Stochastic Models for the Maximum and Minimum Daily Flows of Tigris and Khabur Rivers

Dr. Khalid M. Khadir * Dr. Adil M. Raheem** Mrs. Syran A. Ibrahim***

*Assistant prof., college of engineering- University of Duhok.
**Assistant prof., college of engineering- University of Salahaddin
***Assistant Lecturer, college of engineering- University of Duhok.

ARTICLE INFO

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

Keywords:
Stochastic model
Time series
SARIMA model
Matalas model

*Corresponding Author:
Email: adil.raheem@su.edu.krd

ABSTRACT

In this paper, time series for the available recorded data of Tigris and Khabur Rivers at Tusun and Zakho stations were analyzed, stochastic models were investigated for both stations using several time intervals (maximum and minimum daily flow in a month), normality was tested using Kolmogrov-Smirnov approach, series transformation to normal distribution was carried out using Box-Cox method, Split-Sample method and Kendall's Correlation tests were used to check and remove the jump and trend components from the series of the two rivers, and the periodicity component also was detected.

Stochastic models, SARIMA (Seasonal Autoregressive Integrated Moving Average) and Matalas (multisite), were used for forecasting and generating the time series for the above mentioned rivers. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were used to identify the suitable SARIMA model. The Akaike Information Criteria (AIC) formula was used to find the suitable model for the series with the time intervals as described previously.

SARIMA models of order SARIMA(1,0,0)(0,1,1)\(_{12}\) and SARIMA(1,0,0)(0,1,1)\(_{12}\) were found to be suitable for maximum and minimum daily flow in a month time series for Zakho station respectively, SARIMA(1,1,1)(0,1,1)\(_{12}\) and SARIMA(1,0,0)(0,1,1)\(_{12}\) were found to be suitable for maximum and minimum daily flow in a month time series for Tusun station respectively. Matalas model of first order AR(1) was applied for the time series of both stations.

INTRODUCTION

Stochastic simulation methods were first introduced in hydrology in the reservoir design. The required capacity of a reservoir depends on the sequence of flows, especially a sequence of low flows (Lenseley, 1982).

Stochastic generation techniques can be used to generate time series that differ from the observed time series but retain many properties of the original series. This time series can be divided into different sequences and each one should have the same length of the recorded one. In each sequence, the events will have the same probability of occurrence as the observed sequence.

The technique of time series analysis used for the estimated statistical parameters (from
historical series) is to build a mathematical model capable of describing the evolution of possible sequences of events in time at the site of observations, which have the same statistical properties as the historical sample, this model can then be used to synthesis stream flow sequences. There are varieties of available methods for forecasting stream flows. However, to build a forecasting model is not an easy task to choose a suitable model because none of them is powerful and general enough every model has some degrees of structure and uncertainty parameters (Al-Suhaili, 1985).

Due to lack of the recorded data for Tigris River and its tributaries, two stochastic models were used as SARIMA (Seasonal Autoregressive Integrated Moving Average) and Matalas models for analyzing the available data of a long period, since the two models have wide applications for different types of hydrological time series.

In order to investigate the performance and evaluate the ability and efficiency of these two stochastic models for forecasting and generating different type of series, the time series of different intervals for the two discharges sites was used as maximum and minimum daily flow.

The Study Area:

The present paper deals with the investigation of the two stations, Zakho and Tusun, which are located on Khabur and Tigris Rivers respectively. Khabur River is one of the tributaries feeding Tigris River at the north of Iraq. The total length of Khabur River is about 160 km which feeds Tigris River about 1.6 million cubic meter of water yearly. Tigris River also flows out from the Turkish lands. The total length of Tigris River is about 1900 km where about 1415 km is inside the Iraqi lands, while the remaining length is inside the Turkish lands. Zakho station is located at Lat. $37^\circ 8'$ North and Long. $42^\circ 41'$ East, elevation approximately equals 440m above the mean sea level. The drainage area was equal to 3500 km$^2$ and the available discharge was recorded for the period (1959 to 1982). The maximum daily discharge recorded was 1270 m$^3$/sec on 11 April 1963, while the minimum daily discharge recorded was 8 m$^3$/sec for the 1$^{st}$ to 7$^{th}$ October 1975. The drainage area was about 3500km$^2$. Tusun station is located at Lat $37^\circ 0'$ North and Long. $42^\circ 38'$ East. Its elevation is approximately equal to 330m above the mean sea level. The drainage area was equal to 46700 square kilometers and the available discharge was recorded for the period (1959 to 1975). The maximum daily discharge was 7050 m$^3$/sec on the 2nd April, 1969, while the minimum daily discharge was 53 m$^3$/sec on 7th to 30th September, 1971. The two rivers and stations was shown in Figure 1.

Figure 1 location of Zakho and Tusun stations

SARIMA and Matalas Models:

The general form of seasonal Autoregressive Integrated Moving Average SARIMA (p, d, q)*(P, D, Q)s can be expressed in the following form (Wang, 2006):

$$\varphi(B)D(B^s)w_t = \theta(B)\Theta(B^s)\xi_t \quad (1)$$

Where

$$\varphi = 1 - \sum_{i=1}^{p} \phi_i B^i \quad \text{................. (2)}$$

$$\Theta = 1 - \sum_{i=1}^{q} \theta_i B^i \quad \text{.............. (3)}$$
\[
\Phi = 1 - \sum_{i=1}^{p} \Phi_i B^i \quad \text{......... (4)}
\]

\[
\Theta = 1 - \sum_{i=1}^{Q} \Theta_i s B^s \quad \text{......... (5)}
\]

\[
BX_t = X_{t-1} \quad \text{......... (6)}
\]

\[
w_t = (1 - B)^s (1 - B')^n X_t \quad \text{......... (7)}
\]

With \( \bar{X} \): Average of Recorded time series,
\( \phi \): Coefficient of Autoregressive (AR),
\( \theta \): Coefficient of Moving Average (MA),
\( p \): Order of Autoregressive,
\( q \): Order of Moving Average,
\( \Phi \): Coefficient of seasonal Autoregressive,
\( \Theta \): Coefficient of seasonal Moving Average,
\( \Phi \): Order of seasonal Autoregressive,
\( Q \): Order of seasonal Moving Average,
\( s \): Season length,
\( d \): Order of seasonality difference.

The most common Multisite model used was the first order autoregressive model which was initiated in hydrology by Matalas (1967). This model assumed that the series of the recorded data at each site was normalized and standardized after removing the non homogeneity components (jump and trend) and seasonality. The Matalas model can be built by normalizing each time series by Box-Cox Transformation, checking the stationary and homogeneity of each time series (Jump and Trend), and then standardizing the time series. After the previous steps were satisfied, the Matalas model can be used for any numbers of gauging stations (ns), where (ns) in this study equal to 2.

The form of the first order AR(1) of multisite model was [Al-Mousawi, 2003]:

\[
X_t = A X_{t-1} + B \zeta_t \quad \text{......... (8)}
\]

Where \( X_t, X_{t-1} \) are (ns) - dimensioned column vectors representing the standardized flows corresponding to (ns) sites at time t and t-1. \( \zeta_t \) is the column vector of (ns) serially and mutually uncorrelated independent variables at time t with zero mean and unit variance.

\[
A = M_1 M_0^{-1} \quad \text{................. (10)}
\]

\[
B B^T = M_0 - A M_1^T \quad \text{............. (11)}
\]

\( M_0 \): Lag-zero cross-correlation matrix

\( M_1 \): Lag-one cross-correlation matrix

\( M_0^{-1} \): Inverse matrix of \( M_0 \)

\( M_1^T \): Transposed matrix of \( M_1 \)

The matrix \( A \) can be evaluated directly using equation (10), but the matrix \( B \) can be evaluated from the Matrix \( BB^T \) using equation (11) [Tahir, 1984 as cited in Yong,
assumed the matrix $B$ triangular as shown below:

$$B = \begin{bmatrix}
    b_{11} & 0 & \ldots & 0 \\
    b_{21} & b_{22} & \ldots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    b_{n1} & b_{n2} & \ldots & b_{nn}
\end{bmatrix} \quad (12)$$

The matrix $BB^T$ was assumed to be equal to matrix $C$, where $C(I,J) = C(J,I)$ with number of stations equal to two, then:

$$BB^T = \begin{bmatrix}
    b_{11} & 0 & b_{11} & b_{12} \\
    b_{21} & b_{22} & 0 & b_{22} \\
    b_{11} & b_{12} & b_{11} & b_{12} + b_{22} \\
    b_{21} & b_{22} & b_{12} + b_{22} & b_{22}
\end{bmatrix} = \begin{bmatrix}
    c_{11} & c_{12} \\
    c_{21} & c_{22}
\end{bmatrix}$$

...... (13)

The elements of matrix $B$ can be determined as (Jayarami Reddy, 1997):

$$b_{11} = (c_{11})^{1/2}, \quad b_{21} = c_{21}/b_{11}, \quad b_{22} = (c_{22} - b_{21}^2)^{1/2}$$

...... (14)

After estimating the matrices $A$, $B$, $M_0$ and $M_1$, the model can be applied for generating the required time series for any period.

**Model application and discussion:**

SARIMA model was applied as a single site model by using program MINITAB ver 13.2 for the time series (maximum and minimum daily series). Matalas model was also applied as a multisite model for the time series mentioned above for the two stations by using a developed computer program modified in this study in FORTRAN 2000.

The recorded data for Tigris and Khabur Rivers at Tusun and Zakho stations respectively were used for the available recorded period, which were obtained from the references from the Iraqi publication and documentation, like Iraqi publications of the general directorate of irrigation at 1976, publications of the office of the state organization for maintenance and operation of irrigation projects at 1983 and publications of the directorate of operating the irrigation projects at 1989.

In order to check the normality of the time series of the two stations under study, Kolmogrov-Smirnov test should be used, which was present in the MINITAB computer program ver. 13.2 by comparing the studied time series with a theoretical time series which has the normal distribution. Figure 2 and Figure 3 show the non-normality of the maximum daily time series for Zakho and Tusun stations while Figure 4 and Figure 5 show the non-normality of the minimum daily time series, because the skewness coefficient ($C_s$) for each one was not equal to zero, so the time series for Zakho and Tusun stations must be converted to a normal distribution series as a first step before analyzing the data.
To transform the data to normal distribution, the power transformation of Box-Cox was applied, the value of \( \lambda \) (Box-Cox coefficient) was found to be -0.0685 and -0.03535 for the series of maximum and minimum flow of Zakho station while the value of \( \lambda \) (Box-Cox coefficient) was found to be 0.04491 and -0.1462 for the series of maximum and minimum flow of Tusan station. Applying the \( \lambda \) values in equation 15 the time series of the two station were converted to normal distribution with skewness coefficient equal to zero. (Salvator Grimaldi, 2004):

\[
Y_t = \frac{X_t^\lambda - 1}{\lambda} \quad \cdots \cdots (15)
\]

where

\( Y_t \): Recorded time series after transforming to normal distribution.

\( X_t \): Recorded time series values.

\( \lambda \): Box-Cox coefficient. \( \lambda \) lies within the range (+1,-1) but not equal to zero.

Figures 6 and Figures 7 show the Kolmogrov-Smirnov test after transforming the maximum daily flow time series to the normal distribution. The same procedure were carried out to the minimum daily flow time series for the Zakho and Tusan stations as shown in Figure 8 and Figure 9.

\[
\text{Figures 6 and Figures 7 show the Kolmogrov-Smirnov test after transforming the maximum daily flow time series to the normal distribution. The same procedure were carried out to the minimum daily flow time series for the Zakho and Tusan stations as shown in Figure 8 and Figure 9.}
\]
Test for jump component was made using the split sample method, to certify whether or not the difference between the means and standard deviations of two sub-samples is significantly different from zero at 95% probability level of significant. Each time series for the two stations was divided in two equal sub-samples. The critical t-value (tabulate value) was found to be equal to 2.086. **Tables 1** show the result of the test. The calculated t-values are less than the critical t-value mentioned above for the time series, which indicate that all the time series were free from the jump component.

**Table 1**

t-test results for jump component of Zakho and Tusan Station

| Time series        | t-calculated for Mean | t-calculated for Standard deviation |
|--------------------|-----------------------|-----------------------------------|
| Maximum Daily Flow(Zakho) | 3.617832E-01        | 2.854239E-01                     |
| Maximum Daily Flow(Tusan)  | 5.374522E-01        | 1.355012E-01                     |
| Minimum Daily Flow(Zakho)  | 9.410734E-02        | 1.253649E-01                     |
| Minimum Daily Flow(Tusan)  | 4.878887E-01        | 3.474165E-02                     |

The existence of the trend component in the time series of the two stations must be checked. As far as the correlation between the discharges and time scale are concerned, **Figure 10** to **Figure 13** show that there was no
Table 2 Linear Trend for the Time Series of Zakho and Tusun Stations.

| Time Series                  | Trend Equation               |
|-----------------------------|------------------------------|
| Maximum Daily Flow (Zakho)  | $y_t = 3.74690 - 0.242E-3t^2$ |
| Maximum Daily Flow (Tusun)  | $y_t = 8.01750 - 1.51E-3t^2$  |
| Minimum Daily Flow (Zakho)  | $y_t = 1.49821 + 1.85E-3t^2$  |
| Minimum Daily Flow (Tusun)  | $y_t = 3.17990 + 1.35E-3t^2$  |

The plots of the Autocorrelation Coefficient (ACF) and Partial Autocorrelation Coefficients (PACF) for a series indicate the primary estimation of SARIMA model for the series. Figure 14 to Figure 21 show the plots of ACF and PACF for all the maximum and minimum daily flow time series of Zakho and Tusun stations. All the Autocorrelation Coefficient plots for the mentioned time series indicated that they were non-random series due to the values of ACF which were not within the confidence limits.

After determination of the ACF and PACF of the time series, the parameters of SARIMA for several model were calculated and the parsimony models were selected according to the smallest value of AIC test as (Kadri, 2004):

$$AIC(p,q) = N \cdot \ln(\sigma^2) + 2(M) \ldots \ldots \ldots (9)$$

where

$$M = p + q + P + Q$$

Table 3 and Table 4 shows the SARIMA model for maximum and minimum daily flow time series of Zakho and Tusun station, the seasonal and non-seasonal parameters of the models were evaluated by applying MINTAB program ver 13.2.

Table 3 Parameters of SARIMA Model for Maximum and Minimum Daily Flow Series for Zakho Station and the t-test Results

| Time Series | Seasonal | Non-seasonal |
|-------------|----------|--------------|
|              | t-test   | SAR Coef (θ) | SMA Coef (θ) | AR Coef (φ) | MA Coef (φ) |
| Maximum daily series during the month (1,0,0)(0,1)ₙ | 27.55 | θ₀ = -0.28 | 7.28 | φ₀ = -0.41 | - |
| Minimum daily data during the month (1,0,0)(0,1)ₙ | 24.91 | θ₀ = -0.89 | 29.92 | φ₀ = -0.89 | - |

Table 4 Parameters of SARIMA Model for Maximum and Minimum Daily Flow Series for Tusun Station and the t-test Results
The model is considered suitable for forecasting process when the residuals Autocorrelation Coefficients are not significant i.e., the series of the residuals is random and they are within the confidence limits. **Figure 22** and **Figure 23** show that the ACF of the residuals of SARIMA models for Zakho station were considered random and not significant, the same result was obtained for Tusan station as shown in **Figure 24** and **Figure 25**, which indicate that all models were accepted and can be used for forecasting process.

![Figure 22](image22.png)  
**Figure 22** Residual Autocorrelation Function against Lag for SARIMA Model(1,0,0)(0,1,1)12 of Max.Daily Flow Series for Zakho Station

![Figure 23](image23.png)  
**Figure 23** Residual Autocorrelation Function against Lag for SARIMA Model(1,0,0)(0,1,1)12 of Min.Daily Flow Series for Zakho Station

![Figure 24](image24.png)  
**Figure 24** Residual Autocorrelation Function against Lag for SARIMA Model(1,1,1)(0,1,1)12 of Max.Daily Flow Series for Tusan Station

![Figure 25](image25.png)  
**Figure 25** Residual Autocorrelation Function against Lag for SARIMA Model(1,0,0)(0,1,1)12 of Min.Daily Flow Series for Tusan Station

The best model which described each time series for the two stations was used for forecasting the series for two water years for comparison purpose with the recorded two water years (1981,1982) (which were not entered in any previous calculations). **Figure 26** to **Figure 29** show the recorded and forecasted hydrograph for the two water years mentioned above for the time series of the two stations, generally, these figures display the high convergence between the forecasted and recorded hydrograph for the time series, the acceptability of each model was checked by applying different statistical tests (RMSE, MAE, $\chi^2$ and K-S) ([Wen Wang, 2006](#)) as shown in **Table 5** and **Table 6**, the results show that the series passed the $\chi^2$ and K-S test where the calculated value of each series is less
than the tabulated one. The test and figures show that the best ARIMA model described in each series can be used efficiently for forecasting the flow series, taking into consideration the results of RMSE and MAE which had the largest value for the maximum daily flow time series and smallest value for the minimum daily flow time series.
Table 5 Results of the Statistical Tests for Zakho Station by SARIMA Model

| Time Series | RMSE | MAE | $\chi^2$ | K-S |
|-------------|------|-----|----------|-----|
| Maximum Daily | 35.243 | 24.516 | 1.167 | 9.488 | 0.083 | 0.274 |
| Minimum Daily | 11.955 | 7.766 | 0.313 | 7.815 | 0.042 | 0.274 |

(*Root Mean Square Error(RMSE), **Mean Absolute Error (MAE), ***Kolmogrov-Smirnov (K-S) Test)

Table 6 Results of the Statistical Tests for Tusun Station by SARIMA Model

| Time Series | RMSE | MAE | $\chi^2$ | K-S |
|-------------|------|-----|----------|-----|
| Maximum Daily | 728.572 | 499.578 | 2.405 | 9.488 | 0.083 | 0.274 |
| Minimum Daily | 125.084 | 75.927 | 3.544 | 7.815 | 0.167 | 0.274 |

(*Root Mean Square Error(RMSE), **Mean Absolute Error (MAE), ***Kolmogrov-Smirnov (K-S) Test)

Table 7 and Table 8 shows the matrices ($M_0$), ($M_1$), ($A$) and ($B$) for all time series of the two stations. The first two matrices were calculated by using MINITAB software ver.13.2, and the other matrices were calculated using MathCAD software ver.10.

Matalas model was applied for all the time series in the present study for both Zakho and Tusun stations for forecasting the flow for the two water years 1981 and 1982 by applying a program modified in this study written in FORTRAN 2000. As in the estimation of ARIMA models, the same time series with the same length of records were used to estimate the parameters of the Matalas model and the records of the two years 1981, 1982 were used for comparison purposes between the recorded and forecasted discharges.
Table 7 Matrices for Multisite Model (M₀, M₁, A, B) for Maximum Flow Series of the two Stations

| Type of Matrix | Station | Zakho | Tusun |
|----------------|---------|-------|-------|
| Matrix M₀      | Zakho   | 1.000 | 0.554 |
|                | Tusun   | 0.554 | 1.000 |
| Matrix M₁      | Zakho   | 0.300 | 0.399 |
|                | Tusun   | 0.329 | 0.638 |
| Matrix A       | Zakho   | 0.402 | 0.175 |
|                | Tusun   | 0.035 | 0.658 |
| Matrix B       | Zakho   | 0.854 | 0.000 |
|                | Tusun   | 0.362 | 0.679 |

Table 8 Matrices for Multisite Model (M₀, M₁, A, B) for Minimum Flow Series of the two Stations

| Type of Matrix | Station | Zakho | Tusun |
|----------------|---------|-------|-------|
| Matrix M₀      | Zakho   | 1.000 | 0.402 |
|                | Tusun   | 0.402 | 1.000 |
| Matrix M₁      | Zakho   | 0.722 | 0.350 |
|                | Tusun   | 0.358 | 0.776 |
| Matrix A       | Zakho   | 0.753 | 0.047 |
|                | Tusun   | 0.055 | 0.754 |
| Matrix B       | Zakho   | 0.634 | 0.000 |
|                | Tusun   | 0.151 | 0.610 |

In order to check the adequacy of the model, the hydrographs of the recorded and forecasted series for the mentioned two water years were drawn to notice the degree of approximating between them. Figure 30 to Figure 33 shows the forecasted and recorded hydrograph for the maximum and minimum time series for the two stations. Generally, the above mentioned figures show a high conform between the two hydrographs for the time series. In order to study the effect of the data type on the model adequacy for these two water years, the statistical tests (RMSE, MAE, χ² and K-S) were evaluated for the data of the mentioned two water years. Table 9 and Table 10 show the results of these tests for both stations, the χ² and K-S tests for the time series of the two stations were passed where the calculated value of each series is less than the tabulated one. The RMSE and MAE tests show that the biggest values were for the maximum daily flow time series, while the smallest values were for the minimum daily flow time series, from comparing the results of the RMSE and MAE of the two models (SARIMA and Matalas) for the two station it was noted that the SARIMA model was more accurate in the forecasting than the Matalas model due to the lowest values of the obtained results from the statistical tests in SARIMA model.
Conclusions:

In the present study, the following conclusions may be conducted:

1. The time series of the two station for maximum and minimum daily flow were transformed to a normal distribution using the Box-Cox transformation method with values of $\lambda$ (Box-Cox coefficient) equal to -0.0685 and -0.03535 for the series of maximum and minimum flow of Zakho station while the value of $\lambda$ (Box-Cox coefficient) was equal to 0.04491 and -0.1462 for the series of maximum and minimum flow of Tusun station.

2. The time series were found to be free from the jump and trend components and also it was found to be stationary in mean and variance.

3. The best SARIMA model was found according to the smallest value of Akakie Information Criterion (AIC).
   - For maximum daily flow in a month at Zakho station the best model was SARIMA(1,0,0)(0,1,1)12.
   - For minimum daily flow in a month at Zakho station the best model was SARIMA(1,0,0)(0,1,1)12.
   - For maximum daily flow in a month at Tusun station the best model was SARIMA(1,1,1)(0,1,1)12.
   - For minimum daily flow in a month at Tusun station the best model was SARIMA(1,0,0)(0,1,1)12.

4. In Matalas model, the time series should be standardized then the model was built using first order Markov process.

5. The performance of the two models was calculated according to the results of RMSE, MAE, $\chi^2$ and K-S tests, and the performance of ARIMA model was
found to be more efficient than that the Matalas model.

References:

Al-Suhaili R.H., 1985. "Stochastic Analysis of Daily Streamflow of Tigris River". M.Sc. University of Baghdad.

Al-Mousawi Eman Mehdi, 2003. "Multisite Stochastic Model of Hydrochemical Properties at Selected Region". M.Sc. Thesis, College of Civil Engineering, University of Babylon.

Jayrami Reddy, 1997. "Stochastic Hydrology". Second edition, Delhi.

Kadri Yurekli and Ahmet Kurunc, 2004. "Performance of Stochastic Approach in Generating Low Stream Flow Data for Drought Analysis". Journal of Special Hydrology, Spring Vol.5 No. 1 pp 20-32.

Linsley, Kolher and Paulhus, 1982. "Hydrology for Engineering". McGraw-Hill Series in water Resources and Environmental Engineering, United State of America.

Salvatore Grimaldi, 2004. "Linear Parametric Models Applied on Daily Hydrological Series". Journal of Hydrologic Engineering, pp 1-9.

Wen Wang, 2006. "Stochasticity, Non linearity and Forecasting of Stream flow Processes". Doctorate thesis submitted to the Technical University of Delft.

Wen Wang, Pieter H. A. J. M. Van Gelder, and J. K. Vrijling, 2008. "Measuring Predictability of Daily Streamflow Processes Based on Univariate Time Series Model". International Environmental Modeling and Software Society (iEMSs), Delft, Netherlands, pp 1

Young G.K. and Pisano, W. C, 1968. "Operational of Hydrology Using Residuals". Journal of Hyd. Div ASCE, Vol. 94. HY4, pp 909-923.