ISSN (print):2218-0230**, ISSN (online):** 2412-3986**,DOI: http://dx.doi.org/10.21271/zjpas**

RESEARCH PAPER

Daily Streamflow Prediction for Khazir River Basin Using ARIMA and ANN Models

Abdulwahd A. Kassem^{1*}, Adil M. Raheem², Khalid M. Khidir³

¹Water Resources Engineering Department, College of Engineering, Salahaddin University-Erbil, Kurdistan Region, Iraq ²Surveying Engineering Department, College of Engineering, Alkittab University, Iraq ³Water Resources Engineering Department, College of Engineering, University of Dohuk, Kurdistan Region, Iraq

A B S T R A C T

The present study used both Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) models for Khazir river basin to simulate the daily flow at Asmawa and Khanis gauge stations. Asmawa station lies on Khazir River while Khanis lies on Gomel River as a tributary of Khazir River. In the stochastic ARIMA model, the Autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to determine how robust the ARIMA model is in predicting the streamflow. In this study, the Akaike Information Criterion (AIC) formula and Bayesian information criterion (BIC) were used to evaluate which model is more accurate. The results of this study showed that models of order ARIMA are (2,0,0)(2,1,0) and $(2,0,1)(2,1,0)$ were found much better than the other models for generating and forecasting daily flow time series for aforementioned stations. Coefficients of determination (R^2) were found 0.77 and 0.85 for both Asmawa and Khanis stations, respectively. However, two types of ANN models were used for analyzing the daily flow records of the same two aforementioned stations, Multilayer Perceptron (MLP) and Radial Basis Function (RBF). ANN-MLP model was found to be more accurate than the ANN-RBF for generating and forecasting the daily flow time series as the coefficient of determination provided by ANN-MLP for both stations were 0.83 and 0.85, respectively. In addition, the coefficients of determination produced by the ANN-RBF for both stations were 0.66 and 0.55, respectively. Based on the values of (R^2) and $(RMSE)$ obtained in the current work, one can conclude that the ANN-MLP model is the most accurate model among the others in terms of predicting the streamflow for Asmawa station, whereas the performance of both ARIMA and ANN-MLP models for the Khanis station is the same.

KEYWORDS: Forecasting, Streamflow, ARIMA, and ANN. DOI:<http://dx.doi.org/10.21271/ZJPAS.32.3.4> ZJPAS (2020), 32(3); 30-39 .

1. INTRODUCTION

Many activities associated with the planning and operation of the water resources system, the accuracy and reliability of streamflow forecasting are significant. For the planning and management of the water resources, it is necessary to have an accurate forecasting model for river streamflow.

* **Corresponding Author:** Abdulwahd A. Kassem E-mail: abdulwahid.qassem@su.edu.krd **Article History:** Received: 25/06/2019 Accepted: 15/12/2019 Published: 15/06 /2020

Therefore, in the last decades, many deterministic and stochastic models have been developed, including parametric, nonparametric, linear, and nonlinear models for hydrologic time series data prediction (Marques et al., 2006). In this study, two stochastic models were applied for the Khazir basin to estimate their efficiency and ability for generating the daily streamflow data.

In 1962, Thomas and Fiering introduced a statistical model, which found wide acceptance and can be used for a different interval of time series. Box and Jenkins (1970) developed ARIMA model, which can be used to generate time series

of different time intervals. The progress in developing and finding new ones is ongoing until now. Many researchers have applied the ARIMA model for forecasting streamflow in different basins. Mohammadi et al. (2005) estimated the spring inflow by utilizing ARIMA and ANN models for Amir Kabir reservoir in Iran (Mohammadi K., 2005). Solis et al. (2008) used ARIMA model for forecasting streamflow of a Mexican river (Solis et al., 2008). Singh et al. (2011) forecasted the monthly streamflow of Kangsabati River in India by applying ARIMA and X-12-ARIMA (Singh et al., 2011). Ruqaya (2011) used ARIMA model for forecasting the inflow into Dokan reservoir in Iraq (AlMasudi, 2011). Veiga et al. (2014) developed short-term flow forecasting ARIMA and ANN models in the Bow River in Canada (Veiga et al., 2014). Ghimire (2017) used ARIMA model to predict flow for two hydrological stations in Schuylkill River at Berne and Philadelphia in the USA (Ghimire, 2017). Sameera (2017) compared the performance of both ARIMA and ARIMAX models and found that the ARIMAX model is better for predicting the flow of Balinda River in Iraq (Sameera, 2017). Khalid et al. (2018) applied SARIMA and Matalas models for forecasting the maximum and minimum daily flow of Tigris and Khabur Rivers in Iraq (Khalid et al., 2018).

Artificial Neural Network (ANN) is an empirical model, which has been widely applied to water resources system problems and was found to be a powerful tool for the prediction of streamflow time series. ANN was used for modeling the 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). Werbos (1974) conduct the neural networks as a tool for time series forecasting, based on observational data. Several types of neural network structures were used for forecasting and predicting time series problems such as multilayer perceptron, radial basis function, recurrent, counter propagation, and probabilistic neural networks.

The ANN model to forecast streamflow time series has been increasingly applied over the past two decades. Elena and Armando (2000) applied ANN model in two ways, conceptual type rainfallrunoff models and black-box type runoff simulation for the Sieve River basin in Italy (Toth and Brath, 2000). Sohail et al. (2006) used a new approach of training artificial neural network model (ANN) with a real coded genetic algorithm (GA) named as (GAANN) model (Sohail et al., 2006). Chowdhary and Shrivastava (2009) used the feed-forward neural network (FFNN) and radial basis function (RBF) neural network to forecast the river flow in India (Chowdhary and Shrivastava, 2009). Pandhiani and Shabri (2015) developed new hybrid models by integrating the discrete wavelet transform with an artificial neural network (WANN) model and discrete wavelet transform with least square support vector machine (WLSSVM) model to measure monthly streamflow forecasting for two rivers in Pakistan (Pandhiani and Shabri, 2015). Chu et al. (2018) forecasted runoff for the Yellow River in China by using multiple linear regressions (MLR), radial basis functions neural network (RBFNN) and supports vector regression (SVR) models (Chu et al., 2018). Zhou et al. (2018) forecasted the streamflow of the Jinsha River by using three (ANN) architectures: a radial basis function network, an extreme learning machine, and the Elman network (Zhou et al., 2018).

The main objectives of this study are to investigate the Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) models to forecast the daily flow time series for Khazir and Gomel rivers at Asmawa and Khanis stations respectively.

2. MATERIALS AND METHODS

2.1 Area of study and data collection

The area of this study is Khazir basin, which located in Kurdistan region - Iraq. The basin area is about 3185 km^2 , with a location of 43°14'00" -43°44'25" E longitude and 36°22'00" - 36°52'33" N latitude. The maximum elevation is 2165 meter (AMSL) at the north part of the basin, and the minimum elevation is 216 (AMSL) in the south part of the basin close to the basin outlet (Jassas et al., 2015). The main river in the basin is Khazir River, which started at Asmawa location formed from two side streams, one coming from Chamanke region and the other coming from Bakerman region, as shown in figure (1). Khazir River confluence the Gomel River at the southern part of the basin and then flow into the Greater

ZANCO Journal of Pure and Applied Sciences 2020

32

Zab River, which can be considers as the important tributaries of Tigris River. It is worth mentioning that Khazir River supplies the Tigris River by about 10%, that motivated the authors to select this basin as the case study in this research.

Continuous recorded daily flow time series available from the period $(2004 - 2015)$, which were obtained from two meteorological stations, first at Asmawa location (Asmawa station) which measured daily discharge flow of Khazir river and the second at Khanis location (Khanis station) that measured the daily discharge flow of Gomel River. The statistical description of the obtained data and the location of the aforementioned stations were found in tables (1) and (2).

The first ten years (2004-2013) of the available records data were considered to analyze and calibrate both models (ARIMA and ANN) while the remaining two years (2014-2015) were used to verify both of them.

Figure 1: Khazir basin.

2.2 Application of the Models:

In the current investigation, the ARIMA and ANN models were used to simulate the daily streamflow discharge for the abovementioned stations.

2.3 ARIMA [model:](https://en.wikipedia.org/wiki/Mathematical_model)

Autoregressive Integrated Moving Average (ARIMA) [model](https://en.wikipedia.org/wiki/Mathematical_model) is a generalization of an [Autoregressive Moving Average](https://en.wikipedia.org/wiki/Autoregressive_moving_average) (ARMA) model; both types are fitted to [time series](https://en.wikipedia.org/wiki/Time_series) data to present generalized data and predict future points in the series. (p,d,q) refer to ARIMA [parameters,](https://en.wikipedia.org/wiki/Parameter) which were none negative integers, (p) is referred to the [autoregressive model](https://en.wikipedia.org/wiki/Autoregressive_model) (number of time lags), (d) is the degree of differencing (the number of times the data had past values subtracted) and (q) is the order of the [moving average](https://en.wikipedia.org/wiki/Moving-average_model) model. While, the seasonal ARIMA model, which is denoted by SARIMA (p, d, q) (P, D, Q), in which (S) represents the number of periods in each season, and the uppercase (P,D,Q) stands for the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model. Seasonal Autoregressive Integrated Moving Average SARIMA $(p,d,q)(P,D,0)$ can be expressed in a mathematical form expressed in equation (1) (Wang, 2006):

$$
\varphi(B) * \Phi(B^S) * (W_t - \mu) = \theta(B) * \Theta(B^S) * \zeta_t \tag{1}
$$

Where: φ is coefficient of autoregressive (AR), θ the coefficient of moving average (MA), Φ the coefficient of seasonal autoregressive, Θ is coefficient of the seasonal moving average, ζ is the random value at time t, B is backshift operator and S is season length.

Akaike (1974) suggested a mathematical criterion formula of building the parsimony model as Akaike Information Criterion (AIC) to select an optimal model which fits the time series data among several models. Further, the Bayesian Information Criterion (BIC) is another criterion that has been developed to select an optimal model among a finite set of models (Solis et al., 2008). Akaike mathematical formulation has the form given in equation (2).

$$
AIC(p, q) = N.Ln (\sigma^2) + 2(M)
$$
 (2)

Where
$$
M = p + q + P + Q
$$
 (3)

While Bayesian formula described in equation (4).

BIC (*p*, *q*) = *N.Ln* (
$$
\sigma^2
$$
) + *M** *Ln* (*N*) (4)

Where *σ* is a standard deviation and *N* is the number of available data. The model which possesses least AIC and BIC values will be considered as an optimal model.

In this study, this concept was adopted to determine the more powerful model which can be used for forecasting of daily streamflow in Khazir basin.

2.4 ANN [model:](https://en.wikipedia.org/wiki/Mathematical_model)

Two types of the Artificial Neural Network (ANN) were applied in this research as both have been widely used in water resources engineering applications as indicated by researchers, namely; ANN-MLP and ANN-RBF. The details about both models are presented in the following:

2.4.1 Multilayer Perceptron Neural Networks (MLP)

A multilayer perceptron is a feedforward neural network architecture with uni-directional full connections between successive layers. As it is illustrated in figure (2), the structure of an MLP-ANN consists of three main layers: an input layer, a hidden layer and an output layer of neurons. These three layers were connected by strength called weight. There are two sets of weights: the input-hidden layer weights $(w_{i,j})$ and the hiddenoutput layer weights $(w_{k,i})$. These weights provide the network with high flexibility to freely adapt to the data.

The output results of the multilayer perceptron artificial neural networks can be obtained from equation (5):

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

Where \hat{y}_k is the output variable, x_i is the input variable, *n* is the number of input variables, *m* is

ZANCO Journal of Pure and Applied Sciences 2020

Input layer Hidden laver **Output** layer Bias Compare \mathbf{k} $\overline{\text{Output}}$ Target $\mathbf{w}_{\mathbf{k},\mathbf{j}}$ Error w. Adjusted weights \mathbf{w}_{ki}^* **Back propagation training**

the number of neurons in the hidden layer, $(w_{i,i})$ is the weights of input-hidden and $(w_{k,i})$ is the weight

Figure 2: Structure of multilayer perceptron functions an artificial neural network.

of hidden-output layers, b_j is the bias of the hidden layer and b_k is the bias of the output layer, f_h is the activation function of the hidden layer and f_o is the activation function of the output layer (Dreyfus, 2005). A direct relationship could be obtained using an ANN model, which needs a database of the set of output variables related to the respective input variables. These variables are set in dimensionless terms to obtain a general relationship model (Al Suhaili et al., 2014).

2.4.2 Radial Basis Function Neural Networks (RBF)

The architecture of a radial basis function neural network was shown in figure (3). This type may require more neurons than standard feed-forward backpropagation networks, but often they can be designed with lesser time (Abraham, 2004). The time-series flow data have been entered the

network as an input layer, and these data were transferred to the hidden layer by radial basis function. The response of the network was obtained in the output layer. The mathematical structure of Gaussian activation function is demonstrated in equation (6):

$$
\hat{y}_k = \sum_{j=1}^m \left(w_{k,j} * f_j \left(exp \left(- \frac{\sum_{i=1}^n (x_i - \mu_{j,i})^2}{2\sigma_j^2} \right) \right) \right) + b_k \quad (6)
$$

Where \hat{y}_k is the output variable, *x* is the input variable, *n* is the number of neurons in the inputs layer, μ is the parameter which is the position of the center of the Gaussian while σ is its standard deviation. $w_{k,j}$ is the weight of the connection between the hidden neuron *j* and the output neuron k, b is the bias and m is the number of neurons in the hidden layer.

Figure 3: Structure of typical radial basis functions an artificial neural network.

3. RESULTS AND DISCUSSION

The time series for both stations were found clear from a trend, jump and periodic. The parameters of the ARIMA model were found by applying the three stages of analysis as Identification, Parameters estimation, and Diagnostic. The order of the parameters of ARIMA models was found by applying the Autocorrelation Coefficient (ACF) and Partial Autocorrelation Coefficients (PACF).

A number of ARIMA models were tested, and the best ARIMA (p, d, q) (P, D, Q) parameters model as in equation (1) was found and shown in table (3), based on the least values of AIC and BIC for Asmawa and Khanis stations. Figure (6) shows the Autocorrelation Coefficient (ACF) and Partial Autocorrelation Coefficients (PACF) for the best models of the aforementioned stations.

Table 3: Parameters of the best models of ARIMA for Asmawa and Khanis stations.

Station Name	River	Best ARIMA model	AIC.	BIC.
Asmawa	Khazir	(2,0,0)(2,1,0)	2993.500	3011.872
Khanis	Gomel	(2,0,1)(2,1,0)	2962.986	2985.951

In the present study, the ARIMA model was applied as a single site model by using statistical software (NCSS version 11.0), while the ANN models were applied by (Matlab version 2008) and the package of Statistical Package for the Social Sciences (SPSS version 23.0) for generating and forecasting daily flow time series. However, the linear regression method was used to predict the missing data, especially for the record data for the years 2005, 2006, and 2014 for Khanis station. The stationary test of the data was conducted because the model cannot be built for nonstationary data (Chow, 1988). The normality of the time series data should be checked using the Kolmogorov-Smirnov test by applying for the MINITAB program, which shown in figure (4) with a non-zero skewness coefficient (C_s) not equal to zero. Transformation of the data to a normal distribution was carried out by the Box-Cox method, and the coefficient (λ) , was found to be (-0.4627, -0.225) for Asmawa and Khanis stations respectively. Figure (5) show the normality test of the time series data after transformation with the skewness coefficient

equal to zero.

Figure 4: Testing of the normal distribution for Asmawa and Khanis stations by Kolmogorov-Smirnov test.

ZANCO Journal of Pure and Applied Sciences 2020

Figure 5: Testing after transforming the series to the normal distribution for Asmawa and Khanis stations.

Figure 6: Autocorrelation Function, Partial Autocorrelation Function and Residual against lag for ARIMA model of Average Daily Flow Series for Asmawa and Khanis Stations.

The above ARIMA models were used in forecasting the time series of both stations, the results were demonstrated in figures (7) and (8) for the period (2014-2015) with determination coefficients (R^2) of 0.77 and 0.82 and values of the Root Mean Square Error are 3.48 and 2.19 for Asmawa and Khanis stations respectively.

37

Figure 8: Hydrograph of the forecast and recorded data of daily flow series for Khanis station using the ARIMA model.

Regarding the ANN model, two types, namely, ANN-MLP and ANN-RBF models, were used in this study to forecast the daily streamflow for Khazir and Gomel rivers at Asmawa and Khanis stations, respectively. The best model was obtained by dividing the available recorded data into four seasonal groups (winter, spring, summer, and autumn), so each group was represented by its model. ANN models for Asmawa station were found to be MLP (15,6,1), MLP (15,8,1), MLP $(15,6,1)$ and MLP $(15,6,1)$ for aforementioned seasons, while for Khanis station the best models were found to be MLP (15,9,1), MLP (15,7,1) , MLP $(15,4,1)$ and MLP $(15,8,1)$ respectively.

In ANN model investigations the MLP model was found to be more efficient than the RBF model due to its high value of determination coefficients $(R²)$ which was (0.83, 0.85) and (0.66, 0.57) for Asmawa and Khanis stations respectively, as shown in the table (4).

The architecture structures of both types of ANN models are shown in table (5) and table (6), after several trails the best activation function for MLP type between the input and hidden layers was found to be hyperbolic tangent function, while between the hidden and output layers was found to be the identity function.

		\mathbf{R}^2			RMSE		
River	Station		ANN	ANN		ANN	ANN
		ARIMA	(MLP)	(RBF)	ARIMA	(MLP)	(RBF)
Khazir	Asmawa	0.77	0.83	0.66	9.867	6.542	9.609
Gomel	Khanis	0.851	0.85	0.55	3.449	4.055	6.778

Table 4: Determination coefficient (R^2) and RMSE of ARIMA and ANN models for Asmawa and Khanis stations.

The online type of training was selected, which updates the synaptic weights after every single training data record, while to avoid overtraining, maximum training epochs computed automatically, and to specify the optimization

algorithm, the gradient descent method was selected.

The above ANN models were used in forecasting the time series for both Asmawa and Khanis stations, which shown in figures (9) and (10) respectively for the years (2014-2015).

Figure 9: Hydrograph of the forecast and recorded data of daily flow series for Asmawa station using MLP-ANN and RBF-ANN Models.

4. CONCLUSIONS

The ANN-MLP model was compared with the ARIMA model; the results revealed that the ANN model is more accurate than the ARIMA model in forecasting the daily time series for the years (2014-2015) for Asmawa station due to values of $(R²)$ and (RMSE), while the performance of both ARIMA and ANN-MLP models for the Khanis station is the same. Moreover, the ANN model can further be used to forecast for the stations' understudy, to get a more useful and accurate design of the future proposed hydraulic structures in the area of the basin.

REFERENCES

- ABRAHAM, A. 2004. Artificial Neural Networks. *In:* THORN, P. H. S. A. R. (ed.) *Handbook of Measuring System Design.* John Wiley & Sons, Ltd.
- AL SUHAILI, R. S. H., BAHRAINI, A. & BEHAYA, S. 2014. Artificial Neural Network Modeling for Dynamic Analysis of a Dam-Reservoir-Foundation System. *Journal of Engineering Research and Applications,* 4.
- AL-MASUDI, R. K. 2011. Fitting ARIMA Models for Forecasting to inflow of Dokan Reservoir. *Journal of Babylon University,* 19.
- CHOW, V. T. 1988. *Handbook of Applied Hydrology,* New York, McGraHill Book Co Inc.
- CHOWDHARY, A. & SHRIVASTAVA, R. 2009. RIVER DISCHARGE PREDICTION USING ARTIFICAL NEURAL NETWORK. *Canadian Journal of Pure and Applied Sciences***,** 1275.
- CHU, H., WEI, J., LI, J. & LI, T. 2018. Investigation of the relationship between runoff and atmospheric oscillations, sea surface temperature, and localscale climate variables in the Yellow River headwaters region. *HYP Hydrological Processes,* 32**,** 1434-1448.
- DREYFUS, G. 2005. *Neural networks methodology and applications,* Berlin; New York, Springer.
- GHIMIRE, B. N. S. 2017. Application of ARIMA Model for River Discharges Analysis. *J. Nep. Pnys. Soc. Journal of Nepal Physical Society,* 4**,** 27.
- JASSAS, H., KANOUA, W., MERKEL, B., JASSAS, H. & KANOUA, W. 2015. Actual evapotranspiration in the Al-Khazir Gomal Basin (Northern Iraq) using the surface energy balance algorithm for land (SEBAL) and water balance. *Geosciences (Switzerland),* 5**,** 141-159.
- KHALID, M. K., ADIL, M. R. & SYRAN, A. I. 2018. Stochastic Models for the Maximum and Minimum Daily Flows of Tigris and Khabur Rivers. *ZANCO*

Journal of Pure and Applied Sciences 30(s1)**,** 50- 61.

- MARQUES, C. A. F., FERREIRA, J. A., ROCHA, A., CASTANHEIRA, J. M., MELO-GONCALVES, P., VAZ, N. & DIAS, J. M. 2006. Singular spectrum analysis and forecasting of hydrological time series. *JPCE Physics and Chemistry of the Earth,* 31**,** 1172-1179.
- MOHAMMADI K., E. H. R. A. S. D. D. 2005. Comparison of Regression, ARIMA and ANN Models for Reservoir Inflow Forecasting using Snowmelt Equivalent (a Case study of Karaj). *J. Agric. Sci. Technol,* 7**,** 17-30.
- PALIT, A. K. & POPOVIC, D. 2006. Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications.
- PANDHIANI, S. M. & SHABRI, A. B. 2015. Time series forecasting by using hybrid models for monthly streamflow data. *ams Applied Mathematical Sciences,* 9**,** 2809-2829.
- SAMEERA, A. O. 2017. Comparison Between Forecasting ARIMA and ARIMAX Method. *ZANCO Journal of Pure and Applied Sciences* Vol. 28, no. 6**,** 158- 65.
- SINGH, M., SINGH, R. & SHINDE, V. 2011. Application of Software Packages for Monthly Stream Flow Forecasting of Kangsabati River in India. *IJCA International Journal of Computer Applications,* 20**,** 7-14.
- SOHAIL, A., WATANABE, K. & TAKEUCHI, S. 2006. Streamflow forecasting by artificial neural network (ANN) model trained by real-coded genetic algorithm (GA)-A case study when the role of groundwater flow component in surface runoff is small,48**,** 233-262.
- SOLIS, J. F., ESMERALDA, P. & JAVIER, L. 2008. Shortterm streamflow forecasting: ARIMA vs neural
networks. AMERICAN CONFERENCE ON networks. *AMERICAN CONFERENCE ON APPLIED MATHEMATICS (MATH '08), Harvard, Massachusetts, USA, March 24-26, 2008*.
- TOTH, E. & BRATH, A. 2000 Flood Forecasting Using Artificial Neural Networks in Black-Box and Conceptual Rainfall-Runoff Modeling. *Diaksas*.
- VEIGA, V., HASSAN, Q. & HE, J. 2014. Development of Flow Forecasting Models in the Bow River at Calgary, Alberta, Canada. *Water Water,* 7**,** 99-115.
- WANG, W. 2006. Stochasticity, nonlinearity and forecasting of streamflow processes.
- ZHOU, J., PENG, T., ZHANG, C. & SUN, N. 2018. Data Pre-Analysis and Ensemble of Various Artificial Neural Networks for Monthly Streamflow Forecasting. *Water Water,* 10**,** 628.