Planning and Operational Flow Forecasting of Lesser Zaab River at Dokan Dam, Kurdistan Region, Iraq

Sarfraz Munir, Sathyendar Sreepada, Bryar Bakr Saeed

Department of Natural Resources Engineering and Management,
University of Kurdistan Hewler-Erbil.

ARTICLE INFO

Article History:
Received: 14/05/2017
Accepted: 12/04/2018
Published:01/06/2018

Keywords:
Lesser Zaab River
Markov Model
Forecasting
Water Resources Management
Dokan Dam

*Corresponding Author:
Email: s.munir@ukh.edu.krd

ABSTRACT

Flow forecasting is of paramount importance for efficient management of water resources. Information about availability of flows plays a pivotal role in making efficient decisions in water services operation and planning future water developments. Current study is aimed at understanding the behavior of Lesser Zaab River flows at Dokan Dam and forecast future flows accordingly for planning and operational purposes of water services. Lesser Zaab River is one of the main tributaries in Tigris River Basin. Historical flow series data were collected from dam authorities and it is assumed that data follow normal distribution with no significant trend. The First Order Markov Model for stationary time series was applied for mean annual flow forecasting. The Markov Model with periodicity is also applied for mean monthly flows forecasting. Forecasted data were compared with the observed flow series and found a close match between the two series, both for stationary and periodic flows. Based on the study, it can be recommended that the First Order Markov Model for stationary and periodic hydrologic time series can applied for forecasting with confidence.

INTRODUCTION

Knowledge about availability of future flows is one of the main requirements for all water resources planning, management and modelling projects. It provides basis for the development of any future water project, operation of water facilities, flood control and protection works as well as for other related activities. A variety of hydrologic forecasting models are available which are based on (Box-Jenkin Models, 1970) like Markov Model and Auto Regressive Integrated Moving Average (ARIMA) Models. Other forecasting models are also applied for time series data forecasting like moving average, exponential smoothing, regression analysis and advance computational models using artificial intelligence and data mining techniques.

Time series is defined by (Montgomery et al. 2008) as it is a series of time oriented observations on a variable of interest. Classical techniques of flow forecasting are extensively discussed by (Haan 1977 and Gupta 1989). In flow forecasting, the main purpose is to identify the stochastic process that produce the series of data and then forecast accordingly (Akgun, 2003). Analysis of time series is generally performed by determining the auto/serial correlation, which provides the dependence of successive future events on the past events for different lags. Forecasting problems can be classified into short-term,
medium term, and long term forecasting (Montgomery, 2008). Short term may be in the order of days, weeks and months, whereas medium term forecasts can go for one or two years ahead and the long term for several years. Generally, water system operations are dependent upon the short term forecasts and the long term forecast provides basis for long term planning of water resources. Statistical models are used for water supplies and demand forecasts for short term and long term purposes (Fortin et al. 2004). Autoregressive Model of lag-1 (AR(1)) was applied by (Al-Suhaili 1986) for several locations at Tigris River, Iraq. Abed (2007) applied Box-Jenkins seasonal multiplicative model to monthly records of some physical and chemical properties of river water in Babylon, Najaf. Al-Ta’ee (2009) applied ARIMA models to records of rainfall and evaporation at Babylon Governorate. (Muhammad, 2012) compared Markov Model and ARIMA Model for forecasting flow series and found that ARIMA model provided better results.

This study is focused on flow forecasting for operational as well as for planning purposes of Lesser Zaab River at Dokan Dam. First order Markov Model (Haan, 1977) is applied in this study, which depends on lag-1 autocorrelation and first two moments of a time series data, which is assumed to normally distributed and having significant lag-1 autocorrelation. Flows are forecasted on annual basis for next 25 years for planning purposes. At the same time, mean monthly flows are forecasted for next 15 years, which may help in the operation of Dokan Reservoir. These forecasts provide an insight of long term as well short term water availability in the river.

MATERIALS AND METHODS

Study Area

Study was conducted at Lesser Zaab River at Dokan Dam, which is 65 km northwest of the city of Sulaimani. Coordinates of the gauging station are given as 36° 05’ 33” N and 44° 56’ 09” E. Dokan Dam was commissioned in 1959 at Lesser Zaab River. Dokan Dam is a multipurpose dam, primarily for irrigation purpose and also for hydropower and flood control purposes.

Data Collection

Lesser Zaab River inflow data at Dokan Dam were collected for last 17 years (1999-2015) from Dokan Dam Authority. Data on reservoir inflows, releases and hydropower generation was also collected for last five years’ period. Data were collected on daily basis and then the monthly and annual means, standard deviations and other parameters were calculated.

Generating Missing Data Values

There were some values missing in data in the month of February 2002 for two weeks and then in the whole month of September 2003. The missing data were generated by assuming that data follow normal distribution. Values of normally distributed data can be generated by using Equation (1) (Haan, 1977):

\[ X = \mu_X + R_N \sigma_X \]

Where \( X \) is generated value, \( \mu_X \), \( \sigma_X \) are the mean and standard deviation of variable \( X \) respectively, \( R_N \) is standard normal deviate. Standard normal deviate is a random observation from the standard normal distribution. Numerically generated tables of standard random normal deviates are available from Rand Corporation, 1965 (Haan, 1977). This is a classical method of finding random numbers. Now-a-days several computer programmes, including Excel ® have inbuilt functions of generating random numbers as well as random standard normal deviates.

Random Numbers

A random number is defined as a number selected at random from a population of
numbers in such a fashion that every number in the population has an equal chance of being selected. Many large tables of random numbers are available (Rand Corporation, 1965). Computer routines for generating random numbers are included as a part of the programme libraries for most computers. They generally generate random numbers in interval (0, 1). A random number Y, in the interval (a,b) can be generated from random numbers in the interval (0,1), \(R_u\), by following relationship:

\[ Y = (b - a)R_u + a \]  

**First Order Markov Model**

Markov Model for stationary and periodic flow series is adopted from (Hann, 1977).

Most of the hydrologic time series exhibit significant auto correlation. That is, the value of the random variable under consideration at one time period is correlated with the values of the random variable at earlier time periods. The correlation of a random variable \(X\) at one time period with its value \(k\) time periods earlier is denoted by \(\rho_k\) and is called \(k^{th}\) order auto correlation. If \(\rho_k\) can be approximated by \(\rho_k = \rho_1\), then the time series of the random variable \(X\) might be modelled by a first order Markov Process (Haan, 1977).

The first order Markov Process might also be used for a model if serial correlation for lags greater than 1 are not important.

First order Markov Process is given by (Haan, 1977):

\[ X_{i+1} = \mu_x + \rho_1(X_i - \mu_x)e_{i+1} \]  

Where \(X_i\) is the value of the process at time \(i\), \(\mu_x\) is the mean of \(X\), \(\rho_1\) is the first order serial correlation and \(e_{i+1}\) is a random component with \(E(e) = 0\) and \(Var(e) = \sigma^2_e\). The model states that the value of \(X\) in one time period is dependent only on the value of \(X\) in the preceding time period plus a random component. It is also assumed that \(e_{i+1}\) is independent of \(X_i\). The variance of \(X_i\) is given by \(\sigma^2_x\) and can be shown to be related to \(\sigma^2_e\)

\[ \sigma^2_e = \sigma^2_x[1 - \rho^2_1] \]  

If the distribution of \(X\) is \(N(\mu_x, \sigma^2_x)\) then the distribution of \(\varepsilon\) is \(N(0, \sigma^2_e)\). Random values of \(X_{i+1}\) can now be generated by selecting \(\varepsilon_{i+1}\) randomly from a \(N(0,\sigma^2_e)\).

Thus a model for generating \(X\)'s that are \(N(\mu_x, \sigma^2_x)\) and follow the first order Markov Model is:

\[ X_{i+1} = \mu_x + \rho_1(X_i - \mu_x) + t_{i+1}\sigma_x\sqrt{1 - \rho^2_1} \]  

In order to generate values of \(X_{i+1}\), the \(\mu_x, \sigma_x\) and \(\rho_1\) are estimated by \(\bar{X}, S_x\) and \(r_1\) respectively. Then value of \(t_{i+1}\) is selected at random from a \(N(0, 1)\) distribution. Since \(t\) is \(N(0, 1)\), it is possible to generate values of \(X\) that are less than zero. If this occurs, it is recommended that the negative \(X\) be used to generate the next value for \(X\) and then discarded. If occurrence of negative \(X\)'s is common in the generation process, it may indicate that \(X\) is not normally distributed. In this even some other distribution of \(\varepsilon\) must be used.

In order to generate values using Equation 5, it needs to find autocorrelation for lag-1. Note that if \(\rho_1 = 0\), Equation 5 reduces to the independent process of selecting a random observation from \(N(\mu_x, \sigma^2_x)\). On the other hand if \(\rho_1 = 1\), Equation 5 is completely deterministic in that \(X_{i+1}\) is completely specified by

\[ X_i (X_{i+1} = X_i) \]

Serial correlation with lag-\(k\) is given by:

\[ r_k = \frac{\sum_{i=1}^{n-k}(X_i - \bar{X})(X_{i+k} - \bar{X})}{S_x S_x} \]  

**First Order Markov Process with Periodicity**

First order Markov model of the previous section assumes that the process is stationary in its first three moments (mean, standard...
deviation and skewness). It is possible to generalize the model so that the periodicity in hydrologic data is accounted for to some extent. The main application of this generalization has been in generating monthly stream flow where pronounced seasonality in the monthly flows exist. Looking at the plot of mean monthly flow series of Lesser Zaab River at Dokan Dam it is evident that the data in some months are much higher than the other months, in general. The periodicity may affect all of the moments of data as well as the first order auto correlation.

To generalize the Markov Model, the notations are adopted in such a way that subscript i refers to the year under consideration and the subscript j refers to the season within the year. Thus j may run from 1 to 4, if 4 seasons are being considered, 1 to 12 if monthly data are being considered, 1 to 52 for weekly data, etc. In general j is taken as to run from 1 to m, the number of seasons in the year.

With this notation, \( \mu_{i,j} \) refers to the mean of \( X \) in the \( j^{th} \) season and \( \mu_{x,j} \) is estimated by \( \bar{X}_j \) where

\[
\bar{X}_j = \frac{\sum_{i=1}^{n} X_{i,j}}{n}
\]

With \( n \) equal to the number of years of data and \( X_{i,j} \) the data value in the \( j^{th} \) season of the \( i^{th} \) year. Similarly, \( \sigma^2_{x,j} \) is estimated by \( S^2_{x,j} \), \( \gamma_{x,j} \) is estimated by \( C_{S_{x,j}} \) and \( \rho_{1x,j} \) is estimated by \( r_{1x,j} \). Note that \( \rho_{1x,j} \) is the first order serial correlation between values in successive seasons. If monthly stream flow is being considered, \( \rho_{1x,j} \) would be the first order serial correlation between flows in months 4 and 5. \( \rho_{1x,j} \) would be estimated as:

\[
r_{1x,j} = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_{i,j} - \bar{X}_{i,j})(X_{i,j+1} - \bar{X}_{i,j+1})}{S_{x,j} S_{x,j+1}}
\]

First order Markov Model for normally distributed flows becomes:

\[
X_{i,j+1} = \mu_{x,j+1} + \rho_{1x,j} \frac{\sigma_{x,j+1}}{\sigma_{x,j}} (X_{i,j} - \mu_{x,j}) + \sqrt{1 - \rho^2_{1x,j}} \text{random deviates} \]

In any application the population parameters are estimated by the corresponding sample statistics.

### Validation of Forecasted Values

Forecasted values are validated by comparing their statistical properties like mean and standard deviation of observed data. Furthermore, several trials were conducted and based on the minimum root mean square error (RMSE) and mean absolute error (MAE) the current models parameters were selected. Then the validity of the model is decided based on the minimum values of RMSE, MAE and closeness of mean and standard deviations between observed and forecasted values. RMSE and MAE are given as:

\[
RMSE = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_i)^2
\]

and

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}_i| \]

where \( X_i \) and \( \bar{X}_i \) are the observed and forecasted values of the flow series, \( n \) is the total number of observations.

### RESULTS

### Filling Missing Data

Data during months of February 2002 and September 2003 were missing. The missing values were generated by using Equation 1. Standard normal deviates were generated by using Excel ® Software. Random numbers in excel provides cumulative standard normal distribution values, then by taking the inverse of random numbers, the values of random standard normal deviates, \( R_N \), were generated.
Generated values of flow series were compared with the observed data in the same duration of other hydrologic year and found good proximity in generated and observed values of similar durations.

**Annual Flow Series**

Figure 1 presents mean annual flow of Lesser Zaab River at Dokan Dam for 17 years from 1999 to 2015.

![Figure 1. Mean annual flow series of Lesser Zaab River at Dokan Dam from 1999 to 2015.](image1)

Figure shows at first sight a decreasing trend in the data, but it is quite difficult to establish this statement, based on the limited data. Therefore, it is assumed that the data have no significant trend either decreasing or increasing flows. Furthermore, it is assumed that the mean annual flow series data follow a normal distribution.

**Flow Forecasting for Mean Annual Flow**

Mean annual flows are forecasted using Equation 5 of First Order Markov Model. The model needs lag-1 autocorrelation (AC) in addition to first two moments of the distribution (mean and variance). The lag-1 AC is found to be 0.52 for the given flow series data. It also needs standard normal deviates, which are generated using Excel® Software’s inbuilt function to generate random numbers and then taking inverse standard normal distribution of the random numbers. Figure 2 presents a comparison of mean annual flow’s forecasted values with the observed values. A close proximity is found in the observed and forecasted values.

![Fig. 2. Comparison of observed and simulated annual flow series.](image2)

**Table 1. Mean, standard deviation, MAE and RMSE of the observed and forecasted data.**

| Parameter     | Unit   | Observed Data | Forecasted Data |
|---------------|--------|---------------|-----------------|
| Mean          | m³/s   | 118           | 123             |
| Standard Deviation | m³/s  | 53            | 43              |
| MAE           | m³/s   | 28            |                 |
| RMSE          | m³/s   |               | 55              |

Based on the comparison of validation parameters presented in Table 2, the First Order Markov Model can be used for forecasting flow series for future for long term purposes. Fig. 3 presents the 25 year annual flow series based on the above given parameters.

![Fig. 3. Forecasted flows series for 25 years of Lesser Zaab River at Dokan Dam.](image3)

**Lag-I Autocorrelation for Monthly Flows**
Table 2 presents the values of monthly means, standard deviations and lag-1 autocorrelations for 17 year data (1999 to 2015) of Lesser Zaab River at Dokan Dam.

Table 2. Monthly mean, standard deviation and lag-1 autocorrelation of Lesser Zaab River at Dokan Dam.

| Month     | Mean (m³/s) | Standard Deviation (m³/s) | Lag-1 Autocorrelation |
|-----------|-------------|---------------------------|------------------------|
| October   | 34          | 32                        | -0.143                 |
| November  | 88          | 136                       | -0.026                 |
| December  | 112         | 125                       | -0.483                 |
| January   | 160         | 206                       | -0.232                 |
| February  | 269         | 265                       | 0.361                  |
| March     | 322         | 258                       | -0.425                 |
| April     | 332         | 226                       | -0.157                 |
| May       | 203         | 123                       | 0.379                  |
| June      | 84          | 51                        | 0.980                  |
| July      | 42          | 21                        | 0.965                  |
| August    | 30          | 14                        | 0.890                  |
| September | 25          | 13                        | 0.321                  |

**Forecasting Monthly Flows**

Forecasts about monthly flows help operation managers to take efficient decisions regarding operation of water facilities. In this case, the one month ahead forecasts may be useful for dam authorities to manage reservoir operation. Monthly flows are forecasted using lag-1 autocorrelation. Figure 4 presents a comparison of forecasted flows with observed flows, which provide a satisfactory closeness between the two flow series.

**DISCUSSION**

It is observed from the results that First Order Markov Model provides acceptable long term and short term flow forecasting. The first two moments (mean and variance or standard deviation) of observed flows and forecasted flows have no significant difference. It provides a basis for reliable flow forecasting using this model. In addition, the MAE and RMSE are also found acceptable. The long term forecasting of mean annual flows for 25 years, provides a basis for planning activities for better management of river water.

In addition to planning activities, the operation of water facilities is also one of the main activities in water resources management. The operation decisions are strengthened by having a good idea about water availability in near future. In this case, it is assumed that the decisions are taken on monthly basis, so monthly flows were forecasted. It was observed that Markov Model with periodicity also fits well to the periodic data with lag-1 autocorrelation. Mean monthly flows were forecasted for 15 years in future.

It is found in the study that application of this model can greatly help in better operation and planning of water facilities.
Two main assumptions are made in this study that the data follows normal distribution and have no significant trend.

CONCLUSIONS

Flow forecasting is one of the basic requirement for efficient planning and decision making in water resources management. Keeping this in view, this study has been performed where flows of Lesser Zaab River at Dokan Dam were forecasted for long term and short term purposes. The long term flows are required for long term strategic planning, whereas short term forecasting is required for improved operation of water facilities, like reservoirs and water supply schemes. A variety of techniques are available for time series data forecasting, which required certain number of parameters and somehow computationally intensive. In this study, the First Order Markov Model for stationary data series as well as data with periodicity has been applied. This model needs the first two moments of distribution and depends upon lag-1 autocorrelation for stationary and periodic data. Required parameters for flow forecasting were estimated and the flow series was forecasted accordingly. After that, the flow series were validated by comparing the observed and forecasted series in terms of means, standard deviations, MAE and RMSE. Satisfactory closeness of the observed and forecasted series was found. Based on that the series was forecasted for mean annual flow for next 25 year. Furthermore, the flow series was also forecasted for mean monthly flows for next 15 years. It is concluded that the Markov Model with given parameters can successfully be applied for future flow forecasting for any given duration. The Markov Model was found to be computationally simple and depends only upon lag-1 autocorrelation and at the same time, it provides reasonable good results. The other time series forecasting models may provide better results, but are computationally intensive, especially for data with periodicity.

ACKNOWLEDGEMENTS

We highly acknowledge the support provided by Dokan Dam Authority, Ministry of Agriculture and Water Resources, Kurdistan Regional Government, Iraq for providing the required data of river flow series including reservoir operation. Sincere thanks are extended to Dr. Hamed M. Jassim for his valuable comments and suggestions for the improvement of scientific and presentation quality of this research paper.

REFERENCES

Abed, Z., Al-Ridah, A. (2007). Stochastic Models of Some Properties of water in the middle of Euphrates Region in Iraq. M.Sc Thesis, College of Engineering, University of Babylon.

Akgun, B. (2003). Identification of Periodic Autoregressive Moving Average Models. Middle East Technical University.

Ali, S. (2009). Fitting Seasonal Stochastic Models to Inflow of Bekhme Reservoir. The Iraqi Journal for Mechanical and Material Engineering, Special Issue(A), College of Engineering, University of Babylon. [online] pp. 1-12. Available at: https://www.iasj.net/iasj?func=fulltext&aid=48846. [Accessed at: 2 May. 2017].

Al-Suhaili, R. (1986). Stochastic Analysis of Daily Streamflow of Tigris River. M.Sc. Thesis, College of Engineering, University of Baghdad.

Al-Ta’ee, M. (2009). Analysis of Records of Rainfall and Evaporation in Babylon. M.Sc. Thesis, College of Engineering, University of Baghdad.

Box, G. and Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. Holden Day, San Francisco. Fifth ed. Wiley., pp. 47.

Fortin, V., Perreault, L. and Salas, J. (2004). Retrospective Analysis and Forecasting of Stream flows Using a Shifting Model. Journal of Hydrology, (296), pp.135-163.

Gupta, R. (1989). Hydrology and Hydraulic Systems. Fourth ed. Prentice Hal., pp. 423-467

Haan, C. (1977). Statistical Methods in Hydrology. 1st ed. The Iowan State University Press., pp. 293-298.

Montgomery, D., Jennings, C., Kulahci, M. (2008). Introduction to Time Series Analysis and Forecasting. 1st ed. John Wiley & Sons, Inc., pp.244-261.

Muhammad, M. (2012). Time Series Modelling Using Markov and ARIMA Models. MSc. Thesis.
Faculty of Civil Engineering, University of Technology Malaysia.