



# Annual and Monthly Dam Inflow Prediction Using Bayesian Networks

Parisa Noorbeh<sup>1</sup> · Abbas Roozbahani<sup>1</sup>  · Hamid Kardan Moghaddam<sup>2</sup>

Received: 12 March 2020 / Accepted: 2 June 2020 /  
Published online: 17 June 2020  
© Springer Nature B.V. 2020

## Abstract

Dam inflow prediction is important in terms of optimal water allocation and reduction of potential risks of floods and droughts. It is necessary to select a suitable model to reduce uncertainties in long-term and short-term predictions. In this study a probabilistic model of Bayesian Networks (BNs) was used to evaluate its efficiency in predicting inflow into reservoirs considering the uncertainties. For this purpose, continuous BNs as well as integration of K-means clustering and discrete BNs were applied for predicting magnitude and range of inflows, respectively in terms of annual and monthly prediction scenarios. In this regard, the Zayandehrud Dam reservoir in Iran was selected to test this model. To achieve the best network structure in these scenarios, different patterns were defined based on the combination of predictors. According to the magnitude predictions, the MAPE and  $R^2$  indicators in annual model were respectively 21% and 0.62 and in monthly model were respectively 49% and 0.71. According to the results of the inflow range prediction, the prediction accuracy of the annual and monthly patterns was 75% and 83%, respectively. Modelling results showed that BN performs better in predicting the inflow range than its numerical prediction. The proposed model can improve the decision making of reservoirs operation.

**Keywords** Dam inflow prediction · Bayesian networks · Uncertainty · Clustering · Zayandehrud dam

## 1 Introduction

Operation of dam reservoirs is of great importance due to an increase in the population of the world and consequently the need for development and optimal utilization of water resources.

---

✉ Abbas Roozbahani  
roozbahany@ut.ac.ir

<sup>1</sup> Department of Irrigation and Drainage Engineering, Aburaihan Campus, University of Tehran, Tehran, Iran

<sup>2</sup> Water Research Institute, Ministry of Energy, Tehran, Iran

Hence the access to accurate information is important to predict inflow to the reservoir to plan and control the rule curve of dams (Bae et al., 2007). Short-term (e.g. Monthly) and long-term (e.g. Annual) prediction cannot easily be done, because the pattern of inflow is associated with many complications. Therefore, the development of a model that can take into account this complexity is essential to provide accurate predictions (Allawi et al., 2018). Accurate analysis of short-term predictions is necessary for preventing flood events and water supply. In addition, long-term predictions are necessary for water resources planning (Awan and Bae, 2014).

According to the literature, there are many different approaches to predict dam inflow. Awan and Bae (2013) developed a model based on Adaptive Network-based Fuzzy Inference System (ANFIS) to predict inflow to three dam reservoirs in South Korea. Using this model, they predicted monthly inflow to dams in the next month using predictor parameters of precipitation, temperature and dam inflow. Kumar et al., (2015) used Bootstrap wavelet-based ANN model (BWANN) to predict the daily inflow of the Panchut dam in India. The results of this method were compared with several models such as wavelet-based multiple linear regression (WMLR) and Bootstrap analysis.

Atiqzaman and Kandasamy (2016) studied the accuracy of the Genetic Programming in long-term prediction of inflow to a dam in Australia. In this study, precipitation and inflow to the dam in previous time steps were used in predictions. Li et al., (2016) introduced the Deep Restricted Boltzmann Machine-based Neural Networks (DRBM-NN) and Stack Auto Encoder-based Neural Networks (SAE-NN) models to predict daily inflow of two dams in China. The results of this study were compared with those obtained from ARIMA and a Feed Forward Neural Network. For the operation of Ubonratana Dam reservoir in Thailand, Chiamsathit et al., (2016) predicted inflow to the dam reservoir using a Multilayer Perceptron Artificial Neural Network. Simulations were performed under the role curve of the dam. Esmaeilzadeh et al., (2017) used different combinations of precipitation time series, evaporation and upstream discharge of the river at different time steps to predict inflow to Sattarkhan Dam in Iran. They compared the performance of Artificial Neural Networks (ANN), Support Vector Regression (SVR), Wavelet Neural Networks (WANNs) and M5 tree models.

Different models with different parameters are being used in predictions. However, methods that can reduce uncertainty are more reliable. BNs are one of the efficient probabilistic models in this regard. Bayesian models have been increasingly developed due to high processing speed, graphical representation, no limitation in the number of variables and parameters, a combination of different data sources and management of uncertainties.

According to the literature, this model is a powerful tool for solving complex problems and is able to effectively discuss relationships between them (Leu and Bui, 2016). This model has been used in various areas of water resource management such as water allocation (Ahmadi et al., 2010; Xue et al., 2016), irrigation water management (Rahman et al., 2016; Sherafatpour et al., 2019); supply and demand management (Phan et al., 2016; Asadilour et al., 2012), groundwater management (Mohajerani et al., 2017; Roozbahani et al., 2018), water quality management (Liu et al., 2018; Couture et al., 2018), integrated water resources management (Molina et al., 2010; Xue et al., 2017), urban water management (Anbari et al., 2017; Tabesh et al., 2018) and many other fields of study. In the meantime, this efficient approach has recently received much attention in predicting water resources. For example, it has been used in drought prediction (Madadgar and Moradkhani, 2014; Bae et al., 2017), water consumption prediction (Froelich, 2015; Magiera and Froelich, 2015), runoff prediction (Nagarajan et al., 2010; Humphrey et al., 2016), water environmental risk prediction (Sharifahmadian and Latifi,

2013), water pollution prediction (Hall and Le, 2017; Nodoushan, 2018), pipe failure prediction (Kabir et al., 2015), prediction of pipe leakage (Leu and Bui, 2016), flood prediction (Sikorska and Seibert, 2016; Goodarzi et al. 2019), etc.

According to the literature, this approach has been rarely used to predict dams' reservoirs related backgrounds, in particular, prediction of inflow to dams. For instance, BNs have been used in predicting the optimal utilization of dam reservoir (Mediero et al., 2007), reservoir water dynamics (Das et al., 2017) and the water level in reservoirs (Das et al., 2016), changes in the reservoirs fullness (Ropero et al., 2017) and the seasonal prediction of dam inflow (Kim et al., 2018).

Due to the random and uncertain nature of dams' inflow, the BN is used in the current research. The possibility to enter classified variables numerically is among prominent features of the BNs. Monthly and annual inflow predictions as well as the prediction of inflow range are introduced in this study for the first time. The model is tested for the Zayandehrud Dam as one of the most important dams in Central Iran. The real data were used directly in magnitude predictions and the clustered data were used for predicting the inflow range. The results of this study can be used to help decision-makers in allocating water to various uses or other goals such as long-term water sales contracts, hydraulic power and drought preparedness with the highest degree of reliability.

## 2 Methodology

### 2.1 Study Area

Zayandehrud multi-purpose dam on the Zayandehrud River in Central Iran is located 110 km west of Isfahan. Water is supplied through natural runoff of Zayandehrud River and tunnels that are used for the transfer of inter-basin water including the first, second and third tunnels of Koohrang and Cheshme-Langan. It is noteworthy that the third Koohrang tunnel has been launched, but is not yet in operation. According to the statistics, the average inflow to the Zayandehrud Dam during 1971–72 to 2014–15 was about 44 m<sup>3</sup>/s. The basin upstream of Zayandehrud Dam with an area of 4265 km<sup>2</sup> is located between the northern latitudes of 32° 18' to 33° 10' and eastern longitude of 50° 03' to 50° 40'. The dam was constructed for hydroelectric power generation, seasonal flood control, supplying the agricultural, industrial, drinking and environmental water demands of downstream lands and cities. Figure 1 shows the study area in Iran.

### 2.2 Bayesian Networks (BNs)

Bayesian networks (BNs), also referred to belief networks and Bayesian belief networks was designed by Pearl (1988). This is a graphical model representing probabilistic relationships among different factors in a case study (Pearl, 1988). Probabilistic relationships in this method are estimated according to the Bayes theorem (Roozbahani et al. 2018). If E and F are two events so that P (E) ≠ 0 and P (F) ≠ 0, then we have:

$$P(E | F) = \frac{P(F|E) P(E)}{P(F)} \quad (1)$$

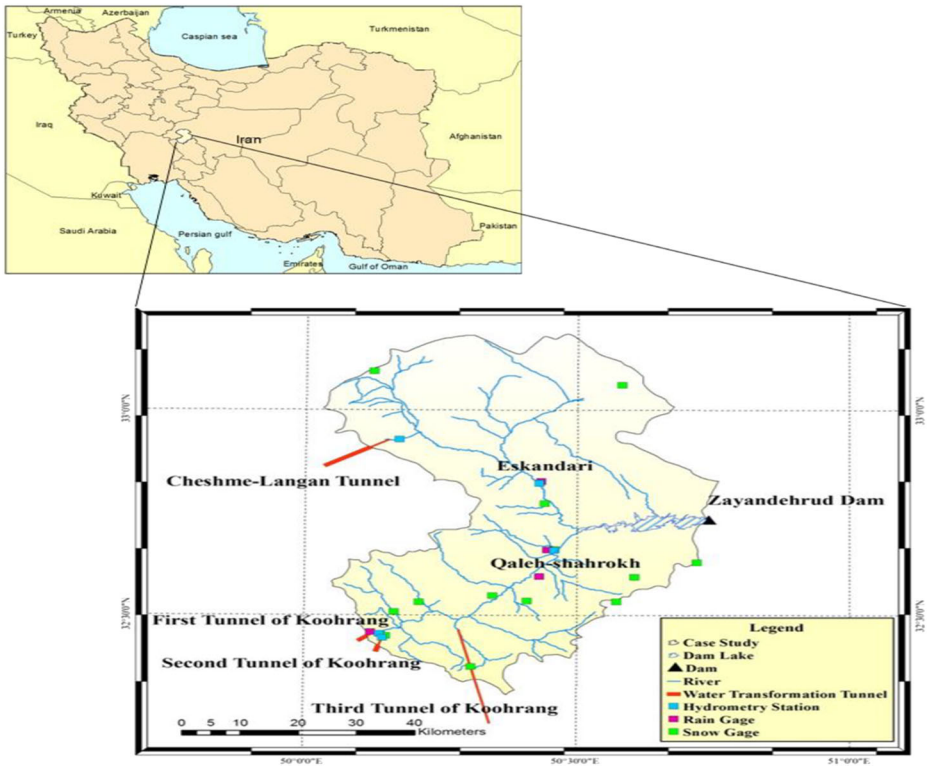


Fig. 1 study area

Similarly, for  $n$  exclusive events of  $E_1, E_2, \dots, E_n$ , we have:

$$P(E_i|F) = \frac{P(F|E_i)P(E_i)}{P(F|E_1)P(E_1) + P(F|E_2)P(E_2) + \dots + P(F|E_n)P(E_n)} \tag{2}$$

Where  $P(E_i)$  is the probability of event  $E_i$ ;  $P(F)$  is the probability of event  $F$ ;  $P(E_i|F)$  and  $P(F|E_i)$  are conditional probability of  $E_i$  given  $F$  and vice versa.

A BN consists of two main quantitative and qualitative components. The qualitative component is a directed linear graph in which each node represents a system variable and edges represent a causal relationship between the variables of the network (Abebe et al., 2018). The quantitative component is represented by a set of probabilistic relationships or probabilistic distributions for each network node. In the absence of any parent for the node (no arc toward the node), the node will have a marginal probability table. If the node has a parent (one or more arcs toward the node), it will have a conditional probability table (Hugin Expert A/S, 2017). Basically there are three types of BN nodes: discrete, continuous and hybrid (discrete and continuous nodes). For discrete nodes, the probabilistic table contains a probability distribution over the states and for continuous nodes, the probabilistic table contains a Gaussian density function (given through mean and variance parameters) for the variables it represents. Once constructed, the network can be used to enter observational data in nodes with known specific conditions to obtain probabilities in other nodes. If the BNs consist

only discrete nodes, then it is called discrete BNs, and if it contains continuous nodes, it is called continuous BNs.

The structure and probabilistic relationships of the BN are unknown in many cases. Learning is done using available observational data referred to as BN learning. This involves two steps of network structure learning and network parameters learning. Network structure learning is to determine dependent and independent variables and to find possible relationships between the variables that their causal relationships can be detected based on observational data. But the learning of parameters means the calculation of conditional probabilities of each node in the network. Among the advantages of the BNs model are risk analysis and uncertainty with greater accuracy than other models, management of missing values from input data, the ability to combine quantitative and qualitative data and providing approximate solutions using simulation techniques of estimation methods in cases where an exact solution is not available (Roozbahani et al. 2018). One of the main advantages of the BNs is development of the network in the case of incomplete data (Anbari et al., 2017). This can be helpful in this study due to the lack of time series of some parameters such as snow which cannot be extended.

### 2.3 Data Clustering

One of the important steps in discrete BN modelling in the predicting inflow range is to provide appropriate numerical intervals for model parameters which play a significant role in the final results of the model. To this end, clustering was used to divide monthly and annually predictors and predictands data to proper classes. Most of the previous researches have determined the classes manually and it cannot guarantee achieving the best results. Clustering is an unsupervised process during which objects are classified into different groups so that objects in a cluster are most similar to each other. The K-means method is one of the most practical clustering methods proposed by Macqueen (1967). This method uses an algorithm to classify objects so that the sum of squares of the distance between the data and the corresponding cluster center is minimized. The K-means clustering algorithm can be summarized as follows:

i: First, an arbitrary value is considered for the number of clusters (K). ii: K points are selected in the space of the objects which are in fact the set of primary centers. iii: Each object is assigned to a group with the shortest distance to its center. iv: When all objects are assigned to clusters, the location of the k centers is recalculated by calculating the average of each cluster's data. v: The steps (iii) and (iv) are repeated until the center of the cluster does not change. vi: At the last step, the objects are divided into separate groups with least error.

This is one of the most popular clustering techniques, but its reliability is influenced by the choice of initial centers for clustering, because the algorithm may stop in local optimums in some cases (Javadi et al., 2017). To obtain a suitable number of clusters (K), clustering validation methods such as Davies-Bouldin Index, Silhouette Width and the newer Gap method (Albalade and Suendermann, 2009; Rendón et al., 2011) have been used in this paper.

### 2.4 The Structure of the Proposed Bayesian Model

Choosing suitable and effective initial input variables improves the performance of the results in smart models. Therefore, it is important to identify the parameters affecting inflow to the dam reservoir. In Bayesian Network modelling, correlation analyses is not necessary and the relations between inputs and outputs are extracted by conditional probabilities. According to

available statistics and information, the discharge to the basin by the first and second tunnels of Koohrang ( $Q_1$ ,  $Q_2$ ) and Cheshme-Langan tunnel ( $Q_3$ ), natural discharge of Zayandehrud River (Zayandehrud River natural discharge is estimated by subtracting the discharge of water transfer tunnels from total dam inflow) ( $Q_z$ ), discharge of two important hydrometric stations (due to suitable positioning and suitable long-term statistical period) of Qaleh-Shahrokh ( $Q_4$ ) and Eskandari ( $Q_5$ ) which measure the inflow to the dam reservoir respectively from the south and north, average Rainfall in the basin (R), average snow height in the basin (S) and total dam inflow ( $Q_d$ ) with a reasonable time delay (monthly/yearly) are the effective predictors in the model. In this research all possible predictors have been used and there is no other variable that can be incorporated in the prediction model due to the lack of data in this region. But before Bayesian Network applying, correlation coefficients between predictors and dam inflow were estimated. As the coefficients were relatively low, therefore applying cause effect and probabilistic models such as BNs is reasonable.

Table 1 shows the basic information of the parameters affecting the inflow into the dam (i.e. years of data, mean annual values and % of missing data). These parameters were identified based on the role of these parameters in the calculations. Mostly, the correlation values between predictors and dam inflow was not considerable and this is one of the main reasons that BN has been chosen for prediction.

Accordingly, four scenarios were designed after introducing effective variables in predictions:

- A). Scenario 1: Prediction of annual inflow magnitude
- B). Scenario 2: Prediction of the annual inflow range
- C). Scenario 3: Prediction of monthly inflow magnitude
- D). Scenario 4: Prediction of monthly inflow range

Magnitude of inflow value is a real value (i.e. 2 m<sup>3</sup>/s) and inflow range is a class of inflow which varies between minimum and maximum values (i.e. 2–3 m<sup>3</sup>/s). According to the modelling scenarios, learning and validation of the proposed Bayesian model in Scenarios 1 and 3 were performed based on predictive numerical data to predict inflow to the dam. In Scenarios 2 and 4, predictor and predicted variables were divided into appropriate intervals with K-means clustering and validation indices mentioned in the methodology section to

**Table 1** Basic information of parameters

Basic information Parameter	mean annual value	Years of data		Missing data (%)
		from	to	
First tunnel of Koohrang discharge( $Q_1$ )	8.95 ( $\frac{m^3}{s}$ )	1971–72	2014–15	0
Second tunnel of Koohrang discharge( $Q_2$ )	10.40 ( $\frac{m^3}{s}$ )	1985–86	2014–15	31.82
Cheshme-Langan tunnel discharge ( $Q_3$ )	4.36 ( $\frac{m^3}{s}$ )	2005–06	2014–15	77.27
Qaleh-Shahrokh hydrometric station discharge( $Q_4$ )	39.37 ( $\frac{m^3}{s}$ )	1971–72	2014–15	0
Eskandari hydrometric station discharge( $Q_5$ )	4.29 ( $\frac{m^3}{s}$ )	1971–72	2014–15	0
Zayandehrud river natural discharge( $Q_z$ )	26.93 ( $\frac{m^3}{s}$ )	1971–72	2014–15	0
Rainfall (R)	550.78 (mm)	1971–72	2014–15	0
Snow (S)	123.51 (mm)	1997–98	2014–15	61.36
Dam inflow ( $Q_d$ )	43.94 ( $\frac{m^3}{s}$ )	1971–72	2014–15	0

predict the range of inflow changes. Also dam inflow is predicted for the next month and next year for monthly and annual prediction models, respectively. Figure 2 shows the modelling flowchart for 4 designed scenarios.

Considering different effects of predictor variables in the modelling structure, different patterns of BN were identified for entering data to identify the best learning structure. Accordingly, 44 patterns were designed for modelling as listed in Table 2.

Three groups were used to define these patterns. In the first group (25 first patterns), inflow to the Zayandehrud Dam was predicted using the average rainfall in the basin, average snow height, discharge of the first and second tunnels of Koohrang and Cheshme-Langan, natural discharge of Zayandehrud River and inflow to the dam with a time delay (monthly/annual).

To define the patterns in the second group (14 patterns), average rainfall in the basin, discharge of the first and second tunnels of Koohrang and Cheshme-Langan, Discharge of Qaleh- Shahrok station, south of the basin and Eskandari station, north of the basin as well as inflow to the dam with a time delay (monthly / yearly) were used. In these two approaches, data from the previous step (last year/month) were used to predict the inflow into the dam at the current time step. Finally, in the third group (5 patterns), predictions were performed only based on the time series of the inflow into the dam in the last one, two and three-time steps (month or year). For monthly and annual prediction scenarios, time step of prediction ( $t$ ) is month and year, respectively.

Hugin Lite is one of the most powerful commercial software for simulating and analyzing BNs. It provides a very good, intelligible and user-friendly interface with practical tools (Phan et al., 2016). Due to the advantages and ease of use, Hugin Lite V. 8.5 was used in this study for modelling the BN to predict annual and monthly inflow into the dam numerically and as intervals (Hugin Expert A/S, 2017).

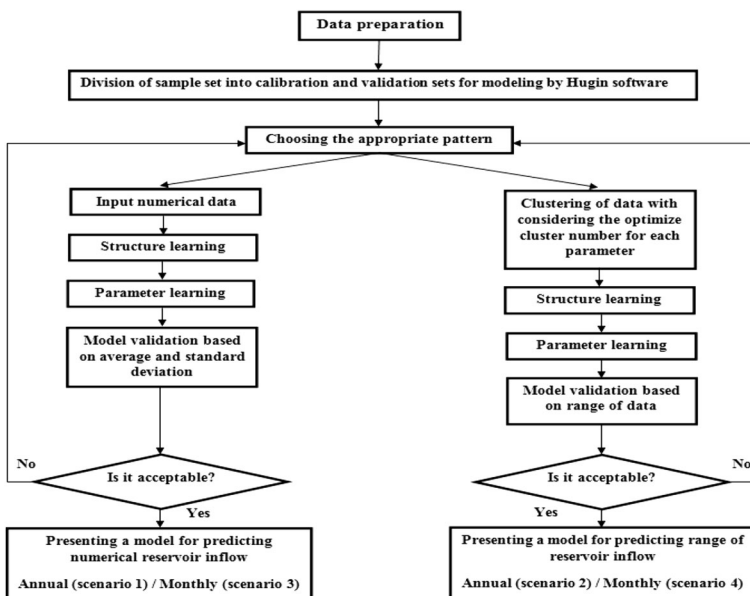


Fig. 2 Flowchart of the proposed model

**Table 2** Patterns defined in the Bayesian network to enter the variables

Pattern*	Q1 (t-1)	Q2 (t-1)	Q3 (t-1)	Q4 (t-1)	Q5 (t-1)	QZ (t-1)	R (t-1)	S (t-1)	QD (t-1)	QD (t-2)	QD (t-3)
1	✓	✓	✓				✓		✓		
2	✓	✓	✓				✓				
3	✓	✓	✓						✓		
4	✓	✓					✓		✓		
5	✓		✓				✓		✓		
6	✓	✓	✓			✓	✓		✓		
7	✓	✓	✓			✓	✓				
8	✓	✓	✓			✓	✓		✓		
9	✓	✓	✓			✓	✓	✓			
10	✓	✓				✓	✓		✓		
11	✓	✓				✓	✓				
12	✓	✓				✓	✓	✓			
13	✓	✓				✓	✓	✓			
14	✓		✓			✓	✓		✓		
15	✓		✓			✓	✓				
16	✓		✓			✓	✓		✓		
17	✓					✓	✓		✓		
18	✓					✓	✓				
19	✓					✓	✓		✓		
20	✓					✓	✓	✓			
21	✓					✓	✓	✓	✓		
22						✓	✓		✓		
23						✓	✓		✓		
24						✓	✓				
25	✓	✓	✓	✓	✓	✓	✓		✓		
26	✓	✓	✓	✓	✓		✓		✓		
27	✓	✓	✓	✓	✓				✓		
28	✓	✓		✓	✓		✓		✓		
29	✓	✓		✓	✓				✓		
30	✓		✓	✓	✓				✓		
31	✓		✓	✓	✓		✓		✓		
32				✓	✓		✓		✓		
33	✓	✓	✓	✓			✓		✓		
34	✓	✓		✓	✓		✓		✓		
35	✓		✓	✓			✓		✓		
36				✓			✓		✓		
37	✓	✓	✓		✓		✓		✓		
38	✓	✓			✓		✓		✓		
39	✓		✓		✓		✓		✓		
40									✓		
41										✓	
42											✓
43									✓	✓	
44									✓	✓	✓

\*Pattern 1 to 25: Group #1; Pattern 26 to 39; Group #2 Patterns 40 to 44;Group #3

### 2.5 Model Evaluation Indicators

The use of statistical indicators for evaluation of results depends on the type of prediction outcomes. After the learning of the network, coefficient of determination ( $r^2$ ), Nash-Sutcliffe coefficient (NS), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were used to investigate the accuracy of predictions of the inflow magnitude



(Ghordoyee Milan et al., 2018; Sherafatpour et al., 2019). The Reliability Percent index (RP) to measure the prediction accuracy of the flow range. In this case, the probability of being located in each interval was calculated and the one with the highest probability was selected and compared with the observational data. PR index is calculated by dividing the number of correct predicted years (or months) to the total number of years (or months).

### 3 Results and Discussion

#### 3.1 Learning Period and Model Validation

The time series of inflow to the Zayandehrud dam is available from 1971 to 2014 which includes 44 years of data for annual modelling and 536 months of data for monthly modelling. Therefore, this period was chosen as the modelling period and the time series of other parameters (except for the snow parameter that cannot be extended) were reconstructed as needed. 80% of data were used for calibration or learning of the BN and the remaining 20% were used for validation to verify the accuracy of the trained network. Figure 3 shows the long-term average of the time series of inflow to the Zayandehrud Dam.

As shown, 20% of the end of the time series is in the hydrological dry period. For validating the model, comprehensive inclusion of a dry period cannot express the accuracy of the model, especially in the future wet periods. Using a moving average of discharge, 20% ending in 2010–2011 (2002–2003 to 2010–2011) distributed in wet and dry periods was selected as period of validation. In other words, to train and test the annual Bayesian model, 36 years were allocated to calibration, 8 years for validation. Calibration period must contain the proportional wet and dry hydrological periods to increase the reliability of the forecast results in the future. Similarly, in the monthly Bayesian model, 432 months were allocated for calibration and 104 months were used to evaluate the accuracy of the trained Bayesian Network.

#### 3.2 Determining the Optimal Number of Clusters for Modelling and Data Clustering

To find suitable numerical intervals in this study, the number of proper clusters for all monthly and yearly predictor and predicted parameters were first calculated with cluster validation

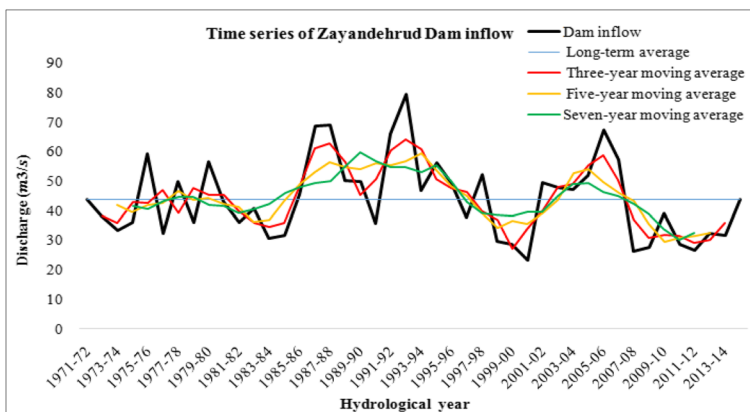


Fig. 3 The time series of inflow to the Zayandehrud Dam

indices. According to Davies-Bouldin, and Silhouette width indices in the annual approach, the optimal cluster number calculated by both indices is the same. However, due to the significant difference between these two validation indices for monthly data and uncertainty about which of the indicators will yield better results, Gap index was used to confirm the results and select the appropriate number of clusters. Finally, the number of clusters with at least two indices was considered as the optimal cluster number. The results on the validation of annual and monthly clusters are given in Tables 3 and 4. The numerical intervals (ranges) obtained for each of the parameters in both annual and monthly approaches are shown in Table 5.

### 3.3 Bayesian Network Learning

BN learning consists of 2 steps of structure learning and parameter learning. Structure of the network refers to the causal relationship between variables. Algorithm of Necessary Path Condition (NPC) is the most well-known for this purpose. Due to the known network structure in terms of causal relationships between the parameters in 44 learning patterns, the model parameters were learned only. Learning the parameters is the fact to find conditional probabilities of nodes using the Expectation-Maximization (EM) algorithm.

### 3.4 Model Validation and Results

Upon completing the BN learning, the patterns were validated under different scenarios.

A) Scenario 1: Table 6 shows the results on validation of the first scenario (prediction of annual inflow magnitude). In this scenario, all predictor and predicted parameters were used numerically to predict the magnitude values of annual inflow. The best results in terms of statistical indicators were obtained from the patterns 15 and 44. Of these two patterns, the pattern 15 was the best in terms of the mean absolute percentage error while the pattern 44 was the best in terms of other statistical indices. After analyzing the results, the pattern 15-b was defined as a specific pattern. In this pattern, the predictors are similar to those in the pattern 15 including the first Koohrang tunnel, Cheshme-Langan tunnel, Zayandehrud natural discharge and rainfall, but with applying lag time of two years for predictors to evaluate its effect on the accuracy of inflow prediction. Accordingly, the results of the evaluation indices of this model (Table 7) show a relatively good improvement in  $r^2$ , NS and RMSE compared to pattern 15. Thus, this pattern can be described as the top model in this scenario.

**Table 3** The optimal number of clusters of annual data

Parameter	Silhouette index	Davies-Bouldin index	Gap index	Optimum Clusters Number
Q <sub>1</sub>	2	3	2	2
Q <sub>2</sub>	3	3	3	3
Q <sub>3</sub>	3	3	3	3
Q <sub>4</sub>	3	3	3	3
Q <sub>5</sub>	3	3	3	3
Q <sub>Z</sub>	3	3	3	3
R	2	3	2	2
S	2	2	2	2
Q <sub>d</sub>	3	3	3	3

**Table 4** The optimal number of clusters of monthly data

Parameter	Silhouette index	Davies-Bouldin index	Gap index	Optimum Clusters Number
Q <sub>1</sub>	3	3	3	3
Q <sub>2</sub>	3	4	4	4
Q <sub>3</sub>	3	5	3	3
Q <sub>4</sub>	3	3	4	3
Q <sub>5</sub>	4	3	3	3
Q <sub>Z</sub>	4	5	3	4
R	3	4	3	3
S	5	3	3	3
Q <sub>d</sub>	4	4	3	4

B) Scenario 2: The BN was modeled in Scenario 2 with an annual approach to predict the range of inflow changes. In this scenario, the range of inflow variation was predicted by clustering predictor and predicted variables. The results in Table 8 indicate that the patterns 22 and 23 provide a higher reliability percent of 75% than other network structure patterns (the ratio of the number of correct prediction intervals to the total number of predicted intervals).

To analyze the difference between the performances of the two top patterns in this scenario, validation results in each year were evaluated in terms of reliability. Since the results obtained

**Table 5** The intervals for predictor and predicted variables

Model Parameters	Symbol	Cluster Number	Annual Range Bound	Monthly Range Bound
First Tunnel of Koohrang Discharge ( $\frac{m^3}{s}$ )	Q <sub>1</sub>	1	Q <sub>11</sub> < 8.72	Q <sub>11'</sub> < 7.7
		2	8.72 ≤ Q <sub>12</sub>	7.7 ≤ Q <sub>12'</sub> < 16.3
		3	–	16.3 ≤ Q <sub>1</sub>
Second Tunnel of Koohrang Discharge ( $\frac{m^3}{s}$ )	Q <sub>2</sub>	1	Q <sub>21</sub> < 4.7	Q <sub>21'</sub> < 7.2
		2	4.7 ≤ Q <sub>22</sub> < 12.09	7.2 ≤ Q <sub>22'</sub> < 21.5
		3	12.09 ≤ Q <sub>23</sub> < 17	21.5 ≤ Q <sub>23'</sub> < 42
		4	–	42 ≤ Q <sub>24'</sub>
Cheshme-Langan Tunnel Discharge ( $\frac{m^3}{s}$ )	Q <sub>3</sub>	1	Q <sub>31</sub> < 2.19	Q <sub>31'</sub> < 4.5
		2	2.19 ≤ Q <sub>32</sub> < 6	4.5 ≤ Q <sub>32'</sub> < 16.53
		3	6 ≤ Q <sub>33</sub>	16.53 ≤ Q <sub>33'</sub>
Qaleh-Shahrokh Hydrometry Station Discharge ( $\frac{m^3}{s}$ )	Q <sub>4</sub>	1	Q <sub>41</sub> < 20	Q <sub>41'</sub> < 31.5
		2	20 ≤ Q <sub>42</sub> < 27.4	31.5 ≤ Q <sub>42'</sub> < 90.5
		3	27.4 ≤ Q <sub>43</sub>	90.5 ≤ Q <sub>43'</sub>
Eskandari Hydrometry Station Discharge ( $\frac{m^3}{s}$ )	Q <sub>5</sub>	1	Q <sub>51</sub> < 3.79	Q <sub>51'</sub> < 5.36
		2	3.79 ≤ Q <sub>52</sub> < 6.66	5.36 ≤ Q <sub>52'</sub> < 16.95
		3	6.66 ≤ Q <sub>53</sub>	16.95 ≤ Q <sub>53'</sub>
Zayandehrud Natural Flow Discharge ( $\frac{m^3}{s}$ )	Q <sub>Z</sub>	1	Q <sub>Z1</sub> < 27.09	Q <sub>Z1'</sub> < 33.2
		2	27.09 ≤ Q <sub>Z2</sub> < 45	33.2 ≤ Q <sub>Z2'</sub> < 81
		3	45 ≤ Q <sub>Z3</sub>	81 ≤ Q <sub>Z3'</sub>
Rainfall (mm)	R	1	R <sub>1</sub> < 602	R <sub>1'</sub> < 39
		2	602 ≤ R <sub>2</sub>	39 ≤ R <sub>2'</sub> < 102.9
		3	–	102.9 ≤ R <sub>3'</sub>
Snowfall (mm)	S	1	S <sub>1</sub> < 175	S <sub>1'</sub> ≤ 45
		2	175 ≤ S <sub>2</sub>	45 ≤ S <sub>2'</sub> < 90
		3	–	90 ≤ S <sub>3'</sub>
Dam Inflow ( $\frac{m^3}{s}$ )	Q <sub>d</sub>	1	Q <sub>d1</sub> < 40.4	Q <sub>d1'</sub> < 47.7
		2	40.4 ≤ Q <sub>d2</sub> < 59.1	47.7 ≤ Q <sub>d2'</sub> < 102.3
		3	59.1 < Q <sub>d3</sub>	102.3 ≤ Q <sub>d3'</sub> < 195
		4	–	195 ≤ Q <sub>d4'</sub>

**Table 6** Results of different prediction patterns in the first scenario (magnitude prediction of annual inflow)

Pattern	MAPE(%)	NS	RMSE ( $\frac{m^3}{s}$ )	r <sup>2</sup>	Pattern	MAPE(%)	NS	RMSE ( $\frac{m^3}{s}$ )	r <sup>2</sup>
1	31	-0.14	15.17	0.07	23	33	0.25	15.04	0.13
2	29	-0.04	14.48	0.11	24	32	0.11	13.39	0.35
3	27	0.02	14.09	0.14	25	59	0.11	26.32	0
4	27	0.18	12.82	0.2	26	57	-2.33	25.92	0.01
5	32	-0.16	15.31	0.02	27	33	0.31	16.36	0.05
6	31	0.28	14.84	0.1	28	36	-0.38	16.7	0.01
7	31	-0.14	15.19	0.07	29	36	-0.38	16.7	0.01
8	27	0.45	13.82	0.17	30	61	-0.2	26.74	0
9	29	-0.08	14.74	0.07	31	59	-0.44	26.71	0.01
10	35	-0.41	16.84	0.01	32	42	0.2	18.71	0
11	26	0.2	12.66	0.21	33	46	-1.23	21.22	0
12	25	0.2	12.74	0.21	34	35	-0.39	16.74	0.02
13	29	-0.06	14.59	0.08	35	56	-2.25	25.6	0.02
14	31	-0.12	15.04	0.08	36	48	-0.32	21.65	0.02
<b>15</b>	<b>21</b>	<b>0.26</b>	<b>12.21</b>	<b>0.32</b>	37	33	-0.23	15.76	0.06
16	54	0.6	28.87	0.08	38	27	0.19	12.78	0.2
17	31	-0.09	14.82	0.09	39	37	-0.51	17.44	0.02
18	22	0.22	12.54	0.25	40	28	0.21	12.65	0.25
19	28	0.4	14.48	0.14	41	34	-0.03	14.54	0
20	21	0.13	13.25	0.2	42	35	-0.02	14.32	0.01
21	26	0.05	13.83	0.13	43	29	0.18	12.88	0.21
22	35	0.06	15.59	0.05	<b>44</b>	<b>28</b>	<b>0.36</b>	<b>11.88</b>	<b>0.36</b>

each year are based on the probability of being located in each cluster, four categories were considered for the results. Three categories were considered in terms of correct prediction probability based on which the probability percentage less than 50% was considered as a low-reliability prediction, 50–100 as a high-reliability prediction and 100% was considered as a decisive prediction. The fourth category was also introduced as the percentage of incorrect prediction probability. The results obtained from the analysis of the patterns 22 and 23 are presented in Fig. 4. The prediction results for the first and third clusters were similar in both patterns, but the pattern 22 in cluster 2 provides high reliable predictions. Analysis of the parameters of these two patterns indicates the significant role of rainfall in selecting the pattern 22. The predictors in this pattern include rainfall, natural discharge of Zayandehrud and runoff into the dam with a one year delay. Figure 5 shows the modelling results in the calibration and validation periods. As seen, 61.1% and 75% of calibration and validation data are correctly predicted, respectively.

C) Scenario 3: In this scenario, prediction of monthly inflow magnitude to the dam was carried out. The statistical indices for different cases are presented in Table 9. As seen, the highest accuracy in this scenario is observed in the pattern 15 with the predictor parameters of discharge of the first Koohrang tunnel and Cheshmeh-Langan, Zayandehrud natural discharge

**Table 7** Validation of Scenario 1 under a specific pattern

Pattern	MAPE(%)	NS	RMSE ( $\frac{m^3}{s}$ )	r <sup>2</sup>
<u>15-b</u>	<u>21</u>	<u>0.47</u>	<u>10.36</u>	<u>0.62</u>

**Table 8** Results of different prediction patterns in the second scenario (prediction of annual inflow intervals)

Pattern	RP(%)	Pattern	RP(%)	Pattern	RP(%)
1	50%	16	50%	31	25%
2	63%	17	63%	32	25%
3	38%	18	63%	33	38%
4	50%	19	63%	34	25%
5	50%	20	63%	35	38%
6	50%	21	63%	36	0%
7	38%	22	75%	37	25%
8	50%	23	75%	38	25%
9	38%	24	38%	39	25%
10	50%	25	38%	40	63%
11	38%	26	13%	41	38%
12	38%	27	38%	42	38%
13	38%	28	13%	43	25%
14	50%	29	25%	44	38%
15	50%	30	25%		

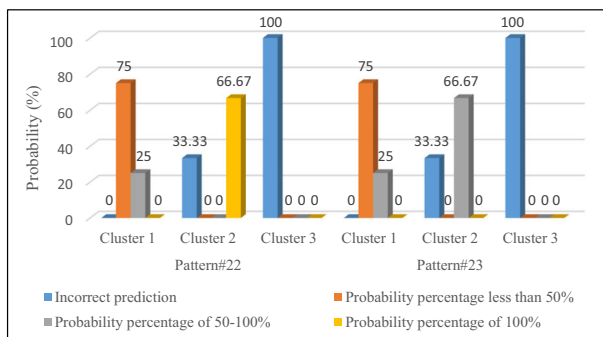
and rainfall. This pattern leads to a mean absolute percentage error rate of 49%, Nash-Sutcliffe of 0.7, RMSE of  $21.82 \frac{m^3}{s}$  with a coefficient of determination of 0.71.

As the best result in scenario 1 has obtained from the specifically defined pattern, similarly, in this scenario, a specific pattern was considered (pattern 15-b). In this pattern, the predictors are similar to those in the pattern 15 with this difference that the time step of predictor parameters has a two-month delay. Accordingly, the result of the statistical indices of this model (Table 10) doesn't show an improvement in accuracy of BN compared to pattern 15.

D) Scenario 4: This scenario determines the reliability percent of the monthly inflow range prediction by defined patterns. Table 11 lists the validation results of this scenario. As seen, the patterns 10, 11, 19, 23 and 24 show the highest reliability percent. The constant parameter in the network structure of these 5 patterns is natural runoff of Zayandehrud. This indicates the importance of this parameter in prediction with more accuracy.

Like scenario 2, the results of the top patterns in this scenario were analyzed with regard to the reliability index (Fig. 6). The data in the fourth clustering group are not available in the validation period and thus will not affect the selection of the top model.

Comparing the results of the top patterns in Fig. 6, one can see that the first cluster in all 5 cases is predicted with a probability of  $50 \leq P \leq 100$  or  $p = 100$ . So this cluster will have the least effect on choosing the best pattern, because the forecast is accurate with a good



**Fig. 4** The performance of the Bayesian model in two top patterns of the scenario 2

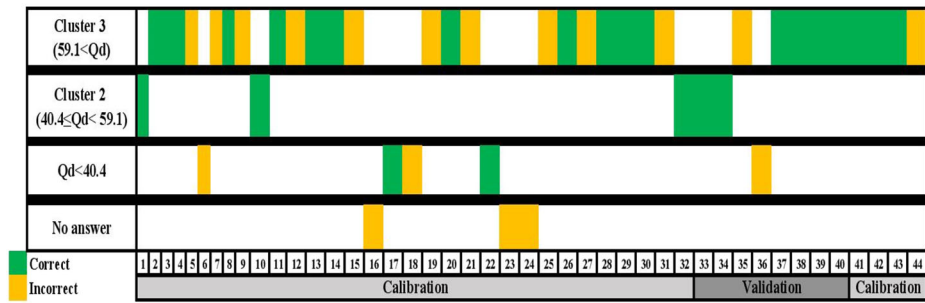


Fig. 5 Results of inflow range prediction for the best pattern in scenario 2

confidence rate in all relevant months. However, the least reliable predictions are seen for the third cluster. So this cluster has the most impact on the selection of best pattern. It seems necessary to select a pattern giving acceptable results from cluster 3. Given that the patterns 11 and 24 were not able to accurately predict the cluster 3 even within a month in the first step, they are removed from the list of top models. In other words, the confidence level of correct prediction in these two patterns is 0%. In the next step, the pattern 10 is removed from three remaining patterns because of the lowest confidence in the prediction of the third cluster.

In the third step, the remaining two patterns, namely 19 and 23, are compared. The confidence level of the correct prediction of the third cluster is equal in these patterns. So the decisive factor in this step is the confidence level obtained from the second cluster. As seen, the percentage of the probability of the correct prediction is higher in the pattern 23. Consequently, the pattern 23 with the predictor parameters of natural discharge of Zayanderroud with a one-month delay and inflow to the dam with a one-month delay can

Table 9 Results of different prediction patterns in the third scenario (magnitude prediction of monthly inflow)

Pattern	MAPE(%)	NS	RMSE( $\frac{m^3}{s}$ )	r <sup>2</sup>	Pattern	MAPE(%)	NS	RMSE( $\frac{m^3}{s}$ )	r <sup>2</sup>
1	53	0.69	22.36	0.7	23	63	0.63	24.24	0.69
2	65	0.48	28.91	0.49	24	57	0.7	21.78	0.71
3	76	0.64	24.15	0.68	25	57	0.7	21.79	0.71
4	53	0.69	22.22	0.7	26	69	0.61	24.89	0.67
5	55	0.68	22.79	0.68	27	78	0.66	23.35	0.68
6	53	0.69	22.36	0.7	28	53	0.68	22.56	0.69
7	53	0.69	22.34	0.7	29	57	0.7	21.82	0.71
8	76	0.64	24.13	0.68	30	73	0.63	24.39	0.66
9	52	0.69	22.2	0.71	31	49	0.69	22.39	0.69
10	53	0.69	22.35	0.7	32	51	0.7	21.83	0.71
11	53	0.66	23.27	0.68	33	53	0.7	21.79	0.71
12	53	0.69	22.42	0.69	34	56	0.68	22.82	0.68
13	52	0.69	22.18	0.7	35	50	0.67	22.89	0.67
14	53	0.69	22.36	0.7	36	53	0.69	22.37	0.69
15	49	0.7	21.82	0.71	37	54	0.69	22.28	0.7
16	76	0.64	24.13	0.68	38	53	0.68	22.54	0.69
17	54	0.69	22.4	0.69	39	69	0.57	26.17	0.58
18	49	0.65	23.79	0.68	40	126	0.13	37.28	0.14
19	76	0.64	23.97	0.68	41	147	0.01	39.94	0
20	48	0.66	23.34	0.7	42	65	0.66	23.46	0.67
21	52	0.7	21.95	0.71	43	63	0.67	23.06	0.69
22	49	0.68	22.49	0.69	44	50	0.69	22.37	0.69

**Table 10** Validation of Scenario 3 under a specific pattern

Pattern	MAPE(%)	NS	RMSE ( $\frac{m^3}{s}$ )	r <sup>2</sup> r <sup>2</sup>
15-b	83	0.45	29.52	0.46

be introduced as the top pattern for predicting the monthly inflow to the dam. Figure 7 shows the prediction results of the calibration and validation periods for this pattern.

Considering the results, 341 out of 432 months (79%) for calibration period and 80 out of 96 months (83%) for validation period, have been correctly predicted. By implementation of 4 scenarios under 44 patterns, one can conclude that the BN model has been able to predict the interval of inflow to the dam with a reasonable accuracy. Predictor parameters of Zayandehrud natural discharge and rainfall are the most important parameters in these four scenarios. Figure 8 shows the preferred pattern in each scenario in the software environment.

## 4 Conclusion

Due to the importance of predicting inflow in reservoirs operational planning and management, the performance of the Bayesian Networks in predicting range and magnitude of monthly and annual inflows, was investigated. Generally, handling of incomplete data sets, facilitating the combination of domain knowledge and available data and probabilistic learning about causal networks are the main benefits of BN modelling. The proposed algorithm for each scenario includes four stages of data preparation, BN learning, BN validation and model prediction. To verify the proposed model, Zayandehrud Dam, one of the most important multi-purpose dams in Central Iran, was selected as a study area. Modelling was performed in each scenario under 44 different patterns of the network structure to find the best combination of predictors. According to the results, inflow prediction ranges obtained by the model, is more realistic and trustworthy in terms of uncertainty consideration. Analysis of the results showed that the Bayesian model has been able to predict the annual inflow range. The reliability percent of inflow range predictions was 75% and 83% for annual and monthly scenarios,

**Table 11** Results of different prediction patterns in the fourth scenario (prediction of monthly inflow ranges)

Pattern	RP (%)	Pattern	RP (%)	Pattern	RP (%)
1	76%	16	77%	31	72%
2	76%	17	80%	32	74%
3	77%	18	80%	33	75%
4	78%	<b>19</b>	<b>81%</b>	34	78%
5	74%	<u>20</u>	<u>76%</u>	35	73%
6	74%	21	74%	36	80%
7	76%	22	79%	37	70%
8	78%	<b>23</b>	<b>83%</b>	38	72%
9	73%	<b>24</b>	<b>83%</b>	39	71%
<b>10</b>	<b>81%</b>	25	72%	40	77%
<u>11</u>	<u>81%</u>	26	72%	41	74%
12	79%	27	72%	42	67%
13	78%	28	73%	43	80%
14	75%	29	74%	44	80%
15	76%	30	71%		





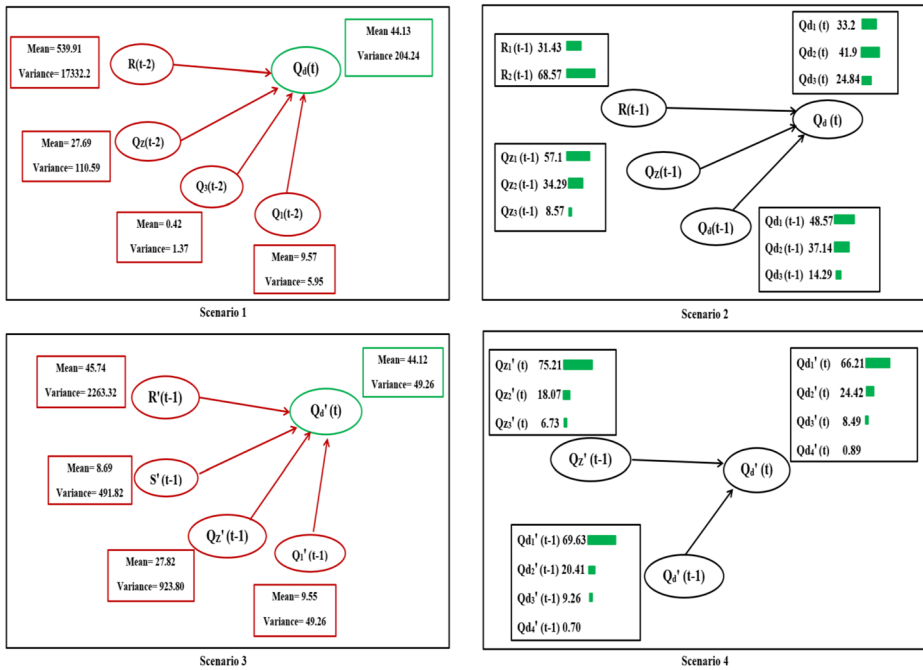


Fig. 8 The top BN patterns of scenarios examined in Hugin software

inflow predictions for a month or year ahead for operators and it can lead to better management and planning for different water users in downstream.

Since this research is one of the first attempts in applying BNs in dam inflow prediction, it is recommended to compare it with other popular machine learning models in this field. In addition, employing the proposed model in other dams with different predictors and clustering methods is suggested. It is also worth noting that in this research due to the availability of long term data (44 years), effect of climate and consumption conditions have been incorporated in BN analysis automatically in terms of calibration and validation phases, but it is suggested to apply the climate change and different human disturbance scenarios in the future researches.

**Acknowledgments** The authors would like to acknowledge the financial support of Regional Water Company of Isfahan in Iran for this research under grant number 95/185.

**Compliance with Ethical Standards**

**Conflict of Interest** None.

**References**

Abebe Y, Kabir G, Tesfamariam S (2018) Assessing urban areas vulnerability to pluvial flooding using GIS applications Bayesian belief network model. *J Clean Prod* 174:1629–1641  
 Ahmadi A, Karamouz M, Moridi A (2010) Robust methods for identifying optimal reservoir operation strategies using deterministic and stochastic formulations. *Water Resour Manag* 24(11):2527–2552

- Albalate A, Suendermann D (2009) A combination approach to cluster validation based on statistical quantiles. International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, Agust 3–5, Shanghai, China: 549–555
- Allawi MF, Jaafar O, Hamzah FM, Abdullah SMS, El-shafie A (2018) Review on applications of artificial intelligence methods for dam reservoir-hydro-environment models. *Environmental Science Pollution Research* 25(14):1–24
- Anbari MJ, Tabesh M, Roozbahani A (2017) Risk assessment model to prioritize sewer pipes inspection in wastewater collection networks. *J Environ Manag* 190:91–101
- Asadilour M, Kaveh F, Manshuri M, Khosrojerdi A (2012) Developing decision support tools for optimum domestic management by Bayesian belief networks in Tehran, Iran. *Adv Environ Biol* 6(1):286–296
- Atiquzzaman M, Kasamy J (2016) Prediction of inflows from dam catchment using genetic programming. *International Journal of Hydrology Science Technology* 6(2):103–117
- Awan JA, Bae DH (2013) Application of adaptive neuro-fuzzy inference system for dam inflow prediction using long-range weather forecast. Digital information management (ICDIM), eighth international conference on IEEE, september 10–12, Islamabad, Pakistan: 247–251
- Awan JA, Bae DH (2014) Improving ANFIS based model for long-term dam inflow prediction by incorporating monthly rainfall forecasts. *Water Resour Manag* 28(5):1185–1199
- Bae DH, Jeong DM, Kim G (2007) Monthly dam inflow forecasts using weather forecasting information and neuro-fuzzy technique. *Hydrol Sci J* 52(1):99–113
- Bae DH, Son KH, So JM (2017) Utilization of the Bayesian method to improve hydrological drought prediction accuracy. *Water Resour Manag* 31(11):3527–3541
- Chiamsathit C, Adeloje AJ, Bankaru-Swamy S (2016) Inflow forecasting using artificial neural networks for reservoir operation. Proceedings of the International Association of Hydrological Sciences 373:209–214
- Couture RM, Moe SJ, Lin Y, Kaste Ø, Haande S, Solheim AL (2018) Simulating water quality and ecological status of Lake Vansjø, Norway, under land-use and climate change by linking process-oriented models with a Bayesian network. *Sci Total Environ* 621:713–724
- Das M, Ghosh SK, Chowdary VM, Saikrishnaveni A, Sharma RK (2016) A probabilistic nonlinear model for forecasting daily water level in reservoir. *Water Resour Manag* 30(9):3107–3122
- Das M, Ghosh SK, Gupta P, Chowdary VM, Nagaraja R, Dadhwal VK (2017) Forward: a model for forecasting reservoir water dynamics using spatial bayesian network (spabn). *IEEE Trans Knowl Data Eng* 29(4):842–855
- Esmailzadeh B, Sattari MT, Samadianfard S (2017) Performance evaluation of ANNs and M5 model tree in Sattarkhan reservoir inflow prediction. *ISH Journal of Hydraulic Engineering* 23(3):283–292
- Froelich W (2015) Forecasting daily urban water demand using dynamic Gaussian Bayesian network. International conference of beyond databases, architectures and structures, may 30 – June 2 Ustron, Poland: 333–342
- Goodarzi R, Banihabib ME, Roozbahani A (2019) A decision-making model for flood warning system based on ensemble forecasts. *J Hydrol* 573:207–219
- Gholamzadeh M, Morid S, Delavar M (2011) Application of drought early warning system for operation of Zayandehrud dam. *Journal of Water and Soil Science* 15(56):35–48
- Ghordoyee Milan S, Roozbahani A, Banihabib ME (2018) Fuzzy optimization model and fuzzy inference system for conjunctive use of surface and groundwater resources. *J Hydrol* 566:421–434
- Hall DC, Le QB (2017) Use of Bayesian networks in predicting contamination of drinking water with *E. coli* in rural Vietnam. *Trans R Soc Trop Med Hyg* 111(6):270–277
- Hugin Expert A/S (2017). Hugin Researcher User Guide, Version 8.5(Aalborg,Denmark)
- Humphrey GB, Gibbs MS, Dy GC, Maier HR (2016) A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. *J Hydrol* 540:623–640
- Javadi S, Hashemy SM, Mohammadi K, Howard KWF, Neshat A (2017) Classification of aquifer vulnerability using K-means cluster analysis. *J Hydrol* 549:27–37
- Kabir G, Tesfamariam S, Sadiq R (2015) Predicting water main failures using Bayesian model averaging and survival modelling approach. *Reliab Eng Syst Saf* 142:498–514
- Kim K, Lee S, Jin Y (2018) Forecasting quarterly inflow to reservoirs combining a copula-based Bayesian network method with drought forecasting. *Water* 10(2):233
- Kumar S, Tiwari MK, Chatterjee C, Mishra A (2015) Reservoir inflow forecasting using ensemble models based on neural networks, wavelet analysis bootstrap method. *Water Resour Manag* 29(13):4863–4883
- Leu SS, Bui QN (2016) Leak prediction model for water distribution networks created using a Bayesian network learning approach. *Water Resour Manag* 30(8):2719–2733
- Li C, Bai Y, Zeng B (2016) Deep feature learning architectures for daily reservoir inflow forecasting. *Water Resour Manag* 30(14):5145–5161
- Liu Y, Engel BA, Flanagan DC, Gitau MW, McMillan SK, Chaubey I, Singh S (2018) Modelling framework for representing long-term effectiveness of best management practices in addressing hydrology and water

- quality problems: framework development and demonstration using a Bayesian method. *J Hydrol* 560:530–545
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA 1(14):281–297
- Madadgar S, Moradkhani H (2014) Spatio-temporal drought forecasting within Bayesian networks. *J Hydrol* 512:134–146
- Magiera E, Froelich W (2015) Application of Bayesian networks to the forecasting of daily water demand. *Intelligent Decision Technologies*, Springer: 385–393
- Mediero L, Garrote L, Martin-Carrasco F (2007) A probabilistic model to support reservoir operation decisions during flash floods. *Hydrol Sci J* 52(3):523–537
- Mohajerani H, Kholghi M, Mosaedi A, Farmani R, Sadoddin A, Casper M (2017) Application of bayesian decision networks for groundwater resources management under the conditions of high uncertainty and data scarcity. *Water Resour Manag* 31(6):1859–1879
- Molina JL, Bromley J, Garcia-Aróstegui JL, Sullivan C, Benavente J (2010) Integrated water resources management of overexploited hydrogeological systems using object-oriented Bayesian networks. *Environ Model Softw* 25(4):383–397
- Nagarajan K, Krekeler C, Slatton KC, Graham WD (2010) A scalable approach to fusing spatiotemporal data to estimate streamflow via a Bayesian network. *IEEE Trans Geosci Remote Sens* 48(10):3720–3732
- Nasri M (2010) Application of artificial neural networks (ANNs) in prediction models in risk management. *World Appl Sci J* 10(12):1493–1500
- Nodoushan EJ (2018) Monthly forecasting of water quality parameters within Bayesian networks: a case study of Honolulu, Pacific Ocean. *Civil Engineering Journal* 4(1):188–199
- Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible reasoning. Morgan Kaufmann Publishers, Los Altos
- Phan TD, Sahin O, Smart JC (2016) System dynamics and Bayesian network models for vulnerability and adaptation assessment of a coastal water supply and demand system. Proceedings of the 8th international congress on environmental Modelling and software: supporting sustainable futures, July 10–14, Toulouse, France :1180–1187
- Rahman MM, Hagare D, Maheshwari B (2016) Bayesian belief network analysis of soil salinity in a peri-urban agricultural field irrigated with recycled water. *Agric Water Manag* 176:280–296
- Rendón E, Abundez I, Arizmendi A, Quiroz EM (2011) Internal versus external cluster validation indexes. *International Journal of computers communications* 5(1):27–34
- Roobahani A, Ebrahimi E, Banihabib ME (2018) A framework for ground water management based on Bayesian network and MCDM techniques. *Water Resour Manag* 32(15):4985–5005
- Ropero RF, Flores MJ, Rumi R, Aguilera PA (2017) Applications of hybrid dynamic Bayesian networks to water reservoir management. *Environmetrics* 28:1–11
- Sharifahmadian E, Latifi S (2013) Water environment risk prediction using Bayesian network. Proceedings of the IEEE SoutheastCon, April 4–7, Jacksonville, Florida, USA: 1–5
- Sherafatpour Z, Roobahani A, Hasani Y (2019) Agricultural water allocation by integration of hydro-economic modeling with Bayesian networks and random Forest approaches. *Water Resour Manag* 33(7):2277–2299
- Sikorska AE, Seibert J (2016) Value of different precipitation data for flood prediction in an alpine catchment: a Bayesian approach. *Journal of hydrology* 556:961–971
- Tabesh M, Roobahani A, Roghani B, Faghihi NR, Heydarzadeh R (2018) Risk assessment of factors influencing non-revenue water using Bayesian networks and fuzzy logic. *Water Resour Manag* 32(11):1–24
- Xue J, Gui D, Lei J, Zeng F, Mao D, Zhang Z (2017) Model development of a participatory Bayesian network for coupling ecosystem services into integrated water resources management. *J Hydrol* 554:50–65
- Xue J, Gui D, Zhao Y, Lei J, Zeng F, Feng X, Shareef M (2016) A decision-making framework to model environmental flow requirements in oasis areas using Bayesian networks. *J Hydrol* 540:1209–1222