https://doi.org/10.26682/sjuod.2023.26.1.30
Journal of University of Duhok., Vol. 26, No.1 (Pure and Engineering Sciences),Pp 305-316, 2023

FORECASTING OF AVERAGE MONTHLY FLOW FOR TWO STATIONS AT KHABOUR RIVER USING ARIMA AND ANN MODELS

AVA S. SAADI, KHALID M. KHIDIR and SHAKER A. JALIL
Dept. of Water Resources Engineering, College of Engineering, University of Duhok, Kurdistan Region-Iraq

(Received: November 10, 2022; Accepted for Publication: April 16, 2023)

ABSTRACT

In this study, autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) models were applied to predict the average monthly flow time series of two stations, named Begova and Chemcermo, for the Khabour River. The analysis of the time series was performed using several criteria and tests. The autocorrelation function (ACF) and partial autocorrelation function (PACF) were applied to check the accuracy of the ARIMA model. The Akaike (AIC) and Bayesian (BIC) equations were performed to determine the optimum model for prediction, which depended on the lowest AIC and BIC values. Applied test results show that the models of order ARIMA (0,0,0)(3,0,0)$_{12}$ and ARIMA((0,0,5)(5,1,4)$_{12}$ have higher acceptance compared to the other models for predicting the average monthly flow for Begova and Chemecermo stations, respectively. An ANN model of type multilayer perceptron method (ANN-MLP) was used for predicting the average monthly flow of Begova and Chemecermo stations, where the best models found are MLP (5,3,1) and MLP (9,7,1), respectively. Different statistical tests were applied and showed that the efficiency of the ANN model was better than the ARIMA model, with deterministic coefficients of 0.914 and 0.876 compared to 0.854, and 0.852 for the ARIMA for the monthly time series of Begova and Chemecermo stations, respectively.

KEYWORDS: Time Series, Stream Flow, Forecasting Models, ARIMA, and ANN.

1. INTRODUCTION

Forecasting streamflow is a technique that has high importance in the design, planning, and management of water resource systems. Stochastic simulation can be used for generating the possible future flows for the rivers, where it is assumed that those rivers' flows are extracted from a common statistical population and that the recorded data form a sample from this population. While the frequency of critical floods and droughts may not be adequately represented within the recorded data, several stochastic models have been developed in the last decades for hydrological time series data prediction and used by several researchers (Mirzavand and Ghazavi, 2015, Al-Saati et al., 2021, Essam et al., 2022).

The ARIMA model, sometimes called the Box-Jenkins approach, is a useful tool for forecasting and predicting data of the time series from the coming years. The Box-Jenkins technique, developed in 1970 and named after statisticians George Box and Gwilym Jenkins, is also known as the ARMA or ARIMA model. Many researchers have applied the ARIMA model for predicting different types of hydrological time series (Kurunç et al., 2005, Tong and Liang, 2005).

Kurunç et al. (2005) evaluated the forecasting performance of two modeling approaches, ARIMA and Thomas–Fiering, for selected water quality constituents and streamflow of the Yeşilırmak River at Durucasu monitoring. The test results for forecast accuracy indicated that, between the two approaches, the Thomas–Fiering model presented more reliable forecasting than the ARIMA model (Kurunç et al., 2005). The monthly streamflow data of the Zayandehrud River in western Isfahan province, Iran, were used to predict by applying ARIMA (1,1,0)(0,1,1), and they found that the data had been best represented by a multiplicative seasonal ARIMA model that satisfied all tests (Modares and Eslamian, 2006).

Frausto Solis et al (2008), compared the performance of two forecasting techniques, ARIMA and multilayer perceptron neural network (ANN-MLP) models, for the San Juan Tetelcingo river in Mexico. They found that the ARIMA model exhibited lower prediction errors than ANN-MLP (Frausto-Solis et al., 2008). Monthly streamflow at the Doyian station in Pakistan was predicted by ARMA and ARIMA models. The results indicated that the ARIMA model's forecasting accuracy is much better than that of the ARMA model (Adnan et al., 2017).

an ANN is an empirical model used for modeling complex hydrological processes by connecting inputs and outputs through mathematical functions without the need to know the relationship between the basin characteristics (Palit and Popovic, 2006). ANN has been widely used and found to be a powerful tool in predicting streamflow time series (Humphrey et al., 2016, Alade et al., 2017, Alquraish and Khadr, 2021).

A stepwise model empowered with genetic programming is developed to predict the monthly flows of the Hurman River in Turkey, the Diyalah River, and the Lesser Zab River in Iraq, where the results based on five different statistical measures show that the proposed stepwise model performed better than the Markovian model and the ARIMA model (Al-Juboori and Guven, 2016). Fashae et al (2019), used ANN and ARIMA models for comparing their efficiency for predicting the flow of the Opeki River station. They found that ARIMA was better than the ANN model according to the results of different tests (Fashae et al., 2019). Al-Saati et al (2021), applied the monthly recorded flow downstream of the Euphrates River (Hindiya Barrage/Iraq) using ARIMA and ANN models. The results show that the standard Box-Jenkins model was more accurate than the ANN model (Al-Saati et al., 2021).

The Khabour River basin has been studied by different researchers for different purposes since 1980 using the recorded data (1958–1982) of the old station at Zakho (Khidir, 1980, R Muhamad and N Hassan, 2005, Saleh, 2010, Khadir et al., 2018). Different techniques were used in the aforementioned research, such as Thomas-Fiering, ARIMA, MATALAS, and others. In this study, the applied data were recorded in two new stations established in 2004 at Begova and Chemcermo. Climate change has great effect on the metrological and hydrological systems in the Kurdistan region and the other neighboring regions since 1997.

The main objective of this research is to investigate ARIMA and ANN models for predicting average monthly flow in the time series at Khabour River at Begova and Chemcermo stations.

2. STUDY AREA AND COLLECTION OF THE DATA

Khabour River is one of the tributaries feeding the Tigris River in the Kurdistan Region of Iraq. The basic source of this river lies in the Turkish lands, which enters the Iraqi lands near Nazori and Begova villages before it empties into the Tigris River near the village of Fishkhabour, as shown in **Fig. (*1***. The area of the Khabour basin is about 3500 km^2, and the total length of the Khabour River is 160 km inside the Iraqi border (Saleh, 2010). Two stations have been established since 2004 for measuring the flow of the Khabour River, named Begova and Chemcermo, which are presented in **Fig. (*1***. Begova station lies near the Iraq and Turkey borders; Chemcermo station lies in a village named Chemcermo near Zakho city. **Table (*1*** shows the topographical characteristics of the two stations that catch up from the Water Resources Directorate of Duhok.

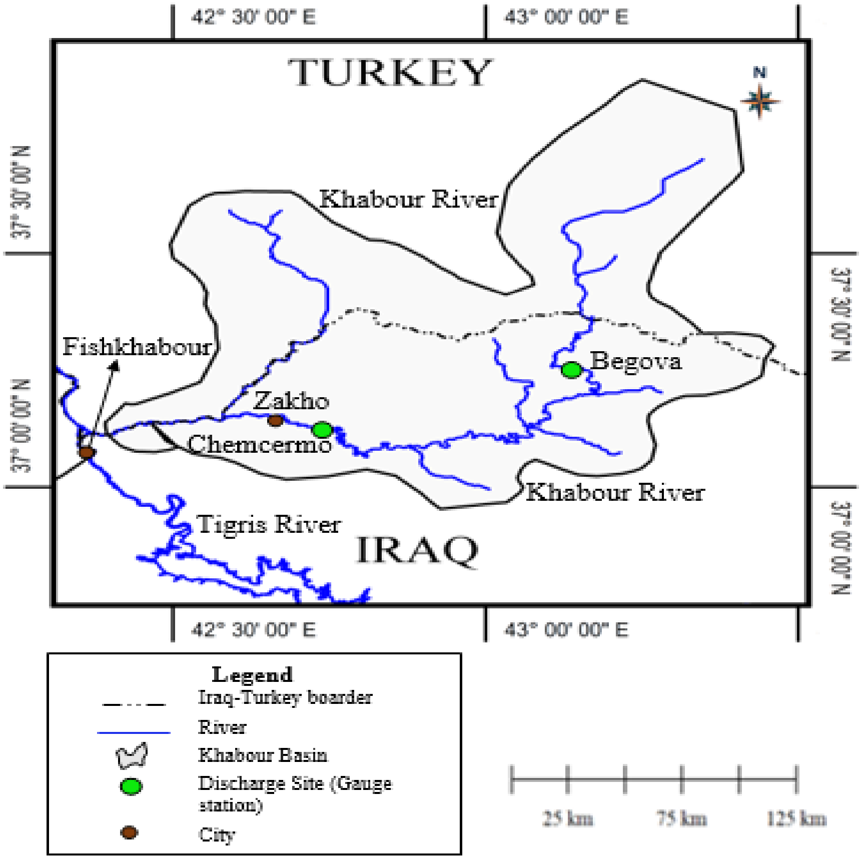


Fig. (1): The basin map of the Khabour River showing the two stations under study.

Table (1): Study area information for two stations, Begova and Chemcermo, in the Khabour River basin.

Stations	Area of the basin upstream of the station (Km^2)	Latitude (0° 0' 0.0")	Longitude (0° 0' 0.0")	Elevation (m) Above MSL
Begova	1517	33° 65' 42"	41° 24' 47"	779
Chemcermo	1983	29° 57' 06"	41° 12' 48"	438

The recorded average monthly flow data for Begova and Chemcermo stations are available for the period from (October 2004 to September 2020), The statistical characteristics of the recorded data for the two stations are shown in **Table 2** and **Table 3**, respectively.

Table 2: **Statistical characteristics of the recorded data for the Begova station.**

Mean (m^3/sec)	Standard deviation (m^3/sec)	Median (m^3/sec)	Minimum Flow (m^3/sec)	Maximum flow (m^3/sec)	Skewness coefficient (m^3/sec)	Kurtosis coefficient (m^3/sec)
80.461	76.380	45.327	12.778	398.811	1.959	6.706

***Table 3:** Statistical characteristics of the recorded data for the Chemcermo station.*

Mean (m^3/sec)	Standard deviation (m^3/sec)	Median (m^3/sec)	Minimum Flow (m^3/sec)	Maximum flow (m^3/sec)	Skewness coefficient (m^3/sec)	Kurtosis coefficient (m^3/sec)
76.167	107.226	35.44	2.7	996.2	2.746	12.255

3. FORECASTING MODELS

3.1. Auto Regressive Integrating Moving Average (Arima) Model

The ARIMA model is an extension of the autoregressive moving average (ARMA) model, which is fitted to time series data to offer generalized data and anticipate potential series points. The ARIMA parameters are (p, d, q), which are all positive integers, with (p) representing the autoregressive model parameter, (d) representing the degree of differencing parameter, and (q) representing the moving average model parameter. Seasonal ARIMA (p, d, q) (P, D, Q), where S specifies the number of seasons and the second part parameters (P, D, Q) signifies the differencing and moving average components for the seasonal member of the model (ARIMA). Mathematically, the seasonal integrated moving average (SARIMA) can be described as an **Error! Reference source not found.** (Tong and Liang, 2005):

$$\phi(B) * \Phi\ (BS) * (Wt - \mu) = \theta(B) * \Theta(BS) * \zeta t \qquad \text{Equation 1}$$

Where:

ϕ: Coefficient (AR).

θ: Moving average coefficient (MA).

Φ: Seasonal coefficient.

Θ: Coefficient of moving average for seasonal.

ζ: Random rate in time (t).

B: Backshift processor.

S: Period of season.

AIC is a useful approach for identifying the statistically best models. Also, it provides a mathematical definition of the parsimony principle in the context of model development, as shown in Equation 2 (Akaike, 1974).

$$AIC(p,q) = N.Ln(\sigma^2) + 2(M) \qquad \text{Equation 2}$$

Where $M = p + q + P + Q$ Equation 3

The Bayesian information criterion (BIC) is a computationally tractable approach represented in Equation 4 (Schwarz, 1978). It can be used to perceive a better labor model between several model trials (Kassem et al., 2020):

$$BIC(p,q) = N.Ln(\sigma^2) + M * Ln(N) \qquad \text{Equation 4}$$

Where:

σ: Standard deviation of the recorded data.

N: Number of recorded data.

The optimal model is the one that has a minimum value for AIC and BIC. This procedure was applied to select the efficient model that can be used to predict the average.

3.2. Artificial Neural Network (Ann) Model

An important aspect of an ANN model is its model design component and structure, input data, and hidden, and output data layers (Arisoy et al., 2012). The input dataset is processed first in the input layer, which is linked to the hidden layers through a neural network. The modeling procedure may uncover one or more hidden levels, depending on the extent of data mining. The number of optimal hidden layers and accompanying neuron weights might then be computed using the input-output dataset during the training phase. Although there are no widely acknowledged guidelines for selecting the ideal amount of input variables, neurons, or hidden layers, it has been demonstrated that data processing improves the effectiveness of ANN models (Nacar et al., 2018).

In this research, MLP-ANN was used. MLP is a feed-forward neural network architecture with uni-directional cell connections between successive layers (Seo et al., 2016), as shown in **Fig. (*2*.**

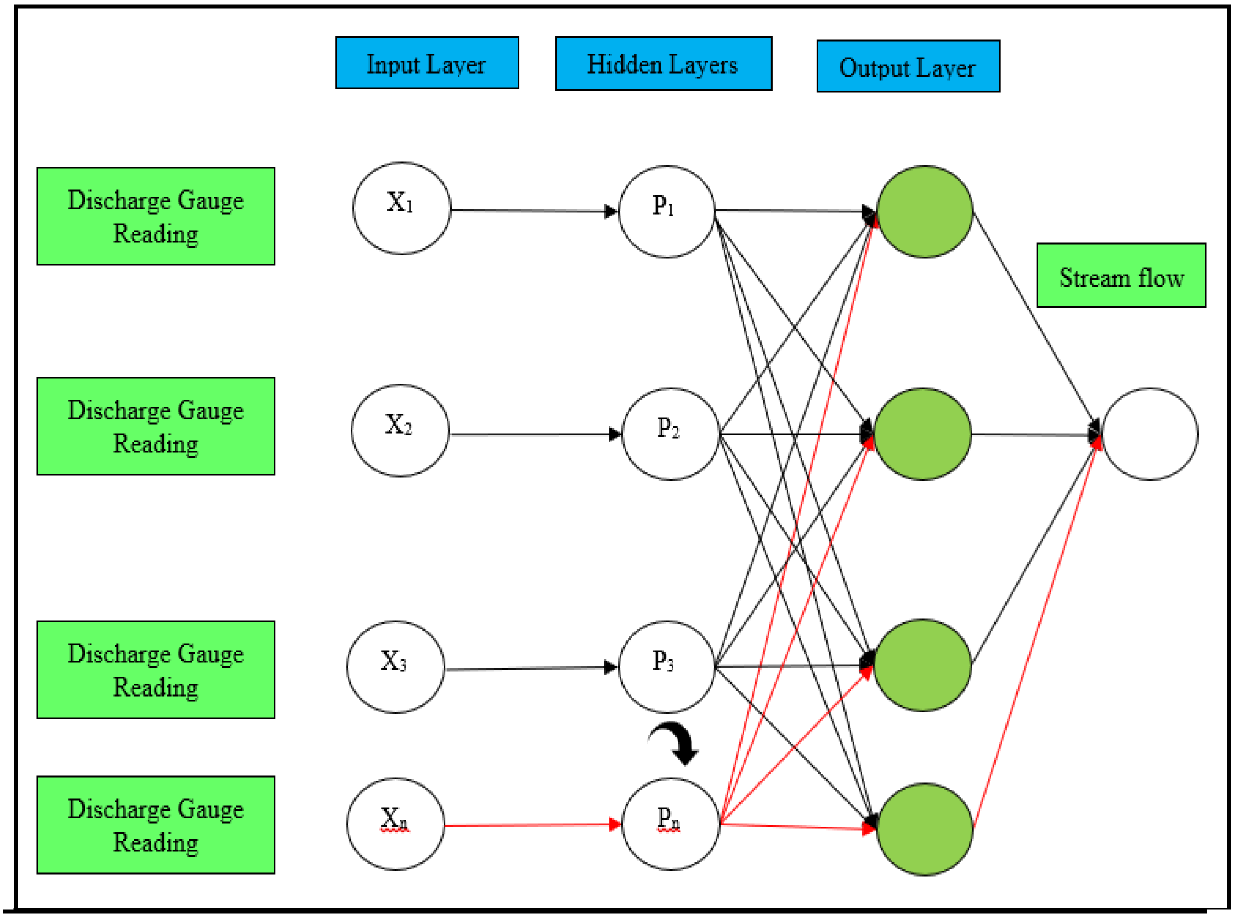


Fig. (2): Three layers of the ANN: input, hidden, and output layers.

The MLP method has one weight for each input and only one output; however, a multilayer network has multiple weights, each of which contributes to more than one output. The output layer results can be obtained from Equation 5 below:

$$\hat{y}_k = f_0\left[\sum_{j=1}^{m}\left(w_{k,j} * f_h\left(\sum_{i=1}^{n}(w_{j,i} + x_i) + b_j\right)\right) + b_k\right] \quad \text{Equation 5}$$

Where:

$w_{k,j}$: Weight of hidden-output.

$w_{j,i}$: Weight of input-hidden layers.

f_h: Activation function of the hidden layer.

f_0: Activation function of the output layer.

x_i: Input variable.

b_j: Hidden layer bias.

b_k: Output layer bias.

m: Hidden layer neurons number.

$\hat{y}_k$: Variable of output.

n: Variable of input.

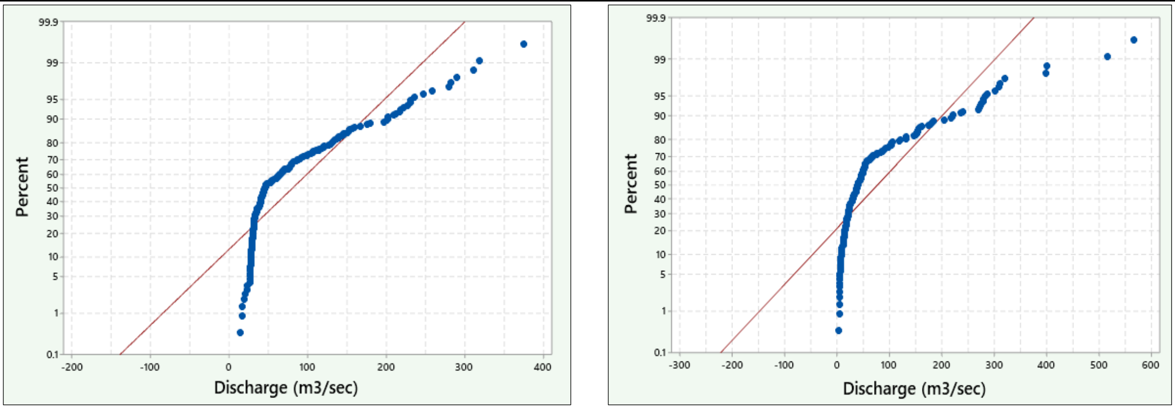
4. APPLICATION OF THE APPLIED MODEL AND DISCUSSION

4.1. Data Analyses

The recorded average monthly flow data of Begova and Chemcermo stations are available for the water years (2004 –2019), which were used in this study. Before analyzing the recorded data, the few missing data points of the year 2006 for Begova were estimated using the linear regression method such as (Haitovsky, 1968, Loh and Wainwright, 2011, Wang et al., 2008) and applying the Microsoft Office Excel version 2016. The available recorded data for each station was divided into two sets. The first set of water years' (2004 –2017) was used for calibration. The last two water years recorded data (2018 and 2019) were used for the verification stage.

Non-stationary time series cannot be used to build the ARIMA model (Chow, 1984), where the stationarity of time series is significant, which means that the statistical characteristics of a process that creates a time series do not change with time (Priestley and Rao, 1969).

Checking the normality of the two series was done using the Kolmogorov-Smirnov test by applying the MINITAB software program (version 19). **Fig. (3** shows the non-normal average monthly flow time series before normalization, where the Box-Cox transformation was used for the transformation of the series to a normal distribution by determining a suitable coefficient (λ) for C_s equaling zero (Box and Cox, 1964).

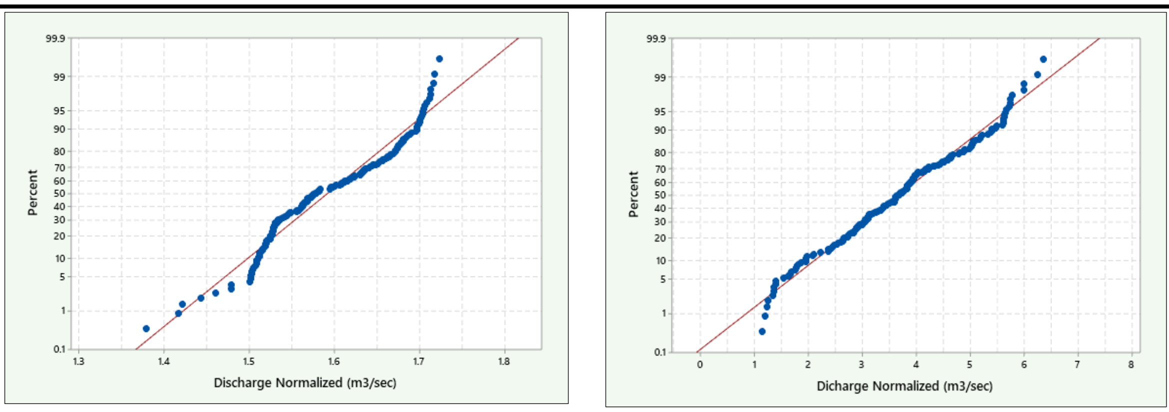


Before normalizing the time series of the Begova station

Before normalizing the time series of the Chemcermo station

Fig. (3): Test of normal distribution before transforming the average monthly flow time series of Begova and Chemcermo stations by Kolmogorov-Smirnov.

Fig. (4 presents the normality of the average monthly flow series after normalization. The skewness coefficient (C_s) was downgraded to zero for the two stations. The coefficients (λ) and (C_s) were found to be equal to (-0.559) and (0.00006) for the Begova station and (0.00045) and (0.00003) for the Chemcermo station, respectively.



After normalizing the time series of the Begova station

After normalizing the time series of the Chemcermo station

Fig. (4): Test of normal distribution after transforming the average monthly time series of Begova and Chemcermo stations by Kolmogorov-Smirnov

The value of (λ) was calculated after transforming the time series of the two stations to a normal distribution using Equation 6 (Grimaldi, 2004):

$$R_t = \frac{S_t - 1}{\lambda} \qquad \text{Equation 6}$$

Where:

R_t: Recorded average monthly discharge after changing to normal dispersion.

S_t: Recorded average monthly discharge before changing to normal dispersion.

λ: Box-Cox coefficient lie between (-1,1).

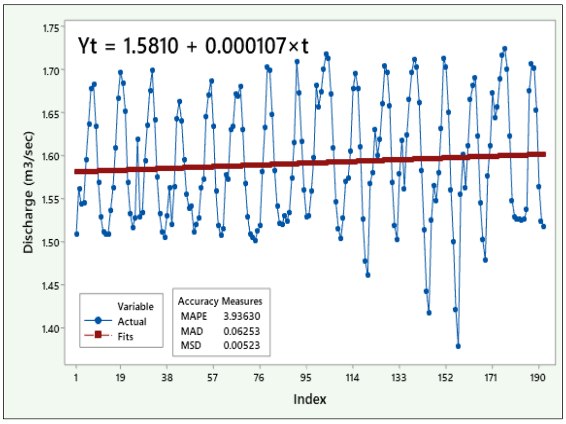
The split sample method was made for testing the jump component by dividing the time series of each station into two equal subsamples (Khadir et al., 2018), at the 95% confidence level, for checking that the difference between the means and standard deviations of two subsamples is significantly different from zero. **Table (4** displays the t-test results, which reveal that the average monthly flow time series for both Begova and Chemcermo stations is significant.

Table (4): Calculated-t the values of the mean and standard deviation of Begova and Chemcermo stations for the average monthly flow time series, respectively.

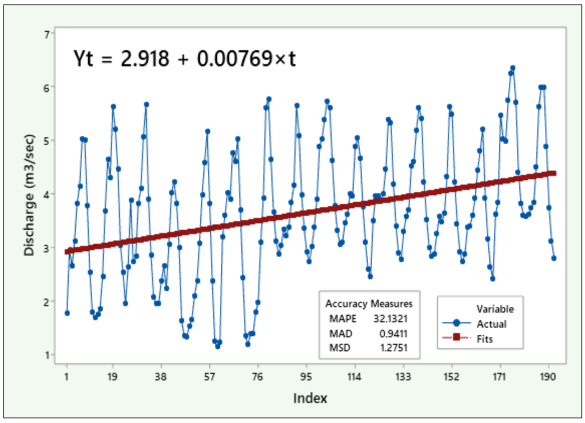
Stations	Average monthly flow time series	
	Calculated t-test for Mean	Calculated t-test for Standard Deviation
Begova	-0.538	-1.764
Chemcermo	-4.162	0.561

The critical value of the t-test was equal to (1.99) for the mean and standard deviation values for the two stations at the 95% probability level of significance, which indicates that the two-time series were free from the jump component.

The existence of the trend component was checked for the time series of the two stations; **Fig. (5** shows the equations of trend and its slopes for Begova and Chemcermo stations, where the Mann-Kendall test was eligible for detecting the trend in the two stations' time series.



Trend component of average monthly flow series at Begova station



Trend component of average monthly flow series at Chemcermo station.

Fig. (5): The Mann-Kendall test of trend components for both stations was applied by NCSS software.

The Mann-Kendall test is non-parametric. It was performed to analyze whether there is a trend in the existing data or not, even if the series contains a seasonal component (Hussain and Mahmud, 2019). The Mann-Tau Kendall's Correlation test results are shown in

Table (5, where Kendall's tau correlation was found to be (0.0689) and (0.2216) for Begova and Chemcermo stations, respectively. This demonstrates that the average monthly data of the two stations is free of trend components.

Table (5): Mann-Kendall trend component test for flow time of both Begova and Chemcermo stations.

Stations	Kendall's Tau Correlation
Begova	0.0689
Chemcermo	0.2216

The correlogram or autocorrelation plot is a helpful method for identifying the periodic component of a time series, as shown in Error! Reference source not found. and **Fig.** (*7* for both stations for testing periodicity in the forecasting model section.

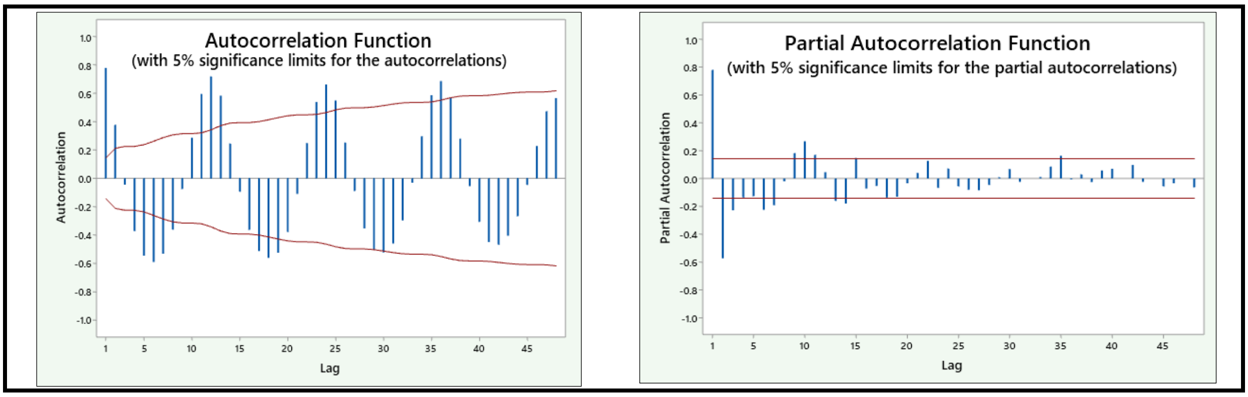


Fig. (6): Autocorrelation (ACF) and partial autocorrelation (PACF) functions of ARIMA models for average monthly flow time series of Begova station

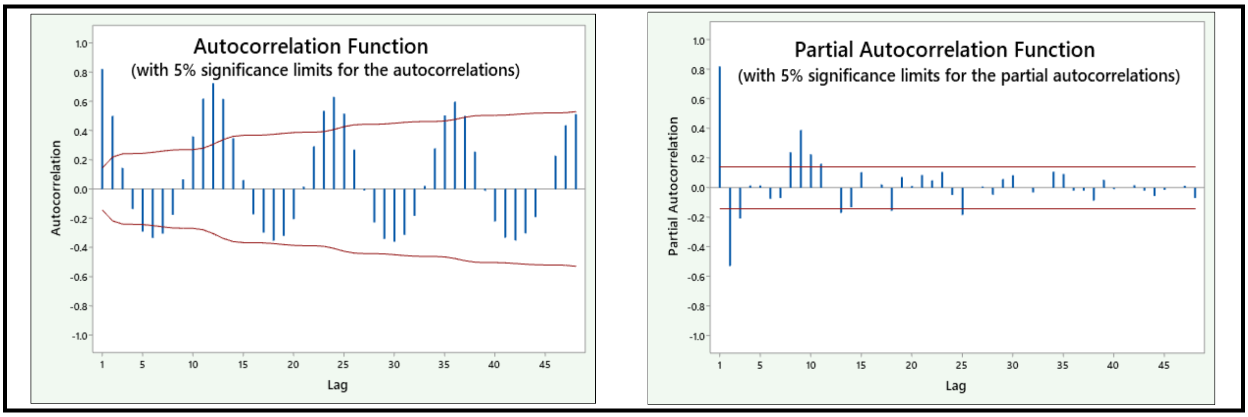


Fig. (7): Autocorrelation (ACF) and partial autocorrelation (PACF) functions of ARIMA models for average monthly flow time series of Chemcermo station.

Both of the two-time series under study ended up being free of trend, jump, and periodicity phenomena.

The recorded average monthly flow data for Begova and Chemcermo stations are available for the water year between 2004 and 2019 obtained from the Directorate of Water Resources of Duhok Governorate. The average monthly flow for 14 water years (2004 –2017) for the two stations, was used in the ARIMA and ANN models in the calibration stage, and that of the last two water years (2018 and 2019) was used in the verification stage.

4.2. Application Of The Arima Model

The components of the ARIMA model were discovered using three stages of analysis: identification, parameter estimation, and diagnostic. ARIMA can be built after testing and converting the average monthly flow time series for the two stations under study to stationery.

The Development of the ARIMA model was applied by NCSS statistical software, Version 12, after normalizing both the monthly flowtime series of Begova and Chemcermo stations and standardizing sixteen years from October 2004 to September 2020 for modeling and forecasting. The primary estimation ARIMA model for the two average monthly flow time series can be indicated by plotting the ACF and PACF. Error! Reference source not found. and **Fig. (***7* show the ACF and PACF for the two average monthly time series. The two figures indicate that the two-time series is random because the values of ACF are outside the confidence limits.

The best two ARIMA models for the two stations were chosen from several ARIMA models and found, as shown in **Table 6**. These two best models were used in forecasting the average monthly flow of the time series in the verification stage for the two water years 2018 and 2019, due to the least values of (AIC) and (BIC) as presented in Equation 2 and Equation 4 for both Begova and Chemcermo stations.

Table 6: **Parameters of the best performing ARIMA models for Begova and Chemcermo stations.**

Station	Best performing (ARIMA) model	(AIC)	(BIC)
Begova	$(0,0,0)(3,0,0)_{12}$	391.479	395.013
Chemcermo	$(1,0,3)(1,0,1)_{12}$	473.064	480.132

In the verification stage, the average monthly flow time series for the two years 2018 and 2019, which were not used in the calibration stage, were applied. This stage was applied to find the most effective model that can be utilized for forecasting and predicting the average monthly flow for the two stations at the Khabour River.

The statistical tests of deterministic coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used to check the acceptability of each model. **Table (***7* shows that the results of predicting average monthly flow by the ARIMA model for Begova are better than that of Chemcermo, as indicated by its high value of R^2 (0.854), compared to that of Chemcermo (0.852). The RMSE and MAE test values for the Begova time series are 63.796 and 40.573, which are less than those of the Chemcermo station 84.937 and 51.544, respectively. The results of the testing indicate that the best ARIMA can be used efficiently for forecasting the average monthly flow time series of the two stations.

Table (7): Statistical test results of average monthly flow time series for Begova and Chemcermo stations.

Stations	Best ARIMA Model	R^2	RMSE	MAE
Begova	$(0,0,0)(3,1,3)_{12}$	0.854	63.129	40.573
Chemcermo	$(0,0,5)(5,1,4)_{12}$	0.852	84.937	51.544

4.3. Application Of Ann Model

The ANN model of the ANN-MLP method was used and applied for forecasting the average monthly flow. The best model was obtained in the verification stage for the years 2018 and 2019 by applying the SPSS statistical software (version 26).

Following multiple trials, in these networks the rescale covariate was standardized in the partition section for training at 70%, testing at 20%, and holdout at 10% nearly, and it was determined that the sigmoid and hyperbolic tangent functions were the optimal activation functions for the MLP method, which linked the two layers one after another, the input and hidden layers together, as well as the hidden and output layers for two stations. The maximum training epochs were estimated automatically to prevent overtraining, and the gradient descent approach was used to describe the optimization procedure. In the training section, the online type was processed, which modified the synaptic weights after each single training data record (Zacharis, 2016). MLP-ANN's architectural information was presented in **Table (***8*.

Table (8): The network information for Begova and Chemcermo stations

Station	Input Layer	Hidden Layer unit	Activation Function	Output Layer unit	Activation Function
Begova	5	3	Sigmoid	1	Sigmoid
Chemcermo	9	7	Hyperbolic tangent	1	Hyperbolic tangent

The best results from the earlier trials for each station were found to be the input layer, hidden layer, and output layer for MLP models (5,3,1) and (9,7,1). **Table (***7* shows that the results of predicting the average monthly flow by the ANN model for the time series of Begova are better than those of Chemcermo, as indicated by its high value of R^2 (0.914), compared to that of Chemcermo (0.876). The RMSE and MAE test values for the Begova time series are 36.953, 24.357 which are lesser than that of Chemcermo 72.796, and 50.236 respectively. The results of the testing indicate that the best ANN model can be used efficiently for forecasting the average monthly flow time series of the two stations.

Table (9): **Statistical tests results by applying the ANN model to the average monthly flow time series of Begova and Chemcermo stations**

Stations	Deterministic coefficient (R^2)	Root Mean Square Error (RMSE)	Mean Average Error (MAE)
Begova	0.914	36.953	24.357
Chemcermo	0.876	72.796	50.236

Fig. (*8* and **Fig. (*9***, show the recorded and predicted average monthly flow hydrograph for the two years of the verification stage by the two models for Begova and Chemecermo stations, respectively. It is clear from these two Figs. that the hydrograph of the predicted average monthly flow by the ANN model is closer to the recorded data than that predicted by applying the ARIMA model for both Begova and Chemcermo stations.

Results tests in **Table (***7*, **Table (*9*,** and **Fig. (*8*** and **Fig. (*9*** show the high acceptance of the ANN model as compared with the result of ARIMA, where the R^2 values of the ANN model (0.914 and 0.876) are higher than those of the ARIMA model (0.854 and 0.852) for Begova and Chemcermo, respectively. RMSE values of the ANN model (36.953, 72.796) are less than those of the ARIMA model (63.129, 84.937) for Begova and Chemcermo, Also, MAE values of the ANN model (24.357, 50.236) are less than those of the ARIMA model (40.573, 51.544) for Begova and Chemcermo, respectively. As a result, the ANN model was more accurate than the ARIMA model for predicting the average monthly flow of the Khabour River for both the Begova and Chemcermo stations.

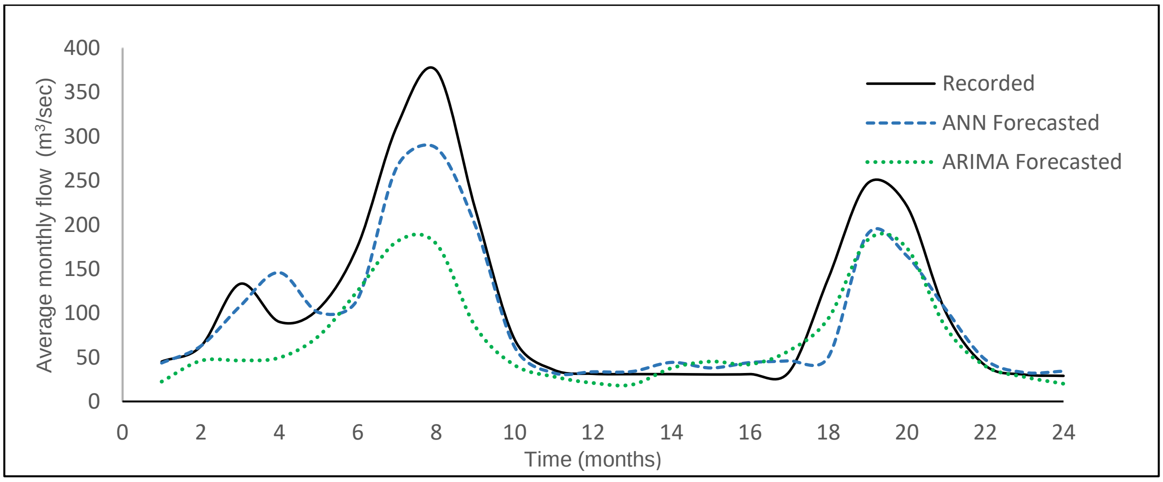


Fig. (8): Recorded and predicted average monthly time series (verification stage) of Begova station by applying ARIMA and ANN models.

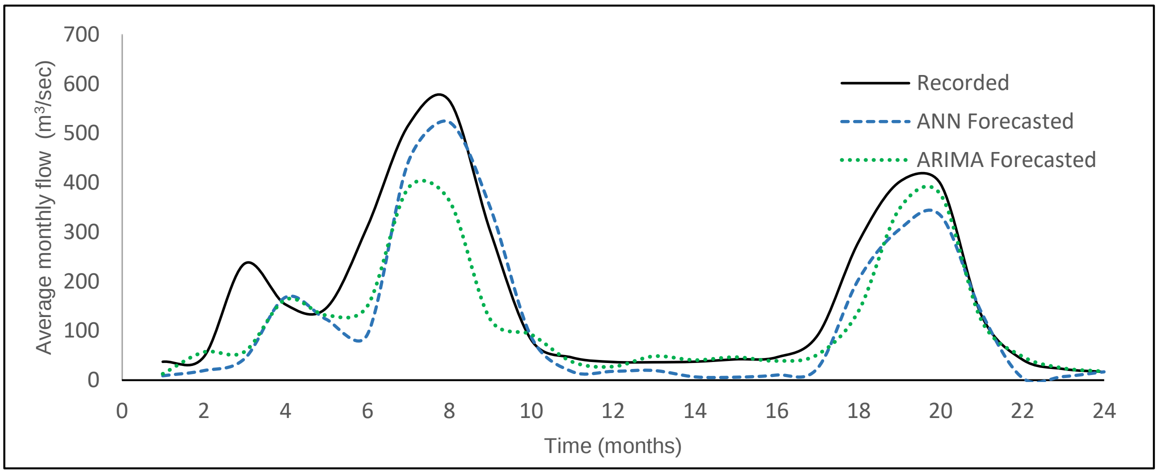


Fig. (9): Recorded and predicted average monthly time series (verification stage) of Chemecermo station by applying ARIMA and ANN models.

5. CONCLUSION AND RECOMMENDATION

The following notes can be concluded from the application and results of ARIMA and ANN models for the average monthly flow time series at Khabour River:

a) The average monthly flow time series of the two stations were exposed to check for normality and stationarity by different methods before applying the ARIMA model. The average monthly time series were found to be free of jumps and trends; the series was also stationary in mean, standard deviation, and variance.

b) The best ARIMA models were found to be ARIMA $(0,0,0)(3,1,3)_{12}$ for Begova station and ARIMA $(0,0,5)(5,1,4)_{12}$ for Chemcermo station which was used for forecasting 24 months of data in the verification stage.

c) The ANN models, ANN-MLP (5,3,1) for Begova station and ANN-MLP (9,7,1) for Chemcermo station, are the best for forecasting the average monthly flow time series.

d) Test results show that the ANN model gives more adequate results for predicting at the aforementioned stations than the ARIMA model.

e) The ANN model can further be used for forecasting and generating average monthly data for any required period to get a more adequate design and management for future proposed hydraulic structures upstream of the two stations at Khabour River.

f) As a recommendation, other stochastic models can be applied to the recorded data of the two discharge sites under study, such as the radial basis function neural networks (RBF) and the K-Nearest Neighbors (KNN) model for predicting flow series.

6. REFERENCES

ADNAN, R. M., YUAN, X., KISI, O. & CURTEF, V. 2017. Application of time series models for streamflow forecasting. *Civil and Environmental Research,* 9**,** 56-63.

AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control,* 19**,** 716-723.

AL-JUBOORI, A. M. & GUVEN, A. 2016. A stepwise model to predict monthly streamflow. *Journal of Hydrology,* 543**,** 283-292.

AL-SAATI, N. H., OMRAN, I. I., SALMAN, A. A., AL-SAATI, Z. & HASHIM, K. S. 2021. Statistical modeling of monthly streamflow using time series and artificial neural network models: Hindiya Barrage as a case study. *Water Practice and Technology,* 16**,** 681-691.

ALADE, O. M., SOWUNMI, O. Y., MISRA, S., MASKELIŪNAS, R. & DAMAŠEVIČIUS, R. A neural network based expert system for the diagnosis of diabetes mellitus. International Conference on Information Technology Science, 2017. Springer, 14-22.

ALQURAISH, M. M. & KHADR, M. 2021. Remote-Sensing-Based Streamflow Forecasting Using Artificial Neural Network and Support Vector Machine Models. *Remote Sensing,* 13**,** 4147.

ARISOY, E., SAINATH, T. N., KINGSBURY, B. & RAMABHADRAN, B. Deep neural network language models. Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT, 2012. 20-28.

BOX, G. E. & COX, D. R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological),* 26**,** 211-243.

CHOW, G. C. 1984. Random and changing coefficient models. *Handbook of econometrics,* 2**,** 1213-1245.

ESSAM, Y., HUANG, Y. F., NG, J. L., BIRIMA, A. H., AHMED, A. N. & EL-SHAFIE, A. 2022. Predicting streamflow in Peninsular Malaysia using support vector machine and deep learning algorithms. *Scientific Reports,* 12**,** 3883.

FASHAE, O. A., OLUSOLA, A. O., NDUBUISI, I. & UDOMBOSO, C. G. 2019. Comparing ANN and ARIMA model in predicting the discharge of River Opeki from 2010 to 2020. *River research and applications,* 35**,** 169-177.

FRAUSTO-SOLIS, J., PITA, E. & LAGUNAS, J. Short-term streamflow forecasting: ARIMA vs neural networks. American Conference on Applied Mathematics (MATH'08), Harvard, Massachusetts, USA, 2008. 402-407.

GRIMALDI, S. 2004. Linear parametric models applied to daily hydrological series. *Journal of Hydrologic Engineering,* 9**,** 383-391.

HAITOVSKY, Y. 1968. Missing data in regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological),* 30**,** 67-82.

HUMPHREY, G. B., GIBBS, M. S., DANDY, G. C. & MAIER, H. R. 2016. A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. *Journal of Hydrology,* 540**,** 623-640.

HUSSAIN, M. & MAHMUD, I. 2019. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software,* 4**,** 1556.

KASSEM, A. A., RAHEEM, A. M. & KHIDIR, K. M. 2020. Daily streamflow prediction for khazir river basin using ARIMA and ANN models. *Zanco Journal of Pure and Applied Sciences,* 32**,** 30-39.

KHADIR, K. M., RAHEEM, A. M. & IBRAHIM, S. A. 2018. Stochastic Models for the Maximum and Minimum Daily Flows of Tigris and Khabur Rivers. *ZANCO Journal of Pure and Applied Sciences,* 30**,** 50-61.

KHIDIR, K. M. 1980. Water Resources Evaluation For The Tigris River Basin North of Mosul.

KURUNÇ, A., YÜREKLI, K. & CEVIK, O. 2005. Performance of two stochastic approaches for forecasting water quality and streamflow data from Yeşilırmak River, Turkey. *Environmental Modelling & Software,* 20**,** 1195-1200.

LOH, P.-L. & WAINWRIGHT, M. J. 2011. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Advances in neural information processing systems,* 24.

MIRZAVAND, M. & GHAZAVI, R. 2015. A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods. *Water resources management,* 29**,** 1315-1328.

MODARES, R. & ESLAMIAN, S. 2006. STREAMFLOW TIME SERIES MODELING OF ZAYANDEHRUD RIVER; RESEARCH NOTE.

NACAR, S., HıNıS, M. A. & KANKAL, M. 2018. Forecasting daily streamflow discharges using various neural network models and training algorithms. *KSCE Journal of Civil Engineering,* 22**,** 3676-3685.

PALIT, A. K. & POPOVIC, D. 2006. *Computational intelligence in time series forecasting: theory and engineering applications,* Springer Science & Business Media.

PRIESTLEY, M. & RAO, T. S. 1969. A test for non-stationarity of time-series. *Journal of the Royal Statistical Society: Series B (Methodological),* 31**,** 140-149.

R MUHAMAD, J. & N HASSAN, J. 2005. Khabur River flow modeling using artificial neural networks. *Al-Rafidain Engineering Journal (AREJ),* 13**,** 33-42.

SALEH, D. K. 2010. *Stream gage descriptions and streamflow statistics for sites in the Tigris River and Euphrates River basins, Iraq*, US Department of the Interior, US Geological Survey Reston, VA, USA.

SCHWARZ, G. 1978. Estimating the dimension of a model. *The annals of statistics***,** 461-464.

SEO, M., KEMBHAVI, A., FARHADI, A. & HAJISHIRZI, H. 2016. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603*.

TONG, L. I. & LIANG, Y. H. 2005. Forecasting field failure data for repairable systems using neural networks and SARIMA model. *International Journal of Quality & Reliability Management*.

WANG, W., VAN GELDER, P. H. & VRIJLING, J. 2008. Measuring predictability of daily streamflow processes based on univariate time series model.

ZACHARIS, N. Z. 2016. Predicting student academic performance in blended learning using artificial neural networks. *International Journal of Artificial Intelligence and Applications,* 7**,** 17-29.