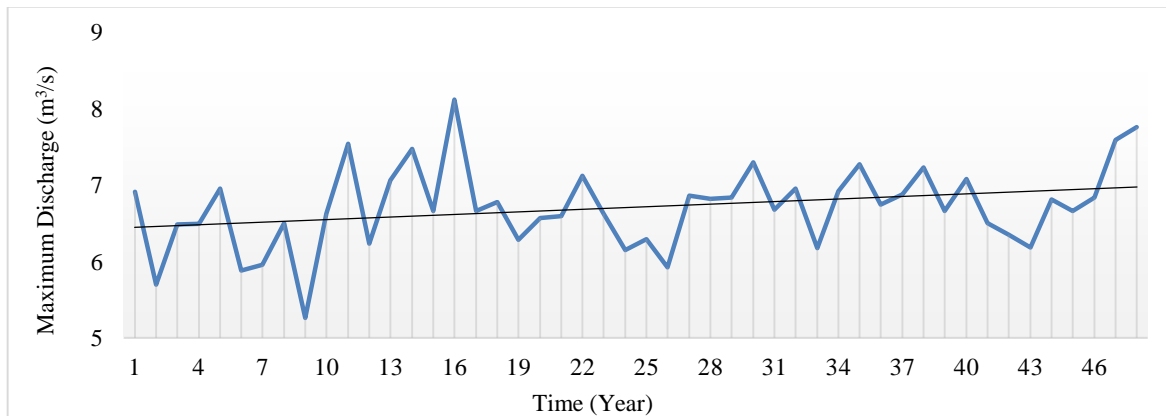


**Figure 6. Ln transformation of annual maximum streamflow (discharge) time series**
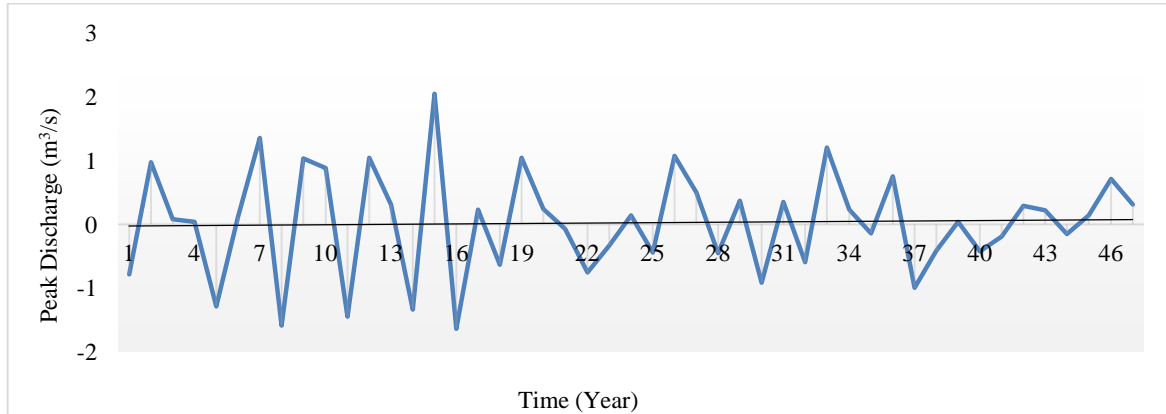


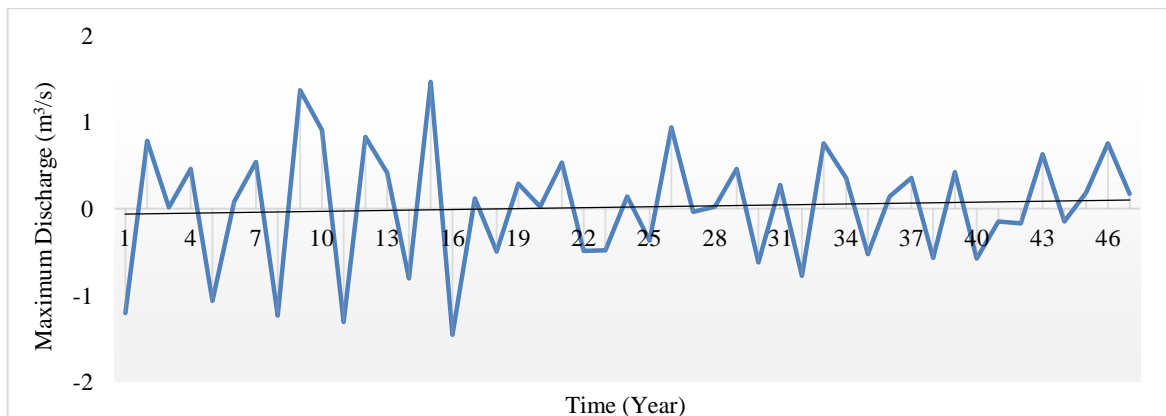**Figure 7. First differencing of annual peak streamflow (discharge) time series**



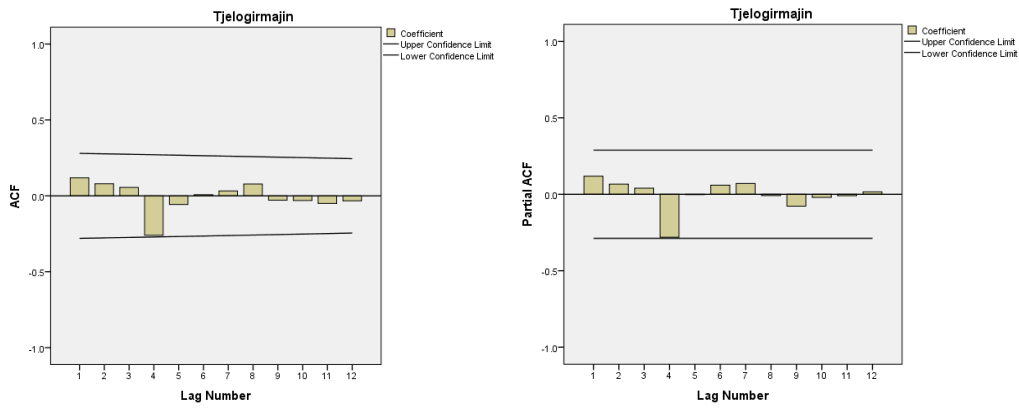**Figure 8. First differencing of annual maximum streamflow (discharge) time series**

**Figure 9. ACF and PACF of Ln original annual peak streamflow (discharge)**
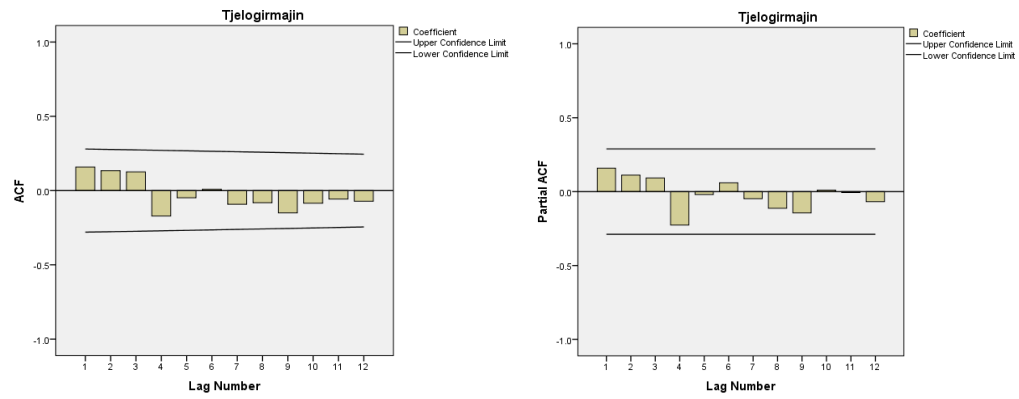


**Figure 10. ACF and PACF of Ln original annual maximum streamflow (discharge)**
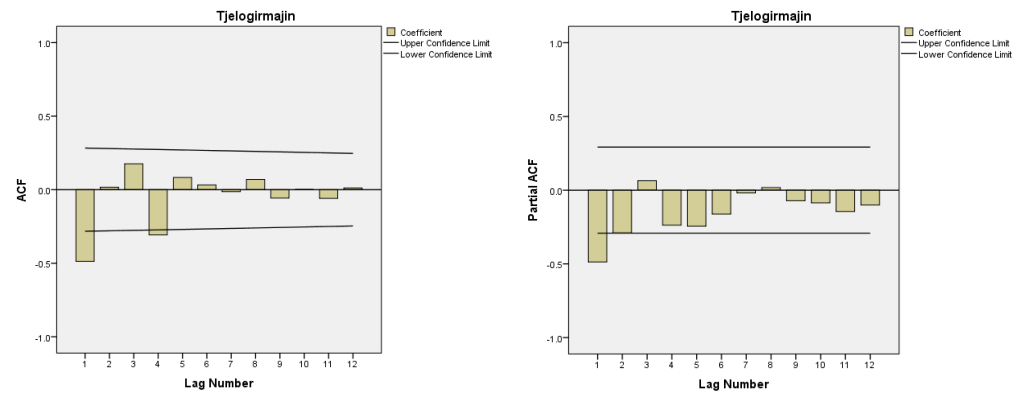


**Figure 11. ACF and PACF of differencing annual peak streamflow (discharge) (d=1)**
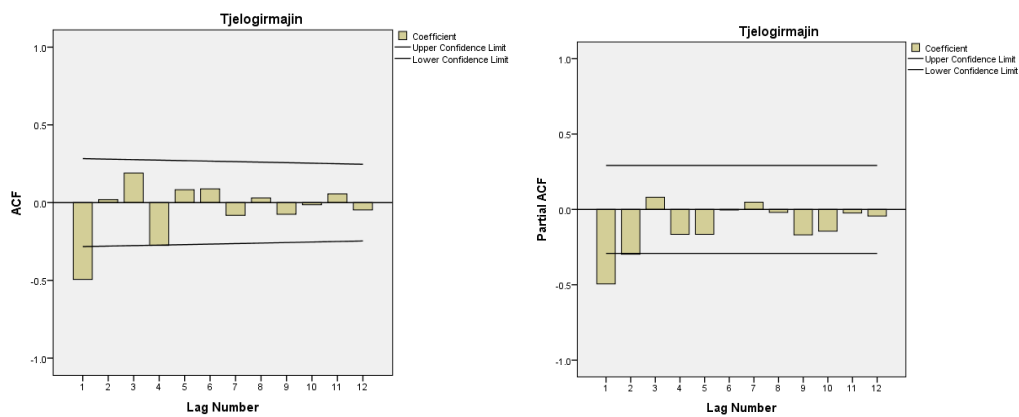


**Figure 12. ACF and PACF of differencing annual maximum streamflow (discharge) (d=1)**

**Table 1. Result of parameter estimation for the selected model**

| Parameter | Estimation Method | Type (Order) and Values of Parameters ARIMA(p,1,q) | Absolute Value of t | Probability of t | Stationary Condition | Invertibility Condition |
|---|---|---|---|---|---|---|
| **Annual Peak Streamflow (Discharge)** | ML | P(1) = -0.48656 Q(0) | -3.81 | 0.0001 | Satisfy | |
| | CLS | P(1) = -0.48713 Q(0) | -3.78 | 0.0005 | Satisfy | |
| | ULS | P(1) = -0.49708 Q(0) | -3.88 | 0.0003 | Satisfy | |
| | ML | P(1) = 0.10744 | 0.66 | 0.5104 | Satisfy | |
| | | Q(1) = 0.93539 | 10.87 | 0.0001< | | Satisfy |
| | CLS | P(1) = 0.11274 | 0.69 | 0.4926 | Satisfy | |
| | | Q(1) = 0.96723 | 15.99 | 0.0001< | | Satisfy |
| | ULS | P(1) = 0.12820 | 0.84 | 0.4072 | Satisfy | |
| | | Q(1) = 0.99998 | 3.35 | 0.0001< | | Not Satisfy |
| | ML | P(4) = -0.3317 | -2.25 | 0.0243 | Satisfy | |
| | | Q(1) = 0.86679 | 9.89 | 0.0001< | | Satisfy |
| | CLS | P(4) = -0.33489 | -2.19 | 0.0339 | Satisfy | |
| | | Q(1) = 0.86679 | 13.48 | 0.0001< | | Satisfy |
| | ULS | P(4) = -0.36065 | -2.4 | 0.0208 | Satisfy | |
| | | Q(1) = 0.89524 | 11.84 | 0.0001< | | Satisfy |

ML: Maximum Likelihood     CLS: Conditional Least Square     ULS: Unconditional Least Square

**Table 2. Result of parameter estimation for the selected model**

| Parameter | Estimation Method | Type (Order) and Values of Parameters ARIMA(p,1,q) | Absolute Value of t | Probability of t | Stationary Condition | Invertibility Condition |
|---|---|---|---|---|---|---|
| **Annual Maximum Streamflow (Discharge)** | ML | P(1) = -0.51759 Q(0) | -4.14 | 0.0001< | Satisfy | |
| | CLS | P(1) = -0.49304 Q(0) | -3.84 | 0.0004 | Satisfy | |
| | ULS | P(1) = -0.52951 Q(0) | -4.23 | 0.0001 | Satisfy | |
| | ML | P(1) = 0.14873 | 0.86 | 0.3922 | Satisfy | Satisfy |
| | | Q(1) = 0.92809 | 8.16 | 0.0001< | | |
| | CLS | P(1) = 0.230274 | 1.46 | 0.1501 | Satisfy | Satisfy |
| | | Q(1) = 0.97512 | 28.55 | 0.0001< | | |
| | ULS | P(1) = 0.17543 | 1.13 | 0.2631 | Satisfy | Not Satisfy |
| | | Q(1) = 0.99999 | 3.35 | 0.0016 | | |
| | ML | P(4) = -0.26820 | -1.71 | 0.087 | Satisfy | Satisfy |
| | | Q(1) = 0.74897 | 6.36 | 0.0001< | | |
| | CLS | P(4) = -0.26824 | -1.7 | 0.0964 | Satisfy | Satisfy |
| | | Q(1) = 0.69253 | 5.76 | 0.0001< | | |
| | ULS | P(4) = -0.27556 | -1.66 | 0.1031 | Satisfy | Satisfy |
| | | Q(1) = 0.85681 | 8.77 | 0.0001< | | |

ML: Maximum Likelihood     CLS: Conditional Least Square     ULS: Unconditional Least Square

**Table 3. Result of autocorrelation check of residuals annual peak streamflow in Jelogir Majin station**

| ARIMA Model | Estimation Method | To Lag | Df | Chi-Square | Pr>Chi Square | Adequacy for Modelling |
|---|---|---|---|---|---|---|
| ARIMA(1,1,0) | ML | 6 | 5 | 9.99 | 0.0754 | Satisfy |
| | | 12 | 11 | 11.25 | 0.4228 | |
| | | 18 | 17 | 14.32 | 0.6447 | |
| | | 24 | 23 | 17.26 | 0.7962 | |
| | CLS | 6 | 5 | 9.68 | 0.0850 | Satisfy |
| | | 12 | 11 | 10.94 | 0.4484 | |
| | | 18 | 17 | 14.02 | 0.6659 | |
| | | 24 | 23 | 16.93 | 0.8128 | |
| | ULS | 6 | 5 | 10.02 | 0.0748 | Satisfy |
| | | 12 | 11 | 11.29 | 0.4190 | |
| | | 18 | 17 | 14.38 | 0.6400 | |
| | | 24 | 23 | 17.33 | 0.7928 | |
| ARIMA(1,1,1) | ML | 6 | 4 | 5.32 | 0.2562 | Satisfy |
| | | 12 | 10 | 5.93 | 0.8212 | |
| | | 18 | 16 | 9.21 | 0.9046 | |
| | | 24 | 22 | 12.45 | 0.9473 | |
| | CLS | 6 | 4 | 4.63 | 0.3269 | Satisfy |
| | | 12 | 10 | 5.20 | 0.8775 | |
| | | 18 | 16 | 8.26 | 0.9409 | |
| | | 24 | 22 | 10.89 | 0.9763 | |
| ARIMA(4,1,1) | ML | 6 | 4 | 1.30 | 0.8608 | Satisfy |
| | | 12 | 10 | 2.77 | 0.9863 | |
| | | 18 | 16 | 7.46 | 0.9653 | |
| | | 24 | 22 | 10.70 | 0.9787 | |
| | CLS | 6 | 4 | 1.36 | 0.8504 | Satisfy |
| | | 12 | 10 | 2.98 | 0.9818 | |
| | | 18 | 16 | 7.56 | 0.9610 | |
| | | 24 | 22 | 10.41 | 0.9822 | |
| | ULS | 6 | 4 | 1.25 | 0.8702 | Satisfy |
| | | 12 | 10 | 2.53 | 0.9904 | |
| | | 18 | 16 | 7.61 | 0.9596 | |
| | | 24 | 22 | 11.47 | 0.9674 | |

ML: Maximum Likelihood      CLS: Conditional Least Square      ULS: Unconditional Least Square

**Table 4. Result of autocorrelation check of residuals annual maximum streamflow in Jelogir Majin station**

| ARIMA Model | Estimation Method | To Lag | Df | Chi-Square | Pr>Chi Square | Adequacy for Modelling |
|---|---|---|---|---|---|---|
| ARIMA(1,1,0) | ML | 6 | 5 | 9.63 | 0.0866 | Satisfy |
| | | 12 | 11 | 10.85 | 0.4559 | |
| | | 18 | 17 | 14.10 | 0.6602 | |
| | | 24 | 23 | 21.24 | 0.5665 | |
| | CLS | 6 | 5 | 8.87 | 0.1145 | Satisfy |
| | | 12 | 11 | 10.14 | 0.5177 | |
| | | 18 | 17 | 13.26 | 0.7185 | |

| | | | | | |
|---|---|---|---|---|---|
| | | 24 | 23 | 20.50 | 0.6118 | |
| | | 6 | 5 | 9.65 | 0.0856 | |
| | ULS | 12 | 11 | 10.89 | 0.4525 | Satisfy |
| | | 18 | 17 | 14.21 | 0.6519 | |
| | | 24 | 23 | 21.26 | 0.5653 | |
| | | 6 | 4 | 3.85 | 0.4262 | |
| | ML | 12 | 10 | 5.64 | 0.8445 | Satisfy |
| | | 18 | 16 | 10.50 | 0.8391 | |
| | | 24 | 22 | 17.76 | 0.7201 | |
| ARIMA(1,1,1) | | 6 | 4 | 4.15 | 0.3856 | |
| | CLS | 12 | 10 | 5.74 | 0.8365 | Satisfy |
| | | 18 | 16 | 10.06 | 0.8637 | |
| | | 24 | 22 | 16.58 | 0.7861 | |
| | | 6 | 4 | 1.22 | 0.8752 | |
| | ML | 12 | 10 | 5.29 | 0.8712 | Satisfy |
| | | 18 | 16 | 10.91 | 0.8152 | |
| | | 24 | 22 | 15.22 | 0.8527 | |
| | | 6 | 4 | 0.97 | 0.9139 | |
| ARIMA(4,1,1) | CLS | 12 | 10 | 6.01 | 0.8144 | Satisfy |
| | | 18 | 16 | 11.58 | 0.7726 | |
| | | 24 | 22 | 15.57 | 0.8364 | |
| | | 6 | 4 | 2.15 | 0.7077 | |
| | ULS | 12 | 10 | 6.91 | 0.7338 | Satisfy |
| | | 18 | 16 | 13.12 | 0.6638 | |
| | | 24 | 22 | 17.33 | 0.7449 | |

ML: Maximum Likelihood      CLS: Conditional Least Square      ULS: Unconditional Least Square

**Table 5. Goodness of fit statistic**

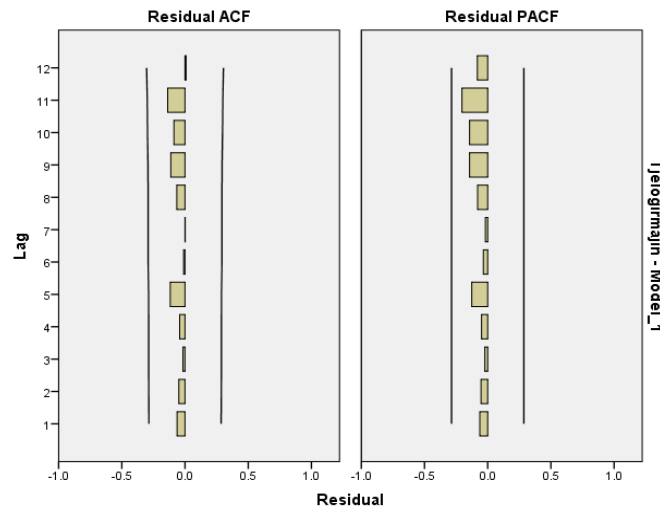| Parameter | ARIMA Model | Estimation Method | Akaikc's Statistic |
|---|---|---|---|
| **Annual Peak Streamflow (Discharge)** | (1,1,0) | ML | 103.4247 |
| | | CLS | 103.4469 |
| | | ULS | 103.4316 |
| | (1,1,1) | ML | 95.2824 |
| | | CLS | 93.1350 |
| | (4,1,1) | ML | 90.8381 |
| | | *CLS* | *88.8680* |
| | | ULS | 91.0387 |
| **Annual Maximum Streamflow (Discharge)** | (1,1,0) | ML | 84.185 |
| | | CLS | 84.9952 |
| | | ULS | 84.1939 |
| | (1,1,1) | ML | 79.7781 |
| | | CLS | 79.6182 |
| | (4,1,1) | *ML* | *77.7478* |
| | | CLS | 78.8943 |
| | | ULS | 78.1438 |

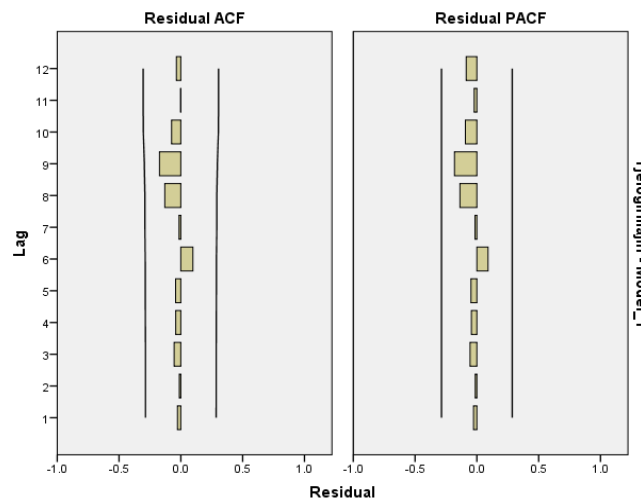**Figure 13. RACF and RPACF for annual peak streamflow (discharge) ARIMA (4,1,1)**



**Figure 14. RACF and RPACF for annual maximum streamflow (discharge) ARIMA (4,1,1)**

**Table 6. Forecasts from period 2006-7 to 2015-16**

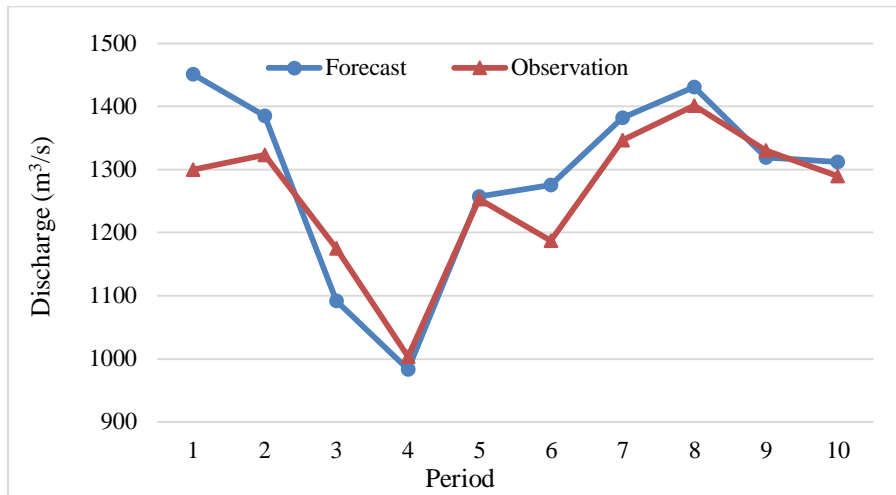| Period | Annual Peak Streamflow (Discharge) | | Annual Maximum Streamflow (Discharge) | |
|---|---|---|---|---|
| | Forecast | Observation | Forecast | Observation |
| 2006-7 (1) | 1451 | 1300 | 1235 | 1161 |
| 2007-8 (2) | 1385 | 1323 | 1177 | 1163 |
| 2008-9 (3) | 1092 | 1175 | 962 | 941 |
| 2009-10 (4) | 984 | 1004 | 920 | 960 |
| 2010-11 (5) | 1257 | 1253 | 1092 | 1070 |
| 2011-12 (6) | 1276 | 1187 | 1106 | 1090 |
| 2012-13 (7) | 1382 | 1346 | 1167 | 1202 |
| 2013-14 (8) | 1431 | 1401 | 1182 | 1210 |
| 2014-15 (9) | 1319 | 1330 | 1129 | 1150 |
| 2015-16 (10) | 1312 | 1290 | 1099 | 1080 |

**Figure 15. Comparison of forecasted and observed annual peak streamflow (discharge) (2006-2015)**
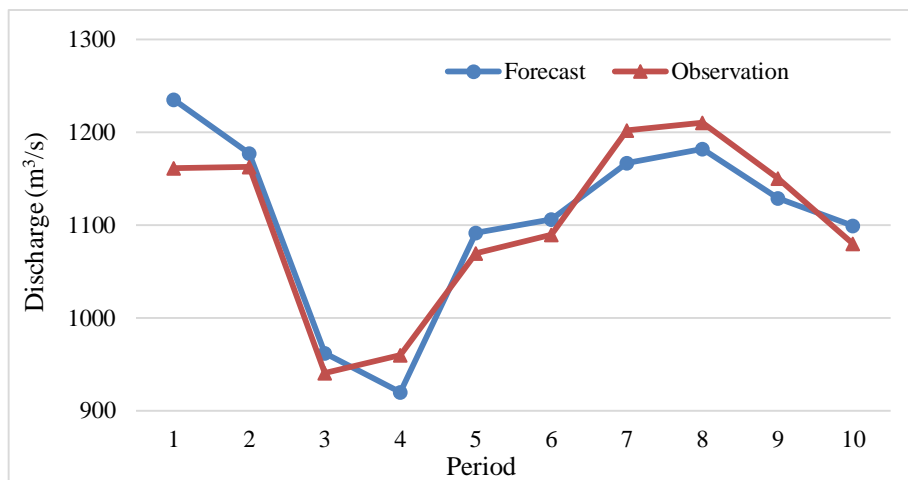


**Figure 16. Comparison of forecasted and observed annual maximum streamflow (discharge) (2006-2015)**
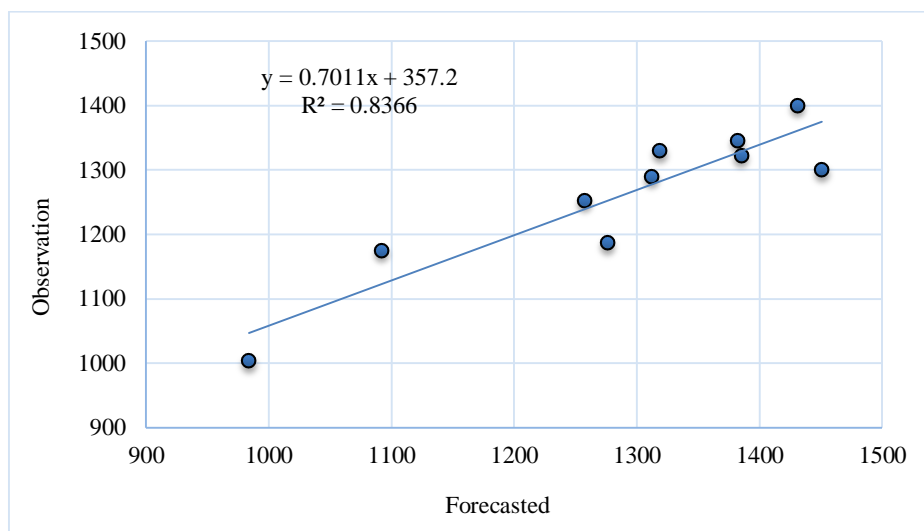


**Figure 17. Correlation between observation and forecasted values of annual peak streamflow (discharge)**
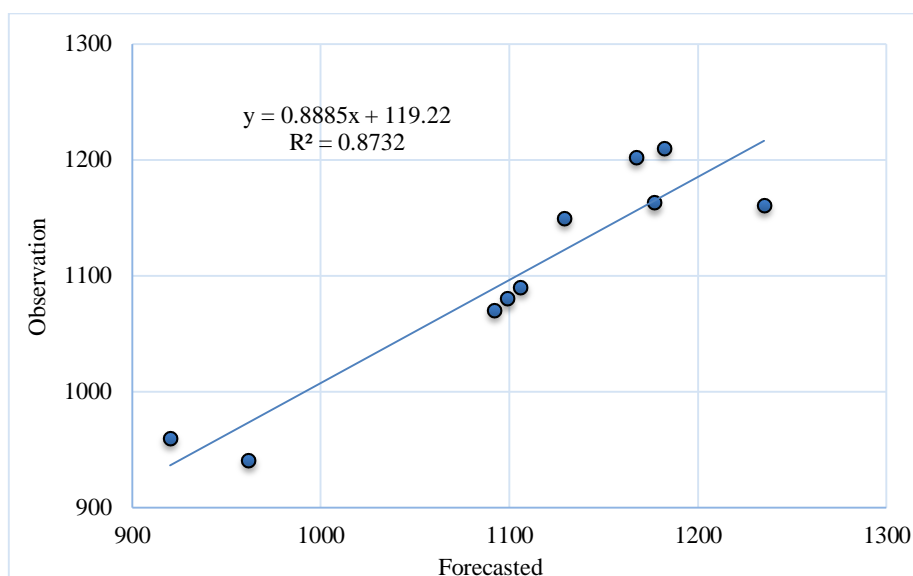
**Figure 18. Correlation between observation and forecasted values of annual maximum streamflow (discharge)**

## 4. Conclusion

The objectives of this study were develop stochastic ARIMA models for the study rivers using Box-Jenkins approach and forecast annual streamflow values in future using the developed ARIMA models. Recognizing and predicting annual peak and maximum streamflow (discharge) of Karkheh River in Jelogir Majin station during statistical period is necessary for flood control and planning the agricultural activities. Results from this reviewing indicated that:

- The best ARIMA model for annual peak streamflow in Jelogir Majin station at Karkheh River was (4,1,1), with their AIC values of 88.9 in CLS estimation method. Also for annual maximum streamflow (discharge) was (4,1,1), with their AIC values of 77.8 in ML estimation method. Forecast series up to a lead time of ten years were generated using the accepted ARIMA models. Model accuracy was checked by comparing the predicted and observation series by coefficient of determination ($R^2$). This coefficient ($R^2$) was 0.84 and 0.87 for annual peak and maximum streamflow respectively. Results show that the ARIMA model was adequate for the river and forecast of the series.

- The study reveals that the Box-Jenkins (ARIMA) model methodology could be used as an appropriate tool to predict the flood in this river for the up-coming 10 years (2006-2015). Also this methodology can help farmers in the area in order to planning the agricultural activities to enlarge the areas of land to be cultivated using supplemental irrigation.

- The significant ACF and PACF functions with order of four can be caused by factors such as area good vegetation and snowmelt. The good vegetation of the region and so the forest causes water retention in the soil surface layer and delay in the rise in surface runoff.

- The ARIMA model is suitable for short term forecasting of series because the ARMA family models can model short term durability very well. The AR model is a finite memory model, thus it does not fare well in long term forecasting.

- The model identification is the critical step in ARIMA modelling. The values of p, q and d had to be determined visually and they depended on the modeler's experience and judgment.

## 5. References

[1] Adeli, A., Fathi, M., Moghadam, Habib Musavi Jahromi, H., "Using Stochastic Models to Produce Artificial Time Series and Inflow Prediction: A Case Study of Talog Dam Reservoir, Khuzestan Province, Iran." International Bulletin of Water Resources and Development, 2015, Issue 5.

[2] Chakraborty, SH., Denis, D.M., and Sherring, A., "Development of Time Series Autoregressive Model for Prediction of Rainfall and Runoff in Kelo Watershed Chhattisgarh". International Journal of Advances in Engineering Science and Technology, 2009, Vol. 2, No. 2, pp. 153-163.

[3] Ghanbarpour, M.R., Abbaspour, K.C., Jalalvand, G., and ashtiani moghaddam, G., "Stochastic Modeling of Surface Stream Flow at Different Time Scales: Sangsoorakh Karst Basin, Iran" Journal of Cave and Karst Studies, 2010, Vol. 72, No. 1, pp. 1–10.

[4] Hamidi machekposhti, K., Sedghi, H., Telvari, A., Babazadeh, H., "Forecasting by Stochastic Models to Inflow of Karkheh Dam

at Iran" Civil Engineering Journal, 2017, Vol. 3, No. 5, pp. 340-350.

[5] Huang Y.F., Mirzaei M., and Yap W.K. "Flood Analysis in Langat River Basin using Stochastic Model" International Journal of GEOMATE, 2016, Vol. 11, No. 27, pp. 2796-2803.

[6] Mohamed, T.M., and Etuk, E.H., "Application of Linear Stochastic Models to Monthly Streamflow Data" Journal of Scientific and Engineering Research, 2017, Vol. 4, No. 4, pp. 322-326

[7] Montgomery D.C., Jennings C.L., and Kulahci M., Introduction to time series analysis and forecasting, John Wiley & Sons, Inc., 2008.

[8] Muhammad A.R., Xiaohui, Y., Ozgur, K., and Valetin, C., "Application of time series models for Streamflow Forecasting" Civil and Environmental Research, 2017, Vol. 9, No. 3.

[9] Mujumdar, P.P., and D. Nagesh Kumar, D., "Stochastic models of streamflow: some case studies". Hydrological Sciences Journal, 2009, Vol. 35, No. 4, pp. 395-410.

[10] Musa J.J., "Stochastic Modelling of Shiroro River Stream flow Process." American Journal of Engineering Research (AJER), 2013, Vol. No.6, pp. 49-54.

[11] Nguyen, V.T.V., Nguyen, T.D., and Cung, A., "A statistical approach to downscaling of sub-daily extreme rainfall processes for climate-related impact studies in urban areas." Water science and technology: water supply, 2007, Vol.7 .No. 2, pp. 183-192.

[12] Nigam, R., Nigam, S., and Mittal, S.K., "The river runoff forecast based on the modelling of time series." Russian Meteorology and Hydrology, 2014, Vol. 39, Issue 11, pp. 750-761.

[13] O'Connel, P.E., ARIMA Model in Synthetic Hydrology: Mathematical Model for Surface Hydrology. John Wiley, NY, USA. 1977.

[14] Otache Y. Martins, M. A. Sadeeq, I. E. Ahaneku, "ARMA Modelling of Benue River Flow Dynamics: Comparative Study of PAR Model." Open Journal of Modern Hydrology, 2011, Vol. 1, pp. 1-9.

[15] Shakeel, A.M., Idrees, A.M., Naeem H.M., and Sarwar, B.M., "Time Series Modelling of Annual Maximum Flow of River Indus at Sukkur" Pakistan Journal of Agricultural Sciences, 1993, Vol. 30, No. 1.

[16] Sherring, A., Hafizishtiyaq, A., Mishra, A.K., and Mohd A.A, "Stochastic Time Series Model for Prediction of Annual Rainfall and Runoff for Lidder Catchment of South Kashmir." Journal of Soil and Water Conservation, 2009, Vol. 8, No. 4, pp. 11-15.

[17] Srikanthan, R., McMohan, T.A., and Irish J.L., "Time series analysis of annual flows of Australian streams." Journal of Hydrology, 1983, Vol. 66, pp. 213-226.

[18] Stojković, M., Prohaska, S., and Plavšić, J., "Stochastic structure of annual discharges of large European rivers." J. Hydrology and Hydromechanics, 2015, Vol. 63, No. 1, pp. 63–70.

[19] Tian, P., Zhao, G.J., Li, J., and Tian, K., "Extreme Value Analysis of Stream flow Time Series in Poyang Lake Basin, China." Water Science and Engineering, 2011, Vol.4, No.2, pp. 121-132.

[20] Vijayakumar, N., Vennila, S., "A comparative analysis of forecasting reservoir inflow using ARMA model and Holt winters exponential smoothening technique." International Jour. of Innovation in Science and Mathematics, 2016, Vol. 4, No. 2, pp. 85-90.