Natural Hazards
and Earth System
Sciences

Bayesian network model for flood forecasting based on atmospheric ensemble forecasts

Leila Goodarzi[1], **Mohammad E. Banihabib**[1], **Abbas Roozbahani**[1], **and Jörg Dietrich**[2]

[1]Department of Irrigation and Drainage, College of Aburaihan, University of Tehran, Tehran, Iran
[2]Institute of Hydrology and Water Resources Management, Leibniz University Hannover, Hanover, Germany

Correspondence: Jörg Dietrich (dietrich@iww.uni-hannover.de)

Abstract. The purpose of this study is to propose the Bayesian network (BN) model to estimate flood peaks from atmospheric ensemble forecasts (AEFs). The Weather Research and Forecasting (WRF) model was used to simulate historic storms using five cumulus parameterization schemes. The BN model was trained to compute flood peak forecasts from AEFs and hydrological pre-conditions. The mean absolute relative error was calculated as 0.076 for validation data. An artificial neural network (ANN) was applied for the same problem but showed inferior performance with a mean absolute relative error of 0.39. It seems that BN is less sensitive to small data sets, thus it is more suited for flood peak forecasting than ANN.

1   Introduction

Floods are the most threatening natural disaster across the world (Hénonin et al., 2010). Studies show that over 80 % of the cities of Iran are at the risk of flooding (Chitsaz and Banihabib, 2015). Flood warning is an efficient way to reduce the flood damage. However, many flood forecasting systems in the world rely on observed rainfall, and thus the lead time of these systems is often short for small basins (Banihabib and Arabi, 2016). Numerical weather prediction (NWP) models can be used to increase the lead time of flood warning by using in advance forecasts of rainfall. Although the combination of NWP and hydrological models can significantly increase the flood warning lead time rather than using observed rainfall, the deterministic weather prediction does not reflect the existing uncertainties. Thus, in the last decades, many operative and research on flood forecasting systems

around the world are increasingly employing ensembles of NWPs instead of single deterministic forecasts, which have considerable uncertainties (Goodarzi et al., 2019). Ensemble methods are considered to be an effective way to estimate the probability of future states of the atmosphere by addressing uncertainties present in initial conditions and in model approximations (Tennant et al., 2007). Various approaches have been developed to produce atmospheric ensemble forecasts including perturbing the initial conditions, perturbing the input parameters of the model, using multi-model ensembles and using different parameterization schemes (Yang et al., 2012).

One of the most important parameterization schemes is the cumulus parameterization. NWP models often use cumulus parameterization schemes (CPSs) to consider the effects of cumulus clouds which are not represented in modelling as they are much smaller than the model grid size (Pennelly et al., 2014). Common CPSs are presented in Table 1.

Kerkhoven et al. (2006) compared various CPSs for a summer monsoon in east China and found that the Kain–Fritsch scheme is the best scheme at simulating moderate rainfall depths. Pennelly et al. (2014) applied the Weather Research and Forecasting (WRF) model with diverse cumulus parameterization schemes for three flood events in Alberta, Canada, and they showed that the Kain–Fritsch and explicit cumulus parameterization schemes were the most accurate for simulating the rainfall. Other studies indicated that ensemble forecasting is promising for predicting heavy rainfall (Deb et al., 2008; El Afandi, 2013; Li et al., 2014).

Ensemble meteorological forecasting is widely coupled with a hydrological model to predict stream flow ensembles. Li et al. (2017) coupled the WRF model with a distributed

Table 1. Common cumulus parameterization schemes.

Model	Reference	Software used
Kain–Fritsch (KF)	Kain and Fritsch (1990)	WRF version 3.8
Betts–Miller–Janjić (BMJ)	Janjić (1994)	WRF version 3.8
Grell 3D ensemble (GR3D)	Grell (1993)	WRF version 3.8
Multi-scale Kain–Fritsch (MSKF)	Zheng et al. (2016)	WRF version 3.8
Grell–Dévényi ensemble (GDE)	Grell and Dévényi (2002)	WRF version 3.8

hydrological model for flood forecasting in a large watershed in southern China. The results suggest that the simulated floods are rational and could benefit the flood management communities due to their longer lead time. Rogelis and Werner (2018) assessed the potential of NWP models for flood early warning in tropical mountainous watersheds. The results showed that the streamflow forecasts resulted from a hydrological model forced by post-processed rainfall using the WRF, and added value to the flood early warning systems.

Only few case studies report how flood hydrographs derived from atmospheric ensemble forecasts (AEFs) can be converted into warning decisions during a flood event. Li et al. (2017) and Abebe and Price (2005) used the exceedance of critical thresholds. Dietrich et al. (2009a) used the quantile of the predicted flow ensemble. Yang et al. (2016) integrated ensemble rainfall forecasts, rainfall thresholds and a real time data assimilation method. Leandro et al. (2019) reduced the ensemble to the upper and lower range of the uncertainty band. Other concepts of deriving a single (deterministic type) warning indicator from ensembles are weighting of ensemble members, e.g., averaging by Bayesian model average (Raftery et al., 2005), by machine learning (Doycheva et al., 2017) or by reduction of members to create a multi-model sub-ensemble (Dietrich et al., 2009b).

According to previous studies, converting the ensemble forecasts into warnings and also deriving a single warning indicator from ensembles are not yet adequately considered and remain a challenging question in ensemble-based flood warning. The main objective of this study is to propose the Bayesian network (BN) model to estimate the flood peak from a meteorological ensemble forecast without employing a hydrological model. BN has been widely used by researchers in many water resources fields. Applications of BN in water resources can be found in studies of Mediero et al. (2007), Sharma and Goyal (2016) and Shin et al. (2016). Phan et al. (2016) reviewed 111 BN applications in water resources management but only four were in the domain of river flow, five were in operational decision making context and none in operational flood warning. BN application in ensemble flood forecasting has not been reported yet to our best knowledge.

In previous studies, meteorological ensemble forecasts are coupled with a hydrological model to predict a set of flood hydrographs with different peak discharge. Ensemble decision making according to a range of possible flood peaks is a challenging issue especially in case of equal likelihood of each ensemble member. In the present study, the hydrological model is replaced by a Bayesian network for deriving a single warning indicator from atmospheric ensemble forecasts.

The purpose of the present study is therefore to predict the flood peak addressing the uncertainties and the probability of occurrence of each ensemble member. Floods are rare extreme events that occur with low frequency in the studied area. Thus, one of the problems in flood modelling is small data size. In the present study, we try to deal with small data size by using Bayesian network, which is less sensitive to small data size (Zhang and Bivens, 2007). As a case study, flood peaks were forecasted in a relatively small mountainous basin, Kan basin, Tehran, Iran. The Weather Research and Forecasting (WRF) model was used to simulate 14 historic precipitation events using five different cumulus parameterization schemes. Then atmospheric ensemble forecasts were coupled to the BN to estimate the flood magnitude for an ensemble forecasting, from which flood warnings could be derived. Forecasting performance of the BN was compared with the results obtained from an artificial neural network (ANN), which is a widely used data-based model.

2 Data and methodology

2.1 Study area

The case study of this research is Kan basin, Tehran, Iran with an area of $197\,\text{km}^2$. The geographical limits lie between $35°46'$ to $35°58'$ N latitudes and $51°10'$ to $51°23'$ E longitudes. Figure 1 shows the location of the study area. Average elevation is 2428.7 m above sea level and the annual rainfall is about 600 mm. The rainfall data were from Emamzadeh-Davood rainfall station and the flow data were collected from Sooleghan hydrometric station that is located downstream of the basin as shown in Fig. 1. The time of concentration (T_c) of the basin is about 3 h, so the NWP models can significantly increase the lead time of flood warning compared to using observed precipitation. Since the increasing of lead time decreases the accuracy of NWP forecasts (Sikder and Hossain, 2016), the forecasting was conducted 1 d before the observed

Table 2. Precipitation and streamflow data.

Event	Observed cumulative precipitation (mm)	Observed peak flow (m^3 s^{-1})	Duration (h)
27.03.2007	25.3	24.2	15
27.04.2007	33.5	57.1	2
07.12.2007	32.3	12.7	17
03.11.2008	37.3	20.9	17
30.04.2009	29	34.4	7
04.02.2010	68.1	11.6	11
08.04.2010	48.8	34.1	29
13.03.2011	32.6	20.9	14
05.04.2011	55.5	24.5	25
29.08.2011	56.4	26.4	11
28.10.2011	55.9	55.1	23
20.11.2011	48	44.7	31
14.04.2012	67.7	67.2	15
13.11.2012	78.9	25	41

event. Long lead time for flood forecasting is very important in large watershed flood mitigation as it provides more time for flood warning and emergency responses (Li et al., 2017). A flow chart of the proposed flood forecast approach is presented in Fig. 2. The cumulative precipitation, peak flow and duration of the events are presented in Table 2.

2.2 The Weather Research and Forecasting model (WRF)

The Weather Research and Forecasting (WRF) model was used to simulate 14 historic heavy precipitation events that caused floods in the study area. In this study, WRF version 3.8 was employed with three domains and 1 h temporal resolution. The horizontal resolutions of the domains are 45, 15 and 5 km, respectively. Figure 3 shows the WRF domain setup using an interactive nested domain inside the parent domain. The outer (the coarsest) domain covers Iran, the middle domain covers the northern part of Iran and the inner domain covers the study area, and only the meteorological information from this domain was used for forecasting of flooding in the study basin.

The NCEP Global Forecast System (GFS) final analysis (FNL) data were used as the initial conditions of the WRF. The model settings were based on the Noah land surface model (Chen and Dudhia, 2001), the Rapid Radiative Transfer Model (RRTM) longwave radiation scheme (Mlawer et al., 1997), the Dudhia shortwave radiation model (Dudhia, 1989), the Yonsei University (YSU) planetary boundary layer scheme (Hong et al., 2006) and the WRF Single-Moment (WSM) three-class microphysics scheme (Hong et al., 2004). Because of the importance of cumulus parameterization for hydrological purpose, an ensemble was created by using five cumulus schemes including KF, BMJ, GR3D,

MSKF and GDE cumulus scheme. The atmospheric ensemble forecasts were fed into the Bayesian network to estimate flood peak flow.

2.3 Bayesian network

This study proposes a probabilistic model to generate the flood forecasts and to estimate the flood magnitude based on Bayesian networks (BN) for an ensemble forecasting. BNs are a class of probabilistic graphical models composed by a set of random variables and directed acyclic graphs (DAGs) to show the potential dependence between variables (Scutari, 2017). The node at the start of an arrow is a casual or preceding event that is called the parent node, and the node at the head is an outcome event that is called the child node. Each node is labelled with a conditional probability table (CPT) based on prior information or statistically observed correlations that show the strengths of the influences of the parent nodes on the child node. In general, assuming random variables with domain size d, the conditional probability table of a child node with n parents needs one to specify $dn + 1$ probabilities (Li et al., 2011).

The goal is to calculate the posterior conditional probability distribution of each of the possible unobserved causes given the observed evidence, i.e. P [Cause | Evidence].

However, in practice we are often able to obtain only the converse conditional probability distribution of observing evidence given the cause, P [Evidence | Cause]. The whole concept of Bayesian networks is built on Bayes' theorem, which helps us to express the conditional probability distribution of cause given the observed evidence using the converse conditional probability of observing evidence given the cause as Eq. (1):

$$P[\text{Cause} | \text{Evidence}] = P[\text{Evidence} | \text{Cause}] \frac{P[\text{Cause}]}{P[\text{Evidence}]}. \quad (1)$$

Any node in a Bayesian network is always conditionally independent of its all non-descendants given that node's parents. The conditional probabilities are represented in the form of conditional probability distribution (CPD) if the nodes represent a continuous variable or a conditional probability table (CPT) if the nodes represent a discrete variable. The joint probability (P_b) can be defined as the product of the local conditional distributions as given in Eq. (2):

$$P_b(x_1 \cdot x_2 \cdot \ldots \cdot x_n) = \prod_{i=1}^{n} P_b(x_i | x_{i+1} \cdot \ldots \cdot x_n). \quad (2)$$

In a BN, a node x_i is independent of all other nodes except its parents (Sharma and Goyal, 2016). A simple example of BN is presented in Fig. 4. The joint probability for this simple network can be defined as Eq. (3):

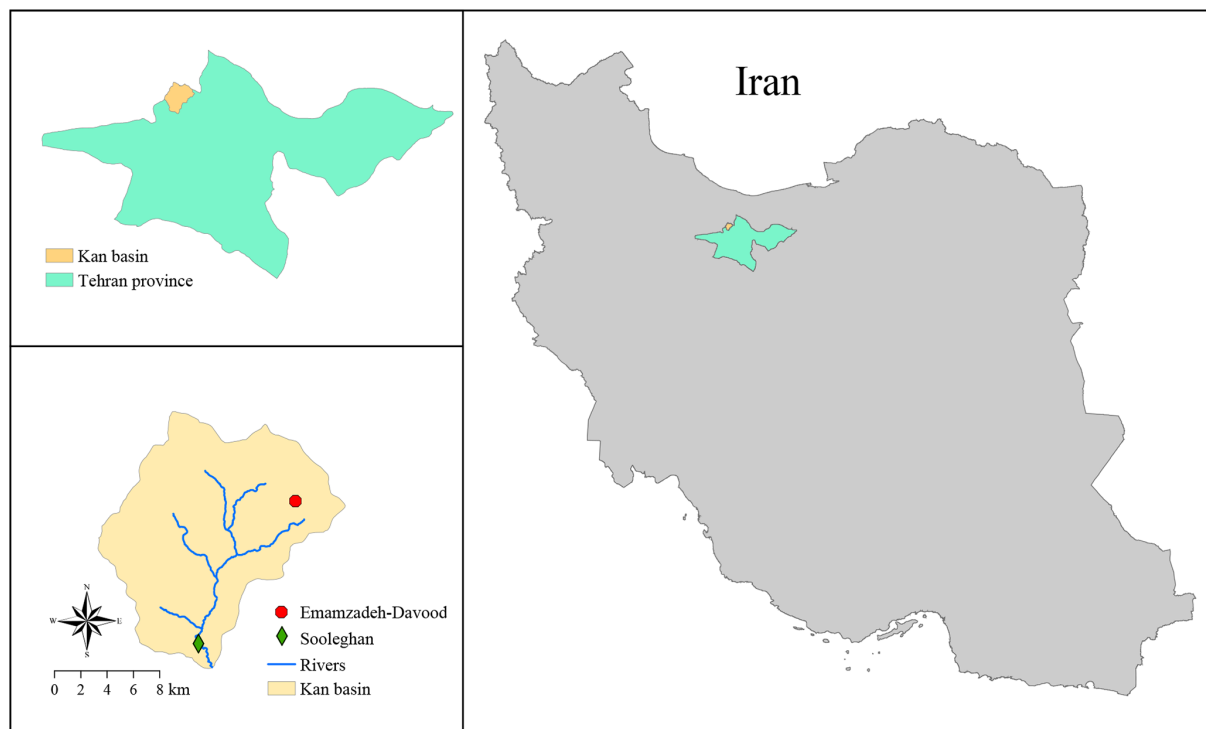$$p(a \cdot b \cdot c) = p(a) \times p(b|a) \times p(c|a \cdot b). \quad (3)$$

Figure 1. Location of study area, rainfall and flow stations.

The graph containing nodes and arrows is called BN structure (BS). Learning a Bayesian network includes two aspects: structure learning and parameter learning.

Structure learning. The purpose of structure learning is to determine the best structure which maximizes the conditional probability $P(\text{BS}|D)$, where BS is the BN structure, and D is the given data (Sharma and Goyal, 2016). Structure learning consists of finding the DAG that encodes the conditional independencies present in the data. This has been achieved in the literature with constraint-based, score-based and hybrid algorithms (Scutari, 2017). Some common structure learning techniques are the K2 algorithm (Cooper and Herskovits, 1992; Amirkhani and Rahmati, 2015) and the Markov chain Monte Carlo (MCMC) algorithm (Madigan et al., 1995). However, BS can be easily defined if the relationship between child nodes and parent nodes is known. In the present study, the flood is influenced by atmospheric ensemble forecasts, base flow of the river and antecedent rainfall, so the BS is known.

Parameter learning. Bayesian network conditional probability tables (CPTs) can be learned when the BN structure is known. Different parameter learning algorithms have been presented, including expectation maximization, Markov chain Monte Carlo methods such as Gibbs sampling and gradient descent methods (Reed and Mengshoel, 2014). In this study, expectation maximization (EM) algorithm was used for Bayesian network parameter learning. The EM algorithm

is an iterative method that performs a number of iterations, each of which calculate the logarithm of the probability of the data given the current joint probability distribution. This quantity is known as the log likelihood, and the algorithm attempts to maximize likelihood estimators (Bergmann and Kopp, 2009). In the HUGIN software (further developed from original work of Lauritzen and Spiegelhalter, 1988), convergence is achieved when the difference between the log likelihoods of two consecutive iterations is less than or equal to the numerical value of a log-likelihood threshold times the log likelihood. Alternatively, the user can specify an upper limit on the number of iterations to ensure that the procedure terminates.

Our proposed ensemble forecasting using a BN model has the following four steps:

1. selecting relevant variables and spatial units,

2. creating training data set for the model,

3. learning the model using the HUGIN software (version 8.3) and

4. evaluating the performance and accuracy of the model.

In the present study, the flood peak is the response variable that is influenced by some predictor variables including ensemble rainfall forecasts, base flow of the river and antecedent soil moisture. Base flow of the river is the normal day to day discharge. Antecedent recharge flow was used as
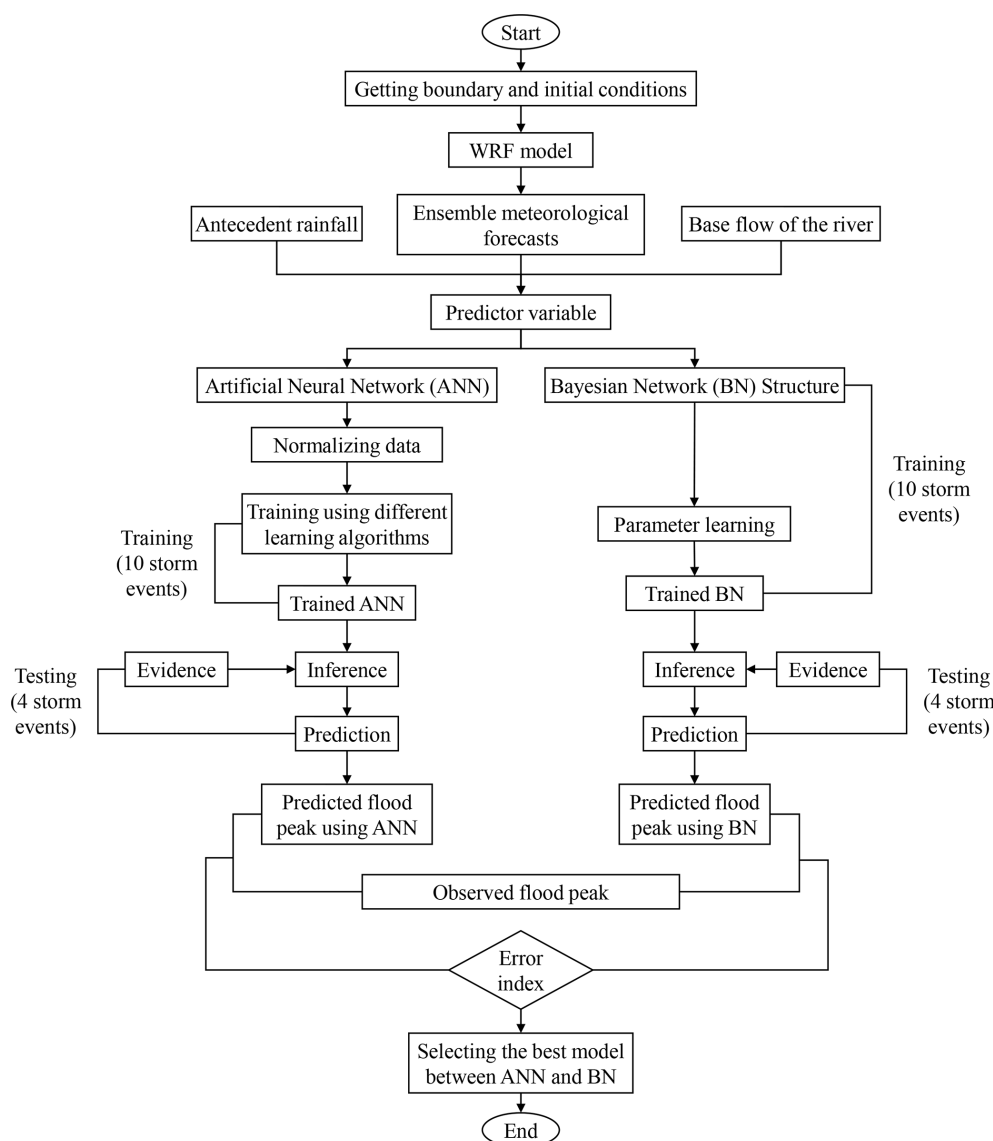
Figure 2. Flow chart of the flood forecast approach in this research.

the base flow of the river. The catchment's antecedent soil moisture represents the relative wetness prior to the flood event and can have an important influence on flood response. Because of the lack of soil moisture data in the Kan basin, antecedent rainfall was used to represent the soil moisture. Antecedent rainfall is the total precipitation amount that occurred in the 24 h before the start of the event. This study was performed on 14 historical storms. It should be noted that approx. 70 % of the available data (10 storm events) are allocated for training and the remaining (4 storm events) data are used for validation. The data sample is relatively small due to the following reasons.

1. NCEP (GFS − FNL) data are not available for some historical storms.

2. During the above-mentioned period, a small number of actual flood events occurred in the study area, since the basin is located in a semi-arid region.

3. There is a lack of flood data because of flood damage to hydrometry equipment in some floods.

Considering the relatively small sample size, we proposed using the BN that is less sensitive to small data set size in comparison with ANN. Some advantages of BN are as follows.

1. *Suitable for small and incomplete data sets.* A very useful advantage of BN is that there are no minimum sample data sizes needed to perform the analysis, and BN takes into account the complete data set (Myllymaki et

Figure 3. WRF domain setup using an interactive nested domain inside the parent domain.
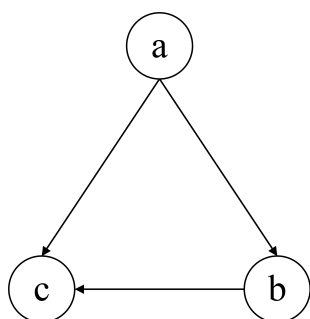

Figure 4. An example of a graphical Bayesian network.

al., 2002). In addition, Kontkanen et al. (1997) demonstrate that BN can show good accuracy of prediction even with a rather small data set. Furthermore, Zhang and Bivens (2007) showed that BN is less sensitive to small data set size in comparison with ANN.

2. *Structural learning possible.* It is possible to use data and also subject matter knowledge to learn the structure of BN. This is an aspect of active research, and although the statistical theory is well understood, the techniques are still under development (Jensen, 2001).

3. *Fast responses.* Since BN is analytically solved, it can provide fast responses to requests once the model is compiled. The compiled form of a BN comprises a conditional probability distribution for each combination of variable values and thus can provide any distribution in-

stantly. This is in contrast to the other simulation models in which the results need to be simulated, which can take very long (Uusitalo, 2007). Thus, BN are recommended for operational ensemble forecasting in particular in fast-reacting basins, where a high number of forecasts must be simulated within a short time.

2.4 Artificial neural networks (ANNs)

Artificial neural networks (ANN) are used as an alternative of statistical models in different aspects including clustering analysis, estimation, sample recognition etc. (Mammadov et al., 2005). An ANN model is basically an engineering method of biological neurons. It is constructed by input, output and hidden layers. ANN consist of a large number of simple processing elements, which are interconnected with each other and also layered (Sharma et al., 2012).

Typically, there are four distinct steps in developing an ANN model. The first step is data transformation or scaling. The input and output variables are first normalized linearly in the range of 0 and 1 using the following equation:

$$\overline{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{4}$$

where $\overline{X}$ is the normalized value of the X, and $X_{\min}$ and $X_{\max}$ are the minimum and maximum of data, respectively. The main purpose for standardizing the data is that the variables are usually measured in different units. By normalizing the variables in dimensionless units, the arbitrary effect of similarity between objects is removed (Aichouri et al., 2015).

The second step is the network architecture definition in which the number of hidden layers, the number of neurons in each layer and the connectivity between the neurons are determined. The number of neurons and hidden layers is problem dependent and is estimated by the trial and error technique or expert experience. A synaptic weight is allocated to each link to represent the relative connectivity strength of two nodes at both ends in predicting the input–output relationship (Raju et al., 2011). A typical ANN architecture is presented in Fig. 5. In this study, the output from the model is the flood peak, and the input variables are atmospheric ensemble forecasts, base flow of the river and antecedent rainfall. The third step is using a learning algorithm to train the network with known data. There are several learning algorithms. In the present study, the most widely used feedforward error back-propagation algorithm was used for training because of the good performance of this algorithm in previous studies (Raju et al., 2011; Banihabib et al., 2015; ASCE, 2000; Sarkar and Kumar, 2012). The success of an ANN application depends on the quality and also the quantity of the available data (Cheng et al., 2017). The final step is the validation, in which the performance of the trained ANN model is evaluated using statistical criteria (Sarkar and Kumar, 2012).
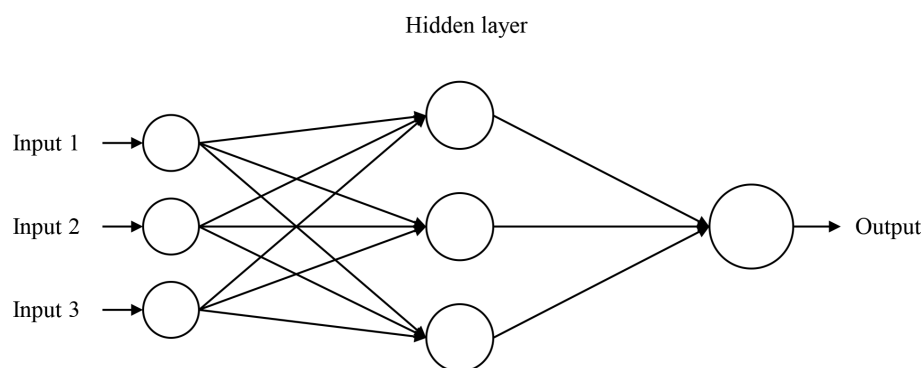
Hidden layer

Figure 5. Typical ANN architecture.

2.5 Statistical criteria for validation

In the present study, mean absolute relative error (MARE), mean relative bias error (MRBE) and the regression coefficient (r) were used for performance evaluation of the model as given in the following equations:

$$\text{MARE} = \frac{1}{n} \sum \frac{|O_i - F_i|}{O_i}, \tag{5}$$

$$\text{MRBE} = \frac{1}{n} \sum \frac{O_i - F_i}{O_i}, \tag{6}$$

$$r = \frac{n(\sum OF) - (\sum O)(\sum F)}{\sqrt{[n \sum O^2 - (\sum O)^2][n \sum F^2 - (\sum F)^2]}}, \tag{7}$$

where O_i is the observed value, F_i is the predicted value and n is the total number of data sets.

3 Results and discussion

3.1 Rainfall verification using the WRF model

In this section, the comparison between the observed and predicted precipitation obtained from the WRF model is addressed. As mentioned earlier, the WRF model was used to simulate 14 historic precipitation events, and the results for some events are presented here. Figure 6 illustrates the predicted cumulative rainfall and the observed cumulative rainfall for these events. In general, the results show that the WRF model was able to capture the heavy rainfall events. The uncertainties in the predicted rainfall lead to a large spread of the ensemble members, and this is why the uncertainty in rainfall forecasting becomes important.

The ensemble precipitation illustrates that both overestimation and underestimation of precipitation occurs using various schemes. Overestimation is very noticeable for the early hours of forecasting, while for the last period of the event underestimation occurs in some schemes.

From the case study, the results of precipitation forecast using different cumulus schemes by the WRF model can be significantly different. Therefore, it is necessary to forecast precipitation by implementing various physics schemes, especially different microphysical schemes. Furthermore, it can be inferred that the difference between observed and predicted rainfall is mainly caused by the initial condition in the NWP models, thus the atmospheric ensemble forecasts can be produced also by perturbing the initial conditions.

3.2 Bayesian network verification

The atmospheric ensemble forecasts were fed into the BN to estimate flood peak flow. Ten various models were developed using various combinations of predictors. In all of the combinations, flood-peak discharge is the predicting variable. Table 3 shows the accuracy of the model for different combinations of predictors to compare the performance of the prediction. The performance of the model was evaluated by MARE and R^2. It is clear from Table 3 that maximum hourly rainfall outperformed accumulated rainfall as the predictor variable (no. 2 in Table 4). It shows for the relatively short concentration-time basin, the Kan basin, that cumulative precipitation is not a good indicator to predict the flood peak and that the maximum hourly rainfall provides better results. Thus maximum hourly rainfall was used in combinations of other predictor variables. This can also be seen by comparing combination no. 5 and no. 9 that there is no considerable decrease in accuracy by deleting the Multi-scale Kain–Fritsch scheme; consequently it can be concluded that MSKF is the least accurate cumulus scheme. It was also found that by deleting the Kain–Fritsch scheme in combination (no. 6 in Table 3), the accuracy is significantly decreased. Thus, the Kain–Fritsch is the most efficient cumulus parameterization scheme in the study area. Other studies on precipitation prediction have also shown similar results. Pennelly et al. (2014) showed that the Kain–Fritsch cumulus parameterization scheme is the most accurate in simulating heavy precipitation across three summer events. Liang et al. (2004) showed that the Kain–Fritsch scheme works better
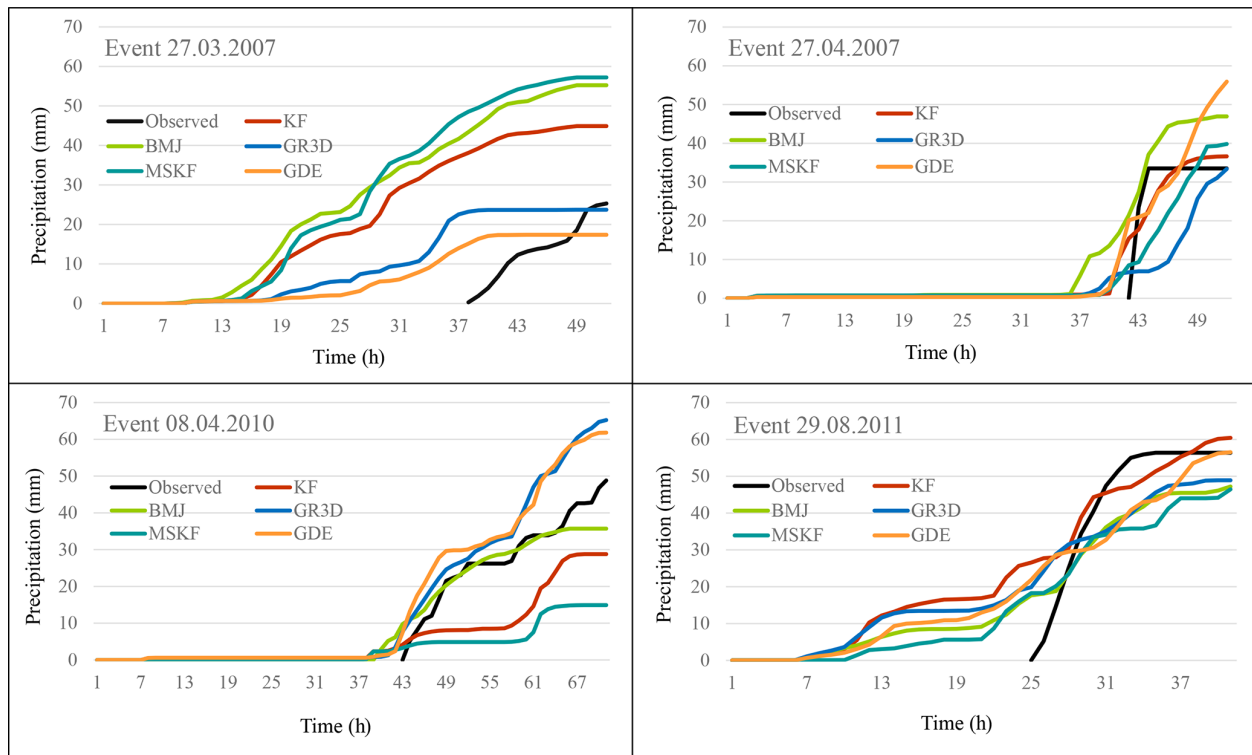
Figure 6. The ensemble forecasted precipitation and the observed cumulative precipitation.

in the southeast of United States, where convection is largely governed by the near-surface forcing.

According to Table 3, the best results were obtained for combination no. 5. The proposed structure of this combination is composed of eight nodes as shown in Table 4. Atmospheric ensemble forecasts, base flow of the river and antecedent rainfall are the parent nodes, and flood peak is the child node. It can also be seen that the base flow is influenced by antecedent rainfall. The mean absolute relative error was calculated at 0.076 for the validation data set in the combination no. 5. The coefficient of determination (R^2) is another criterion for testing, and it is seen from Table 3 that it's values are close to unity. We should compare our study to similar studies to determine whether our R^2 is in the right ballpark. Khan and Coulibaly (2006) used a Bayesian learning approach to train a multilayer feed-forward network for daily river flow and reservoir inflow simulation. Their result also showed a high R^2 value. The results showed that the BN is an efficient method for modelling and combining the ensemble flood forecast prediction. The proposed BN approach in this study predicts flood peak flow. Since the Kan River in the studied reach is a mountainous river without any flood plain storage, the peak discharge is almost not reduced by flood routing along the river, and so we can use the peak flood instead of routing the flood hydrograph. However, in our study, we consider peak flow as the variable of interest. In other fields of application, flow volume or time to peak might be of interest.

Moreover, Bayesian cluster analysis could also provide probabilistic results for flood early warning, but since the data sample is relatively small in this study, cluster analysis cannot be achieved. This method can be also tested in basins with sufficient historical hydrological data in future works.

The performance of the BN model is compared with the results obtained from an ANN model as a benchmark. The comparison is conducted using the same data set for training and validation. These results are presented in Sect. 3.3.

3.3 Artificial neural network verification

The first step in developing an ANN model is to determine the input and output variables. The output of the model is the magnitude of flood peak discharge. The input variables are the same as those used for the BN with the best performing combination of predictor variables (Table 3, combination no. 5). The feed-forward error back-propagation algorithm has been employed as the training algorithm in this study. A difficult task in working with ANN is the selection of parameters such as the number of hidden nodes. There is no established algorithm until now to determine how many hidden nodes are required to approximate any given function. Here, we use the common trial and error method to choose the number of hidden nodes, which are varied from two to six according to previous studies (Banihabib et al., 2015).

Table 3. Performance of the Bayesian network for different combinations of predictor variables.

Combination no.	predictor variables	R^2	MARE
1	Maximum hourly rainfall	0.99	0.16
2	Accumulated rainfall	0.74	1.06
3	Maximum hourly rainfall, base flow of the river	0.99	0.18
4	Maximum hourly rainfall, antecedent rainfall	0.99	0.12
5	Maximum hourly rainfall, base flow of the river, antecedent rainfall	0.99	0.076
6	Maximum hourly rainfall (deleting KF), base flow of the river, antecedent soil moisture	0.58	0.46
7	Maximum hourly rainfall (deleting BMJ), base flow of the river, antecedent rainfall	0.99	0.23
8	Maximum hourly rainfall (deleting GR3D), base flow of the river, antecedent rainfall	0.99	0.15
9	Maximum hourly rainfall (deleting MSKF), base flow of the river, antecedent rainfall	0.99	0.087
10	Maximum hourly rainfall (deleting GDE), base flow of the river, antecedent rainfall	0.99	0.10

Table 4. The cause–effect relationships among the variables in the proposed structure of the Bayesian network.

Parent ID	Entity name	Child ID	Child name
n1	Predicted rainfall using KF cumulus parameterization scheme	n8	Flood peak
n2	Predicted rainfall using BMJ cumulus parameterization scheme	n8	Flood peak
n3	Predicted rainfall using GR3D cumulus parameterization scheme	n8	Flood peak
n4	Predicted rainfall using MSKF cumulus parameterization scheme	n8	Flood peak
n5	Predicted rainfall using GDE cumulus parameterization scheme	n8	Flood peak
n6	Base flow	n8	Flood peak
n7	Antecedent rainfall	n6 and n8	Base flow and flood peak

Table 5. MARE and R^2 of the artificial neural network in the verification phase.

Number of nodes in hidden layer	MARE	R^2
2	1.14	0.44
3	0.74	0.92
4	0.39	0.77
5	0.51	0.93
6	1.23	0.12

Error index is usually used to select the best performance of the network model compared to observed data. The accuracy of the model for different numbers of nodes in the hidden layer is presented in Table 5. It was found that four hidden nodes give the best results. The mean absolute relative error (MARE) was calculated as 0.39 for the validation data set while this index was calculated 0.076 in BN. The comparison shows that BN offers better accuracy. Although our data set was relatively small, the result of BN model was accurate enough. Therefore, it seems that BN is less sensitive to small data set size, so it is more suited for rare events such as floods, where the available data are limited due to the high return period of such events.

4 Conclusions

This study proposed a probabilistic model to address the uncertainties of flood forecasts using the Bayesian networks (BNs) and to estimate the flood peak in an ensemble flood forecasting. This is the first attempt to use BN in ensemble flood forecasting. The Weather Research and Forecasting (WRF) model was used to simulate some historic precipitation rainfall events using five various cumulus parameterization schemes. The results showed that there is no considerable decrease in accuracy by deleting the Multi-scale Kain–Fritsch scheme, thus it can be concluded that is the least accurate cumulus scheme. It also was found that Kain–Fritsch is the most efficient cumulus parameterization scheme. Atmospheric ensemble forecasts were coupled with the Bayesian network to estimate the flood magnitude in an ensemble forecasting. Results of the BN are compared with the results obtained from an artificial neural network as a widely used model to show the performance of BN. The comparison is conducted using the same data set for validation and training. The results showed that the BN is an efficient method for flood forecasting based on ensemble rainfall forecasts and offers better accuracy than ANN. We showed that BN is less sensitive to small data set size in comparison with other models, thus it is more suited for rare events such as floods. The results of this study indicate that BN might be a suitable tool for a fast computation of peak flow and flood warnings from numerical ensemble weather predictions. Our study is

a proof of concept at the current stage that flood warnings can be done by evaluating hydrological pre-conditions and meteorological ensembles by a trained BN instead of a hydrological model. However, further studies are required to confirm the applicability of BN. The present study was conducted with a lead time of 1 d before the observed event in a small basin. Future studies may test BN for other catchments and for larger lead times.

References

Abebe, A. and Price, R.: Decision support system for urban flood management, J. Hydroinform., 7, 3–15, https://doi.org/10.2166/hydro.2005.0002, 2005.

Aichouri, I., Hani, A., Bougherira, N., Djabri, L., Chaffai, H., and Lallahem, S.: River flow model using artificial neural networks, Energy Proced., 74, 1007–1014, https://doi.org/10.1016/j.egypro.2015.07.832, 2015.

Amirkhani, H. and Rahmati, M.: Expectation maximization based ordering aggregation for improving the K2 structure learning algorithm, Intell. Data Anal., 19, 1003–1018, https://doi.org/10.3233/ida-150755, 2015.

ASCE: Task Committee on Application of Artificial Neural Networks in Hydrology: Artificial neural networks in hydrology. II: hydrologic applications, J. Hydrol. Eng., 5, 124–137, 2000.

Banihabib, M. and Arabi, A.: The impact of catchment management on emergency management of flash-flood, International Journal of Emergency Management, 12, 185–195, https://doi.org/10.1504/ijem.2016.076618, 2016.

Banihabib, M. E., Arabi, A., and Salha, A. A.: A dynamic artificial neural network for assessment of land-use change impact on warning lead-time of flood, International Journal of Hydrology Science and Technology, 5, 163–178, https://doi.org/10.1504/ijhst.2015.070093, 2015.

Bergmann, K. and Kopp, S.: GNetIc–Using bayesian decision networks for iconic gesture generation, International Workshop on Intelligent Virtual Agents, Springer, Berlin, Heidelberg, 76–89, 2009.

Chen, F. and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity, Mon. Weather Rev., 129, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0587:caalsh>2.0.co;2, 2001.

Cheng, K.-S., Lien, Y.-T., Wu, Y.-C., and Su, Y.-F.: On the criteria of model performance evaluation for real-time flood forecasting, Stoch. Env. Res. Risk A., 31, 1123–1146, 2017.

Chitsaz, N. and Banihabib, M. E.: Comparison of different multi criteria decision-making models in prioritizing flood management alternatives, Water Resour. Manag., 29, 2503–2525, 2015.

Cooper, G. F. and Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data, Mach. Learn., 9, 309–347, https://doi.org/10.1002/0470091754.ch2, 1992.

Deb, S., Srivastava, T., and Kishtawal, C.: The WRF model performance for the simulation of heavy precipitating events over Ahmedabad during August 2006, J. Earth Syst. Sci., 117, 589–602, https://doi.org/10.1007/s12040-008-0055-5, 2008.

Dietrich, J., Schumann, A. H., Redetzky, M., Walther, J., Denhard, M., Wang, Y., Pfützner, B., and Büttner, U.: Assessing uncertainties in flood forecasts for decision making: prototype of an operational flood management system integrating ensemble predictions, Nat. Hazards Earth Syst. Sci., 9, 1529–1540, https://doi.org/10.5194/nhess-9-1529-2009, 2009a.

Dietrich, J., Denhard, M., and Schumann, A.: Can ensemble forecasts improve the reliability of flood alerts?, J. Flood Risk Manag., 2, 232–242, https://doi.org/10.1111/j.1753-318x.2009.01039.x, 2009b.

Doycheva, K., Horn, G., Koch, C., Schumann, A., and König, M.: Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application

to runoff simulations and flood warning, Adv. Eng. Inform., 33, 427–439, https://doi.org/10.1016/j.aei.2016.11.001, 2017.

Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, J. Atmos. Sci., 46, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:nsocod>2.0.co;2, 1989.

El Afandi, G., Morsy, M., and El Hussieny, F.: Heavy rainfall simulation over sinai peninsula using the weather research and forecasting model, International Journal of Atmospheric Sciences, 2013, 241050, https://doi.org/10.1155/2013/241050, 2013.

Goodarzi, L., Banihabib, M. E., and Roozbahani, A.: A decision-making model for flood warning system based on ensemble forecasts, J. Hydrol., 573, 207–219, https://doi.org/10.1016/j.jhydrol.2019.03.040, 2019.

Grell, G. A.: Prognostic evaluation of assumptions used by cumulus parameterizations, Mon. Weather Rev., 121, 764–787, https://doi.org/10.1175/1520-0493(1993)121<0764:peoaub>2.0.co;2, 1993.

Grell, G. A. and Dévényi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett., 29, 38-1–8-4, https://doi.org/10.1029/2002gl015311, 2002.

Hénonin, J., Russo, B., Roqueta, D. S., Sanchez-Diezma, R., Domingo, N., Thomsen, F., and Mark, O.: Urban flood real-time forecasting and modelling: a state-of-the-art review, Proceedings, MIKE by DHI conference, https://doi.org/10.2166/hydro.2013.132, 2010.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, Mon. Weather Rev., 132, 103–120, https://doi.org/10.1175/1520-0493(2004)132<0103:aratim>2.0.co;2, 2004.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, https://doi.org/10.1175/mwr3199.1, 2006.

Janjić, Z. I.: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes, Mon. Weather Rev., 122, 927–945, https://doi.org/10.1175/1520-0493(1994)122<0927:tsmecm>2.0.co;2, 1994.

Jensen, F. V.: Bayesian Networks and Decision Graphs, Springer-Verlag, New York, ISBN 0-387-95259-4, 2001.

Kain, J. S. and Fritsch, J. M.: A one-dimensional entraining/detraining plume model and its application in convective parameterization, J. Atmos. Sci., 47, 2784–2802, https://doi.org/10.1175/1520-0469(1990)047<2784:aodepm>2.0.co;2, 1990.

Kerkhoven, E., Gan, T. Y., Shiiba, M., Reuter, G., and Tanaka, K.: A comparison of cumulus parameterization schemes in a numerical weather prediction model for a monsoon rainfall event, Hydrol. Process., 20, 1961–1978, https://doi.org/10.1002/hyp.5967, 2006.

Khan, M. S. and Coulibaly, P.: Bayesian neural network for rainfall-runoff modelling, Water Resour. Res., 42, W07409, https://doi.org/10.1029/2005WR003971 2006.

Kontkanen, P., Myllymaki, P., Silander, T., and Tirri, H.: Comparing predictive inference methods for discrete domains, Proceedings

of the sixth International Workshop on Artificial Intelligence and Statistics, Ft. Lauderdale, USA, 311–318, 1997.

Lauritzen, S. L. and Spiegelhalter, D. J.: Local computations with probabilities on graphical structures and the application to expert systems, J. Roy. Stat. Soc. B, 50, 157–224, 1988.

Leandro, J., Gander, A., Beg, M. N. A., Bhola, P., Konnerth, I., Willems, W., Carvalho, R., and Disse, M.: Forecasting upper and lower uncertainty bands of river flood discharges with high predictive skill, J. Hydrol., 576, 749–763, https://doi.org/10.1016/j.jhydrol.2019.06.052, 2019.

Li, J., Chen, Y., Wang, H., Qin, J., Li, J., and Chiao, S.: Extending flood forecasting lead time in a large watershed by coupling WRF QPF with a distributed hydrological model, Hydrol. Earth Syst. Sci., 21, 1279–1294, https://doi.org/10.5194/hess-21-1279-2017, 2017.

Li, L., Li, W., and Jin, J.: Improvements in WRF simulation skills of south eastern United States summer rainfall: physical parameterization and horizontal resolution, Clim. Dynam., 43, 2077–2091, https://doi.org/10.1007/s00382-013-2031-2, 2014.

Li, W., Poupart, P., and van Beek, P.: Exploiting structure in weighted model counting approaches to probabilistic inference, J. Artif. Intell. Res., 40, 729–765, 2011.

Liang, X. Z., Li, L., Dai, A., and Kunkel, K. E.: Regional climate model simulation of summer precipitation diurnal cycle over the United States, Geophys. Res. Lett., 29, L24208, https://doi.org/10.1029/2004gl021054, 2004.

Madigan, D., York, J., and Allard, D.: Bayesian graphical models for discrete data, International Statistical Review/Revue Internationale de Statistique, 63, 215–232, https://doi.org/10.2307/1403615, 1995.

Mammadov, M., Yazici, B., Yolaçan, Ş., Aslanargun, A., Yüzer, A., and Ağaoğlu, E.: Statistical Methods and Artificial Neural Networks, Journal of Modern Applied Statistical Methods, 5, 495–512, 2005.

Mediero, L., Garrote, L., and Martin-Carrasco, F.: A probabilistic model to support reservoir operation decisions during flash floods, Hydrolog. Sci. J., 52, 523–537, https://doi.org/10.1623/hysj.52.3.523, 2007.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res.-Atmos., 102, 16663–16682, https://doi.org/10.1023/b:asys.0000017241.50767.32, 1997.

Myllymaki, P., Silander, T., Tirri, H., and Uronen, P.: B-Course: a web-based tool for Bayesian and causal data analysis, Int. J. Artif. Intell. T., 11, 369–387, https://doi.org/10.1142/s0218213002000940, 2002.

National Centers for Environmental Prediction/National Weather Service/NOAA/US Department of Commerce: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colorado, https://doi.org/10.5065/D6M043C6, 2000.

Pennelly, C., Reuter, G., and Flesch, T.: Verification of the WRF model for simulating heavy precipitation in Alberta, Atmos. Res., 135, 172–192, https://doi.org/10.1016/j.atmosres.2013.09.004, 2014.

Phan, T. D, Smart, J. C. R., Capon, S. J., Hadwen, W. L., and Sahin, O.: Applications of Bayesian belief networks in water resource management: A systematic, Environ. Modell. Softw., 85, 98–111, 2016.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian model averaging to calibrate forecast ensembles, Mon. Weather Rev., 133, 1155–1174, https://doi.org/10.21236/ada459828, 2005.

Raju, M. M., Srivastava, R. K., Bisht, D., Sharma, H. C., and Kumar, A.: Development of artificial neural-network-based models for the simulation of spring discharge, Adv. Art. Int., 1–11, 2011.

Reed, E. and Mengshoel, O. J.: Bayesian network parameter learning using EM with parameter sharing, Proceedings of the Eleventh UAI Conference on Bayesian Modeling Applications Workshop-Volume, 27 July 2014, Quebec, Canada, 1218, 48–59, 2014.

Rogelis, M. C. and Werner, M.: Streamflow forecasts from WRF precipitation for flood early warning in mountain tropical areas, Hydrol. Earth Syst. Sci., 22, 853–870, https://doi.org/10.5194/hess-22-853-2018, 2018.

Sarkar, A. and Kumar, R.: Artificial neural networks for event based rainfall-runoff modelling, Journal of Water Resource and Protection, 4, 891–897, https://doi.org/10.4236/jwarp.2012.410105, 2012.

Scutari, M.: Bayesian network constraint-based structure learning algorithms: Parallel and optimised implementations in the bnlearn R package, J. Stat. Softw., 7, 1–20, 2017.

Sharma, V., Rai, S., and Dev, A.: A comprehensive study of artificial neural networks, International Journal of Advanced Research in Computer Science and Software Engineering, 2, 278–284, 2012.

Sharma, A. and Goyal, M. K.: Bayesian network for monthly rainfall forecast: a comparison of K2 and MCMC algorithm, International Journal of Computers and Applications, 38, 199–206, https://doi.org/10.1080/1206212x.2016.1237131, 2016.

Shin, J. Y., Ajmal, M., Yoo, J., and Kim, T.-W.: A Bayesian network-based probabilistic framework for drought forecasting and outlook, Adv. Meteorol., 2016, 9472605, https://doi.org/10.1155/2016/9472605, 2016.

Sikder, S. and Hossain, F.: Assessment of the weather research and forecasting model generalized parameterization schemes for advancement of precipitation forecasting in monsoon-driven river basins, J. Adv. Model. Earth Sy., 8, 1210–1228, https://doi.org/10.1002/2016ms000678, 2016.

Tennant, W. J., Toth, Z., and Rae, K. J.: Application of the NCEP ensemble prediction system to medium-range forecasting in South Africa: New products, benefits, and challenges, Weather Forecast., 22, 18–35, 2007.

Uusitalo, L.: Advantages and challenges of Bayesian networks in environmental modelling, Ecol. Model., 203, 312–318, 2007.

Yang, B., Qian, Y., Lin, G., Leung, R., and Zhang, Y.: Some issues in uncertainty quantification and parameter tuning: a case study of convective parameterization scheme in the WRF regional climate model, Atmos. Chem. Phys., 12, 2409–2427, https://doi.org/10.5194/acp-12-2409-2012, 2012.

Yang, T.-H., Hwang, G.-D., Tsai, C.-C., and Ho, J.-Y.: Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts, Hydrol. Earth Syst. Sci., 20, 4731–4745, https://doi.org/10.5194/hess-20-4731-2016, 2016.

Zhang, R. and Bivens, A. J.: Comparing the use of Bayesian networks and neural networks in response time modeling for service-oriented systems, Proceedings of the 2007 workshop on Service-oriented computing performance: aspects, issues and approaches, 67–74, https://doi.org/10.1145/1272457.1272467, 2007.

Zheng, Y., Alapaty, K., Herwehe, J. A., Del Genio, A. D., and Niyogl, D.: Improving high-resolution weather forecasts using the Weather Research and Forecasting (WRF) Model with an updated Kain–Fritsch scheme, Mon. Weather Rev., 144, 833–860, https://doi.org/10.1175/mwr-d-15-0005.1, 2016.