CrossMark

ORIGINAL ARTICLE

Assessment of Stochastic Models and a Hybrid Artificial Neural Network-Genetic Algorithm Method in Forecasting Monthly Reservoir Inflow

Hamid Moeeni[1] · Hossein Bonakdari[1,2] · Seyed Ehsan Fatemi[3] · Amir Hossein Zaji[1]

Abstract In this study, the performance of Seasonal Autoregressive Integrated Moving Average (SARIMA) models and hybrid Artificial Neural Network-Genetic Algorithm (ANN-GA) method in forecasting the monthly inflow to a dam is examined and compared. The number of parameters required for SARIMA models is determined using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) diagrams. In terms of these diagrams, 14 input combinations are considered for the ANN-GA models. An examination of the model's performance in prediction utilizing ANN-GA indicates that applying discharges with 1, 2, 6, and 12-month lag time leads to the best prediction. The model results are compared using mean absolute relative error (MARE) and correlation coefficient (R) indexes. These index values for the SARIMA model are 0.388 and 0.76 and for the ANN-GA model they are 0.815 and 0.8, respectively. Moreover, to evaluate the ability of the models to make short-term and long-term predictions, relative error (E_i) and average of the cumulative relative error (F_i) are computed for the forecasted discharges. The results indicate that the SARIMA model is more capable in forecasting monthly inflow, especially for low values, than the ANN-GA model. The SARIMA model is also much more accurate than the ANN-GA model in short-term and long-term forecasting.

✉ Hossein Bonakdari
  bonakdari@yahoo.com

1   Department of Civil Engineering, Razi University, Kermanshah, Iran

2   Water and Wastewater Research Center, Razi University, Kermanshah, Iran

3   Department of Water Engineering, Razi University, Kermanshah, Iran

Introduction

There are several stochastic models such as stochastic and computational intelligence models for hydrological modelling and prediction purposes used by different researchers (Irvine and Eberhardt 1992; Mohan and Vedula 1995; Ahmad et al. 2001; Yurekli and Kurunc 2005; Yurekli et al. 2005; Abebe and Foerch 2008; Nirmala and Sundaram 2010, Zhang et al. 2011; Ali 2013). Damle and Yalcin (2007) attempted to predict floods in the Mississippi River, USA, by utilizing time series. Their results showed that time series are capable of generating daily discharge data and the resulting predictions are accurate. Mombeni et al. (2013) used SARIMA model for seasonal and non-stationary time series to estimate residential water consumption. Their results allow practitioners and planners to explore realistic decision making scenarios for designing effective water demand management. Valipour (2015) studied the performance of SARIMA and ARIMA models in forecasting long-term runoff in the United States. The results indicated that the SARIMA model outperformed ARIMA with a relative error of less than 5% for all states. Wang et al. (2015) combined the ARIMA model with ensemble empirical mode decomposition (EEMD) to forecast annual runoff time series. It was found that the new model can significantly improve ARIMA time series approach prediction.

On the other hands, Artificial Neural Network (ANN) application in hydrology has also been developed in recent years (Jain et al. 1999; Coulibaly et al. 2000; Xu and Li 2002; Anctil et al. 2004). Kilinç and Cigizoglu (2005)

Springer

predicted the monthly inflow, storage and evaporation in the Canak Valley area using by the ANN. They used both radial and sigmoid function types in ANN. Their study results signified that the sigmoid function is more accurate than the radial function. Kisi and Cigizoglu (2007) examined different ANN techniques, including Feed-forward Backpropagation (FFBP), Radial Basis Function-based Neural Networks (RBF) and Generalized Regression Neural Networks (GRNN) along with AR modeling in predicting daily discharge flow. The results showed that the RBF technique performs better than the other models.

Stochastic models are usually categorized by linear models and also consider the random component of data whereas ANN models are capable to model the linear and nonlinear parameters in hydrological modelling. Therefore, several researchers have examined stochastic linear models and a variety of methods based on artificial intelligence in predicting hydrological time series (Kisi 2004; Mohammadi et al. 2005; Landeras et al. 2009; Khatibi et al. 2014). Fereydooni et al. (2012) employed the multilayer perceptron using a backpropagation model to predict rainfall, temperature and evaporation data, and compared the results with the ARIMA method. They concluded that ANN performs better than ARIMA. Valipour et al. (2013) used ARMA, ARIMA and ANN models to forecast Dez reservoir monthly inflow. According to the results, the ARIMA model has less error compared with the ARMA model and the ANN model has less error than ARIMA. Golabi et al. (2013) predicted seasonal rainfall using artificial neural network (ANN) and ARIMA models. It was concluded that the ANN model is much more accurate for their datasets. Pektaş and Cigizoglu (2013) modelled the monthly runoff coefficients of 7 southern large basins by using univariate autoregressive integrated moving average (ARIMA), multivariate ARIMA (ARIMAX) and ANN models. The authors indicated that multivariate models are superior to univariate models in predicting monthly runoff coefficients.

According to these studies, the large and comprehensive stochastic and computational intelligence models were used to forecast some hydrological parameters that the models performance was compared by together. The key question is 'the results would be constant in critical condition?' In the other words, if the ANN models have the better performance rather than SARIMA models to forecast the river flows, would they be better in forecasting river flows in droughts and floods conditions or not? The second question is 'how many important to choose the type of algorithm in the computational intelligence models in compare of stochastic models to find the best model performance?' In the current study, the performance of the Seasonal Autoregressive Integrated Moving Average (SARIMA) models and hybrid Artificial Neural Network-

Genetic Algorithm (ANN-GA) method in predicting the input monthly discharge flow to the Jamishan Reservoir located in the west of Iran is compared and evaluated. Therefore, 14 and 80 different input combinations with different time lags are determined and used in the ANN-GA and SARIMA models, respectively.

Materials and Method

Study Area and Data

Data from Pirsalman Hydrometric Station (46°40′N, 34°14′E) are used in this research for modeling and forecasting the monthly discharge of the Jamishan River, which is located in the west of Iran. The data include 30 water years worth of data from October 1981 to September 2011. The initial 21 years are used as the calibration period and the 9 remaining years are used as the validation period.

The Jamishan Reservoir dam is currently under construction on the Jamishan River located in southwest Songor Province. This reservoir dam will have a normal volume of 62.8 million cubic meters and is aimed at providing the essential agricultural water for this area, controlling floods and preventing damages caused by flooding.

Data Pre-Processing

In hydrologic time series modelling, time series duration should include droughts and wet periods. Thus, the number of data duration must be adequate for modelling. In this regard, the Hurst coefficient (Hurst et al. 1965) is applied in this study. This coefficient represents the adequacy of time duration. If this ratio is greater than 0.5, the time series will adequate for modelling. This coefficient is as follow:

$$H = \frac{\text{Log}\left(\frac{S_{max} - S_{min}}{S_x}\right)}{\text{Log}\left(\frac{N}{2}\right)}, \tag{1}$$

where H is the Hurst coefficient, N is the number of data, S_x is the standard deviation of the data, S_{max} is the maximum cumulative mean difference and S_{min} is the minimum cumulative mean difference. In this study, this ratio is obtained by 0.711.

In the ARIMA model, data should follow normal distribution (Salas et al. 1988) The Box-Cox transformation (Box and Cox 1964) is used to stabilize data and for standard deviation normalization. Jarque–Bera testing (Jarque and Bera 1980) serves to investigate discharge flow data normalization. Figure 1 shows the ACF and PACF diagrams for the normalized inflow in the calibration period for delays 1–48. In this figure, LL and UL are the upper and lower limit in the 95% confidence level, respectively.

Fig. 1 ACF and PACF graphs for the normalized data in the calibration period
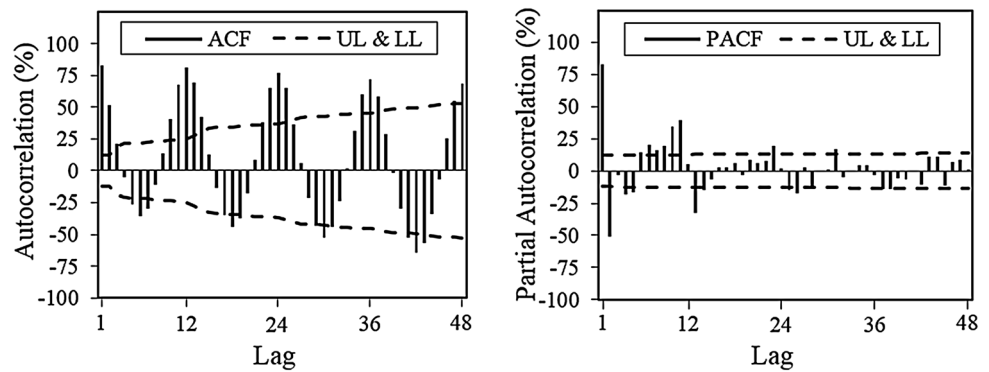
As seen in the figure, the values of both diagrams in the first two lags is high. As a result, maximum two, autoregressive and moving average parameters are needed for the SARIMA model (Cryer and Chen 2008).

These diagrams also used to determine the seasonal differencing order (D), non-seasonal differencing order (d), and data periodicity (ω) to make the time series stationary. It can be understood that the ACF and PACF diagrams are not damped. The PACF model also changes its sign (while it is high) immediately after the first lag time, which indicates that the time series is non-stationary. The ACF diagram has intervals of 12, signifying that the time series data have seasonal fluctuations (periodical component) with intervals of 12. Thus, one seasonal differencing is required. On the other hand, the ACF diagram is around the symmetrical vertical axis with a mean of nearly 0. This indicates the lack of a "trend component" in the time series. As a result, seasonal differencing is not necessary. Thus, we have ω = 12, d = 0 and D = 1.

Modeling Process with Stochastic Approaches

SARIMA model (p,d,q) (P,D,Q)$_\omega$ has seasonal and non-seasonal autoregressive and moving average components that are expressed as follows:

$$
\begin{aligned}
&(1 - \Phi_1 B^\omega - \Phi_2 B^{2\omega} - \cdots - \Phi_P B^P)(1 - \phi_1 B^\omega - \phi_2 \\
&\quad \times B^{2\omega} - \cdots - \phi_p B^p)(1 - B^\omega)^D (1 - B)^d x_t \\
&= (1 - \Theta_1 B^\omega - \Theta_2 B^{2\omega} - \cdots - \Theta_Q B^Q) \\
&\quad \times (1 - \theta_1 B^\omega - \theta_2 B^{2\omega} - \cdots - \theta_q B^q)\varepsilon_t,
\end{aligned} \tag{2}
$$

where x_t is the observed data in time t, ε_t is the random variable, ω is the periodicity, B is the difference operator B $(x_t) = x_{t-1} (1-B^\omega)^D$ is the Dth seasonal difference measure ω, $(1-B)^d$ is the dth non-seasonal difference, p is the non-seasonal autoregressive order, q is the order of the non-seasonal moving average, P is the seasonal autoregressive order, Q is the order of seasonal moving average, φ is the non-seasonal autoregressive parameter, θ is the parameter of non-seasonal moving average, Φ is the seasonal autoregressive parameter and Θ is the seasonal moving average.

After fitting the models, those that have parameters with a significant difference from zero and that also have independent residuals that will be accepted. In this study, the t-student test is used to examine the significance of the model parameters. This test statistic is expressed as:

$$\tau = \text{Pr/Se} \tag{3}$$

where τ, Pr and Se are the test statistic, estimated value of each parameter (including Φ, Θ, ϕ and θ) and estimated standard error for each parameter, respectively. By considering 5% as the significance level, the intended parameter is significant when the value of the corresponding probability of τ statistic level is smaller than this value in the t-student distribution ($P_\tau < 0.05$). This means that the parameter effectively participates in modelling.

The Box-Pierce test (Box and Pierce 1970; McLeod 1978) is used to examine the independency of the residuals and this test statistic is as follows:

$$Q^* = n(n+2) \sum_{l=1}^{L} r_l^2(\varepsilon)/(n-1) \tag{4}$$

where Q^*, $r_l(\varepsilon)$, n and L are the test statistic, the residual's autocorrelation coefficient with lag l, the number of non-missing months after differencing (n = N−d−Dω) and the maximum time lag, respectively. In addition, the number of model parameters is defined as k = p + q + P + Q. In case the level of the corresponding probability of Q* in the Chi-square distribution with a degree of freedom DF = L−K−1 is larger than the intended level of significance, which is 5% ($P_{Q*} > 0.05$), the residuals series will be independent.

The periodical component is one of the most effective factors that cause dependency in time series. Since monthly discharge data have this component, the model residuals may have the periodical component and they will consequently not be independent. The cumulative periodogram is also used to ensure that the periodical component is

completely removed from the residuals. The cumulative periodogram for residuals is expressed as:

$$MSD(h_j) = 2\left[\left(\sum_{t=1}^{n}\varepsilon_t\cos(2\pi h_j t)\right)^2 + \left(\sum_{t=1}^{n}\varepsilon_t\sin(2\pi h_j t)\right)^2\right]$$
$$/n^2, \quad h_j = j/n \tag{5}$$

$$P_i = \sum_{j=1}^{i} MSD(h_j)/\sigma_\varepsilon^2, \quad i = 1, 2, \cdots, n/2, \tag{6}$$

where P_i, h_j, MSD, ε_t, and σ_ε^2 are the cumulative periodogram of residuals, the frequency, the mean squared deviation, the values of residuals in time t and the variance of ε_t, respectively. If the residual time series is independent, the P_i graph, in terms of h_i, will be close to the line that connects the (0, 0) and (0.5, 0) points. The Kolmogorov–Smirnov confidence limits are away from the mentioned lines by $\pm K_\alpha/\sqrt{n'}$. $n' = (n - 2)/2$, $n' = (n - 1)/2$ for even and odd numbers, respectively. The value of K_α is also equal to 1.36 at the 95% confidence level.

Hybrid Artificial Neural Network-Genetic Algorithm (ANN-GA) Method

The multi-layer perceptron (MLP) neural network is one of the most common artificial intelligence methods. An MLP is formed from three types of layers: an input layer, one or more hidden layers and an output layer, all of which consist of neurons. The neurons in each layer evaluate the weighted summation of the previous layer's neurons and transfer the result to the next layer. The numbers of input and output layer neurons are equal to the numbers of input and output variables of the problem, respectively. However, there is no specific rule to determine the number of hidden layer neurons and is thus a perplexing problem in MLP simulations. In this study, the novel, hybrid ANN-GA method is employed. The ANN-GA procedure is shown in Fig. 2, which presents the steps in evaluating the optimum ANN-GA model. To begin with, the input variables are presented to the model. Secondly, the random populations of different MLP models with various numbers of hidden layer neurons are constructed. These MLPs perform as chromosomes for the modified GA employed. In the present models, MLPs with two hidden layers are used. Next, each MLP generated in the previous step is run and its cost is evaluated. The calculated costs are then sorted and the ANN-GA termination criteria are checked. If the criteria are fulfilled, the best chromosome is saved as the optimum ANN-GA model. Otherwise, the procedure continues. Next, by using GA operators such as crossover, mutation and the elite process, the next generation is constructed.

Lastly, the termination criteria are checked; if fulfilled, the process stops, otherwise the previous step is run again.

To train the MLP models, the Levenberg–Marquardt (LM) Algorithm (Levenberg 1944) is used. Thus, the MLP models are trained in a random manner. As a result, a high-performance chromosome may be turned off by the GA due to bad luck in the LM training algorithm. In order to solve this defect, a modification is done on the GA. As seen in the flowchart of the modified GA, the elite population runs several times and the best cost of the repeated runs is saved. With this modification, the probability of elite chromosome elimination (due to training defects) is reduced significantly.

According to the ACF graph in Fig. 1, the autocorrelation values take large values in lags 1, 2, 3, 6, 12, and 24, such that they intersect at UL and LL. This indicates that discharges with the above-mentioned lags dramatically affect each other. Therefore, combinations of these discharges are used to select the proper input for the hybrid Artificial Neural Network-Genetic Algorithm (ANN-GA) method. The 14 input combinations considered for the ANN-GA model are presented in Table 1.

Evaluation Criteria of the Best Models

The Mean Squared Error (MS) and Corrected Akaike's Information Criterion (AICc) are used to select the best SARIMA model (the model with the minimum error and minimum number of parameters):

$$AICc = n.Ln(\sigma_\varepsilon^2) + 2kn/(n - k - 1) \tag{7}$$

$$MS = \sum_{i=1}^{n} (Q_{ni} - \hat{Q}_{ni})^2/(n - k - 1), \tag{8}$$

where Q_{ni} is the normalized value of the observed discharge, Q_{ni} is the normalized fitted discharge value, k is the number of model parameters and n is the number of non-missing months after differencing in the data sample for the calibration period. Moreover, to ensure the best SARIMA model is selected to subsequently select the best ANN-GA model and compare these two models, Mean Absolute Relative Error (MARE), correlation coefficient (R), Root mean squared errors (RMSE) and Nash-Sucliffie criteria are separately calculated for the calibration and validation periods between the observed and fitted, or forecasted data.

$$MARE = \sum_{i=1}^{n} \left|\frac{Q_i - \hat{Q}_i}{Q_i}\right|/n \tag{9}$$

$$R = \sum_{i=1}^{n} (Q_i - \bar{Q}_i)(\hat{Q}_i - \bar{\hat{Q}}_i)/\sqrt{\sum_{i=1}^{n}(Q_i - \bar{Q}_i)^2 \sum_{i=1}^{n}(\hat{Q}_i - \bar{\hat{Q}}_i)^2} \tag{10}$$

Fig. 2 ANN-GA flowchart
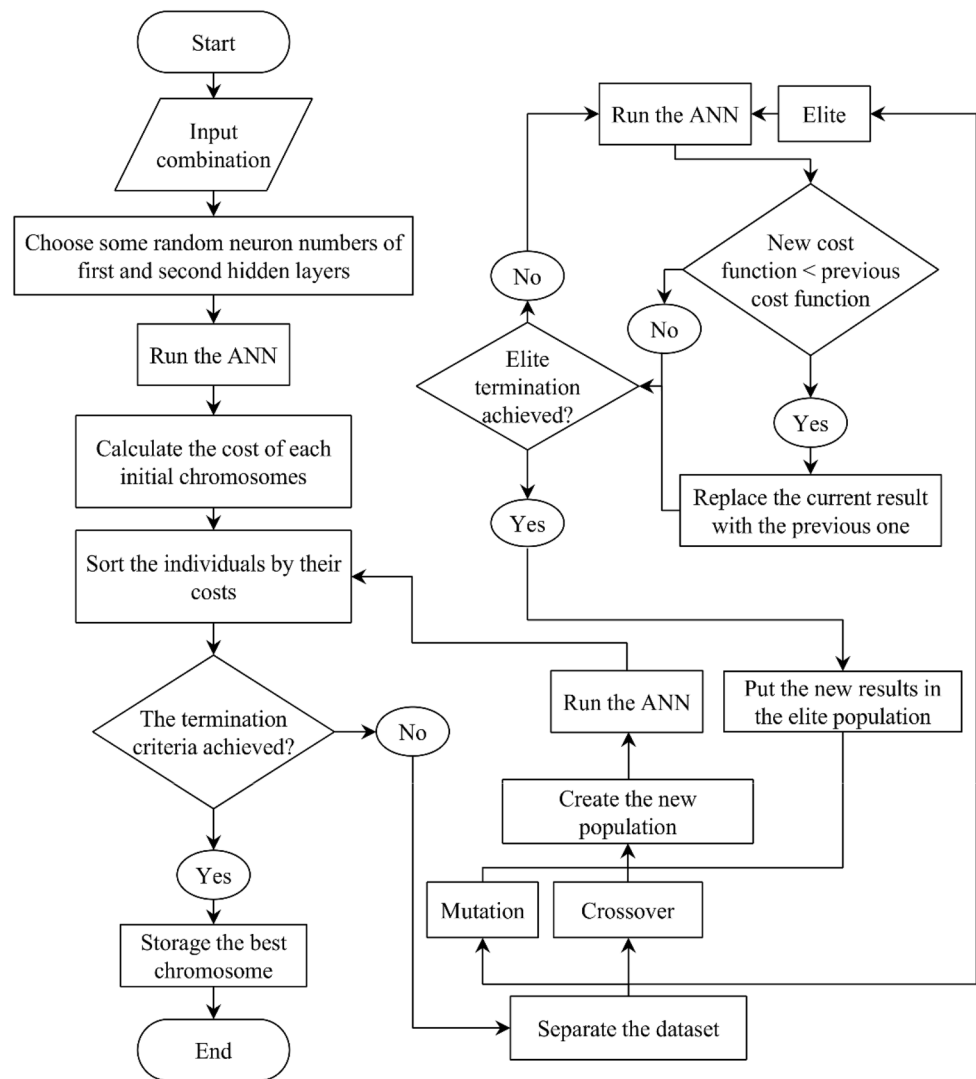
$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (Q(t) - \hat{Q}(t))^2}{n}} \tag{11}$$

$$\text{Nash-Sucliffe} = 1 - \frac{\sum_{t=1}^{n} (Q(t) - \hat{Q}(t))^2}{\sum_{t=1}^{n} (Q(t) - \bar{Q}(t))^2}, \tag{12}$$

where Q_i is the observed discharge in the ith month, $\hat{Q}_i$ is the computed discharge in the ith month, $\bar{Q}_i$ is the mean of the observed discharges, $\bar{\hat{Q}}_i$ is the mean of the computed discharges, n is the number of non-missing months after differencing for calibration or the validation period, and i is the month number. The mean values of the observations obtained are 1.85 and 1.54 m^3/s in the calibration and validation periods, respectively. In addition, determining the time error and best time of forecasting, and comparing the performance of the best SARIMA and ANN-GA models, the absolute relative error in month i (E_i), average of cumulative relative errors in month i (F_i), and coefficient of relative error variation (C_V) are calculated as follows:

$$E_i = \left| Q_i - \hat{Q}_i \right| / Q_i \tag{13}$$

$$F_i = \sum_{j=1}^{i} E_j / i \tag{14}$$

$$C_v = \sqrt{\sum_{i=1}^{n} \frac{(E_i - E)}{n}} / E \tag{15}$$

where $\bar{E}$ is the relative error average.

Results and Discussion

Selecting the Best Models

Different SARIMA and ANN-GA models were applied to the monthly inflow data collected from the Jamishan Reservoir dam. The results obtained by utilizing different statistical indexes for both calibration and validation

Table 1 Different ANN-GA models with different input combinations

Models	Input combinations
ANN-GA-1	Q_{t-1}
ANN-GA-2	Q_{t-1}, Q_{t-2}
ANN-GA-3	Q_{t-1}, Q_{t-12}
ANN-GA-4	Q_{t-1}, Q_{t-6}
ANN-GA-5	Q_{t-1}, Q_{t-6}, Q_{t-12}
ANN-GA-6	Q_{t-1}, Q_{t-2}, Q_{t-12}
ANN-GA-7	Q_{t-1}, Q_{t-2}, Q_{t-6}
ANN-GA-8	Q_{t-1}, Q_{t-2}, Q_{t-3}
ANN-GA-9	Q_{t-1}, Q_{t-2}, Q_{t-6}, Q_{t-12}
ANN-GA-10	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-12}
ANN-GA-11	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-6}
ANN-GA-12	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-6}, Q_{t-12}
ANN-GA-13	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-12}, Q_{t-24}
ANN-GA-14	Q_{t-1}, Q_{t-2}, Q_3, Q_6, Q_{12}, Q_{t-24}

periods were presented in Table 2. As seen in this table, to select the best model using SARIMA, 5 superior models were fitted from amongst 80 different models that were evaluated with the AICc and MS indexes. These indexes not only consider the difference between the observed and fitted values but also take into account the effects of number of parameters. The SARIMA$(1,0,2)(0,1,1)_{12}$ model is clearly more accurate than other models. To select the best model presented using ANN-GA, the MARE and R indexes are used for the validation and calibration periods. The maximum value of R calculated is 0.86 in the calibration period for the ANN-GA-14 model. This model has the maximum discharge as input with 1, 2, 6, 12 and 24-month lag times. However, it does not present good results in the validation periods and is amongst the weakest models. The ANN-GA-3 model with the best MARE in calibration and that uses 1 and 12-month lag times in discharge forecasting, is fairly accurate in the validation period. It presents quite similar results in both periods, which is an indication of this model's flexibility in forecasting monthly discharge. However, since the final objective of modeling is forecasting and the model's forecasting error is of utmost importance, the ANN-GA-9 model with the best result in the validation period (MARE = 0.715, R = 0.802) as Q = f(Q_{t-1}, Q_{t-2}, Q_{t-6}, Q_{t-12}) is selected as the best model presented using ANN-GA.

The t-student test serves as a significance test in examining the significance of parameters. The Box-Pierce test is used as an independence test to examine the independency of the selected SARIMA model's residuals. The results are presented in Table 3.

The significance test results indicate that the P_τ values are smaller than the intended significance level of 5% for

each parameter. The estimated value of each parameter is therefore within this level of significance. In fact, there is a substantial difference between all parameters and 0. Therefore, the parameters effectively participate in modelling and forecasting. Regarding the independence test, since the value of the corresponding probability of Q* in the Chi-square distribution with degree of freedom (DF) per lag L, meaning the P_{Q*} value, is larger than the intended significance level of 5% in different lags, the residuals of the selected SARIMA model are independent based on the results of this independence test and they are not dependent on time. The large difference between the P_{Q*} values and 5 indicates that the residuals are independent with a high degree of confidence.

A cumulative periodogram was employed to make sure that the periodical component was completely removed from the model residuals. Figure 3 shows the cumulative periodogram for the residuals of the best SARIMA model. The vertical axis represents the frequency in this graph and the horizontal axis shows the cumulative periodogram for the residuals. UL and LL are the Kolmogorov–Smirnov confidence limits at the 95% level of confidence. Since the graph of P_i in terms of h_i, is close to the line that connects the (0, 0) and (0.5, 0) points and it is within the range between UL and LL, it can be concluded that the periodical component has been perfectly removed from the residuals. This is another reason for the independency of residuals in the SARIMA model selected.

Model's Ability to Forecast Inflow

Figure 4 represents the monthly discharge time series forecasted by the ANN-GA-9 and SARIMA$(1,0,2)(0,1,1)_{12}$ models against the observed discharge values. The models identified the monthly discharge changes with regards to time to an acceptable degree. By increasing the observed discharge values, the two model's furcated values increased and vice versa. This shows that both models identified the monthly mean discharge changes for these datasets very well. These are, however, less accurate as the discharge value increases due to the intense seasonal changes in peak discharge. For instance, the monthly discharge was 0.17 m³/s in September 1994 and 13.34 m³/s in December 1994. These dramatic changes decreased the peak-discharge forecasting accuracy.

Figure 5 shows the performance of the ANN-GA-9 and SARIMA$(1,0,2)(0,1,1)_{12}$ models in modeling and forecasting the monthly inflow to the dam in the calibration and validation periods. As shown in this figure, the results for low-flow discharge in both validation and calibration periods are closer to the best fit line than the peak discharge, which indicates that these models perform better for low-value discharge than peak discharge. The SARIMA

Table 2 Evaluation criteria results for determining the best SARIMA and ANN-GA models

Model	MS	AICc	Calibration		Validation	
			MARE	R	MARE	R
SARIMA(1,0,2)(0,1,1)$_{12}$	0.224	−356.48	0.38	0.826	0.388	0.761
SARIMA(1,0,1)(0,1,1)$_{12}$	0.227	−354.05	0.381	0.827	0.399	0.759
SARIMA(1,0,0)(0,1,1)$_{12}$	0.229	−353.02	0.391	0.83	0.398	0.751
SARIMA(1,0,2)(3,1,0)$_{12}$	0.267	−311.27	0.405	0.803	0.5	0.748
SARIMA(1,0,0)(3,1,0)$_{12}$	0.272	−309.49	0.423	0.813	0.493	0.738
ANN-GA-1	–	–	1.399	0.705	1.134	0.674
ANN-GA-2	–	–	1.284	0.768	0.943	0.741
ANN-GA-3	–	–	0.801	0.787	0.877	0.765
ANN-GA-4	–	–	1.466	0.846	1.093	0.685
ANN-GA-5	–	–	0.995	0.849	1.045	0.75
ANN-GA-6	–	–	0.946	0.828	0.819	0.774
ANN-GA-7	–	–	2.635	0.781	0.973	0.775
ANN-GA-8	–	–	1.421	0.829	0.787	0.752
ANN-GA-9	–	–	1.183	0.795	0.715	0.802
ANN-GA-10	–	–	1.243	0.824	0.914	0.756
ANN-GA-11	–	–	1.087	0.827	1.057	0.687
ANN-GA-12	–	–	2.658	0.769	1.125	0.764
ANN-GA-13	–	–	1.8	0.809	1.168	0.648
ANN-GA-14	–	–	1.245	0.861	0.903	0.69

Table 3 Independence test results for the residuals and significance test results for the selected SARIMA model parameters

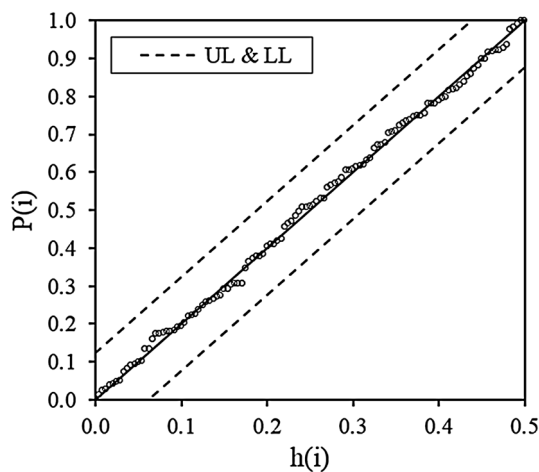| Significance test					Independence test				
Pr	φ_1	θ_1	θ_2	Θ_1	L	12	24	36	48
Estimated Pr	0.9	0.195	0.149	0.92	Q^*	2.5	19.5	27.3	41.2
τ	21.71	2.46	1.98	20.76	DF	7	19	31	43
P_τ	0	1.4	4.8	0	P_{Q^*}	93	42.4	65.7	55

Fig. 3 Cumulative periodogram for the best SARIMA model residuals

model outperforms ANN-GA in this respect. The results obtained by SARIMA are underestimated and overestimated for all discharges (low and peak discharges); although, the underestimated, fitted and forecasted discharges have a relative error greater than that of the overestimated ones. The results obtained with ANN-GA are inconsistent with SARIMA as the discharge increases and especially at peak discharges. ANN-GA underestimates most of the discharges as discharge increases. ANN-GA estimates the monthly inflow to the dam with a smaller relative error compared with the validation state where the peak discharge has a smaller value than the calibration state. Regarding the incorrect estimation and large difference from the actual values, ANN-GA will not perform properly for major floods.

Evaluation of the Models in Drought and Wet Conditions

Along with MARE criterion based on the relative errors, Nash-Sucliffe and RMSE criteria based on absolute errors, were calculated only for the best models. These criteria calculated for the best ANN-GA model and then it was considered with the same settings for the simple ANN

Fig. 4 SARIMA and ANN-GA
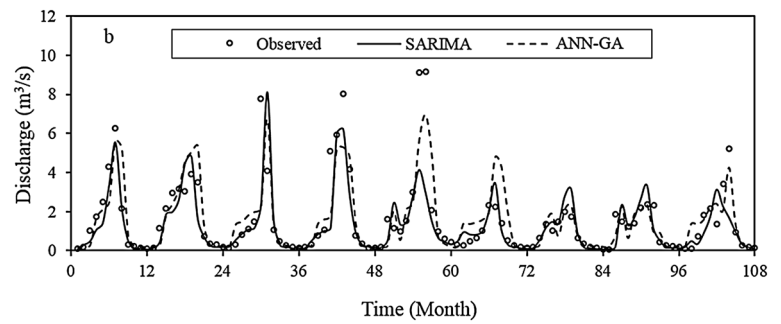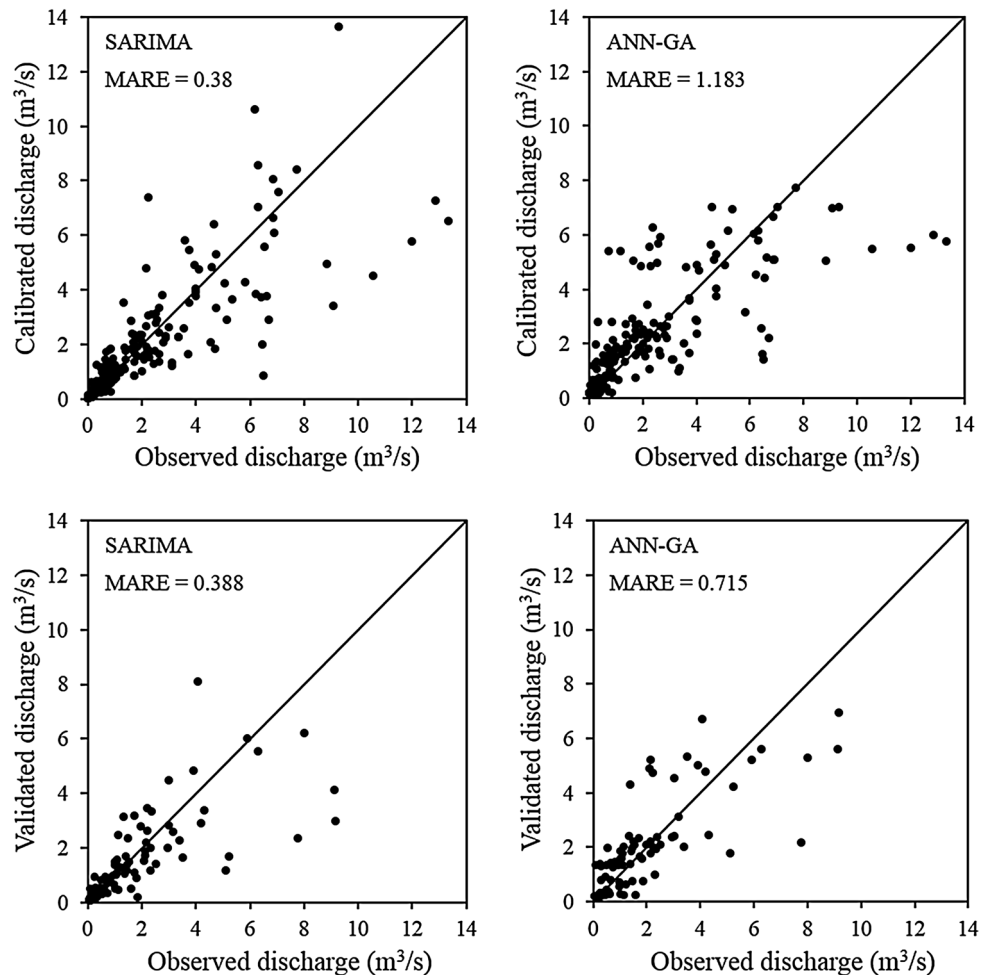model results for the validation
period

Fig. 5 Observed and modeled
monthly discharge by the
SARIMA and ANN-GA models
in the calibration and validation
periods

model presented in Table 4. The results show that these criteria are so closely in ANN-GA and ANN models and also using the genetic algorithm in ANN model cause to occur insignificantly changes in prediction accuracy. Therefore, using advanced algorithm in ANN model always not to be useful in increasing of forecasting model accuracy. In addition, Table 4 shows that in comparison of ANN and ANN-GA models, the MARE criterion is significantly improved and the other criteria are also decreased in SARIMA model. Since MARE is obtained based on relative errors, it's very sensitive in the small values and it changes suddenly in the small values. On the other hand, Nash-Sucliffe and RMSE criteria are obtained based on the absolute errors are more sensitive in large values. As a result, SARIMA model is the best model to forecast the base flows and in drought conditions and also ANN and ANN-GA models are the best models to predict the peak flows and in the flood conditions in this study.

Table 4 Evaluation criteria results for best models

Criteria	ANN-GA	ANN	SARIMA
R^2	0.64	0.61	0.58
MARE	0.72	0.75	0.40
Nash-Sucliffe	0.64	0.58	0.56
RMSE	1.16	1.24	1.27

Table 5 Minimum F and E index values along with the mean and standard deviation of the relative error

|	E_{min}	$\bar{E}$	C_v	F_{min}	Month
SARIMA$(1,0,2)(0,1,1)_{12}$	0.007	0.398	1.243	0.137	First
ANN-GA-9	0.004	0.715	1.940	0.424	Seventh

Time Changes of the Model's Forecast Error

Equations 10–12 are used to examine the values forecasted by the models in order to study the forecasting changes with time. Table 5 shows the values of E_{min}, C_v and $\bar{E}$ along with F_{min} and the month in which this value occurred for the validation period. The C_V and $\bar{E}$ values are smaller for SARIMA, indicating that the error changes are lower and this model is superior to ANN-GA, since the smaller the relative errors, the smaller their mean will be. C_V being smaller indicates the fact that the errors are closer to the mean for the SARIMA model, and as a result, the error changes in time and the peak error points are lower in this model. Figure 6 shows the changes in the relative error and average cumulative relative error in time for SARIMA and ANN-GA in the forecast, or validation period. It can be understood from the E index graph that the ANN-GA model has more and much larger peak error points than SARIMA. It is clear from the index F graph in this same figure that this index value is smaller for the SARIMA

model than the ANN-GA model. The maximum F value is 0.44 for the SARIMA model and the minimum value of this index is equal to 0.43 for the ANN-GA model ($F_{max(SARIMA)} \simeq F_{min(ANN-GA)}$). The F index value for the SARIMA model hardly reached the index value of the ANN-GA model. This also shows the superiority of SARIMA over ANN-GA. Two points can be noted from this graph. First, the minimum error of the SARIMA model occurs in predicting the following month (1 month later). The F value of ANN-GA is 10 times the value of the F index for the SARIMA model in this month, meaning that SARIMA forecasts the discharge of the following month 10 times more precisely than ANN-GA. The second point is that the SARIMA model graph of F index changes become almost horizontal with time and as the forecast horizon increases, such that it reaches a sort of stagnation that does not exist with the ANN-GA model. This graph displays an upward trend for the ANN-GA model, meaning that the ANN-GA model error increases as the forecast horizon increases. These two points indicate that the SARIMA model is capable of making short-term forecasts with much lesser error and long-term forecasts without an increase in error—two abilities lacking in the ANN-GA model. Therefore, the SARIMA model most definitely performs better than ANN-GA in short-term planning such as exploitation and consumption management and in long-term planning such as designing and constructing hydraulic structures.

Conclusion

In this study the abilities of SARIMA and the hybrid Artificial Neural Network-Genetic Algorithm (ANN-GA) method in forecasting the monthly inflow to the Jamishan Dam in the west of Iran were analyzed and their results were compared with simple ANN model. Data from

Fig. 6 Graphs of E and F index changes over time for the discharges forecasted by the selected models
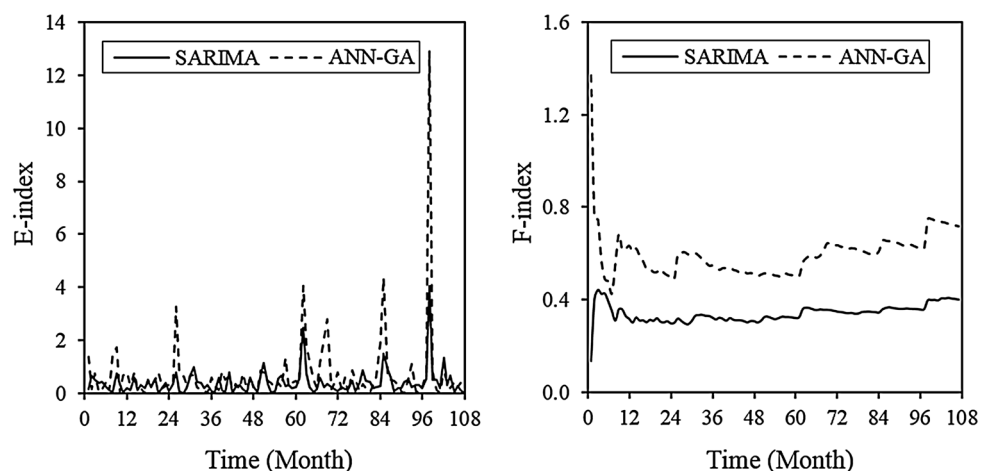
Pirsalman hydrometric station were used, including 30 water years. The initial 21 years served as the calibration period and the 9 final years were used as the validation period. The best SARIMA model had significance parameters and independence residuals. Both models identified the process of monthly discharge changes very well. Forecasting the base flow and low flows was done precisely by the SARIMA model precisely in comparison of ANN and ANN-GA models and also these models are more suitable to forecast the peak flows and flood flows in comparison of SARIMA model. The results show that there were no significant changes between ANN and ANN-GA models while the results of ANN-GA were a little bit better than ANN model. So, SARIMA model is more suitable in the drought years and low flow forecasting and ANN-GA model is more suitable in the wet years and flood flows forecasting. The SARIMA model also had lower error changes over time during the prediction period, such that it had lower error peak points in the E index graph than ANN-GA. Analysing the graph of the F index over time leads to concluding that the SARIMA model had an error less than one tenth of the ANN-GA error in forecasting the following month discharge. On the other hand, the forecast error did not increase much with increasing forecast horizon, meaning the SARIMA model is able to make much more precise short-term forecasts than the ANN-GA model and is also able to make long-term forecasts without a noticeable error increase. Therefore, the SARIMA model definitely outperforms the ANN-GA model in short-term plans such as exploitation and consumption management and in long-term plans such as designing and constructing hydraulic structures.

References

Abebe A, Foerch G (2008) Stochastic simulation of the severity of hydrological drought. Water Environ J 22(1):2–10

Ahmad S, Khan IH, Parida B (2001) Performance of stochastic approaches for forecasting river water quality. Water Res 35(18):4261–4266

Ali SM (2013) Time series analysis of Baghdad rainfall using ARIMA method. Iraqi J Sci 54(4):1136–1142

Anctil F, Perrin C, Andréassian V (2004) Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. Environ Model Softw 19(4):357–368

Box GE, Cox DR (1964) An an alysis of transformations. J R Stat Soc B Methodol 26(2):211–252

Box GE, Pierce DA (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Assoc 65(332):1509–1526

Coulibaly P, Anctil F, Bobee B (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. J Hydrol 230(3–4):244–257

Cryer J, Chen K (2008) Time series analysis with applications in R. Springer Texts in Statistics, New York

Damle C, Yalcin A (2007) Flood prediction using time series data mining. J Hydrol 333(2–4):305–316

Fereydooni M, Rahnemaei M, Babazadeh H, Sedghi H, Elhami MR (2012) Comparison of artificial neural networks and stochastic models in river discharge forecasting, (case study: Ghara-Aghaj River, Fars Province, Iran). Afr J Agric Res 7(40):5446–5458

Golabi MR, Radmanesh F, Akhondali AM, Kashefipoor M (2013) Simulation of seasonal precipitation using ANN and ARIMA Models: a case study of (Iran) Khozestan. Elixir Comput Sci Eng 55:13039–13046

Hurst H, Black R, Simaika Y (1965) Long-term storage: an experimental study. Constable, London

Irvine K, Eberhardt A (1992) Multiplicative, seasonal ARIMA models for lake Erie and lake Ontario water levels. Water Resources Bulletin, New York, pp 385–396

Jain S, Das A, Srivastava D (1999) Application of ANN for reservoir inflow prediction and operation. J Water Resour Plan 125(5):263–271

Jarque CM, Bera AK (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econ Lett 6(3):255–259

Khatibi R, Ghorbani M, Naghipour L, Jothiprakash V, Fathima T, Fazelifard M (2014) Inter-comparison of time series models of lake levels predicted by several modeling strategies. J Hydrol 511:530–545

Kilinç I, Cigizoglu K (2005) Reservoir management using artificial neural networks. In: Proceedings of the 14th Reg. directorate of DSI (State Hydraulic Works), CiteSeerx, Istanbul, Oct 2014

Kisi Ö (2004) River flow modeling using artificial neural networks. J Hydrol Eng 9(1):60–63

Kisi Ö, Cigizoglu HK (2007) Comparison of different ANN techniques in river flow prediction. Civ Eng Environ Syst 24(3):211–231

Landeras G, Ortiz-Barredo A, López JJ (2009) Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. J Irrig Drain Eng 135(3):323–334

Levenberg K (1944) A method for the solution of certain non–linear problems in least squares. Q Appl Math 2(2):164–168

McLeod A (1978) On the distribution of residual autocorrelations in Box-Jenkins models. J R Stat Soc B Methodol 40(3):296–302

Mohammadi K, Eslami H, Dardashti SD (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(1–2):17–30

Mohan S, Vedula S (1995) Multiplicative seasonal ARIMA model for longterm forecasting of inflows. Water Resour Manag 9(2):115–126

Mombeni HA, Rezaei S, Nadarajah S, Emami M (2013) Estimation of water demand in Iran based on SARIMA models. Environ Model Assess 18(5):559–565

Nirmala M, Sundaram S (2010) A seasonal ARIMA model for forecasting monthly rainfall in Tamilnadu. Natl J Adv Build Sci Mech 1(2):43–47

Pektaş AO, Cigizoglu HK (2013) ANN hybrid model versus ARIMA and ARIMAX models of runoff coefficient. J Hydrol 500:21–36

Salas J, Delleur J, Yevjevich V, Lane W (1988) Applied modeling of hydrologic time series. Water Resour Publ, Colorado

Valipour M (2015) Long-term runoff study using SARIMA and ARIMA models in the United States. Meteorol Appl 22(3):592–598

Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J Hydrol 476:433–441

Wang J, Hu J, Ma K, Zhang Y (2015) A self-adaptive hybrid approach for wind speed forecasting. Renew Energy 78:374–385

Xu Z, Li J (2002) Short-term inflow forecasting using an artificial neural network model. Hydrol Process 16(12):2423–2439

Yurekli K, Kurunc A (2005) Performances of stochastic approaches in generating low streamflow data for drought analysis. J Spat Hydrol 5(1):20–32

Yurekli K, Kurunc A, Ozturk F (2005) Application of linear stochastic models to monthly flow data of Kelkit Stream. Ecol Model 183(1):67–75

Zhang Q, Wang B-D, He B, Peng Y, Ren M-L (2011) Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. Water Resour Manag 25(11):2683–2703