

Forecasting by Stochastic Models to Inflow of Karkheh Dam at Iran

Karim Hamidi Machekposhti ^a, Hossein Sedghi ^{b*}, Abdolrasoul Telvari ^c,
Hossein Babazadeh ^d

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

^b Department of Water Sciences and Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

^c Department of Civil Engineering, Islamic Azad University, Ahvaz, Iran.

^d Department of Water Sciences and Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Received 14 April 2017; Accepted 30 May 2017

Abstract

Forecasting the inflow of rivers to reservoirs of dams has high importance and complexity. Design and optimal operation of the dams is essential. Mathematical and analytical methods use for understanding estimating and prediction of inflow to reservoirs in the future. Various methods including stochastic models can be used as a management tool to predict future values of these systems. In this study stochastic models (ARIMA) are applied to records of mean annual flow Karkheh river entrance to Karkheh dam in the west of Iran. For this purpose we collected annual flow during the period from 1958/1959 to 2005/2006 in Jelogir Majin hydrometric station. The available data consists of 48 years of mean Annual discharge. Three types of ARIMA (p, d, q) models (0, 1, 1), (1, 1, 1) and (4, 1, 1) suggested, and the selected model is the one which give minimum Akaike Information Criterion (AIC). The Maximum Likelihood (ML), Conditional Least Square (CLS) and Unconditional Least Square (ULS) methods are used to estimate the model parameters. It is found that the model which corresponds to the minimum AIC is the (4, 1, 1) model in CLS estimation method. Port Manteau Lack of fit test and Residual Autocorrelation Function (RACF) test are applied as diagnostic checking. Forecasting of annual inflow for the period from 2006 to 2015 are compared with observed inflow for the same period and since agreement is very good adequacy of the selected model is confirmed.

Keywords: Conditional Least Square; ARIMA Models; Box-Jenkins Model; Karkheh Dam.

1. Introduction

The application of statistical hydrology in earlier days was restricted to surface water problems, especially for analyzing the hydrologic extremes such as floods and droughts. However, during past three decades, the statistical domain of hydrology with the advent of fast computing technology has broadened to encompass the problems of both surface water and groundwater systems. With such a broad domain, statistics has emerged as a powerful tool for analyzing hydrologic time series. The main aim of time series analysis is to detect and describe quantitatively each of the generating processes underlying a given sequence of observations. In hydrology, time series analysis is used for building mathematical models to generate synthetic hydrologic records, to forecast hydrologic events, to detect trends and shifts in hydrologic records, and to fill in missing data and extend records.

Forecasting the inflow of rivers to reservoirs of dams has high importance and complexity. Design and optimal operation of the dams is essential. Mathematical and analytical methods use for understanding estimating and inflow to reservoirs. The analysis of inflow to reservoirs of dams based on such methods has the relative error and uncertainty

* Corresponding author: h.sedghi1320@gmail.com

➤ This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

in engineering science. One of the main reasons for the failure of such systems, there is uncertainty in estimating the parameters or design criteria water structures. To quantify the main sources of uncertainty can be a first step in the analysis performance of a system. Due to the limited water resources, optimum management of water resources is the most important task policy-makers and engineers. One of the most important elements of water resources management is prediction of these resources in the future. The important of these predictions are in the decisions in order to manage water resources. So far, many efforts have been made to design predictive models, including ARMA statistical models and another models based regression. Over the last few decades, predicting future values of hydrologic systems have come to the focus of the researchers consideration for planning and management of water resources. For this purpose, various methods including stochastic models can be used as a management tool to predict future values of these systems.

Stochastic methods are based on time series, and in some areas, we are faced with a lack of historical data. The methods proposed by Box-Jenkins approach (1976) have been used more widely, which are based on the combination of autoregressive moving average (ARMA) methods. Stochastic models are widely used in forecasting river flow, water quality parameters, rainfall and other hydrologic phenomena [4]. Time series analysis approach has been used to develop parametric models for (daily, weekly, monthly, yearly and etc.) data by many researchers like [15]. We think the Hydrologists do not like to use time series models for forecasting annual values of river flow, especially for annual peak and maximum values. Because they think that these models do not appropriate method for forecasting. And they love used to the hydrology distributions.

Shakeel et al. (1993) applied Time Series Modeling of annual maximum flow of River Indus at Sukkur India. They find that ARIMA (2, 1, 1) was appropriate for this series [13]. AlMasudi (2011) fitted seven seasonal multiplicative models for forecasting of monthly inflow to Dokan reservoir [3]. Srikanthan et al. (1983) used time series models to analyze annual flow of Australian streams. Auto-correlation and partial auto-correlation functions were applied to determine the appropriate form of Box-Jenkins time series models [14]. O'Connell (1977) studied the Auto-regressive moving average (ARMA) models to generate synthetic flow series and found that ARMA (1, 1) was better to preserve the characteristics of stream flow series [11]. Stojković et al. (2015) studied stochastic structure of annual discharges of large European rivers. They suggest that the stochastic flows simulated by the model can be used for hydrological simulations in river basins [16]. Valent et al. (2011) used nonlinear time series models for analysis the nitrate concentrations in River Ouse and Stour situated in the eastern England [17]. Mahmood (2000) applied AR (1), AR (2) and ARIMA (1, 1, 1) models in analyzing monthly time series of water quality data on the Euphrates river at Kufa city [7]. Nguyen et al. (2007) suggested a statistical approach to downscaling of sub daily extreme rainfall processes for climate related impact studies in urban areas. They used annual maximum precipitation data at 15 raingage stations in Quebec (Canada) for the 1961-1990 Period [9]. Gargano et al. (2016) used a stochastic model for daily residential water demand [5]. Also the river runoff forecast based on the modeling of time series used to by Nigam et al. (2014) [10]. Pisarenko et al. (2005) used the Statistical methods for river runoff prediction and the result showed that, in the case of prediction for one time step for time series of mean monthly water discharge, the simpler auto-regression model is more efficient [12].

Vijayakumar and Vennila (2016) suggested that ARMA (2, 4) is the best model for generated annual inflow of Krishnagiri Reservoir in the state of Tamilnadu, India [18]. Abd Saleh (2013) studied the mean flow to Haditha reservoir in the middle west of Iraq using Box-Jenkins modeling (ARIMA models) for period from water year 1999/2000 to water year 2008/2009. It is found that the ARIMA (0, 1, 2) (0, 1, 1)₁₂ model is suitable for forecasting mean flow in future [1]. Lilly (2014) applied time series analysis to generate a monthly synthetic stream flow for Krishna River at Srisaillam dam site, India [6].

Musa (2013) studied flow discharge from the Shiroro River (about 22 years (1990-2011)) by an Autoregressive Moving Average model (ARMA) model and analyzed with three different models namely; Autoregressive (AR) model, Autoregressive Moving Average (ARMA) model and Autoregressive Integrated Moving Average (ARIMA) model. Based on the model analysis and evaluations, proper predictions for the effective usage of the flow from the river for farming activities and generation of power for both industrial and domestic us were made. It also highlights some recommendations to be made to utilize the possible potentials of the river effectively [8].

Adeli et al. (2015) used stochastic models to produce artificial time series and inflow prediction in Talog Dam Reservoir, Khuzestan Province, Iran. The results of Modeling showed that ARMA (2, 3) model was the best model in comparison with the other models [2].

In this study, we used statistical models such as ARMA and ARIMA models to predict Karkheh river flow at the entrance to Karkheh dam in jelogir majin hydrometric station. Similar research results indicate the ability and advantage of statistical models to predict the flow of the river. We tried to demonstrate the ability of these models for predict the annual flow of a river.

1. Materials and Methods

1.1. Study Area and Data Collection

Karkheh Dam is a multi-purpose hydro-development designed to control the Karkheh river flow in the interests of irrigation, electric power generation and for partial accumulation of extreme Karkheh river inflows into Karkheh dam. Karkheh Dam was constructed on the Karkheh River in the West of Iran 24 kilometer upstream from Andimeshk town in Khuzestan province at northwest Iran. In 2001 the project was completed. The project generates (400 Mw) of electrical power from performing its flood control function. Khuzestan province of Iran gets the benefit of irrigation water from its reservoir. Figure 1. shows the study area location. The means of inflows to Karkheh dam for the entire period of record of 48 years from 1958 to 2005 in jelogir majin hydrometric station are plotted as shown in Figure 2. This data were taken from Iran Water Resources Management Organization (IWRMO).

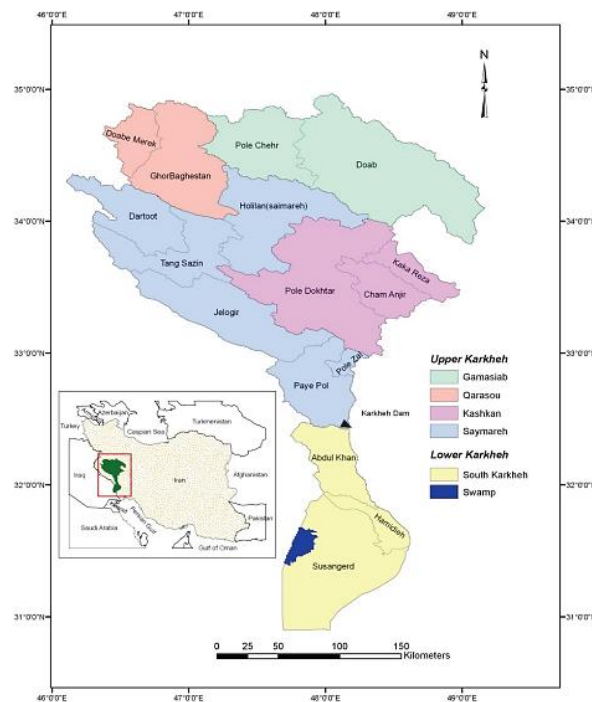


Figure 1. The study area location

1.2. Time Series Modeling

A series of random variables ordered in time is called a stochastic process. By this we mean that observation $X_t, t=1, 2, \dots$ is presumed to be a realized value of some random variable X_t . If the time series $\{X_1, \dots, X_t\}$ is a discrete equidistant univariate series of consecutive observations (real data observed at one place), then x_t is the observed value of the time series at time t . So this series is a single realization of a stochastic process $\{X_1, \dots, X_t\}$. We will use the term "time series" to refer both to the observed data and the stochastic process. The length n of the time series is the total number of observations. It is natural to assume that the generating mechanism is probabilistic and to model time series as stochastic processes. In order to make any kind of statistical inference from a single realization as a random process, the stationarity of the process is often assumed. Intuitively, a process $\{X_t\}$ is stationary if its statistical properties do not change over time. A stochastic process $\{X_t\}$ is called a Gaussian white noise process when each random variable X_t has a normal distribution with a mean value $E(X_t) = \mu = 0$, a constant variance $D(X_t) = \delta^2$ for all t and $\text{CORR}(X_t, X_s) = 0$ for $t \neq s$.

A time series can be viewed as a realization of a stochastic process. Many problems related to water resources and environmental systems deal with temporal data which need to be analyzed by means of a time series analysis, which has become a major tool in hydrology. It is used for building mathematical models to describe hydrological data, forecast hydrologic events, detect trends, provide missing data, etc.

A time series often exhibits trends, sometimes shifts (jumps), seasonality and Periodicity. These attributes are referred to as components.

1. *Trend (T)*: In general, natural, human, economic and other processes produce gradual trends. A trend is a long-term component that represents a growth or a decline of a time series over an extended period of time.

2. *Seasonal Component (S)*: This term of seasonality is used for time series defined at time intervals which are fractions of a year. It is a pattern of change that repeats itself from year to year.

3. *Cyclical Component (C)*: Changes in time series sometimes show a wavelike fluctuation around a trend, which shows the possible existence of periodicity with longer intervals.

4. *Irregular Component (e)*: This is a part of a time series represented by residuals, after the above-mentioned components have been removed.

The following additive model of these components was applied:

$$X_t = T_t + S_t + C_t + e_t \quad (1)$$

The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past. Due to the indispensable importance of time series forecasting in numerous practical fields such as business, economics, finance, science and engineering, etc., proper care should be taken to fit an adequate model to the underlying time series. It is obvious that a successful time series forecasting depends on an appropriate model fitting. A lot of efforts have been done by researchers over many years for the development of efficient models to improve the forecasting accuracy. As a result, various important time series forecasting models have been evolved in literature.

In this study we wanted to examine the ability of annual data instead daily or monthly or other measures such as weekly. If we can demonstrate that annual data of hydrologic phenomenon such as river flow is appropriate for prediction then we can easily predict flooding and reduce flood damages. In order to controlling of data in this station, we must have checks three conditions adequacy, accuracy and relevance. We use run test method for homogeneity and Water Resources Council (WRC) test for outlier detection of data. Our discharges data have all above conditions and there is no outlier data. Therefore this data are suitable for estimation by ARIMA models.

1.3. Autoregressive Moving Average Models

The basic assumption of modelling a time series is that the value of the time series at time t , X_t , depends only on its previous values (the deterministic part) and on a random disturbance (the stochastic part). Furthermore, if this dependence of X_t on the previous p values is assumed to be linear, we can write:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + D_t \quad (2)$$

Where $\{\phi_1, \phi_2, \dots, \phi_p\}$ are real coefficients and D_t is a disturbance at time t ; it is usually modeled as a linear combination of the zero mean, uncorrelated random variables or a zero-mean white noise process $\{Z_t\}$:

$$D_t = Z_t + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_q X_{t-q} \quad (3)$$

($\{Z_t\}$ is a white noise process with the mean 0 and the variance δ^2 if and only if $E(Z_t) = 0, E(Z_t^2) = \delta^2$ for all t , and $E(Z_s, Z_t) = 0$ if $t \neq s$, where E denotes the expectation).

Z_t is often referred to as a random error or a noise at time t . The constants $\{\phi_1, \phi_2, \dots, \phi_p\}$ and $\{\theta_1, \theta_2, \dots, \theta_q\}$ are called autoregressive (AR) coefficients and moving average (MA) coefficients respectively, for the obvious reason that (2) resembles a regression model and (3) a moving average. Combining (2) and (3), we get

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_q X_{t-q} \quad (4)$$

This defines a zero-mean autoregressive moving average (ARMA) process of p and q orders, or ARMA(p, q). One of the deficiencies of ARMA model is that it is not able to reproduce in a satisfactory manner the trend, seasonality, and periodicity component of a time series. For these reason we used the Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA model integrates autoregressive (AR), moving average (MA) models and also integrated component (I), that allows using this methodology for modeling non-stationary time series. According to Box and Jenkins time series is sequence of random variable value that indicates time series have a stochastic or random character. We can try to find in this stochastic character time patterns. The basic assumption made to implement this model is that the considered time series is linear and follows a particular known statistical distribution, such as the normal distribution. ARIMA model has subclasses of other models, such as the Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) models. The popularity of the ARIMA model is mainly due to its flexibility to represent several varieties of time series with simplicity as well as the associated Box-Jenkins methodology for optimal model building process. But the severe limitation of these models is the pre-assumed linear form of the associated time series which becomes inadequate in many practical situations. To overcome this drawback, various nonlinear stochastic models have been proposed in literature; however from implementation point of view these are not so straight-forward and simple as the ARIMA models.

1.4. Long Memory Models

In the natural sciences time series generated by non-stationary stochastic processes are often analyzed. The non-stationarity can be caused either by an in time changing mean value or an in time changing variance. Process y_t is called the integrated of order d , $I(d)$

$$(1 - b)^d y_t = a_t \quad (5)$$

Where d is the integer, B is the backward operator ($B y_t = y_{t-1}$) and a_t is a stationary and ergodic process with a positive limited spectrum. Process y_t which verifies (5) for $0 < d < 1$, is called a fractional integrated process of order d . The autocorrelation function (ACF) of stationary processes ARMA falls exponentially with a rising time-lag \Rightarrow correlation of outlying stochastic variables is statistically nonsignificant. But in hydrology we know about time series made of stationary processes, where outlying variables are strongly correlated. In 1951, Hurst first reminded us of this phenomenon. A time series with this attribute is called a series with long memory and is generated by stochastic processes called processes with a long memory. It is characterized by a decrease of their ACF values with a growing lag, which is hyperbolic rather than exponential. A time series with a long memory is modeled with Autoregressive Fractional Integrated Moving Average (p, d, q) models (ARFIMA), which are a combination of fractional differentiation and a linear ARMA (p, q) process. An important attribute is that the effect of parameter d on outlying variables falls hyperbolically but the influence of the AR and MA processes falls exponentially. The correlation structure for large time-lags is characterized by d , and all the other parameters characterize the correlation structure in small time-lags. Even though ARFIMA (p, d, q) for $0.5 < d < 1$ is a non-stationary process, time series is generated tightened to the process mean value. By testing a time series for a long memory the null hypothesis H_0 is tested: the process has a short memory. Next, the Hurst coefficient H containing the parameter d is estimated to standardize the time series. At first the series is adapted to a $12 \times N$ matrix (N is the number of years) with the elements Y_{ij} , $i=1, \dots, 12$, $j=1, \dots, N$. After accounting for the mean value \bar{X}_i and the standard deviation \hat{S}_i for each month or year, the normalized values of the standardized series are calculated as

$$M_{ij} = \frac{Y_{ij} - \bar{X}_i}{\hat{S}_i} \quad (6)$$

The Hurst coefficient H is the slope of the regression curve

$$\ln[E(R_T/S_T)] \approx \ln c + H \ln(T) \quad (7)$$

Where the left side of the equation presents the observed data and the right side presents the regression curve. From the estimation of H we get an estimation of the fractionalization parameter $\hat{d} = \hat{H} - 0.5$. If d is in the interval $(0, 0.5)$, the standardized time series is generated by stationary and invertible long memory processes. Next, the time series is differenced and modeled with a linear ARMA.

1.5. Parameter Estimation

Several methods are available for estimating parameters including Maximum Likelihood (ML), Conditional Least Squares (CLS) and Unconditional Least Squares (ULS). So there are many methods and criteria for selecting the order of an AR or ARMA model available. One of them is based on the so-called information criteria. The idea is to balance the risks of under-fitting (selecting an order smaller than the true order) and over-fitting (selecting an order larger than the true order). The order is chosen by minimizing a penalty function. The two most commonly used functions are:

1. The Akaike's information criterion:

$$AIC(p, q) = \ln \hat{\sigma}^2 + \frac{2(P + q)}{n} \quad (8)$$

2. The Bayesian information criterion:

$$BIC(b, q) = \ln \hat{\sigma}^2 + \frac{2(P + q) \ln(n)}{n} \quad (9)$$

($\hat{\sigma}^2$ Is a sample covariance function) The most commonly used approach for checking the model's adequacy is to examine the residuals Calculated from the estimated model. If the fitted model is the "true" model, the residuals should behave like a white noise process with a zero mean and constant variance. We used the Portmanteau test, which is based on the following statistic,

$$Q_h = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_z(k)}{n-k} \quad (10)$$

Which it has an asymptotic X^2 distribution with h-p-q degrees of Freedom. If $Q_h > X_{1-\alpha}^2(h-p-q)$ the adequacy of the model is rejected at Level α .

2. Results and Discussion

2.1. Fitting Box-Jenkins

The non-seasonal ARIMA (p, d, q) model, which is called "non-seasonal autoregressive integrated moving average model" or Box-Jenkins non-seasonal model, is fitted to a time series by using a three stages procedure. These stages are model identification, estimation of model parameters and diagnostic checking of the estimated parameters. First of all the observations are plotted against time this will show up important features such as trend, seasonality, discontinuities and outliers. Figure 2. show that there is little increase trend for annual inflow. The procedure of fitting is summarized by the following steps:

1. Transform data using natural log transformation which was found the most appropriate.
2. Removing trend component by using the first order differencing.
3. Model identification by plotting ACF and PACF of annual observations.

2.2. Identification of Representative Models

By identification it is meant the use of the data, and of any information on how the series was generated, to suggest a subclass of models from the general Box- Jenkins family, i.e., Equation 1 for further examination. In other word, identification provides clues about the choice of the order of p, d and q. However, in practice the degree of differencing d is assumed 1 while ACF and PACF are plotted to guess the order of p and q. The ACF and PACF of the natural mean annual flow show in Figure 3. The suggest 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. The estimation autocorrelation and partial autocorrelation function as shown in Table 1. The results estimation of parameters in Table 1. shows that all selected models are suitable for modeling and then go to the next stage.

2.3. Estimation of the Model Parameters

The maximum likelihood (ML), conditional least square (CLS) and unconditional least square (ULS) methods are used to estimate the model parameters. The result of values for the parameters of three models (1, 1, 0), (1, 1, 1) and (4, 1, 1) show in Table 1. Table 1 show that all three models are suitable for modeling this set data. Therefore we enter the next stage that is diagnostic check stage.

2.4. Diagnostic Check

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:

$$Q = (N - d - DS) \sum_{k=1}^M r_k^2 a_t \quad (11)$$

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 $p > \chi^2$ is greater than 0.05 where 0.05 is the level of significant. The results of this test indicate in Table 2. show that all three models are adequate for annual data.

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) are shown in Figure 4. The figure show that the most of computed lags lie inside the tolerance interval ($\pm 2/\sqrt{N}$, at 95% confidence limits). Hence, the Suggested model can be considered as appropriate model because of its capability of removing the dependency from data.

The goodness of fit statistic is shown in Table 3. Therefore, the ARIMA (4, 1, 1) model in CLS estimation parameter method is the best model for mean annual inflow to Karkheh dam.

2.5. Forecasting

Forecasted annual data are computed for 10 years a head of original data for the period from 2006 to 2015 by applying the ARIMA (4, 1, 1) model in CLS estimation parameter method. After obtaining the forecasted series (Z_t for $t=1, 2, 3 \dots 10$), then the final series (X_t) is determined by reversing (ln) transformation. The forecasting actuals were brought in Table 4. Figure 5. shows the forecasted series for these data. The corresponding observed values are also shown in the Figure 6. and since agreement between observed and forecasted values is very good ($R^2=0.72$), it is confirmed that the model is adequate.

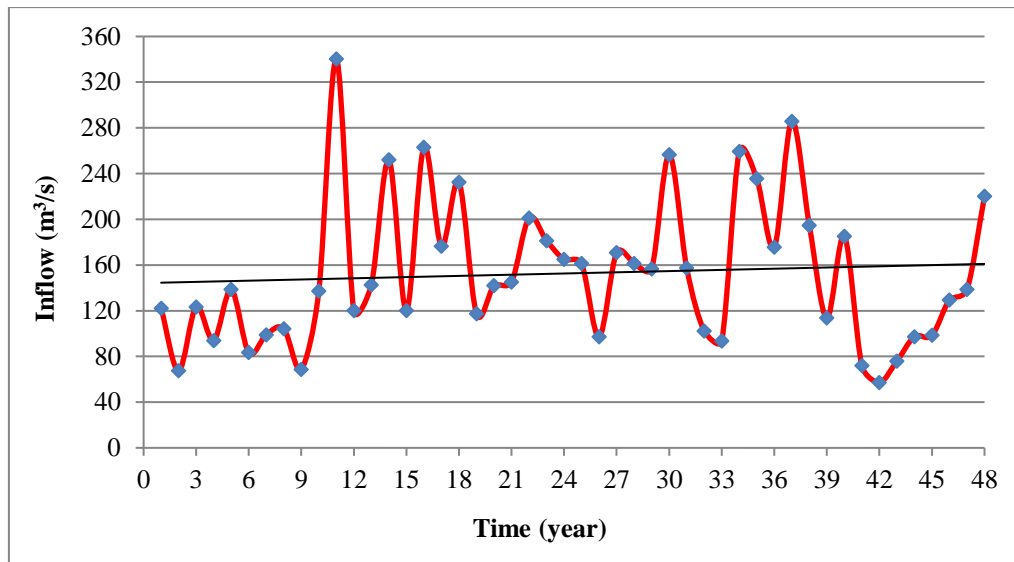


Figure 2. Annual Observed Inflow to Karkheh Reservoir from 1958 to 2005

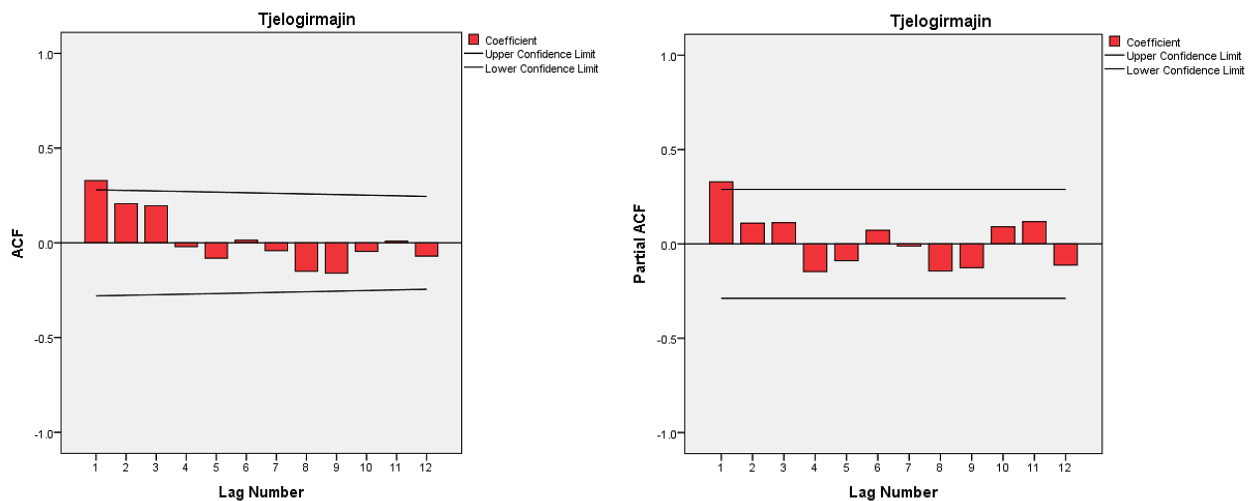


Figure 3. Autocorrelation and Partial Autocorrelation Function for the Series

Table 1. Result of parameter estimation for the selected model

| Parameter | Estimation Method | Type (Order) and Values of Parameters ARIMA(p,1,q) | Std. Error | Absolute Value of t | Probability of t | Stationary Condition | Invertibility Condition |
|------------------|-------------------|----------------------------------------------------|-------------------|---------------------|------------------|----------------------|-------------------------|
| Mean Annual Flow | ML | p(1) = -0.41932 q(0) | 0.13457 | -3.12 | 0.0018 | Satisfy | |
| | CLS | p(1) = -0.41473 q(0) | 0.13576 | -3.05 | 0.0037 | Satisfy | |
| | ULS | p(1) = -0.42872 q(0) | 0.13480 | -3.18 | 0.0026 | Satisfy | |
| | ML | p(1) = 0.36414 q(1) = 0.99994 | 0.15423 52.247 | 2.36 0.02 | 0.0182 0.9847 | Satisfy | Not Satisfy |

| | | | | | | |
|-----|-----------------------------------|--------------------|---------------|-------------------|---------|-------------|
| CLS | p(1) = 0.34486 q(1) = 0.96630 | 0.15246 0.04271 | 2.26 22.62 | 0.0286 0.0001< | Satisfy | Satisfy |
| ULS | p(1) = 0.35777 q(1) = 0.99998 | 0.14375 0.29826 | 2.49 3.35 | 0.0166 0.0016 | Satisfy | Not Satisfy |
| ML | p(4) = -0.1636 q(1) = 0.66032 | 0.15570 0.12124 | -1.5 5.45 | 0.2933 0.0001< | Satisfy | Satisfy |
| CLS | p(4) = -0.17507 q(1) = 0.68987 | 0.15914 0.11569 | -1.1 5.96 | 0.2771 0.0001< | Satisfy | Satisfy |
| ULS | p(4) = -0.18146 q(1) = 0.68935 | 0.15870 0.11617 | -1.14 5.93 | 0.2589 0.0001< | Satisfy | Satisfy |

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

Table 2. Result of Autocorrelation Check of Residuals

| Parameter | ARIMA Model | Estimation Method | To Lag | Df | Chi-Square | Pr>Chi-Square | Adequacy for Modelling |
|------------------|----------------|-------------------|--------|--------|------------|---------------|------------------------|
| Mean Annual Flow | ARIMA (1, 1,0) | ML | 6 | 5 | 8.04 | 0.1540 | Satisfy |
| | | | 12 | 11 | 11.80 | 0.3792 | |
| | | | 18 | 17 | 17.84 | 0.3990 | |
| | | | 24 | 23 | 26.29 | 0.2875 | |
| | | CLS | 6 | 5 | 7.87 | 0.1636 | |
| | | | 12 | 11 | 11.83 | 0.3764 | |
| | | | 18 | 17 | 17.88 | 0.3963 | |
| | | | 24 | 23 | 26.24 | 0.2898 | |
| | | ULS | 6 | 5 | 8.08 | 0.1517 | |
| | | | 12 | 11 | 11.87 | 0.3733 | |
| | | | 18 | 17 | 17.97 | 0.3906 | |
| | | | 24 | 23 | 26.52 | 0.2771 | |
| | ARIMA (1,1,1) | CLS | 6 | 4 | 3.17 | 0.5296 | Satisfy |
| | | | 12 | 10 | 4.99 | 0.8920 | |
| | | | 18 | 16 | 8.33 | 0.9386 | |
| | | | 24 | 22 | 11.96 | 0.9583 | |
| | | ML | 6 | 4 | 2.98 | 0.5618 | |
| | | | 12 | 10 | 6.31 | 0.7883 | |
| | | | 18 | 16 | 11.02 | 0.8082 | |
| | | | 24 | 22 | 16.88 | 0.7699 | |
| | | CLS | 6 | 4 | 3.1 | 0.5416 | |
| | | | 12 | 10 | 7.19 | 0.7074 | |
| | | | 18 | 16 | 11.74 | 0.7617 | |
| | | | 24 | 22 | 17.32 | 0.7455 | |
| ULS | 6 | 4 | 3.01 | 0.557 | | | |
| | 12 | 10 | 6.34 | 0.7858 | | | |
| | 18 | 16 | 10.94 | 0.8134 | | | |
| | 24 | 22 | 16.57 | 0.7866 | | | |

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

Table 3. Goodness of fit statistic

| Parameter | ARIMA model | Estimation Method | Akaike's Statistic |
|------------------|-------------|-------------------|--------------------|
| Mean Annual Flow | (1,1,0) | ML | 103.4247 |
| | | CLS | 103.4469 |
| | | ULS | 103.4316 |
| | (1,1,1) | CLS | 93.1350 |
| | | ML | 90.8381 |
| | (4,1,1) | CLS | 88.8680 |
| | | ULS | 91.0387 |

Table 4. Forecasts from period 2006-2007 to 2015-2016 for mean annual flow

| Period | Forecast | Observation |
|---------|----------|-------------|
| 2006-7 | 139 | 135 |
| 2007-8 | 133 | 131 |
| 2008-9 | 131 | 128 |
| 2009-10 | 131 | 130 |
| 2010-11 | 127 | 125 |
| 2011-12 | 132 | 127 |
| 2012-13 | 132 | 131 |
| 2013-14 | 134 | |
| 2014-15 | 132 | |
| 2015-16 | 132 | |

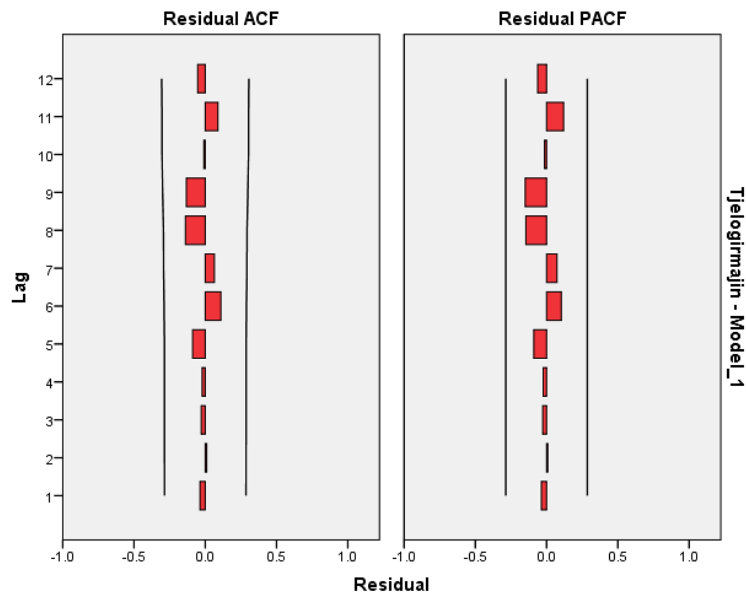


Figure 4. Auto-correlogram of Residual Series Parameter

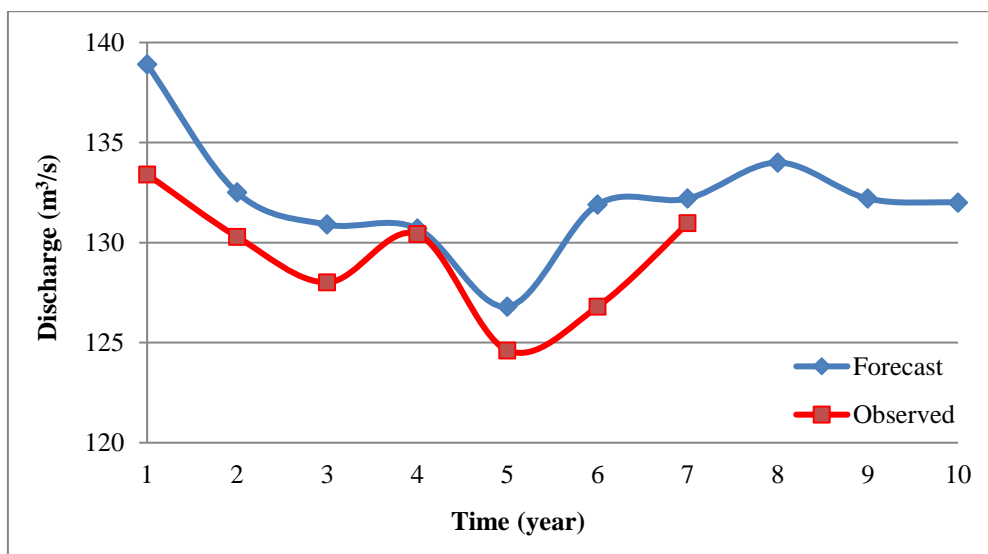


Figure 5. Comparison of Forecasted and Observed Inflow (2006-2016)

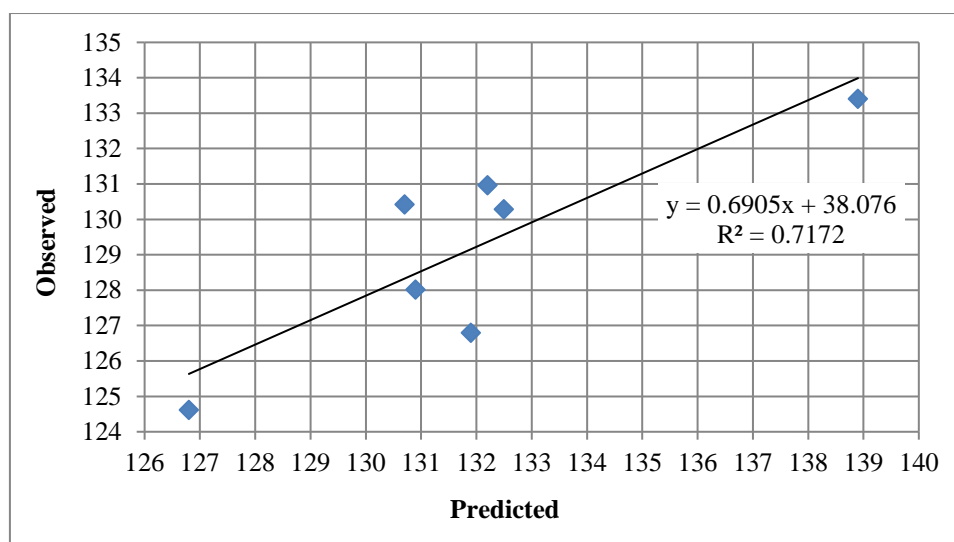


Figure 6. Correlation between actual values and predicted values of inflow to dam reservoir

3. Conclusions

The time series (1958 till 2005) of mean annual inflow to Karkheh Dam is a non-periodic series and therefore, the stochastic model which represents it is a non-seasonal one. The model (4, 1, 1) by CLS estimation parameter method is fitted to the series and it is found that Akaike Information Criterion (AIC) 88.8680 is less than the other models. The diagnostic checking show that model (4, 1, 1) is adequate. Forecast using the model for the period from 2006 to 2015 agrees with the observed values well. The significant ACF and PACF with 4 Order can be caused by factors such as area good vegetation and water from snowmelt. The good vegetation of the region and the forest causes water retention in the soil surface layer and delay in the rise in surface runoff. As well as vegetation reduces the power and erodibility destroyed by a severe storm (intense rain events happening across the region) and runoff from the storm drainage system seems to cause significant delays.

The ARIMA model is suitable for short term forecasting because the ARMA family models can model short term persistence very well. The autoregressive model is a finite memory model, thus it does not fare well in long term forecasting. At the end, the study shows that the Box-Jenkins (ARIMA) model methodology could be used as an appropriate tool to forecast the inflow in Karkheh dam for the up-coming ten years (2006-2015).

4. References

- [1] Abd Saleh, Z., "Forecasting by Box-Jenkins (ARIMA) Models to Inflow of Haditha Dam." Journal of Babylon University, Engineering Sciences, 2013, Vol. 21, No. 5.
- [2] 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.
- [3] AlMasudi, R.K.M., "Fitting ARIMA Models for Forecasting to inflow of Dokan Reservoir." Journal of Babylon University, 2011, Vol. 19, No. 4.
- [4] Box G.E.P. and Jenkins G.M., Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, California, 1976.
- [5] Gargano, R., Tricarico, C., Giudice, G.D., and Granata, F., "A stochastic model for daily residential water Demand." Water science and technology: water supply, available online ws 2016102, 2016.
- [6] Lilly Padmaja Joshi, "Generation of Monthly synthetic flow Sequences for Krishna River at Srisailam Dam Site.", International Conference on Hydrology and Watershed Management, 2014, 2: 28-34
- [7] Mahmood, H.S., Time Series Analysis of Hydrochemical Pollutant: Application of Box-Jenkins Models on Euphrates River at Kufa, M. Sc. Thesis College of Engineering, University of Babylon, 2000.
- [8] Musa, J.J., "Stochastic Modelling of Shiroro River Stream flow Process." American Journal of Engineering Research (AJER), 2013, Vol. 2, No. 6, pp 49-54.
- [9] 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.

- [10] 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.
- [11] O'Connell, P.E., *ARIMA model in synthetic hydrology: Mathematical model for surface hydrology*, John Wiley, NY, USA, 1977.
- [12] Pisarenko, V.F., Lyubushin, A.A., Bolgov, M.V., Rukavishnikova, T.A., Kanyu, S., Kanevskii, M.F., Savel'eva, E.A., Dem'yanov, V.V., Zalyapin, I.V., "Statistical methods for river runoff prediction." *Water Resources*, 2005, Vol. 32, Issue 2, pp. 115–126.
- [13] 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.
- [14] 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.
- [15] Stedinger, J.R., "Estimation correlations in multivariate stream flow models." *Journal of Water Resource Research*, 1981, Vol. 17, No. 1.
- [16] 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.
- [17] Valent, P., Howden, J.K.N., Szolgay J., and Komorníková, M., "Analysis of Nitrate Concentrations Using Nonlinear Time Series Models." *J. Hydrology and Hydromechanics*, 2011, Vol. 59, No. 3, pp. 157–170.
- [18] Vijayakumar, N., Vennila, S., "A comparative analysis of forecasting reservoir inflow using ARMA model and Holt winters exponential smoothing technique." *International Jour. of Innovation in Science and Mathematics*, 2016, Vol. 4, No. 2, pp 85-90.