

Application of Software Packages for Monthly Stream Flow Forecasting of Kangsabati River in India

Manjushree Singh
Research Scholar
AgFE Dept, IIT Kharagpur
WB- 721302, INDIA

R. Singh
Professor
AgFE Dept, IIT Kharagpur
WB- 721302, INDIA

Vipul Shinde
Research Scholar
AgFE Dept, IIT Kharagpur
WB- 721302, INDIA

ABSTRACT

India has made considerable progress as far as creation of irrigation potential is concerned. The gap between irrigation potential created and utilized is a matter of concern. The success of irrigation system operation and planning depends on the quantification of supply and demand and equitable distribution of supply to meet the demand if possible, or, to minimize the gap between the supply and demand. Hence, it is essential to forecast reservoir inflow for proper planning and management of canal irrigation projects. Autoregressive Integrated Moving Average (ARIMA) and X-12-ARIMA are one of the extensively used software packages for time series forecasting. This study focused on the Application of these software packages for Monthly Stream Flow Forecasting of Kangsabati River of India. Here, ARIMA (2, 1, 1) (2, 1, 2) and ARIMA-X-12 (2, 1, 1) (2, 1, 2) models were found to have less Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) and many other statistical values, selected for mean monthly forecasting. In the comparison of ARIMA and X-12-ARIMA models, the X-12-ARIMA model is found more accurate than the ARIMA model for monthly stream flow forecasting. This study suggests that the selected models can be used successfully for monthly stream flow forecasting of Kangsabati river.

Keywords: ARIMA, X-12-ARIMA, Stream flow forecasting, Time series analysis, Diagnostic checks.

1. INTRODUCTION

The pace of advancement of economic super structure of a nation primarily depends on the strength of its agricultural base. India continues to be an agriculture-incentive country with over 70% of population living in rural areas. The agriculture sector provides food to population exceeding a billion, livelihood to nearly two-third of them, raw materials to country's agro-based industries and contributes nearly one sixth of the total earnings of the country. Steering the overall growth of the economy, agriculture sector contributes 22% to the Gross Domestic Product (GDP) [1] (AIC, 2006). Irrigation, the single largest user of the water resources, accounts for about 84% of all withdrawals in India [9] (Planning Commission, 2002). However, with increasing municipal and industrial needs, its share of water is likely to go down. Thus, in future, irrigation has to become efficient and produce more with less water. For this purpose, watershed management with focus on reduction of runoff by designing engineering structures is essential [8] (Nandgude et.al. 2011). The success of irrigation system

operation and planning depends on the quantification of supply and demand and equitable distribution of supply to meet the demand if possible, or, to minimize the gap between the supply and demand. Hence, for proper planning, it is essential to have forecast of inflow to the reservoir.

There are basically two types of model, i.e., process driven and data driven, for estimating reservoir inflow. Process-driven models are based on physical facts of the problem and constituted with combination of some experimental equations. The data driven models are more useful as they can be applied easily and stay away from complicated mathematical models. Common models of this group are regression models, time series models, artificial neural network (ANN) and fuzzy logic (FL). Regression models are frequently used to forecast stream flows as these are simple and easy to use. Graphical techniques developed by [6] Linsley et al. (1975) can be regarded as among the first regression models. Later on, multiple regression model developed by [12] Zuzel et al. (1975) and non-parametric regression model developed by [11] Smith (1975) are notable studies carried out in this subject. Other important data-driven models are time serial models, which are different forms of autoregressive integrated moving average (ARIMA) model. Most frequently used ones of this kind are autoregressive moving average (ARMA), autoregressive (AR), autoregressive integrated moving average (ARIMA), partial autoregressive moving average (PARMA) and seasonal autoregressive integrated moving average (SARIMA) models. Among these models, ARIMA model is the most suitable model for forecasting the inflow. ARIMA model is an extrapolation method for forecasting and like any other such method, it requires only the historical time series data on the variable under forecasting. Several improved version of ARIMA are also available. The most widely used variant X-11 is developed by [10] Shiskin et al. (1967) was further modified as X-11-ARIMA by [4] Dagum(1988). Later on the improved version of X-11-ARIMA i.e., X-12-ARIMA was developed by [5] Findley et al. (1997). X-12-ARIMA provides facility to extend the original series with forecast to ensure that more of the observations are adjusted using the full weighted averages. The initial values can also be forecast backward in time. These forecasts are obtained using ARIMA time series model or regression model with ARIMA errors.

2. STUDY AREA

Kangsabati Irrigation Project, situated in the Western part of West Bengal, India has been chosen as the study area for this research study. Kangsabati dam, built just above the confluence

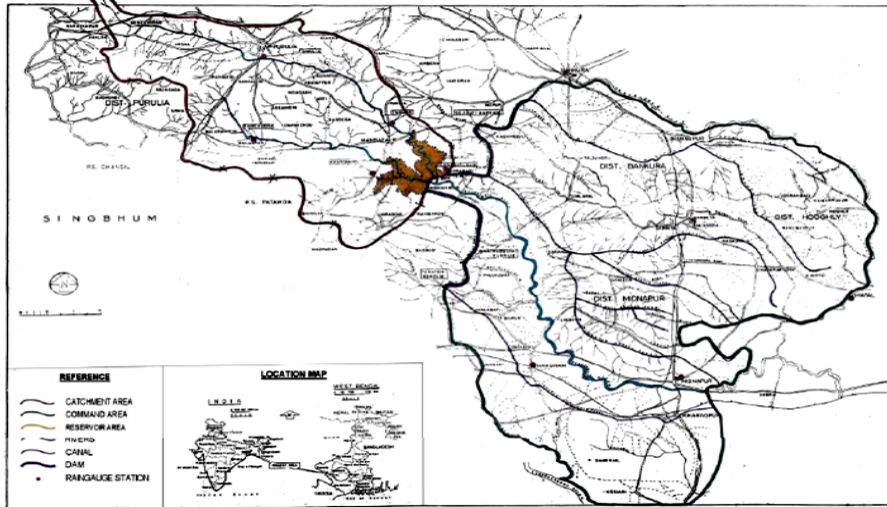


Fig 1: Catchment, reservoir and command area of Kangsabati Irrigation Project.

of Kangsabati and its tributary Kumari, is located at 22° 57' 30" N latitude and 86° 45' 30" E longitude. Kangsabati dam was constructed in two phases. The dam was first constructed on Kangsabati river in 1965. Subsequently, the dam over tributary Kumari was constructed in 1973 and both dam were connected to form a single reservoir, viz., "Kangsabati reservoir". Figure 1 shows both the catchment and command area of Kangsabati reservoir. Total catchment area and gross command area of Kangsabati reservoir are about 3428 sq. km and 5568 sq. km, respectively. The reservoir supplies water to two main canal systems, namely, Right Bank Main Canal (RBMC) system and Left Bank Feeder Canal (LBFC) system. The design discharge of the head regulator at RBMC system, LBFC system and the spillway of the Kangsabati dam are 79.10, 199.55 and 6372 cumec, respectively. The present study considers whole catchment of Kangsabati reservoir for reservoir inflow prediction. Daily inflows to the reservoir from the catchment were collected for the period of 17 years (1987-2003) from the Office of the Superintending Engineer, Irrigation and Waterways Department, Bankura, Govt. of West Bengal, India.

3. METHODOLOGY

3.1 ARIMA Modelling

ARIMA models are the most general class of models for time series forecasting which can be stationarized by transformations such as differencing and logging. It was introduced by [2] Box and Jenkins (1970) which includes autoregressive as well as moving average parameters, and explicitly includes differencing in the formulation of the model. A general ARIMA model is summarized as ARIMA (p, d, q) where, p is the autoregressive parameters, d is the number of differencing passes and q is the moving average parameters. A time series is a set of values of a continuous variable Y (Y₁, Y₂, ..., Y_n), observed over time period t (1, 2, ..., n). In general, in a given time series the following can be recognized and separated.

- (1) **Trend**- It is a regular long-term component of variability that represents the whole evolution pattern of the series.
- (2) **Seasonality**- It is a regular short-term component whose shape occurs periodically at intervals of s lags of the index variable.

- (3) **AR (p)**- An autoregressive component of p order, which relates each value Z_t = Y_t - (trend and seasonality) to the p previous Z values, according to the following linear relationship

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t \quad (1)$$

where $\phi_i (i = 1, \dots, p)$ are parameters to be estimated and ε_t is a residual term.

- (4) **MA(q)**- A moving average component of q order, which relates each Z_t value to the q residuals of the q previous Z estimates

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p} \quad (2)$$

where $\theta_i (i = 1, \dots, p)$ are parameters to be estimated.

The stationary series can be then modelled by an ARMA (p, q) process. The combined use of the ∇ operator and the ARMA (p, q) process, results in an ARIMA (p, d, q) model. Further, ARIMA can account for the seasonal component of s lag period, by using both correlations between Z_t and Z_{t-s} values and those between the corresponding residuals ε_t and ε_{t-s} . Therefore, a seasonal ARIMA model is an ARIMA (p,d,q) model whose residuals ε_t can be further modelled by an ARIMA(P,D,Q)s structure with linear operators (P,D,Q) being functions of the B^s operator.

The operators of a seasonal ARIMA model, defined as ARIMA (p,d,q)(P,D,Q)s, can be expressed by:

AR (p) nonseasonal operator of p order, $\phi(B) = 1 - \phi_1(B) - \phi_2 B^2 - \dots - \phi_p B^p$;

AR (P) seasonal operator of P order, $\phi(B) = 1 - \phi_1(B)^s - \dots - \phi_p B^{ps}$;

MA (q) nonseasonal operator of q order, $\theta(B) = 1 - \theta_1(B) - \theta_2 B^2 - \dots - \theta_q B^q$;

MA (Q) seasonal operator of Q order, $\theta(B) = 1 - \theta_1(B)^s - \theta_2(B)^{2s} - \dots - \theta_Q B^{Qs}$; and $\nabla^d = (1 - B)^d$

3.2 Box-Jenkins methodology for Time-Series modelling

The Box-Jenkins methodology (Box and Jenkins 1970) for time series modelling is characterized by three phases via., model

identification, parameter estimation and diagnostic testing and forecasting (Figure 2). The first two phases of model are

repeated several times to obtain satisfactory model. This model is then used for prediction purposes.

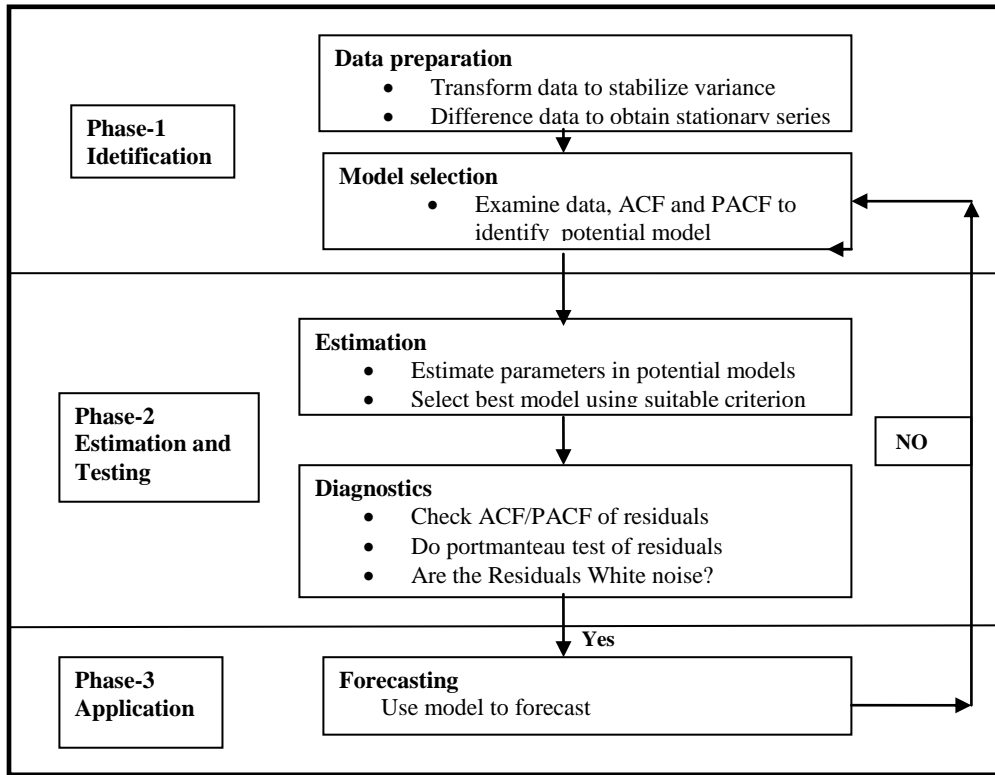


Fig 2: Schematic representation of the Box-Jenkins methodology for time series modelling.

3.2.1 Phase I (Identification)

The input series for ARIMA needs to be stationary. A stationary series should have a constant mean, variance, and autocorrelation through time. The purpose of identification phase is to determine the differencing required for producing stationarity and also the order of nonseasonal AR and MA operators for a given series. When the observed time series presents trend, differencing and transformation are often applied to the data to remove the trend and stabilize variance before an ARIMA model can be fitted.

The autocorrelation function (ACF $\rho(K)$) at lag k of the Z_t series is the linear correlation coefficient between Z_t and Z_{t-k} , calculated for $k=0, 1, 2, \dots$, as

$$\rho_k = \frac{\text{cov}(Z_t - Z_{t-k})}{\sqrt{\text{var}(Z_t) \text{var}(Z_{t-k})}} \quad (3)$$

The major tools used in the identification phase are plots of the series, correlograms of auto correlation (ACF), and partial autocorrelation (PACF). The PACF is defined as the linear correlation between Z_t and Z_{t-k} , controlling for possible effects of linear relationships among values at intermediate lags. Theoretically, both an AR (p) process and an MA (q) process should be associated with well-defined patterns of ACF and PACF. These are usually decreasing exponential or alternate in sign or decreasing sinusoidal patterns. A precise correspondence between ARMA (p, q) processes and defined ACF and PACF

patterns is more difficult to recognize. When the order of at least one of the two components (AR or MA) is clearly detectable, the other can be identified by attempts in the following step of parameter estimation. Finally, the existence of a seasonal component of length s is underlined by the presence of a periodic pattern of period s in the ACF.

3.2.2 Phase II (Estimation and Testing)

After identifying the suitable ARIMA (p, d, q)(P,D,Q)_s structure, subsequent steps of parameter estimation and testing are performed. Estimation stage consists of using the data to estimate and make inferences about parameters of tentatively identified model. The parameters are estimated such that an overall measure of residuals is minimized. The last stage of model building is the testing or diagnostic checking of model adequacy. This stage determines whether residuals are independent, homoscedastic and normally distributed. Several diagnostic statistics and plots of the residuals are used to examine the goodness of fit. After identifying tentative model, the process is again followed by the stage of parameter estimation and model verification. Diagnostic information may help to suggest alternative model(s).

Validation of the goodness of fit of an ARIMA model can be developed according to the following steps:

1) Evaluation of statistical significance of parameters by the usual comparison between the parameter value and the standard deviation of its estimate. For a test statistic that is valid only

asymptotically, a parameter whose value exceeds twice its standard error can be considered significant.

2) Analysis of the ACF of residuals: In this step, residuals ε_t are considered as a new time series, and ACF and PACF are estimated to be sure that values at lag $k > 0$ are not statistically different from zero.

3) Calculation of BIC: In this step, BIC is calculated for all models and according to the lowest BIC value, the appropriate model is selected.

$$BIC(\hat{\sigma}^2) = -2 \ln \times L \left(\hat{\beta}, S \left(\frac{\hat{\beta}}{T} \right) \right) + (p+q+1) \ln T \quad (4)$$

For prediction purposes, ARIMA models are different from the analytical functions of time: $Z_t = f(t)$, because ARIMA forecasting uses previous values of the series and errors in the previous estimates. Actually, this peculiarity of ARIMA forecasting is valid in the short term because parameters of the model cannot account, in the long term, for changes in the dynamics of the series.

3.2.3 Phase III (Application)

After suitable form of ARIMA (p, d, q)(P,D,Q)s is selected and its parameter are estimated, the model is ready for forecasting future events at different lead time. The following equation is used for forecasting.

$$(1-B) \times (1-B^{12}) Y_t = (1-\theta_1 B) (1-\theta_1 B^{12}) e_t \quad (5)$$

3.3 X-12-ARIMA

X-12-ARIMA is the successor to X-11-ARIMA. It is the latest in the family of seasonal adjustment methods that have been developed over several decades by the US Census Bureau and Statistics Canada, with contributions from others. The method is based on a moving-average technique, but is more sophisticated and readily able to provide adjustments tailored to each series. Key characteristics of X-12-ARIMA are:

- i. It contains a time-series-modelling component, which aids the identification of outliers, shifts in the level of a series and calendar effects and the estimation of seasonality at the start and end of the series.
- ii. Filters of several different lengths are readily available to identify the seasonality appropriate to each series.
- iii. For an individual series, it can provide multiplicative – proportionate – seasonal adjustments.
- iv. It includes systematic estimation and removal of calendar effects, for the differing lengths of months, public holidays etc.
- v. It has wide-ranging statistical diagnostics, available graphically where appropriate, enabling the nature, robustness and stability of the seasonal adjustments to be easily monitored.

3.4 Forecasting

For a given regARIMA model with parameters estimated by the X-12-ARIMA program, the forecast spec will use the model to compute point forecasts, and associated forecast standard errors and prediction intervals. The point forecasts are minimum mean squared error (MMSE) linear predictions of future y_t s based on the present and past y_t s assuming that the true model is used—

which means we assume the regARIMA model form is correct, that the correct regression variables have been included, that no additive outliers or level shifts will occur in the forecast period, that the specified ARIMA orders are correct, and that the parameter values used (typically estimated parameters) are equal to the true values. These are standard assumptions, though obviously unrealistic in practical applications. What is more realistically hoped is that the regARIMA model will be a close enough approximation to the true, unknown model for the results to be approximately valid. Two sets of forecast standard errors are produced. One assumes that all parameters are known. The other allows for additional forecast error that comes from estimating the regression parameters, while still assuming that the AR and MA parameters are known. For a reasonably long time series, [2] (Box and Jenkins 1970) the contribution to forecast error of the error in estimating the AR and MA parameters is generally small, thus providing a justification for ignoring this source of error when computing the forecast standard errors. If there are any user-defined regression variables in the model, X-12-ARIMA requires that the user supply data for these variables for the forecast period. For the predefined regression variables in X-12-ARIMA, the program will generate the future values required. If user-defined prior adjustment factors are specified, values for these should also be supplied for the forecast period.

3.5 ARIMA Model for Stream Flow forecasting of Kangsabati Reservoir

To fit an ARIMA and X-12-ARIMA models, a sufficiently large historical data set is required. In this study, inflow data for the period 1986 to 2003 were used. SPSS (Statistical Package for the Social Sciences), which has become a leader in predictive analytics technologies, is used here to implement the ARIMA model. All the steps discussed in previous section are performed on the flow data using the SPSS package. For X-12-ARIMA model, however, software package *gretl* was used. The collected daily inflow data of Kangsabati reservoir (1986 to 2003) was converted into mean monthly data for monthly stream flow forecasting and mean weekly data for weekly stream flow forecasting.

4. RESULTS AND DISCUSSION

4.1 Identification

Stationarity is a necessary condition in building an ARIMA model that is useful for forecasting. Most common method to check stationarity in data series is examining the graph or time plot of the data and ACF and PACF. To test whether the monthly data are stationary or non-stationary, ACF and PACF are determined against different lags. Figure 3 present ACF and PACF plots with upper and lower confidence limits. For a stationary series, all the ACF and PACF should lie within the confidence limit. It is evident from the Figure that both ACF and PACF have high values at certain lags. This shows that differencing is needed for the data sequence.

4.2 Estimation and Testing

ACF and PACF after differencing are plotted in Figure 4. In this case difference of order 1 was sufficient to achieve stationarity in mean. However, at periodic lag of 12, there are peaks, which

suggest seasonality in the data. This shows that, a more complicated mixture ARIMA model is required here.

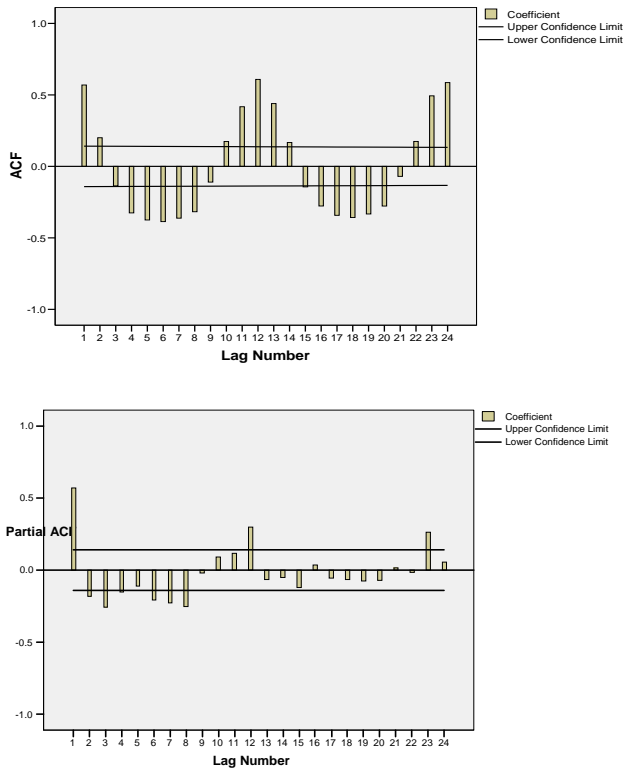


Fig 3: ACF & PACF of the monthly data without differencing.

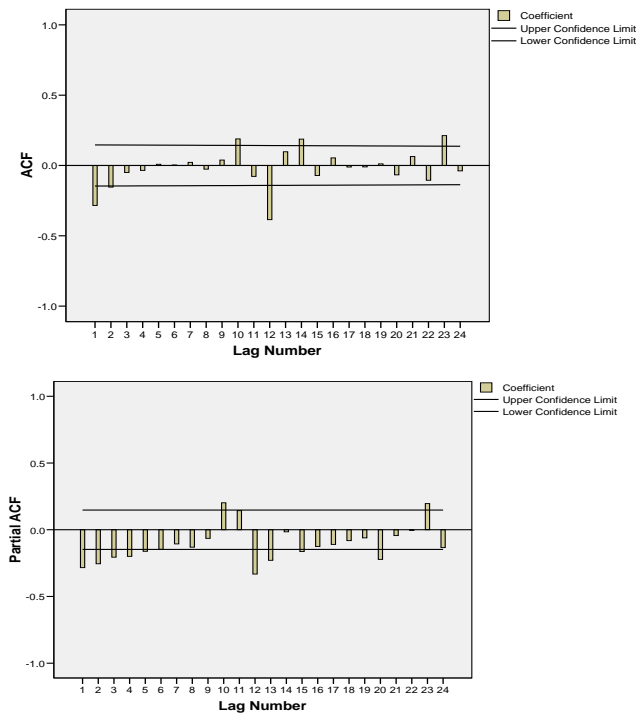


Fig 4: ACF & PACF of the monthly differenced data.

Alternative ARIMA models were estimated by considering the ACF and PACF graphs for the monthly data series. Here, altogether one hundred twenty six ARIMA and X-12-ARIMA models were analyzed. According to the minimum AIC and BIC criteria, twenty two models were selected. AIC and BIC value of the selected twenty two models are presented in Table 1. From the table it is seen that ARIMA (2,1,1)(2,1,2); ARIMA (1,1,1)(2,1,2); ARIMA(2,0,1)(2,2,2) and X-12-ARIMA (2,1,1)(2,1,2); X-12-ARIMA (1,1,1)(2,1,2); X-12-ARIMA (2,0,1)(2,2,2) models have comparatively lower BIC and AIC values. Hence, these models were selected and further diagnostic checks were performed to determine the most suitable model from amongst these six selected models (Table 2). Referring to Table 2, ARIMA (2, 1, 1) (2, 1, 2) and X-12-ARIMA (2, 1, 1) (2, 1, 2) are found to be the most suitable models. Though all the parameter values of the selected models are not less than other models, e.g., RMSE value of X-12-ARIMA (1,1,1) (2,1,2) is less than selected X-12-ARIMA model or R-squared value of ARIMA(2,0,1) (2,2,2) is less than selected ARIMA model and so on, but altogether ARIMA (2,1,1)(2,1,2) and X-12-ARIMA (2,1,1)(2,1,2) models show comparatively lower values for all statistical parameters. To further establish, this conclusion, model verification was performed by examining the autocorrelations and partial autocorrelations of the residuals of various orders. Figure 5 show ACF and PACF of residuals of ARIMA (2, 1, 1) (2, 1, 2) and X-12-ARIMA (2, 1, 1) (2, 1, 2) respectively.

Table 1. ARIMA and X-12-ARIMA models with their corresponding AIC & BIC values

Model (p,d,q)(P,D, Q)	ARIMA		X-12-ARIMA	
	AIC	BIC	AIC	BIC
(2,1,2)(2,1,2)	1876.914	1911.472	1830.604	1900.322
(2,2,2)(2,2,1)	1831.803	1862.429	1821.703	1860.309
(2,1,1)(2,1,2)	1822.768	1853.457	1804.803	1830.703
(1,1,1)(2,1,2)	1822.503	1850.123	1803.604	1825.164
(2,1,2)(2,2,2)	1824.279	1858.037	1812.753	1830.860
(3,1,1)(2,2,2)	1824.340	1858.097	1813.111	1828.612
(2,0,1)(2,2,2)	1821.819	1846.420	1810.529	1830.300
(2,0,2)(2,2,2)	1822.703	1856.530	1820.436	1850.197
(0,0,0)(2,2,2)	1840.061	1861.587	1832.060	1858.295
(2,2,1)(1,2,0)	1938.696	1960.134	1919.135	1958.600
(1,0,1)(1,1,1)	1875.458	1875.458	1864.268	1897.490
(1,0,1)(0,1,1)	1873.665	1892.550	1866.625	1880.200
(1,0,0)(1,1,1)	1873.945	1892.830	1865.196	1886.439
(0,0,1)(1,1,1)	1874.713	1893.598	1863.500	1890.598
(0,0,1)(0,1,1)	1872.875	1888.612	1860.400	1855.612
(0,0,0)(1,1,1)	1891.930	1907.668	1881.490	1907.667
(0,0,0)(0,1,1)	1890.164	1890.164	1850.150	1902.753
(1,1,1)(1,1,0)	1911.321	1930.171	1900.350	1923.560
(1,1,0)(1,1,1)	1916.152	1931.860	1905.543	1913.741
(1,1,0)(0,1,1)	1917.715	1936.565	1905.985	1929.500
(0,1,1)(1,1,1)	1889.293	1908.143	1879.085	1900.350
(0,1,1)(0,1,1)	1887.642	1903.350	1865.250	1892.112

Table 2. Estimates of the fitted ARIMA model

Model statistics	Value (2,1,1) (2,1,2)		Value (1,1,1) (2,1,2)		Value (2,0,1) (2,2,2)	
	ARIMA	ARIMA-X-12	ARIMA	ARIMA-X-12	ARIMA	ARIMA-X-12
Stationary R ²	0.753	0.810	0.672	0.774	0.507	0.772
R-squared	0.753	0.825	0.612	0.773	0.702	0.752
RMSE	42.837	30.625	50.697	29.001	43.832	30.870
MAE	23.376	17.248	28.351	15.118	22.572	17.574
Box-Ljung Statistic	8.454	5.424	9.793	8.974	12.042	9.03
AIC	1822.76	1804.80	1822.50	1803.60	1821.81	1818.52
BIC	1846.42	1830.300	1850.12	1825.16	1853.45	1830.31

Table 3. Comparison of observed and forecasted mean monthly stream flow during December -02 and December-03

Month	Mean Inflow, m ³ /s		
	Observed	Forecasted ARIMA	Forecasted X-12-ARIMA
Dec-02	4.22	12.12164	3.23743
Jan-03	0.00	2.15339	6.2963
Feb-03	1.86	3.410783	1.3065
Mar-03	6.04	5.711706	4.8433
Apr-03	2.60	3.945493	1.7312
May-03	0.00	6.36742	2.94422
Jun-03	16.99	64.7452	12.2429
Jul-03	76.62	147.8672	93.8437
Aug-03	120.98	234.4775	100.6419
Sep-03	52.99	77.35611	44.81356
Oct-03	171.97	200.3414	99.0086
Nov-03	42.46	50.97314	60.2022
Dec-03	3.33	21.11285	18.2366

It is evident from Figure 5, the values of the ARIMA (2, 1, 1) (2, 1, 2) and ARIMA-X-12 (2, 1, 1) (2, 1, 2) residuals lie within the upper and lower confidence limits. Moreover, these values lie between -0.5 to 0.5. In case of other two models, all the values however, do not lie within upper and lower confidence limits. Thus, the ACF and PACF of the residuals also indicate 'good fit' of the ARIMA (2, 1, 1) (2, 1, 2) and X-12-ARIMA (2, 1, 1) (2, 1, 2) models. Further analysis was done with the selected models ARIMA (2, 1, 1) (2, 1, 2) and X-12-ARIMA (2, 1, 1) (2, 1, 2) to check whether the residuals of the model are independent, homoscedastic and normally distributed. The autocorrelation and partial autocorrelation upto 24 lags were computed and their significance was tested using Box-Ljung test. The results indicate that none of these correlations are significantly different from zero at a 95% confidence level. This shows that the selected ARIMA (2, 1, 1) (2, 1, 2) and X-12-

ARIMA (2, 1, 1) (2, 1, 2) models are appropriate models for the monthly mean stream flow forecasting.

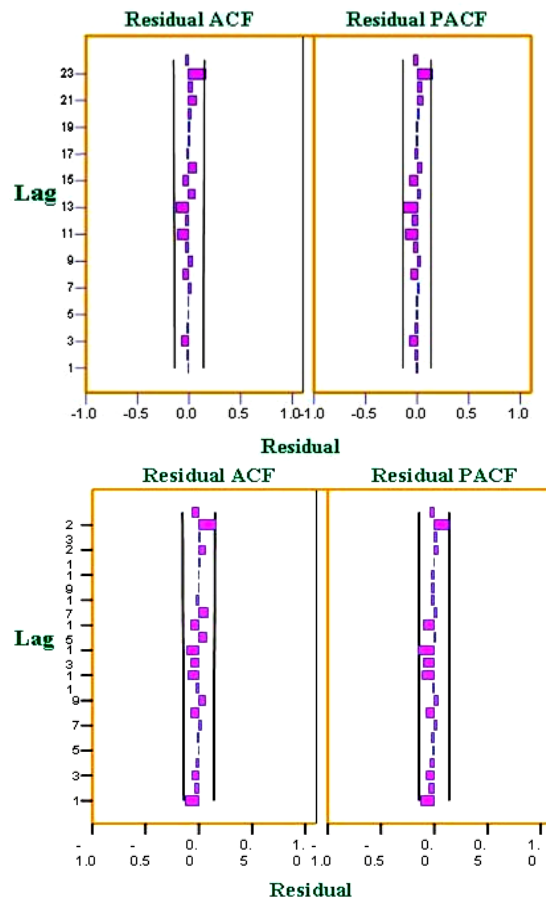


Fig 5: ACF and PACF of residuals of ARIMA (2, 1, 1) (2,1,2) and X-12-ARIMA (2, 1, 1) (2, 1, 2) model.

4.3 Stream flow forecasting with ARIMA and X-12-ARIMA models

The selected ARIMA (2, 1, 1) (2, 1, 2) and X-12-ARIMA (2, 1, 1) (2, 1, 2) were used to forecast the mean monthly stream flow for the period December-02 to December-03 by using the observed data of the period January-86 to December-02. The forecasted data were compared with the observed data (Table 3 and Figure 6). As evident from Figure 8, though both ARIMA (2, 1, 1) (2, 1, 2) and X-12-ARIMA (2, 1, 1) (2, 1, 2) are able to capture the flow trend, X-12-ARIMA (2, 1, 1) (2, 1, 2) performs better than ARIMA (2, 1, 1) (2, 1, 2) in capturing the flow magnitudes. Figure 7 present the scatter plots of forecasted and observed stream flow for ARIMA and X-12-ARIMA respectively. The scatter plots also show that X-12-ARIMA (2, 1, 1) (2, 1, 2) forecasted values are spread uniformly around 1:1 line; whereas ARIMA (2, 1, 1) (2, 1, 2) results in overestimation.

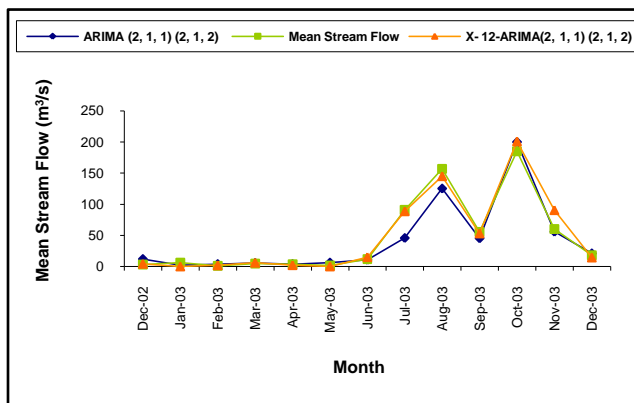


Fig 6: Observed and forecasted mean monthly stream flow value during December -02 to December-03.

After obtaining satisfactory forecasting results over a short period, the selected ARIMA and X-12-ARIMA models were employed to forecast stream flow over a longer period, i.e., January-86 to January-03 (Figure 8). As evident, the X-12-ARIMA model does reasonably well in long-term forecasting except for a few peak flows.

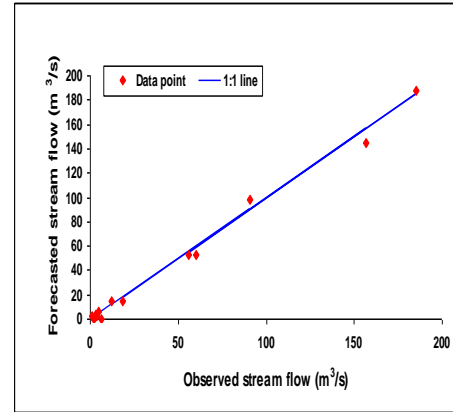
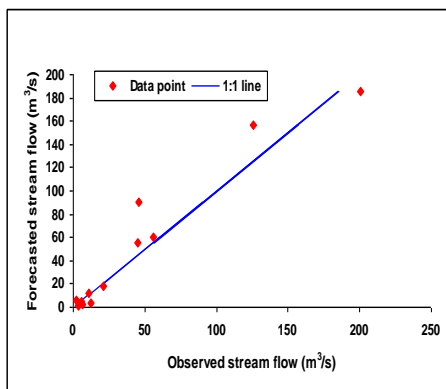


Fig 7: Scatter plot for ARIMA (2, 1, 1) (2, 1, 2) and X-12-ARIMA (2, 1, 1) (2, 1, 2).

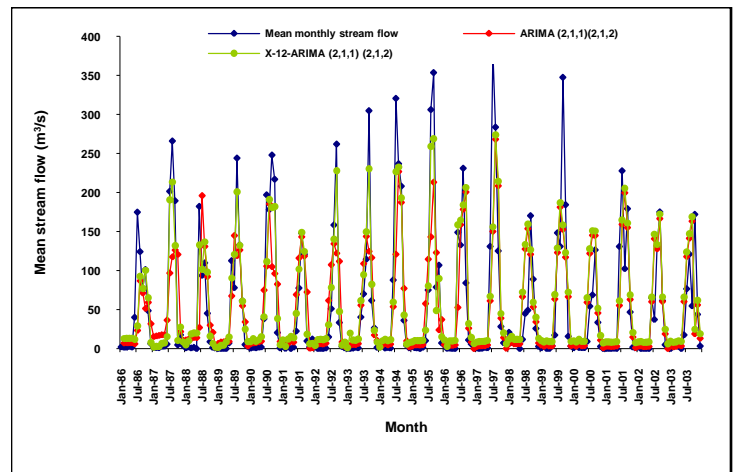


Fig 8: Observed and forecasted mean stream flow forecasting during January -86 and January-03.

5. SUMMARY AND CONCLUSION

ARIMA and X-12-ARIMA are most sophisticated extrapolation method for forecasting. ARIMA model has been popularized by George Box and Gwilym Jenkins in the early 1970's, and there names have frequently been used synonymously with general ARIMA models applied to time series analysis and forecasting. Box and Jenkins (1970) effectively put together in a comprehensive manner the relevant information required to understand and use univariate time series ARIMA model. X-12-ARIMA was developed by Findley et al. (1997). X-12-ARIMA provides facility to extend the original series with forecast to ensure that more of the observations are adjusted using the full weighted averages. The initial values can also be forecast backward in time. These forecasts are obtained using ARIMA time series model or regression model with ARIMA errors.

ARIMA and X-12-ARIMA models offer a sound technique for predicting the magnitude of any variable. The strength of ARIMA lies in the fact that the method is suitable for any time series with any pattern of change and it does not require the forecaster to choose, a priori, the value of any parameter. Its limitations include its requirement of a long time series. Often it is called a 'Black Box' model. Like any other method, this

technique also does not guarantee perfect forecasts. Nevertheless, it can be successfully used for forecasting long time series data. X-12-ARIMA provides facility to extend the original series with forecast to ensure that more of the observations are adjusted using the full weighted averages. In this study, several ARIMA and X-12-ARIMA models were developed and tested for monthly and weekly stream flow forecasting. The models were subjected to various diagnostic tests to evaluate their performance.

The specific conclusions are drawn from this study are.

1. ACF and PACF achieved stationarity after first difference.
2. ARIMA (2, 1, 1) (2, 1, 2) and X-12 ARIMA (2, 1, 1) (2,1,2) are found to be the most suitable models for mean monthly forecasting.
3. X-12 ARIMA (2, 1, 1) (2, 1, 2), however, performs better than ARIMA (2, 1, 1) (2, 1, 2) as ARIMA tends to overestimate the flow magnitude.

6. REFERENCES

- [1] AIC (2006). From the Chairman's desk. Agriculture Insurance Company of India Ltd. <http://aicofindia.nic.in/cmd2.html> (accessed on July 21, 2006).
- [2] Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control, San Francisco, Holden-Day (revised edition, 1976).
- [3] Box, G.E.P., Reinsel, G.C. and Jenkins, G.M. (1994). Time Series Analysis: Forecasting and Control, Prentice Hall, Englewood Cliffs, NJ, 3rd edition, 1994.
- [4] Dagum, E.B. (1988). The X-11-ARIMA/88 Seasonal Adjustment Method – Foundations and Users' Manual. Statistics Canada.
- [5] Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen, B.C. (1997). New capabilities and methods of the X-12-ARIMA seasonal adjustment program (with discussion). *Journal of Business and Economic Statistics*, 16: 127-176.
- [6] Linsley, R.K., Kohler, M.A. and Paulhus, J.L.H. (1958). *Hydrology for Engineers*. Second Edition, McGraw-Hill, New York. (revised edition,1975).
- [7] Lothian, J. and Morry, M. (1978). A set of quality control statistics for the X-11-ARIMA seasonal adjustment method. Research Paper, Seasonal Adjustment and Time Series Staff, Statistics Canada.
- [8] Nandgude, S.B., Solaskar, N.N., Kulkarni, V.U., Bandekar, P.A., Mahale, D.M. (2011). Development of Software for Design of Soil and Water Conservation Structures in Watershed. *International Journal of Computer Applications*, 14 (5): 33-39
- [9] Planning Commission. (2002). Tenth Five Year Plan, Vol. II, Sectoral Policies and Programmes, Section VIII, Infrastructure, Chapter 8.1, Irrigation, Flood Control And Command Area Development. Govt. of India, New Delhi.
- [10] Shiskin, J., Young, A. and Musgrave, J. (1967). The X-11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper 15, Washington, D.C., U.S. Department of Commerce, Bureau of the Census.
- [11] Smith, J.A. (1975). Long-Range Streamflow Forecasting Using Nonparametric Regression. *Water Resources Bulletin*, 27(1): 39-46.
- [12] Zuzel, J.F., Robertson, D.L. and Rawls, W.J. (1975). Optimizing long term streamflow forecast. *Journal of Soil and Water Conservation*, 30(2):76-78.