

Drought Forecasting using Markov Chain Model and Artificial Neural Networks

Mehdi Rezaeianzadeh¹ · Alfred Stein² · Jonathan Peter Cox³

Received: 3 July 2015 / Accepted: 1 March 2016 /

Published online: 11 March 2016

© Springer Science+Business Media Dordrecht 2016

Abstract Water resources management is a complex task. It requires accurate prediction of inflow to reservoirs for the optimal management of surface resources, especially in arid and semi-arid regions. It is in particular complicated by droughts. Markov chain models have provided valuable information on drought or moisture conditions. A complementary method, however, is required that can both evaluate the accuracy of the Markov chain models for predicted drought conditions, and forecast the values for ensuing months. To that end, this study draws on Artificial Neural Networks (ANNs) as a data-driven model. The employed ANNs were trained and tested by means of a statistically-based input selection procedure to accurately predict reservoir inflow and consequently drought conditions. Thirty three years' data of inflow volume on a monthly time resolution were selected to enable calculation of the standardized streamflow index (SSI) for the Markov chain model. Availability of hydro-climatic data from the Doroodzan reservoir in the Fars province, Iran, allowed us to develop a reservoir specific ANN model. Results demonstrated that both models accurately predicted drought conditions, by employing a randomization procedure that facilitated the selection of the required data for the ANN to forecast reservoir inflow close to the observed values over a validation period. The results confirmed that combining the two models improved short-term prediction reliability. This was in contrast to single model applications that resulted into substantial uncertainty. This research emphasized the importance of the correct selection of data or data mining, prior to entering a specific modeling routine.

✉ Mehdi Rezaeianzadeh
mzr0031@auburn.edu

¹ School of Forestry and Wildlife Sciences, Auburn University, 602 Duncan Drive, Auburn, AL 36849, USA

² Faculty of Geo-Information Science and Earth Observation (ITC), Twente University, Enschede, The Netherlands

³ Caribbean Institute for Meteorology and Hydrology, West Indies, Barbados

Keywords Reservoir inflow · Markov chain · Data-driven models · Drought forecasting · Reservoir operation · ANN

1 Introduction

Reservoirs are important sources of water especially in many developing countries, and as such, forecasting their monthly inflow and the related drought conditions are vital for achieving optimal reservoir performance and reliability. Moreover, predicting inflow can help to better address droughts, flood risk assessments and the allocation of potable water, simultaneously with agricultural and industrial uses. In recent times, demands for water use in Iran have increased significantly due to a rapid population growth in an arid and semi-arid climate (Sattari et al. 2012). Global warming and climate change have also amplified the pressure on reservoir storage and supply guarantee, consequently, there is a strong need to develop accurate models to forecast monthly reservoir inflow.

A common and functional method of forecasting inflow to reservoirs is provided by artificial neural networks (ANNs). ANNs have been used extensively in modeling the non-linear behavior of hydrological processes. Applications of ANNs in hydrology and water resources includes rainfall-runoff processes (Hsu et al. 1995; Shamseldin 1997; Tokar and Johnson 1999; Kumar et al. 2005; Rezaeianzadeh et al. 2010), streamflow forecasting (Kisi 2007; Isik et al. 2013), water level prediction (Emamgholizadeh et al. 2014; Rezaeianzadeh et al. 2015), water quality (Singh et al. 2009; Kalin et al. 2010) and drought forecasting (Mishra and Desai 2006; Morid et al. 2007; Keskin et al. 2011; Rezaeianzadeh and Tabari 2012). Likewise, the application of ANNs in forecasting inflow to reservoirs has been addressed both on a daily scale (Coulibali et al. 2000; Coulibali et al. 2001; Coulibali et al. 2005; Sattari et al. 2012; Krishna 2014) and a monthly scale (Jain et al. 1999; Valipour et al. 2013).

Markov chains are useful to stochastically model a time series composed of discrete variables. When applied to the Palmer index time series, they have demonstrated the ability to be utilized effectively for predicting wet and dry periods (Lohani and Loghanthan 1997). Paulo and Pereira (2007) used a Markov chain to understand the stochastic characteristics of droughts by means of analyzing the probabilities for each severity class of drought based on the standardized precipitation index (SPI) values in Alentejo, southern Portugal. A two-state, first-order Markov chain was efficient for describing wet and dry weather patterns based on daily rainfall data in Colombo Sri Lanka when evaluating wet and dry spells (Sonnadara and Jayewardene 2015).

In a Markov chain, monthly volumes of inflow to reservoirs are the key inputs, with the evaluation and forecasting of drought conditions for the subsequent month serving as the output. A major drawback of such modeling is that it provides the user with the predicted wet, dry, or normal drought conditions, together with the probability of the occurrence of that condition, however, no discharge is predicted as the forecasted inflow to the reservoir. Both ANNs and Markov chains are capable of forecasting drought conditions; the ANN provides a single value and with a comparison to threshold values, the drought condition can be evaluated, whereas a Markov chain is based on transition probabilities.

Tabari et al. 2015 proved that first-order Markov chain models are adequate to reproduce the statistical structure of the streamflow drought index (SDI)-based hydrological droughts. In the current study, a three-state, first-order Markov chain model was applied to standardized

stream flow index (SSI) data to predict drought conditions for the succeeding month. An ANN model was also developed to be used in conjunction with the Markov chain model, which to the best of our knowledge, is adequate in evaluating the suitability of such an ANN model compared with a Markov chain model for forecasting drought conditions based on inflow volumes. Hydrological drought and Markov chains have been studied in the past by Nalbantis and Tsakiris (2009). Also, Tsakiris and Vangelis (2004), Tsakiris and Vangelis (2005), Tsakiris et al. (2006), Araghinejad (2011), Tabari et al. 2013; Tsakiris et al. 2013, provide a solid background in a variety of drought indices.

The aim of the current study was to develop an ANN model based on hydro-climatic inputs to (i) forecast the monthly inflow volume for the subsequent month and to (ii) evaluate the Markov chain results for forecasting drought conditions 1 month ahead as transition probabilities, controlling the next state of the system based only on the current state. The latter purpose can be achieved by comparing values forecasted by ANN with the threshold values of wet, dry and normal states from the Markov chain. In addition to the main purposes of the study, different training algorithms in ANN models for forecasting reservoir inflow were evaluated where the outcome could be used as a preliminary guideline for future studies. The model was applied to the watershed of the Doroodzan reservoir in Iran. To the best of the authors' knowledge, there is no further study reported to develop an ANN model as a complementary method for evaluating drought conditions forecasted by a Markov chain. This simple but efficient methodology can be exploited to forecast drought conditions and inflow to reservoirs. In doing so, it is able to flag potential short-term future water shortage in reservoir systems, whereas it helps to improve operational water management in the presence of drought over a sub annual to multi annual time frame.

2 Study Area and Data set

The Doroodzan watershed within the Fars Province in Iran, (29°50'N, 51°53'E), (30°15' N, 52°22'E), was selected as a case study catchment (Fig. 1). Construction of the Mollasadra reservoir dam as a regulating dam upstream of the Doroodzan dam has caused a reduction in inflow volume to the cited reservoir during the past several years. For this motive, the combined watershed upstream of both the Doroodzan and Mollasadra reservoirs was chosen as the study area. The drainage area totals 4116 km² with terrain ranging in height from a maximum of 3677 m to a minimum of 1626 m, and an average land slope approximately 26 %. Daily precipitation data from the five weather stations Chamriz, Jamalbeig, Chobkheleh, Doroodzan, and Morozeh were used for this study. The mean annual rainfall over the watershed is estimated as 443 mm. Doroodzan dam is a homogenous earth filled dam with a 57 m crest height and approximately 700 m crest length. Pre-construction studies and investigations were carried out between 1963 and 1966 with the dam construction initiated in 1970. The dam was completed in 1974 to a capacity of 994×10^6 m³. It was designed to mitigate the effects of severe flood events in the Karbal region, and to provide a reliable potable water supply, as well as accommodate agricultural and industrial uses. The facility also supports a hydroelectric plant with an installed electricity generating nominal capability of 10 MW.

The Mollasadra dam is an earth filled dam with a clay core (2006) and was designed to meet local needs for agriculture and potable water supply demand coupled with hydroelectric energy production. The height of the crest is 72 m with a length of 630 m. The capacity of the reservoir was calculated at: 440×10^6 m³ when it entered into exploitation.

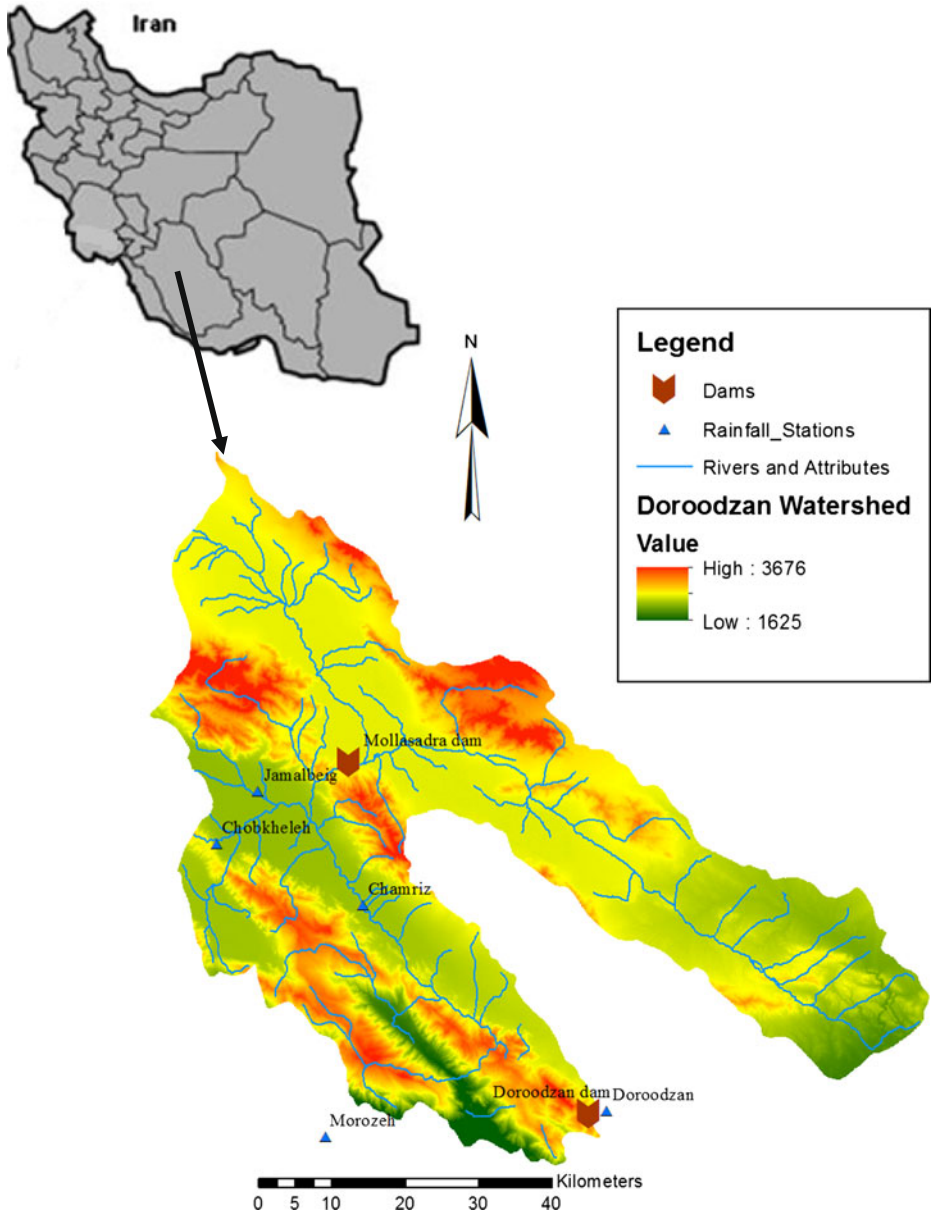


Fig. 1 Map of the study area

In total, over ninety six thousand daily values were processed from five rain gauges, two stream gaging stations; Doroodzan and Chamriz, and one type A evaporation pan. This data set spanned thirty three water years and comprised of 396 monthly values. Precipitation data from the 1976 to 2009 water years were used as input to the neural networks. Additionally, an analysis was initiated to emphasize the importance of encountering the optimum input combinations and distributing the data between training and testing datasets prior to modeling. This was particularly important in the case of long term time series precipitation data and corresponding stream flow data. Increasing the

length of the time series augments the non-stationarity of the data, and is important for the distribution of the extreme and normal values of data (here inflow to the reservoir), within the calibration (training) and validation (testing) phases. To this end, a randomization procedure based on evaluation of the equality of variances and means of the two samples, training and testing datasets, was followed so that the datasets displaying no statistically significant difference based on *t*-test were chosen to be the optimal datasets to be considered as final training and testing portions. This showed to be a very efficient method, especially when various drought conditions had been seen in a long term time series of data, as was the case for this study. As Chen et al. (2014) noted, knowledge discovery processes on the data includes data recording, data filtering and analysis, and constituted a significant segment of this study.

Approximately 70 % of the data, (277 monthly values), were used for training the ANN, with 30 % of the data (119 monthly values) employed for testing. Investigating the area-weighted precipitation and inflow to the Doroodzan reservoir suggested that periods with zero or negligible precipitation have correspondingly increased in the reservoir inflow data. Accordingly, the analysis of the available data showed that inflow to the Doroodzan reservoir arises partially and dominantly from runoff registered at the Chamriz station located upstream along the Kor River. Rezaeianzadeh et al. (2013b), reported that the area-weighted precipitation was superior when applied as an input to ANNs compared to spatially varied precipitation inputs; consequently, area-weighted precipitation values were considered in this study. Mean diurnal precipitation on the watershed area was estimated using Thiessen polygons, accordingly, the area-weighted precipitation over the catchment was determined by calculating each station's rainfall amount in proportion to its area of influence (Rezaeianzadeh et al. 2013a). Weights equal to 0.31, 0.22, 0.18, 0.2, and 0.09 were assigned to the Chamriz, Jamalbeig, Chobkheleh, Doroodzan, and Morozeh precipitation stations, respectively. For the sake of clarity, the variables are named here and the procedures for their selection will be discussed in detail in the results section. The target (dependent) variable for the ANN model, is the inflow volume of the subsequent month ($V(t+1)$), whereas predictor variables are the inflow volume from the current ($V(t)$) and antecedent (previous) time steps ($V(t-1)$), area-weighted precipitation of the current month ($P(t)$), antecedent precipitation with one month lag ($P(t-1)$), evaporation of both the current ($E(t)$) and the antecedent month ($E(t-1)$).

3 Methodology

3.1 Standardized Precipitation Index (SPI)

The standardized precipitation index (*SPI*) was developed by McKee et al. (1993), as a means to define and monitor drought events. Computation of the *SPI* involves fitting a probability density function (PDF) to total precipitations for the stations of interest. In this study, the gamma distribution is applied, and defined by its frequency or PDF as:

$$G(x) = \int_0^x g(x)dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{(\alpha-1)} e^{(-x/\beta)} dx \quad \text{for } x > 0 \quad (1)$$

Where x is the precipitation amount, α and β are shape and scale parameters and $\Gamma(\alpha)$ is the Gamma function. The α and β parameters have to be estimated, to each time scale of interest

(1, 2, 3...months) and for each month of the year. The maximum likelihood estimation was also employed. The resulting parameters were used to find the cumulative probability of an observed precipitation event for a specific month and timescale. This was then used in turn to obtain *SPI* values classified into different ranges of above and below normal values, in this way indicating the severity of the drought or non-drought event (Table 1). Several characteristics of droughts such as magnitude, duration or intensity can be derived based on the *SPI* values. In order to account for the probability q of zero rainfall to occur, the cumulative distribution function (CDF) for the Gamma distribution is modified as:

$$H(x) = q + (1 - q) G(x) \tag{2}$$

The calculated precipitation probabilities were transformed into the corresponding standard normal values, from which the *SPI* values were subsequently calculated. Additional descriptions can be found in Edwards and McKee (1997). A discussion of the advantages and disadvantages of using the *SPI* to characterize drought severity has been offered by Hayes et al. (1999). Table 1 provides a drought classification based on the *SPI* (McKee et al. 1993). Since monthly volumes of streamflow values were engaged in this study, the *SPI* was replaced with a new definition, standardized streamflow index (*SSI*). In the Doroodzan watershed, sources of water include direct overland flow, snowmelt, and spring discharges. Much of the so called spring discharges are actually delayed flows from rainfall or snowmelt, which may take several weeks or months to materialize in the hydrographic network. As a result, while a continuous streamflow regime is maintained throughout the year, the rainy season spans just seven months (October to April). As a result, the so called annual *SPI*, would only include data from the rainy season (Tabrizi et al. 2010) and consequently it would affect the streamflow values, consequently, a 12-month time scale was adopted.

3.2 Three-State, First-Order Markov Chain

A common class of stochastic models to represent a time series of discrete variables is known as the Markov chain or *MC*. A *MC* is based on a collection of system states, with the first-order *MC* as the most common form depending only on the current system state, and not on preceding states. On the contrary, a first-order Markov chain is a stochastic

Table 1 The *SPI* drought category classification (McKee et al. 1993)

SPI value	Category	Three-state classification
2.0 and above	Extremely Wet	Wet
1.5 to 1.99	Very Wet	
1.00 to 1.49	Moderately Wet	
-0.99 to 0.99	Near Normal	Normal
-1.0 to -1.49	Moderately Dry	Dry
-1.5 to -1.99	Severely Dry	
-2.0 and less	Extremely Dry	

process (random variable), such that X_{t+1} is conditionally independent on $X_0, X_1, X_2, \dots, X_{t-1}$, given X_t , for any time t . The probability that X_{t+1} takes a particular value j is then obtained as (Çinlar 1975):

$$\Pr \{X_{t+1} = j|X_0, X_1, \dots, X_t\} = \Pr \{X_{t+1} = j|X_t = i\} \quad \forall i, j \in S, \quad t \in T \quad (3)$$

A Markov chain is thus characterized by a set of states, S , and by the transition probability, p_{ij} , between states. The transition probability p_{ij} is the probability that the Markov chain is at the next time point in state j , given that it is at the present time point in state i (Paulo and Pereira 2007).

The transition probabilities of a Markov chain are conditional probabilities. A conditional probability distribution therefore pertains to each possible state that specifies the probabilities for the states of the system at the next time period. These conditional probability distributions allow for different transition probabilities that depend upon the current state. A three state-, first-order Markov chain is illustrated schematically in Fig. 2. The three states wet (W), dry (D) and normal (N) were considered in this study as at each time t , the random variable X adopts one state. First-order time dependence implies that there are $3^2=9$ transition probabilities, p_{ij} , with $p_{i1} + p_{i2} + p_{i3} = 1$ each $i = 1, 2, 3$. Estimation of the transition probabilities for multiple-state Markov chains are obtained from the conditional relative frequencies of the transition counts (n_{ij}):

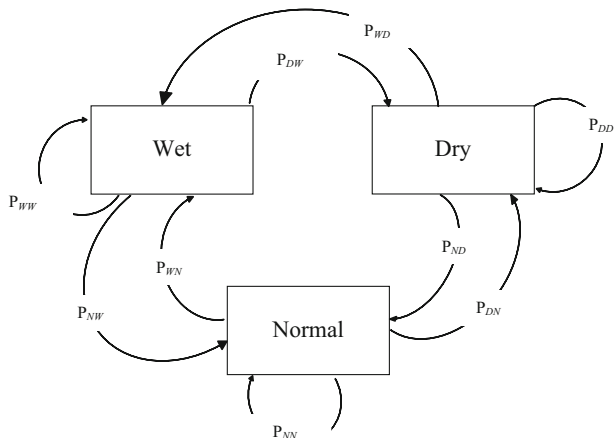
$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i+}}; \quad i, j = 1, 2, 3 \quad (4)$$

For the 3-state the Markov chain (4) can be written as:

$$\hat{p}_{DW} = \frac{n_{DW}}{(n_{WW} + n_{DW} + n_{NW})} \quad (5)$$

Where: n_{DW} indicates the number of changes from dry to wet. The three-state Markov chain has been used to characterize transitions between below-normal, near-normal and above-normal months as defined by Wilks (1995).

Fig. 2 Schematic illustration of a three-state, first-order Markov chain. Note that, P_{DW} is the probability of a dry day given that the previous day is wet



3.3 Artificial Neural Networks

An artificial neural network (ANN) is made up of a number of interconnected nodes (called neurons) arranged into three basic layers (input, hidden and output) (Dawson and Wilby 1998). Developing a multilayer feed forward back-propagation network is a common practice in a range of hydrology and water resources projects. Various training algorithms, including, resilient backpropagation (*rp*), scaled conjugate gradient (*scg*), variable learning rate (*gdx*), and Levenberg-Marquardt (*lm*) were considered to optimally train the MLPs. Readers are referred to Rezaeianzadeh et al. (2010), Rezaeianzadeh et al. (2013c) and Rezaeianzadeh et al. 2013a for information about MLP. After determining the optimal network architecture, the ANN model was trained and tested using the procedure later discussed. To validate the forecasted drought conditions by the Markov chain, threshold values for transition from one state to another (in three-state classification) were considered and according to those thresholds, the ANN forecasted value expressed as a drought condition.

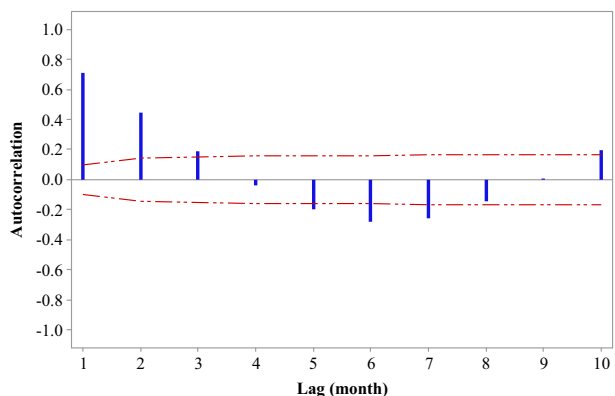
3.4 Data Preprocessing and Input Selections

To discover the optimum input combinations for the ANNs, an autocorrelation analysis (Kisi 2007; Rezaeianzadeh et al. 2010) of the streamflow volume data was carried out. Figure 3 shows the autocorrelation results for the streamflow volume to Doroodzan dam. The function indicates a significant correlation up to a lag of two months for the time series of the inflow volume, and then drops within the confidence limits. Therefore, the application of antecedent inflow volumes as input to the ANNs can be determined to be effective inputs. For this motive, ($V(t)$) and ($V(t-1)$) variables were selected as inputs to the ANN models.

Subsequently, a multiple regression analysis was used to model the relationship between the predictor variables: inflow volume, precipitation, evaporation and their antecedents (values with time lags), and the inflow volume one month ahead, ($V(t+1)$) as the dependent variable. Plainly, the antecedent inflow volume and precipitation with one month lag ($V(t)$ and $P(t)$) respectively, and precipitation with two months lag ($P(t-1)$) showed a significant correlation (p -value < 0.05) with inflow volume one month ahead. Although there was no significant correlation between evaporation and $V(t+1)$, its inclusion into the models improved the ANN.

A significant correlation between Chamriz streamflow values and inflow volume into the Doroodzan reservoir ($r=0.9$) confirmed the importance of Chamriz streamflow values for

Fig. 3 Autocorrelation functions for streamflow volume at 95 % confidence level



predicting inflow to the reservoir, even though a considerable amount of inflow to the reservoir comes from the watershed upstream of the Chamriz station. Therefore, discharge (monthly streamflow) data of the Chamriz station were selected as a major input to the developed models. Following this, an autocorrelation analysis of monthly stream flowdata from Chamriz station confirmed that the application of up to two months lag can play a considerable role in forecasting inflow volume to Doroodzan reservoir. All the discussed statistical analysis accompanied by simple trial and errors resulted in the best input combinations for the training and testing phases of the ANN model:

$$V(t + 1) = f\left(P(t), P(t-1), V(t), V(t-1), E(t), E(t-1), Q_{Ch}(t), Q_{Ch}(t-1)\right) \quad (6)$$

where P, V, E refer to the Doroodzan data and Q_{Ch} are streamflow data at Chamriz.

After confirming the input combinations, four singular training algorithms, including resilient back propagation (*rp*), scaled conjugate gradient (*scg*), variable learning rate (*gdx*), and Levenberg–Marquardt (*lm*) were worked to establish which was optimal. Numerous epochs were considered to confirm the ability of the optimal training algorithm to predict reservoir inflow. The results specifically relating to fifty epochs are presented in Table 2. Performance of the models was evaluated in terms of the root mean square error (*RMSE*) and coefficient of determination (R^2). All the models had one neuron in output layer.

4 Results and Discussion

Having endured drought events in recent years, there was an obvious reduction in the values of inflow volume to the Doroodzan reservoir; hence, in an attempt to distribute the wet and dry conditions to training and testing datasets for developing ANN models, the monthly stream

Table 2 Optimization of the number of neurons in the hidden layer accompanied by the related results for various training algorithms

Topology of Layers	GDX		RP		LM		SCG	
	R^2	<i>RMSE</i> (MCM)	R^2	<i>RMSE</i> (MCM)	R^2	<i>RMSE</i> (MCM)	R^2	<i>RMSE</i> (MCM)
3	0.24	74.78	0.53	48.12	0.45	59.45	0.64	34.60
4	0.22	75.76	0.59	48.68	0.49	49.14	0.61	42.06
5	0.53	49.53	0.58	46.05	0.06	140.08	0.61	48.47
6	0.32	60.65	0.60	47.84	0.30	96.49	0.65	45.71
7	0.52	83.01	0.52	50.32	0.31	86.65	0.60	47.05
8	0.42	55.75	0.58	49.3	0.28	92.83	0.54	49.92
9	0.25	71.03	0.50	51.7	0.29	99.22	0.59	45.5
10	0.33	60.03	0.58	45.53	0.22	72.47	0.55	52.26
11	0.30	79.42	0.57	48.76	0.31	75.82	0.52	51.41
12	0.20	96.23	0.46	59	0.30	81.91	0.57	49.51
13	0.21	84.2	0.48	55.34	0.15	95.04	0.60	44.37
14	0.41	117.68	0.50	51.71	0.47	100.187	0.61	44.44
15	0.21	127.96	0.49	51.21	0.21	135.49	0.58	46.23

flow data were randomly selected. This procedure was progressively undertaken to achieve optimal datasets for training and testing which would include all the aspects of extremes and normal values. The Levene test and t -test (Rezaeianzadeh et al. 2010, 2015) were used to obtain two sets of data for different flow regimes (mean, high and low flows) in training and testing the models. To apply the t -test, there is a need to specify the equality of variance from two groups of data (training and testing datasets). To that end, the p -value of the Levene test was equal to 0.679, thus, the hypothesis of unequal variances was rejected, and the t -test was executed with the assumption of equal variances for the training and testing datasets. Accordingly, the p -value for t -test was equal to 0.373 declaring that there is no statistically significant difference between these two datasets and those datasets that were finally selected as optimal.

Table 2 shows that the most useful results were related to the application of the *scg* training algorithm. Optimum architecture was achieved using 8 input vectors and 3 hidden neurons, with R^2 and $RMSE$ values attained of 0.64 and $34.6 \times 10^6 \text{ m}^3$, respectively. This conclusion is in sound agreement with the study by Rezaeianzadeh et al. (2013c). One of the major drawbacks of that study was that the monthly discharge volume and the precipitation (as major input) were at the same time step, and as such it could not be considered valid for a real world project requiring a forecast. Nonetheless, in this study a successful endeavor was made to build upon that study in that all data from current (t) or previous time steps ($t-1$) were considered to predict the inflow volume for the next time step ($t+1$).

Table 3 presents the threshold values of drought conditions (wet, normal and dry conditions) using SSI values, and includes the predicted drought conditions for the 2009–2010 water year. Inputs to ANN and (real) observed and ANN predicted inflow volumes to the Doroodzan reservoir are presented in Table 4. To clarify the use of threshold values, the predicted MC condition for September 2009 is ‘dry’. Table 4 shows that the predicted value for the inflow volume to the reservoir equals, $9.81 \times 10^6 \text{ m}^3$, which is less than $24.14 \times 10^6 \text{ m}^3$. This signifies that the forecasted value from the ANN is placed in the class ‘dry’ drought condition. The drought conditions of values predicted by the ANN using the threshold values detailed in Table 4 shows mainly dry conditions, with the exception of the month of May. It is interesting to note that the predicted drought conditions by the Markov chain model are all placed in the dry conditions, thus confirming a good agreement between the two models. To further elucidate, the predicted value in May, 2010 equals $27.55 \times 10^6 \text{ m}^3$, whereas, the threshold value equals $24.52 \times 10^6 \text{ m}^3$. Therefore, the ANN-predicted inflow volume for May is showing a normal condition for May, whereas the MC model predicted dry conditions. Predicted inflow volume by the ANN hence confirmed the ability of a three-state, first-order Markov chain to be utilized for drought forecasting.

A closer look at the predicted inflow volumes by the ANN can be obtained from Fig. 4. Nevertheless, there are some discrepancies especially in Dec 2009, Feb 2010 and May 2010. Figure 5 displays the box-plots of inflow to the Doroodzan reservoir for various months. There are considerable variances for Dec 2009 and Feb 2010 compared to the other months, accompanied by some significant outliers. Accordingly these variances and outliers disturb the exactness of the ANNs, and the predicted inflow volume is considerably dissimilar from the observed values. For May 2010, the number of outliers dominates the variance of the data, and consequently the result was not acceptable. The standard deviation (StD) for Dec 2009 and Feb 2010 were equal to $23.0 \times 10^6 \text{ m}^3$ and $133.3 \times 10^6 \text{ m}^3$, respectively, and are listed among the highest standard deviations in all the months of a year.

Table 3 Threshold values of wet, dry and normal conditions in monthly and annual scales for 2009–2010 water year using three state, first-order Markov chain. Note that the values related to wet, dry and normal conditions must be multiplied with 10^6 m^3

State	Condition	Condition											
		Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Wet	Greater than or equal to	47.83	57.62	102.57	151.23	171.31	207.49	218.18	129.49	57.57	45.25	40.96	38.36
Normal	Between	47.83	57.62	102.57	151.23	171.31	207.49	218.18	129.49	57.57	45.25	40.96	38.36
Dry	Less than or equal to	24.14	29.88	44.54	57.59	76.41	87.49	100.76	56.11	24.52	20.37	20.33	16.81
Probability (%)		24.14	29.88	44.54	57.59	76.41	87.49	100.76	56.11	24.52	20.37	20.33	16.81
Predicted drought condition		100.0	71.4	80	80	100	90	45	100	87.5	75	62.5	71.4
		Dry	Dry	Dry	Dry	Dry	Dry	Dry	Dry	Dry	Dry	Dry	Dry

Table 4 Input data from 2009 to 2010 water year (accompanied by antecedent 2 months since those values have been considered as input to ANN), observed and forecasted inflow volumes to Doroodzan reservoir. Note that Sep to Dec refer to 2009 and the rest refer to 2010

Date	Area-weighted Precipitation (mm)	Evaporation (mm)	Chamriz discharge (m ³ /sec)	Observed inflow volume (10 ⁶ m ³)	Predicted inflow volumes (10 ⁶ m ³)
Jul	0	298.8	6.7	3.94	-----
Aug	0	253.9	5.71	4.32	-----
Sep	0	178.4	3.75	7.27	9.81
Oct	40.92	100.2	3.194	11.44	14.84
Nov	186.49	52.6	8.346	40.47	28.04
Dec	20.65	65.7	4.988	23.46	55.81
Jan	97.18	72.8	13.321	50.11	44.38
Feb	48.08	111.6	27.066	81.54	48.99
Mar	52.51	120.8	8.365	33.00	48.69
Apr	27.77	162.7	5.736	18.87	32.86
May	0.225	274.6	4.695	3.93	27.55
Jun	0	325.3	7.536	3.87	18.57
Jul	0	272.2	5.701	3.76	11.70
Aug	2.17	235.0	9.119	4.67	7.69

This study focused on finding a way of capturing more information from their inputs i.e., targeted data mining, before inputting them into the models (here ANNs). As a matter of detail, ANNs were used to evaluate and confirm the MC estimations. Since we cannot disturb the time series structure of MC, the data were used by MC in their original order. By using a randomization procedure, the top input combinations among all the available inputs were chosen to be distributed into training and testing phases of ANNs which significantly facilitated help the training of the ANNs and the confirmation of the MC results. Although the randomization procedure was introduced and applied by Rezaeianzadeh et al. (2010), the importance of this input selection procedure has been realized recently by the researchers of this study. A graphical user interface (GUI) for the ANN to predict future inflow to the reservoir will be available upon request, as well as the raw and analyzed datasets.

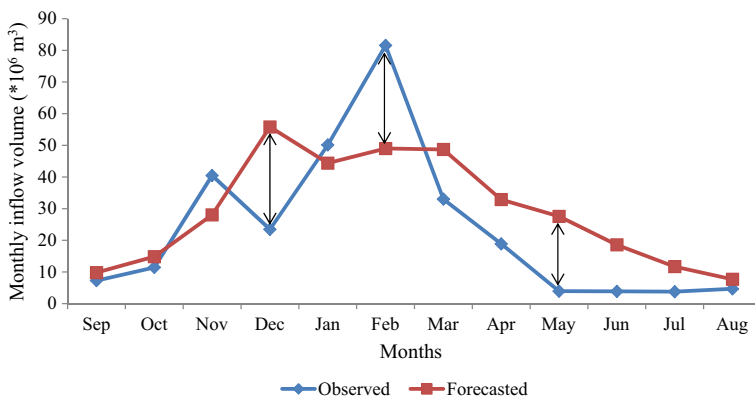
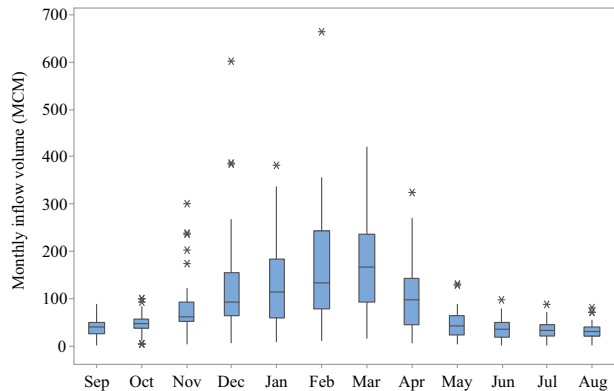


Fig. 4 Forecasted versus observed inflow volumes for 2009–2010 water year using ANN model

Fig. 5 Box plots for monthly inflow volumes to Doroodzan reservoir



5 Conclusions

In this study, stochastic and data-driven models were used to predict drought conditions 1 month ahead and the inflow volume using an ANN at the Doroodzan reservoir dam in the Fars province, Iran. The ANN model was developed, trained and tested using hydro-climatic variables. Although the construction of the Mollasadra reservoir as a regulating dam upstream the Doroodzan dam caused a reduction in the inflow volume to this dam over recent years, the proposed ANN model showed satisfactory results in forecasting both the moisture condition and reservoir inflow. The scaled conjugate gradient (*scg*) training algorithm produced superior results as compared to other applied training algorithms in predicting the inflow volume. Application of correlation/autocorrelation analysis accompanied by multiple regression analysis identified the most important input vectors to the ANNs. To evaluate all possible inputs to establish a more robust model in the application of ANNs, a randomization procedure was established. It derived the most informative sections for the training and testing phases. The study concluded that the ANN model provided an effective alternative to the MC model, and their simultaneous application reduced both the uncertainty and the error as compared to separate application.

Acknowledgments This project was funded by Fars Regional Water Authority with the contract number of FAW-88028. The first author acknowledges the support of Ms. Armina Soleymani for her help in manuscript preparation.

Compliance with Ethical Standards

Ethical Statement This manuscript entitled “Drought Forecasting using Markov Chain Model and Artificial Neural Networks” conforms to all ethical rules listed below:

- The manuscript has not been submitted to more than one journal for simultaneous consideration.
- The manuscript has not been published previously (partly or in full), unless the new work concerns an expansion of previous work (please provide transparency on the re-use of material to avoid the hint of text-recycling (“self-plagiarism”).
- A single study is not split up into several parts to increase the quantity of submissions and submitted to various journals or to one journal over time (e.g., “salami-publishing”).
- No data have been fabricated or manipulated (including images) to support your conclusions
- No data, text, or theories by others are presented as if they were the author’s own (“plagiarism”). Proper acknowledgements to other works must be given (this includes material that is closely copied (near verbatim), summarized and/or paraphrased), quotation marks are used for verbatim copying of material, and permissions are secured for material that is copyrighted.

- Consent to submit has been received explicitly from all co-authors, as well as from the responsible authorities - tacitly or explicitly - at the institute/organization where the work has been carried out, before the work is submitted.
- Authors whose names appear on the submission have contributed sufficiently to the scientific work and therefore share collective responsibility and accountability for the results.

References

- Araghinejad S (2011) An approach for probabilistic hydrological drought forecasting. *Water Resour Manag* 25: 191–200
- Chen CLP, Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. *Information Sciences* 275:314–347
- Çınlar E (1975) Introduction to stochastic processes. Prentice-Hall, New Jersey, p 402
- Coulibali P, Ancil F, Bobe'e B (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J Hydrol* 230(3–4):244–257
- Coulibali P, Ancil F, Bobe'e B (2001) Multivariate reservoir inflow forecasting using temporal neural network. *J Hydrol Eng* 6(5):367–376
- Coulibaly P, Hache' M, Fortin V, Bobe'e B (2005). Improving daily reservoir inflow forecasts with model combination. *J Hydrol Eng* 10(2):91–99
- Dawson CW, Wilby R (1998). "An artificial neural network approach to rainfall–runoff modeling." *Hydrol Sci J* 43(1):47–66
- Edwards CD, McKee TB (1997). Characteristics of 20th century drought in the United States at multiple time scales. Atmospheric Science Paper No. 634, Climatology Report, No. 97-2, Department of Atmospheric Sciences, Colorado State University
- Emamgholizadeh S, Moslemi K, Karami G (2014) Prediction the groundwater level of bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). *Water ResourManag* 28(15):5433–5446
- Hayes MJ, Svoboda MD, Wihite DA, Vanyarhko OV (1999). Monitoring the 1996 drought using the standardized precipitation index, *Bulletin of American Meteorological Society* 80(3):429–438
- Hsu KL, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of rainfall-runoff process. *Water Resour Res* 31(10):2517–2530
- Isik S, Kalin L, Schoonover J, Srivastava P, Lockaby B (2013) Modeling effects of changing land use/cover on daily streamflow: an artificial neural network and curve number based hybrid approach. *J Hydrol* 485:103–12
- Jain SK, Das A, Srivastava DK (1999) Application of ANN for reservoir inflow prediction and operation. *J Water Resour Plann Manag* 125(5):263–271
- Kalin L, Isik S, Schoonover JE, Lockaby BG (2010). "Predicting water quality in unmonitored watersheds using artificial neural networks." *J Environ Qual* 39(4):1429–1440
- Keskin ME, Terzi O, Taylan ED, Küçükyaman D (2011) Meteorological drought analysis using artificial neural networks. *Sci Res Essays* 6(21):4469–4477
- Kisi O (2007) Streamflow forecasting using different artificial neural network algorithms. *J Hydrol Eng* 12(5): 532–539
- Krishna B (2014) Comparison of wavelet-based ANN and regression models for reservoir inflow forecasting. *J Hydrol Eng* 19(7):1385–1400
- Kumar APS, Sudheer KP, Jain SK, Agarwal PK (2005) Rainfall-runoff modeling using artificial neural networks: comparison of network types. *Hydrol Process* 19:1277–1291
- Lohani VK, Loganathan GV (1997) An early warning system for drought management using the Palmer drought index. *J Am Water Resour Assoc* 33(6):1375–1386
- McKee TB, Doesken NJ, Kleist J (1993). The relationship of drought frequency and duration to time scales. In: Proc. 8th Conf. on Applied Climatology, January 17 – 22, 1993. American Meteorological Society, Massachusetts, pp. 179 – 184
- Mishra AK, Desai VR (2006) Drought forecasting using feed-forward recursive neural network. *Ecol Model* 198: 127–138
- Morid S, Smakhtin V, Bagherzadeh K (2007) Drought forecasting using artificial neural networks and time series of drought indices. *Int J Climatol* 27:2103–2111
- Nalbantis I, Tsakiris G (2009) Assessment of hydrological droughts revisited. *Water Resour Manag* 23:881–897
- Paulo AA, Pereira LS (2007) Prediction of SPI drought class transitions using Markov chains. *Water Resour Manag* 21(10):1813–1827

- Rezaeianzadeh M, Tabari H (2012) MLP-based drought forecasting in different climatic regions. *Theor Appl Climatol* 109(3–4):407–414
- Rezaeianzadeh M, Amin S, Khalili D, Singh VP (2010) Daily outflow prediction by multi-layer perceptron with logistic sigmoid and tangent sigmoid activation functions. *Water Resour Manag* 24(11):2673–2688
- Rezaeianzadeh M, Stein A, Tabari H, Abghari H, Jalalkamali N, Hosseiniour EZ, Singh VP (2013a) Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting. *Int J Environ Sci Technol* 10(6):1181–1192
- Rezaeianzadeh M, Tabari H, ArabiYazdi A, Isik S, Kalin L (2013b) Flood flow forecasting using ANN, ANFIS and regression models. *Neural Comput & Applic* 25(1):25–37
- Rezaeianzadeh M, Tabari H, Abghari H (2013c) Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions. *Arab J Geosci* 6(7):2529–2537
- Rezaeianzadeh M, Kalin L, Anderson C (2015) Wetland water-level prediction using ANN in conjunction with base-flow recession analysis. *J Hydrol Eng* doi:10.1061/(ASCE)HE.1943-5584.0001276, **D4015003**
- Sattari MT, Yurekli K, Pal M (2012) Performance evaluation of artificial neural network approaches in forecasting reservoir inflow. *Appl Math Model* 36(6):2649–2657
- Shamseldin AY (1997) Application of a neural network technique to rainfall-runoff modelling. *J Hydrol* 199: 272–294
- Singh KP, Basant A, Malik A, Jain G (2009). “Artificial neural network modeling of the river water quality: A case study.” *Ecol Modell* 220:888–895
- Sonnadara DUJ, Jayewardene DR (2015) A Markov chain probability model to describe wet and dry patterns of weather at Colombo. *Theor Appl Climatol* 119(1–2):333–340
- Tabari H, Nikbakht J, Hosseinzade P (2013) Hydrological drought assessment in Northwestern Iran based on streamflow drought index (SDI). *Water Resour Manag* 27:137–151
- Tabari H, Zamani R, Rahmati H, Willems P (2015) Markov Chains of different orders for streamflow drought analysis. *Water Resour Manag* 29:3441–3457
- Tabrizi AA, Khalili D, Kamgar-Haghighi AA, Zand-Parsa S (2010) Utilization of time-based meteorological droughts to investigate occurrence of streamflow droughts. *Water Resour Manag* 24:4287–4306
- Tokar AS, Johnson A (1999) Rainfall–runoff modeling using artificial neural networks. *J Hydrol Eng* 4(3):232–239
- Tsakiris G, Vangelis H (2004) Towards a drought watch system based on spatial SPI. *Water Resour Manag* 18:1–12
- Tsakiris G, Vangelis H (2005) Establishing a drought index incorporating evapotranspiration. *Eur Water* 9(10):3–11
- Tsakiris G, Pangalou D, Vangelis H (2006) Regional drought assessment based on the reconnaissance drought index (RDI). *Water Resour Manag* 21(5):821–833
- Tsakiris G, Nalbantis I, Vangelis H, Verbeiren B, Huysmans M, Tychon B, Jacquemin I, Canters F, Vanderhaegen EG, Poelmans L, De Becker P, Batelaan O (2013) A system-based paradigm of drought analysis for operational management. *Water Resour Manag* 27(15):5281–5297
- 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
- Wilks DS (1995) *Statistical methods in the atmospheric sciences*. Academic Press