BEST ARIMA MODELS FOR FORECASTING INFLOW OF HIT STATION

Adnan K. Shathir
College of Engineering, Civil Department
University of Karbala, Iraq.
Adnanshathir@gmail.com

Layla A. Mohammed Saleh
College of Engineering, Civil department.
University of Karbala, Iraq.
Lazu200919.la72@gmail.com

***Abstract*-Time series analysis for hydrological phenomena has an important role in water resources engineering. In this study, seven models of ARIMA family are tested for forecasting the monthly discharge at Hit station on Euphrates river in Iraq. The statistical analyses were done for models with help of IBM SPSS statistics 21 software, The number of observations used is equal to 480 reading, start from October 1932 and end at September 1972, this period represents the near-natural stream flow of the river before the construction of dams in Syria and Turkey. Statistical tests such as T-test and F-test were used to detect any change in Mean and Variance at 95% significant probability level. Results showed that the best model is $(2,0,1)\times(0,1,1)_{12}$ which gives minimum error and good agreement between observed and forecast discharge.**

***Key words*- ARIMA, Box and Jenkins, Forecasting, Hit station, Time series analysis**.

I. INTRODUCTION

Forecasting and time series analysis are very useful in planning, development, design, operation and maintenance of the available water resources. In hydrological forecasting, past data must be analyzed to find the best model which can be adopt to estimate the future water events. There are two types of forecasting models, deterministic or physical models and statistical or stochastic models. In the first approach, theoretical or empirical physical relationships is used to described the hydrological system, there is always a unique coincidence between input and output [1], while in the stochastic approach, the forecasting techniques is the most popular method, it depend upon the time series data which formed from measurements of variables taken at regular intervals over time, the hydrologic data of stream flows fall under the category of time series [2].

The most famous techniques used to forecast the time series phenomena are the Box-Jenkins method 1976, which is based on examining a wide range of models for forecasting a time series. Some of these models are moving average process (MA), Autoregressive process (AR), Autoregressive moving average process (ARMA) and Autoregressive integrated moving average process (ARIMA). The most commonly used stochastic time series models is (ARIMA) model, generally, there are three steps to selecting an appropriate model from a general class of ARIMA models [3] first, preparing initial model from historical data, estimating the tentative model parameters, and test the accuracy of selected model. For more investigation and parameter estimation, the model that appears to represent the behavior of the series is examined through Autocorrelation ACF and partial autocorrelation functions PACF [4]. Many important models have been proposed in literature for studying the accuracy and efficiency of time series modeling and forecasting. P.P. Mujumdar etal. 1990 observed ten models of the (ARMA) family for representing and forecasting monthly and ten-day stream flow in three rivers in Indian, the models selected, based on the min. mean square error and max. likelihood criteria [5]. Jain etal. 2003 examined two types of regression models namely, a linear multi-regression and nonlinear multi-regression models for modeling rainfall-runoff process[6]. Momani 2009 was developed ARIMA (1,0,0) (0,1,1) to forecasting the monthly rainfall of Amman airport station for 10 years upcoming [7]. Ranjbar etal 2014 used ARIMA (2,0,0) model to forecast qualitative parameters (TSS, NO_3 and DO) in two stations of Sefidrud River in Iran [8]. The aim of this study is to forecast the monthly inflow of Hit station on Euphrates in Iraq by using the Box and Jenkins technique with help of IBM SPSS statistics 21 software.

II. STUDY AREA

Euphrates River is the longest river in Western Asia, about 2700 km, it has three riparian countries, Turkey, Syria and Iraq. Three important measurement stations were constructed on the Euphrates River in Iraq, Husaibah, Hit and Al- Hindya barrage. Hit station fig.[1] located at latitude 33° 36' 23" N and longitude 42° 50' 14" E with drainage area about 264,100 square kilometers[9]. After the construction and operation of Keban Dam in Turkey in 1974 and Tabqa Dam in Syria in 1975, the inflow to Hit station is decreased and the natural hydrological regime of the river is change. In 1985 Haditha dam was operated in Iraq, thus, inflow to Hit station

Fig.1 Hit station on Euphrates river in Iraq

became represent the water release from downstream the dam pulse the valley runoff between the dam and the station.

In this paper, the period 1932-1973 was selected because it represents the near-natural stream flow of the river, and there was no effect of human activities on river morphology.

III. METHODOLOGY

ARIMA models are the most well-known of models for time series forecasting. It was introduced by Box and Jenkins (1970). Ageneral ARIMA model is expressed as (p,d,q) where, p is the autoregressive parameters, d is the number of differencing operators and q is the moving average parameter. The general Stochastic models according to Box and Jenkins are [3] -[10] :

A. Autoregressive model AR of order p is :

$$Z_i = \emptyset_i Z_{i-1} + \emptyset_2 Z_{i-2} + \cdots + \emptyset_p Z_{i-p} + a_i \ \ldots\ldots(1)$$

By using the backshift operator B, which defines ($BZ_t = Z_{t-1}$), equation (1) can be written as :

$$(1 - \emptyset_1 B - \emptyset_2 B^2 - \ldots . \emptyset_p B^p) Z_i = a_i$$

where : $\emptyset_1,\ldots,\emptyset_p$ are AR(p) parameters, a_i is the random shock which is independent of z_{i-1} and normally distributed with zero mean and σ^2 variance.

B. Moving Average MA model of q is :

$$Z_i = -\theta_1 a_{i-1} - \theta_2 a_{i-2} \ldots - \theta_q a_{i-q} + a_i \ \ldots\ldots(2)$$

MA model can be written in equivalent form as :

$$Z_i = (1 - \theta_1 B - \theta_2 B^2 \ldots\ldots - \theta_q B^q) a_i$$

where q is the order of MA(q), and θ coefficients are MA(q) model parameters.

C. ARMA (p,q) model :

In this model, both autoregressive and moving average operators are combined.

$$Z_i = \emptyset_i Z_{i-1} + \emptyset_2 Z_{i-2} + \cdots + \emptyset_p Z_{i-p} + a_i - \theta_1 a_{i-1} - \theta_2 a_{i-2} \ldots - \theta_q a_{i-q} \ \ldots\ldots(3)$$

or:

$$(1 - \emptyset_1 B - \ldots . \emptyset_p B^p) Z_i = (1 - \theta_1 B - \cdots . \theta_q B^q) a_i \ldots\ldots(4)$$

D. Autoregressive integrated moving average models ARIMA(p,d,q)

The first of these conditions implies that the series Z_i following eq.(3) is stationary. In practice Z_i may well be non stationary, but with stationary first difference,

$$Z_i - Z_{i-1} = (1 - B) Z_i$$

If (1-B) Z_i is non stationary, the second difference must be taken,

$$Z_i - 2Z_{i-1} + Z_{i-2} = (1 - B)[Z_i(1 - B)] = (1 - B)^2 Z_i$$

By taking the d^{th} difference $(1\text{-}B)^d Z_i$ (although rarely is d larger than 2). substituting $(1\text{-}B)^d Z_i$ for Z_i in eq. (4) yields the general simple ARIMA (p,d,q) model:

$$(1 - \emptyset_1 B - \ldots\ldots - \emptyset_p B^p)(1 - B)^d Z_i = (1 - \theta_1 B \ldots - \theta_q B^q) a_i \ldots . (5)$$

Or : $\emptyset(B)(1 - B)^d Z_i = \theta(B) a_i$

A multiplicative ARIMA (p,d,q) $(P,D,Q)_s$ model is used for time series which exhibit potential seasonal variation such as monthly, where (p,d,q) represent simple part and (P,D,Q) for seasonal part, which can be expressed as [11]:

$$\emptyset_{sp}(B^s)(1 - B^s)^D Z_i = \theta_{sQ}(B^s) a_i$$

where:
D the order of seasonal differencing.
$\emptyset_{sp}(B^s)$ the seasonal autoregressive operator of order P.
$\theta_{sQ}(B^s)$ the seasonal moving average operator of order Q.

A multiplied seasonal ARIMA model can be expressed as[1] :

$$\emptyset_{sp}(B^s)\emptyset_p(B)(1 - B^s)^D (1 - B)^d Z_i = \theta_{sQ}(B^s)\theta_q(B) a_i$$

In non seasonal model, only the notation (p,d,q) is needed.

IV. RESULTS AND DISCUSION

The practical application of stochastic techniques to hydrologic time series may be divided into three steps. First involves data preparation, the second is building the form of the mathematical model and the third is the application or using the adequate model for forecasting. Time series analysis is performed for the historical monthly discharge of the Euphrates river in Iraq at Hit station (IRAQ-E2). The number of observations used in this study is equal to 480 reading start from October 1932 and end at September 1972. The study of the historical man-made activities upstream of the Hit station shows significant events after year 1973 due to construction of new dams in Syria and Turkey. So, the time series before 1973 may be considered no changes in hydrologic characteristics of the time series and the data is homogeneous and the process is time invariant. Plotting the whole time series for this period as shown in Fig.2 doesn't indicate any sudden change which means no trend component in the time series as a first assumption. Statistical tests such as T-test were used to detect any change in mean of the two sub-sample of the series and F-test to detect any differences in variance of the two sub-sample of the series. These two tests show no change in mean and variance of the two sub-sample of the series which is observed at 95% significant probability level.

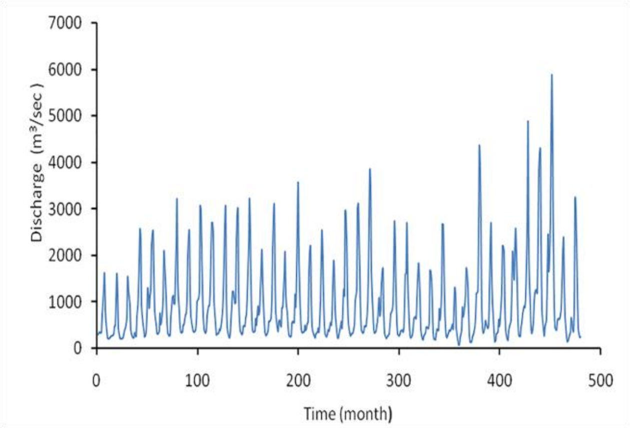


Fig. 2 Monthly discharge to Hit station for the period (Oct.1932 - Sep. 1972).

The Log transformation is adopted to the original raw skewed series to achieve normality, this process may be considered as a step in data preparation.

The multiplicative ARIMA model was selected to model the data because this model is adequate for both stationary and no stationary time series and it is quite suitable for forecasting future values of seasonal series. The series mean, variance, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are the principal tools in model identification procedure.

Figures from 3 to 6 indicate the estimated ACF and PACF for different order of differencing sample and seasonal for lag time equal 48 months.

From fig.3 it is clear that (d = 0, D = 0) does not remove the seasonality in the series. Since the seasonal differences reduces the seasonality from the series as shown in Fig.4 which implies to use seasonal differencing instead of indifferences series.

As mentioned above, the series mean, variance ,Autocorrelation Function and Partial Autocorrelation Function are the powerful tools in identification techniques. So the intelligent inspection of the ACF and PACF will take the main role in decide proper models and suggested the best model to the series. From table (1), it is clear that all seven models give good results for forecasting the discharge of year 1972 with lag time equal 12 months as shown in Figures 7 to 13 .

To choose the proper model it must take the forecasting values for different years to sure the results are good in both flood and draught years. Also the residual must indicate no relations and considered as white noise (random shock).

So the models are estimated also for subsample from October 1932 to September 1958 and forecasted for year 1959 as shown in figs.14 to 20.

The results show there are good agreement between observed and forecasted values, also the parameters have the same values which mean the stability is very good.

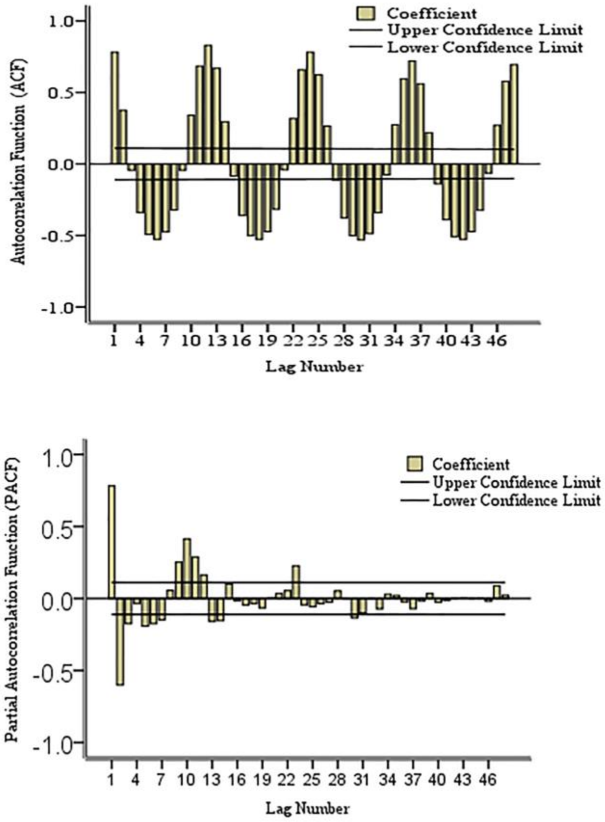


Fig.3 ACF and PACF for monthly discharge without differencing (D=0, d=0)

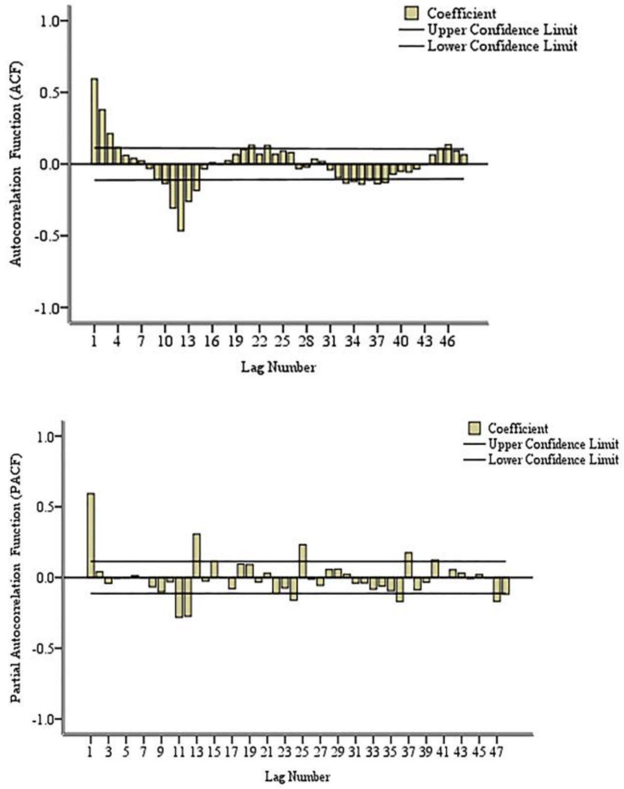


Fig.4 ACF and PACF for monthly discharge with differencing (D=1, d=0).

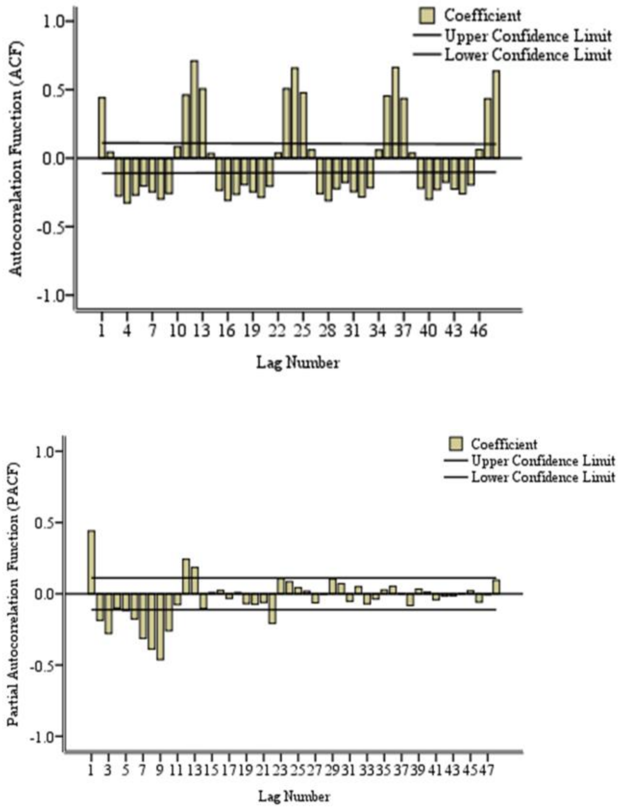


Fig.5 ACF and PACF for monthly discharge with differencing (D=0, d=1).

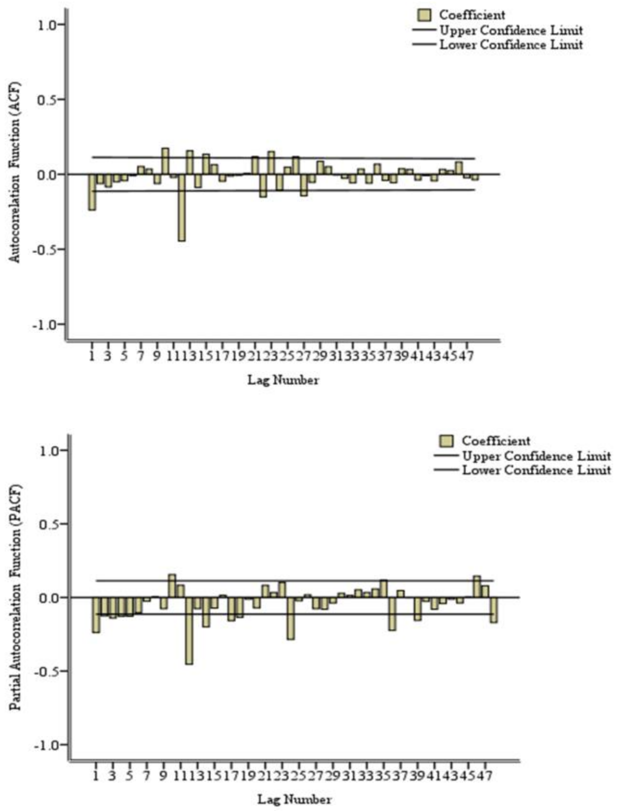


Fig.6 ACF and PACF for monthly discharge with differencing (D=1, d=1).

TABLE 1
THE ESTIMATED PARAMETERS FOR SUGGESTED MODELS

No.	ARIMA Model	Simple AR		Simple MA		Seasonal AR		Seasonal MA	
		Φ1	Φ2	Θ1	Θ2	φs1	φs2	Θs1	Θs2
1	$(1,0,1) \times (0,1,1)_{12}$	0.812	-----	0.133	-----	-----	-----	0.905	-----
2	$(2,0,1) \times (0,1,1)_{12}$	1.556	- 0.576	0.864	-----	-----	-----	0.912	-----
3	$(1,1,1) \times (0,1,1)_{12}$	0.655	-----	0.939	-----	-----	-----	0.915	-----
4	$(1,1,1) \times (1,1,1)_{12}$	0.655	-----	0.939	-----	- 0.020	-----	0.913	-----
5	$(1,1,1) \times (1,2,1)_{12}$	0.678	-----	0.978	-----	- 0.427	-----	0.990	-----
6	$(1,0,2) \times (1,2,1)_{12}$	0.710	-----	0.057	- 0.028	- 0.420	-----	0.998	-----
7	$(2,0,0) \times (1,2,2)_{12}$	0.653	0.082	------	-----	- 0.255	-----	1.360	- 0.380

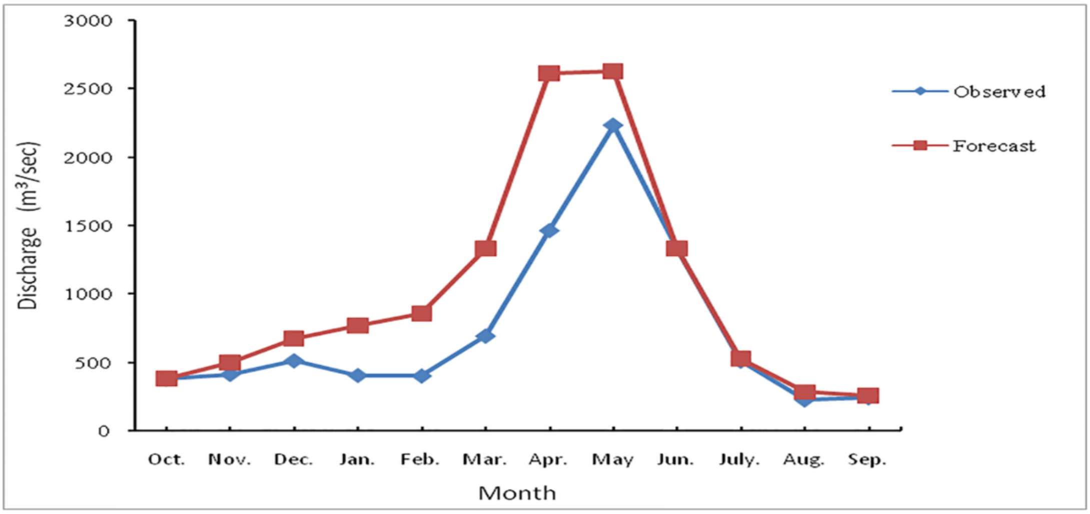
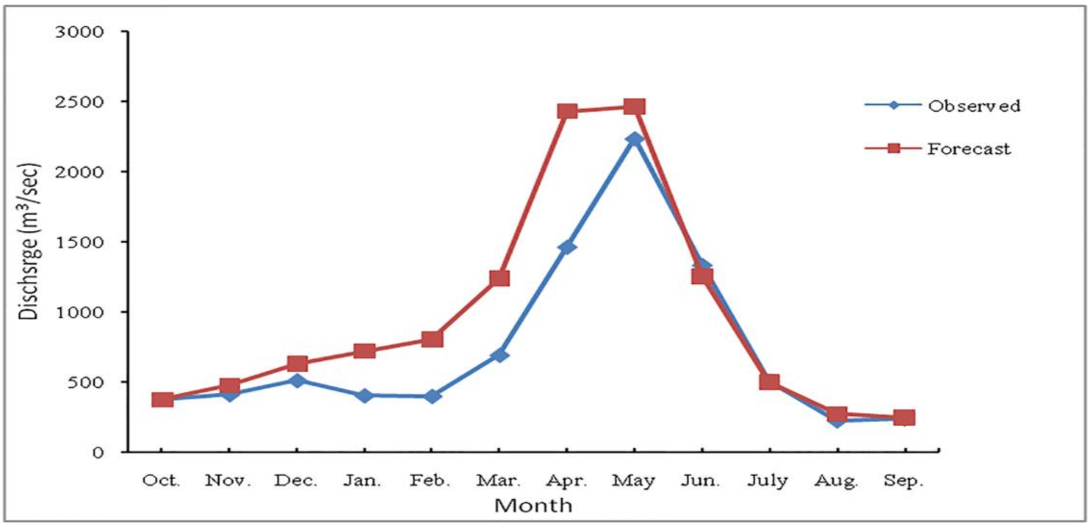


Fig.7 Observed and forecasted discharge of year 1972 for ARIMA (1,0,1) × (0,1,1) model



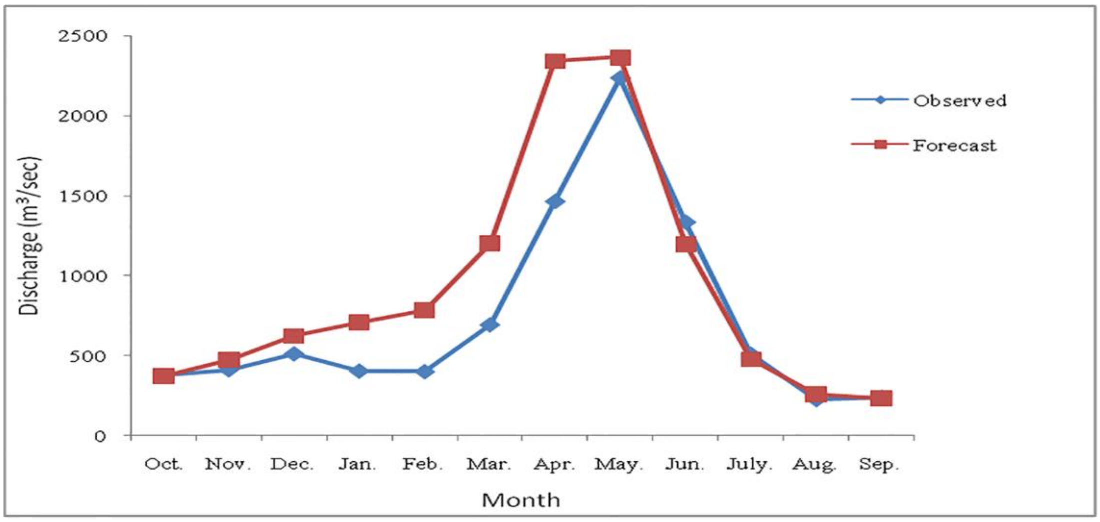


Fig. 9 Observed and forecasted discharge of year 1972 for ARIMA (1,1,1) x (0,1,1) model .

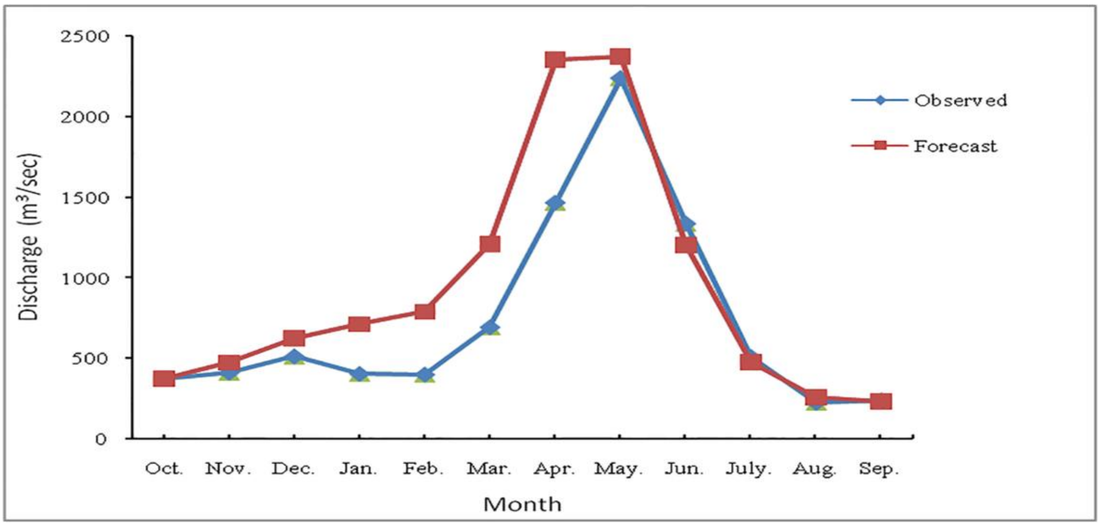


Fig. 10 Observed and forecasted discharge of year 1972 for ARIMA (1,1,1) x (1,1,1) model .

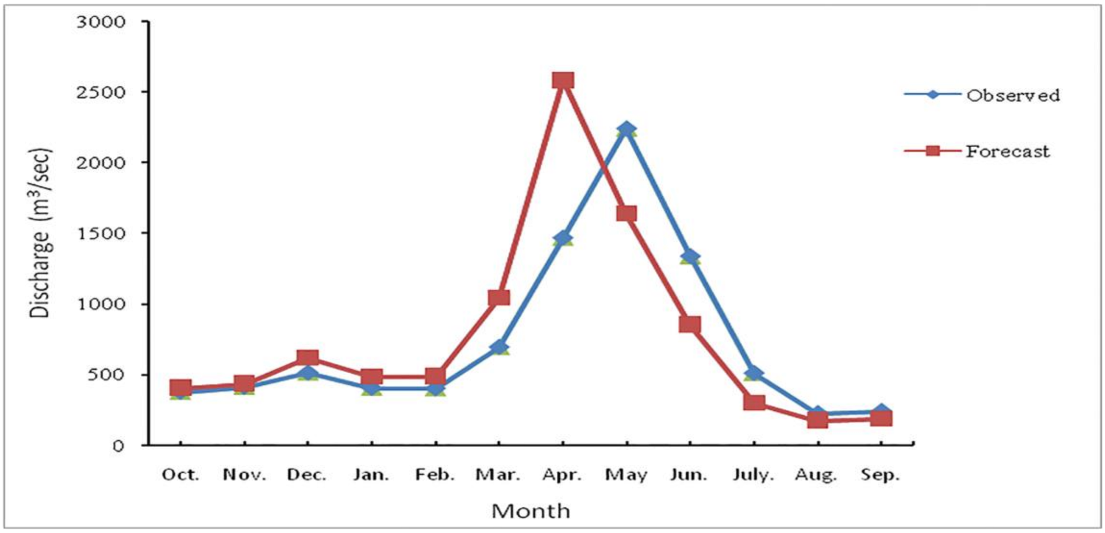


Fig. 11 Observed and forecasted discharge of year 1972 for ARIMA (1,1,1) x (1,2,1) model.

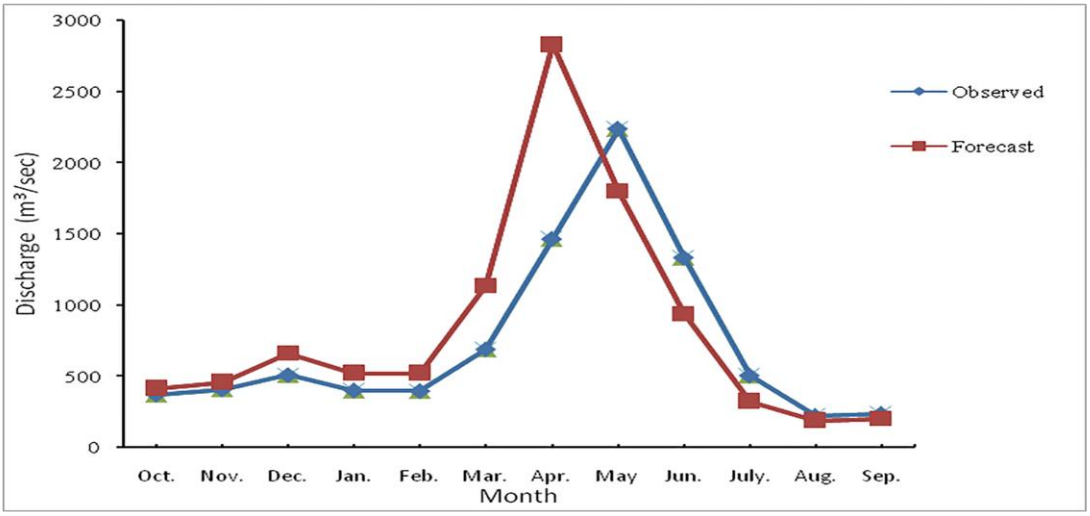


Fig. 12 Observed and forecasted discharge of year 1972 for ARIMA (1,0,2) x (1,2,1) model.

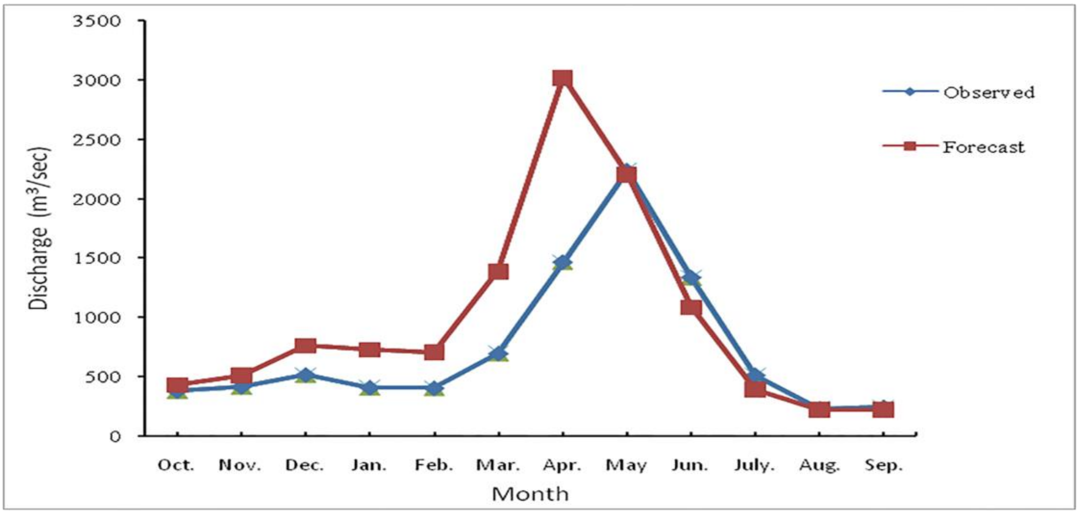


Fig. 13 Observed and forecasted discharge of year 1972 for ARIMA (2,0,0) x (1,2,2) model.

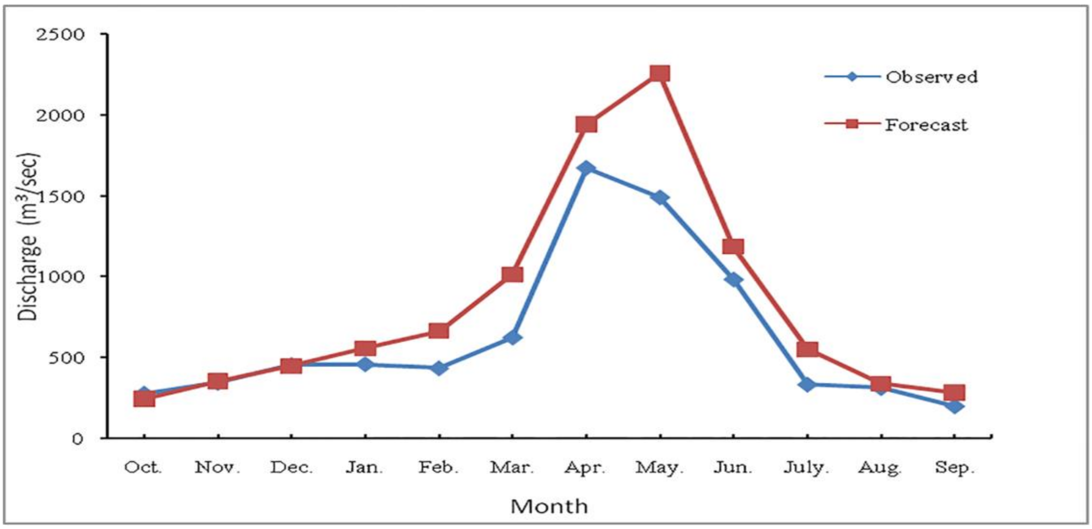


Fig. 14 Observed and forecasted discharge of year 1959 for ARIMA (1,0,1) x (0,1,1) model.

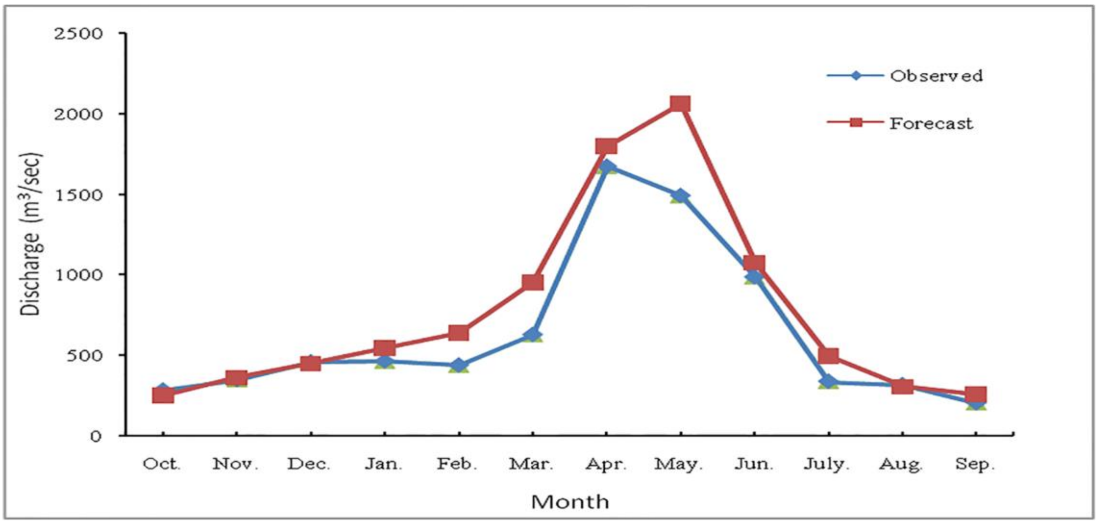


Fig. 15 Observed and forecasted discharge of year 1959 for ARIMA (2,0,1) x (0,1,1) model.

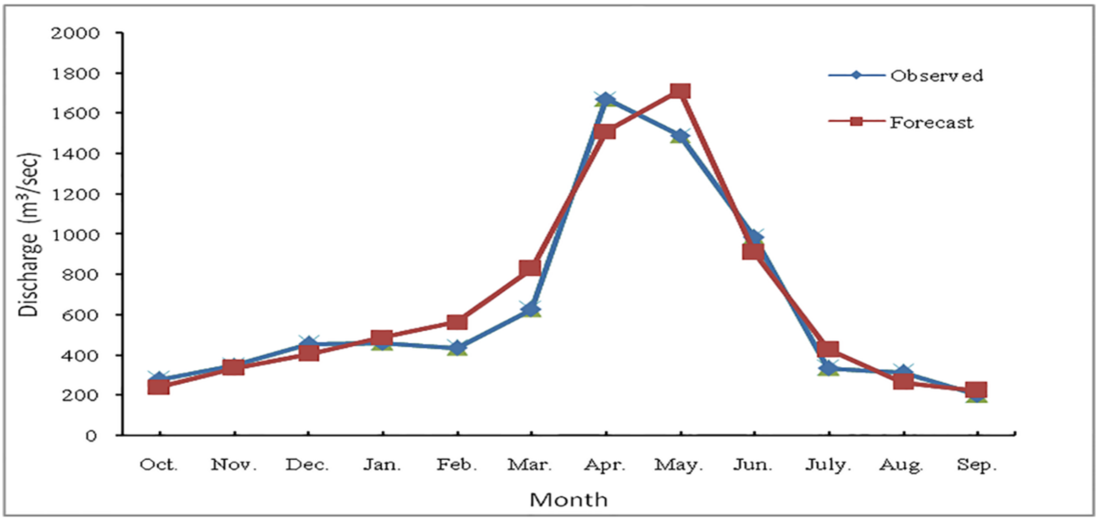


Fig.16 Observed and forecasted discharge of year 1959 for ARIMA (1,1,1) × (0,1,1) model

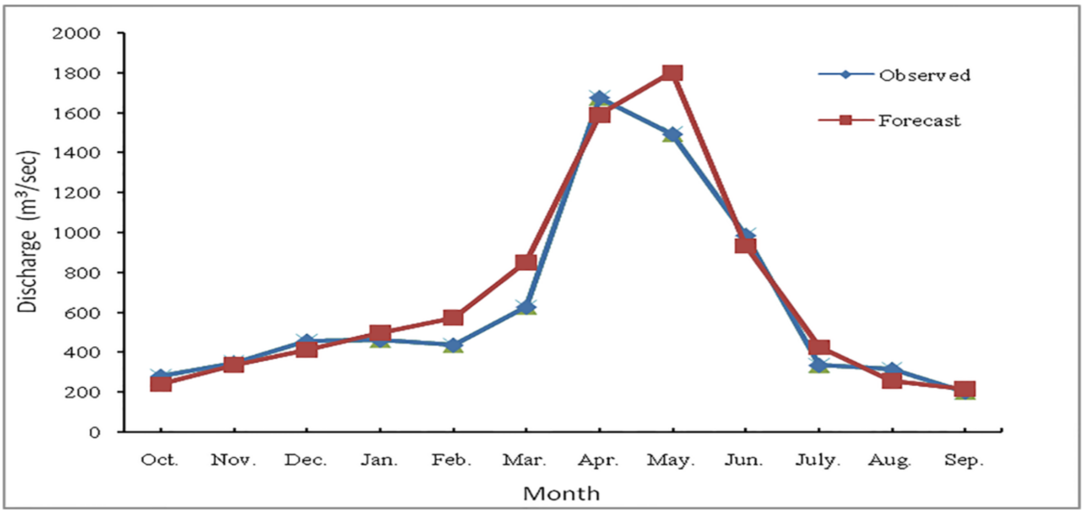


Fig.17 Observed and forecasted discharge of year 1959 for ARIMA (1,1,1) × (1,1,1) model

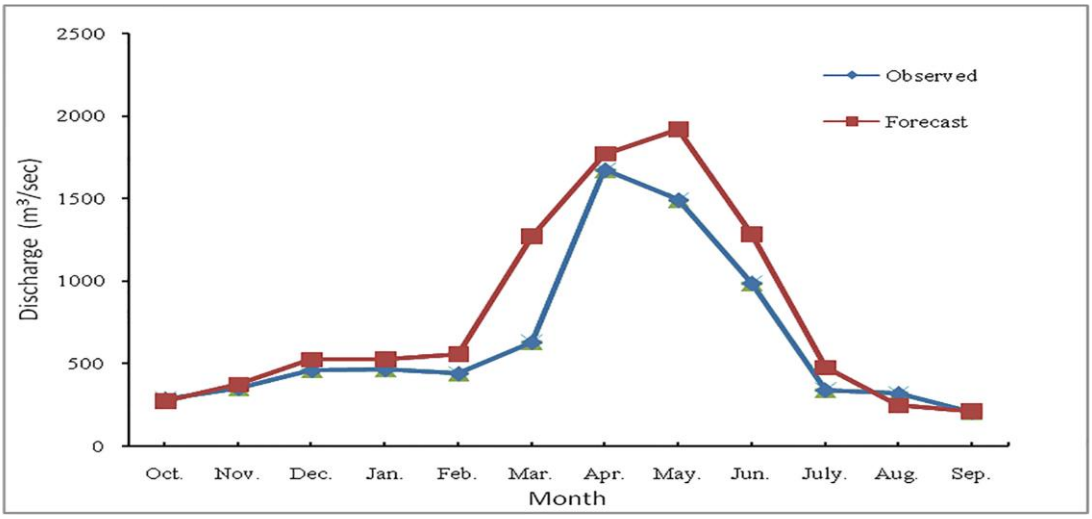


Fig.18 Observed and forecasted discharge of year 1959 for ARIMA (1,1,1) × (1,2,1) model

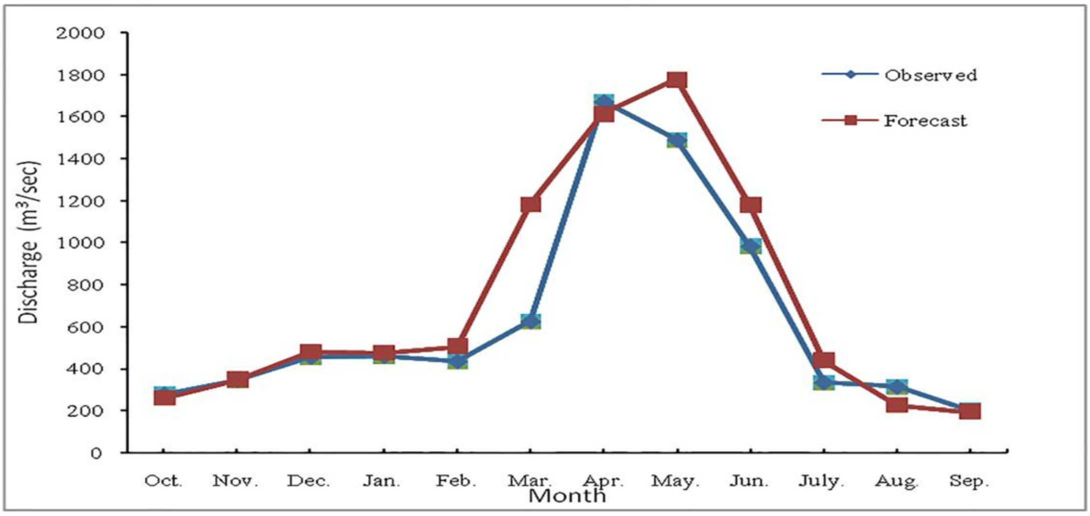


Fig.19 Observed and forecasted discharge of year 1959 for ARIMA (1,0,2) × (1,2,1) model

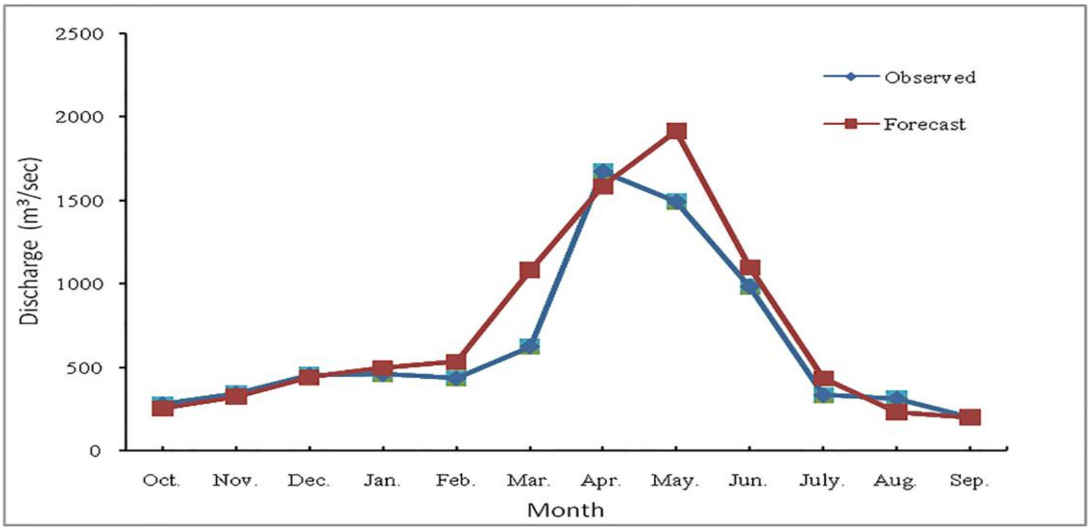


Fig..20 Observed and forecasted discharge of year 1959 for ARIMA (2,0,0) × (1,2,2) model

As shown in table (1), the three models (4, 6, 7) have parameters significantly approach to zero which can be neglected as Akaike suggested [12]. Also the seasonal differencing is too much in model (5), which mean to neglect it. Three models 1,2 and 3 must be taken under focus inspection, so the mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) for three models are listed in table (2).

TABLE 2
ARIMA MODEL STATISTICS

ARIMA Model	RMSE	MAPE	MAE
$(1,0,1)\times(0,1,1)_{12}$	0.254	2.834	0.188
$(2,0,1)\times(0,1,1)_{12}$	0.254	2.802	0.186
$(1,1,1)\times(0,1,1)_{12}$	0.256	2.834	0.188

It is clear from table (2) that ARIMA (2,0,1)×(0,1,1) model is more appropriate and can adopt to explain the natural characteristic properties of the flow due to moving average which explain the fluctuation of the hydrograph, while the recession part can explain the auto regressive model.

V. CONCLUSION

1- ARIMA $(2,0,1)\times(0,1,1)_{12}$ is most suitable model to forecast the monthly discharge at Hit station for recent years with taken in account the trend in flow due the construction of dams upstream the hit station .

2- Nomo graph chart can be draw instead of forecasting equation to account future discharge directly .

VI. REFERNCES

[1] J.D. Salas, J.W. Delleur, V. Yevjevich and W. Lane, "Applied modeling of hydrologic time series," Water Resources Publications, Littleton, Colorado, 1980, p.484.

[2] R.S. Gupta, "Hydrology and hydraulic systems," Prentice Hall, pp. 343-350, 1989.

[3] G. E. P. Box, and G. M. Jenkins, "Time series analysis, forecasting and control," Holden Day, San Francisco, 1976.

[4] A.E. McLeod, K.W. Hipel, and W. C. Lennox, "Advances in Box- Jenkins modeling, applications," Water Resources Research, vol.13, pp.577-586, 1977.

[5] P.P. Mujumdar, and D.N. Kumar, "Stochastic model of stream flow: some case studies," Hydrological Sciences Journal, UK, vol. 35, pp. 395-410, 1990.

[6] A. Jain, and S.K.V.P. Indurthy, "Comparative analysis of event based rainfall-runoff modeling techniques-deterministic, statistical, and artificial neural networks," Journal of Hydrol. Eng, ASCE, vol.8, pp. 93-98, 2003.

[7] P.E. Naill and M. Momani, " Time series analysis model for rainfall data in Jordan; case study for using time series analysis," American Journal of Environmental Sciences, vol. 5, pp. 599-604, 2014.

[8] M. Ranjbar, and Khaledian," Using Arima time series model in forecasting the trend of changes in qualitative parameters of Sefidrud river," International Research Journal of Applied and Basic Sciences, vol. 8, pp. 346-351, 2014.

[9] D.K., Saleh, " Stream gage descriptions and stream flow statistics for sites in the Tigris river and Euphrates river basins, Iraq," Data series 540, US Geological Survey, Reston, Virginia, 154 pp., 2010.

[10] A.KH., Al- Zairjawi, " forecasting model for inflow to Saddam reservoir," Master thesis, irrigation and drainage dep., university of Baghdad, 1988.

[11] C. Lee, and Ko, C., "Short-term load forecasting using lifting scheme and ARIMA models. Expert systems with applications," vol. 38, pp. 590-591, 2011.

[12] A.R. Al-Dabbagh, "Stochastic modeling of the upper Rio Grande system," Ph.D. dissertation, dept. of civil eng, New Mexico state university, 1986.