CrossMark

ORIGINAL ARTICLE

Comparison of methods used for quantifying prediction interval in artificial neural network hydrologic models

K. S. Kasiviswanathan[1,2] · **K. P. Sudheer**[2]

Abstract The application of artificial neural network (ANN) has gained significant interest while modeling various hydrologic processes. The main reason is the ANN models have produced promising results without the detailed information of watershed characteristics as required in physics based models. Still, the uncertainty in ANN models is a major issue that cannot be ignored. There could be different forms to represent model uncertainty, in which quantification of prediction interval for the model output has been mostly reported. In this paper, three different methods [i.e. Bootstrap method, Bayesian Approach and Prediction Interval (PI) method] were employed for quantifying the prediction interval in ANN models. The modeling procedure presented in this paper, is illustrated through river flow forecasting using the data collected from Kolar basin, India. The prediction interval was quantified using the measures such as percentage of coverage and average width. The comparison between these methods indicated that PI method has resulted in relatively less prediction and parameter uncertainty, besides the improved model performance. In addition, the PI method produced accurate prediction of hydrograph peak, which is a general concern in ANN models.

Keywords Artificial neural network · Bayesian · Bootstrap · Prediction interval · Uncertainty

✉ K. S. Kasiviswanathan
  vishwaiitr@gmail.com

1 Department of Civil Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary T2N 1N4, Canada

2 Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600 036, India

Introduction

Hydrologic modeling and simulation techniques are powerful tools for conceptualizing the hydrologic processes. The model output is mainly used for assessing the influence/impact on water resources systems and for decision making. The hydrologic models in general are classified into theory driven and data driven models. The advantage of using theory driven models lies in representing the system completely or partially with physics based equations. In contrast, the data driven models do not consider the underlying physics of the system. Hence, they model the hydrologic processes usually with historically observed time series data. Nevertheless, the results of data-driven models have been proved to be promising and comparable with theory-driven models. The major reasons can be attributed to robustness and non-linear approximation of data driven models, which has a flexibility to bring the functional relationship among hydrologic variables. In this context, the ANN based data driven models have gained significant interest among hydrologists (Hsu et al. 1995; Thirumalaiah and Deo 2000; Sharma et al. 2015). It may be noted that the main objective of majority of earlier studies pertained to train the ANN model for the point prediction/forecasting of variables of interest. However, the point prediction often fails to explain the inherent variability of the system. Therefore, it is hard to trust the reliability of ANN models. In addition, some of the other reasons could be (1) the stochastic nature of ANN models do not produce identical results (Elshorbagy et al. 2010a, b), (2) difficulty in assigning the confidence interval to the output (Shrestha and Solomatine 2006) and (3) lack of transparency (Abrahart et al. 2010). In general, the variability/uncertainty associated with hydrologic model prediction is mainly influenced by the measurement error in model input

🖄 Springer

and output variables, model parameters and the model structure (Schoups and Vrugt 2010). The benefits of uncertainty quantification in hydrologic models not only help in selecting the appropriate model, but it also gives an indication of best and worst scenarios, weak and strong points of modeling (Srivastav et al. 2007).

Numerous methods have been reported in literature for quantifying the uncertainty in hydrologic models. In general, the methods are classified (Shrestha and Solomatine 2006) into (1) analytical methods, (2) approximation methods, (3) simulation and sampling based methods, (4) Bayesian methods, (5) methods based on analysis of model errors, (6) fuzzy theory based methods. The assumption and difficulty of applying these methods require the information of probability distribution of parameters, statistical information of normality and homoscedasticity of residuals, knowledge of membership functions.

The quantification of uncertainty in ANN models is computationally a challenging task, possibly due to parallel computing architecture and large degrees of freedom exist in their development. Therefore, currently employed uncertainty analysis methods in physics and conceptual models cannot be applied directly to ANN models without suitable modifications. Still, different methods for quantifying the uncertainty of ANN models have been reported such as ensemble based approach (Boucher et al. 2010), Bayesian approach (Zhang et al. 2009), bootstrap approach (Srivastav et al. 2007), heuristic method (Han et al. 2007), fuzzy based approach (Alvisi and Franchini 2011). Despite various methods to quantify uncertainty of neural network based hydrologic model, each method has its own merits and demerits while representing the uncertainty.

In this paper, comparisons between three different uncertainty methods applied in ANN models have been investigated. The methods include (1) Bootstrap method, (2) Bayesian approach and (3) Prediction Interval (PI) method. The rationale in selecting these methods is that they use different mathematical formulation, principle and assumption. It is demonstrated through river flow forecast models using the data collected from Kolar basin, India. The uncertainty in ANN models was assessed quantitatively using the indices called average width (AW) of prediction interval and percentage of coverage (POC). Based on the estimate, the effective method was identified which produce narrow prediction interval, which also includes more number of observed values within the prediction interval of ANN model output. In addition, the qualitative comparison was also carried out based on (1) the difficulty of implementation, (2) computational efficiency, (3) fulfilment of statistical and probabilistic assumption, (4) parameter convergence and (5) meaningful quantification and accuracy in predicting the peak flows.

Methodology

An ANN is a mathematical tool inspired by biological neurons. It can be characterized as massively parallel connections of neurons called nodes arranged in a layer. The typical ANN architecture has input, hidden and output layer. The input and output layer is problem dependent where as the hidden layer is responsible for bringing suitable relationship between the inputs and outputs. There could be one or several hidden layers depend on the complexity of the problem. The connections are linked through weights and biases, which are numerical values estimated from training/calibration of ANN models (i.e. parameters of the ANN). More details about the functioning of ANN are available in various literatures and are not presented herein for brevity. In this section, the methods employed for carrying out uncertainty analysis is explained as follows.

In addition to that, a brief description is presented for the ANN model development that includes input determination, data division and identification of ANN architecture. It may be noted that this study considers only a parameter uncertainty as a source of uncertainty while estimating the model prediction interval; hence other sources are not included in the analysis. However, other sources can also be included, which is in general computationally intensive (Zhang et al. 2011). Figure 1 shows the flowchart describing the overall methodology.

Bootstrap method

It has been emphasized that bootstrap is a simple method to quantify the parameter and prediction uncertainty of neural network (Srivastav et al. 2007). The advantage of using the method does not require the assumption of probability distribution of parameters or complex derivatives of non-linear function. The quantification of uncertainty is carried out training independent ANN models through sampling subset of input–output patterns with replacement (Srivastav et al. 2007) from the wholedata set. Suppose, if 'B' such random samples are bootstrapped each time from total available dataset, the simple arithmetic average of prediction $\hat{y}_i$ can be considered as model output which corresponds to the ith input data point 'x'.

$$\hat{y}_i = \frac{1}{B} \sum_{b=1}^{B} f(x_i; p_b) \tag{1}$$

where p_b denotes the parameter obtained from b^{th} bootstrap sample and f denotes the functional form of ANN model.

Fig. 1 *Flow chart* describing
methodology
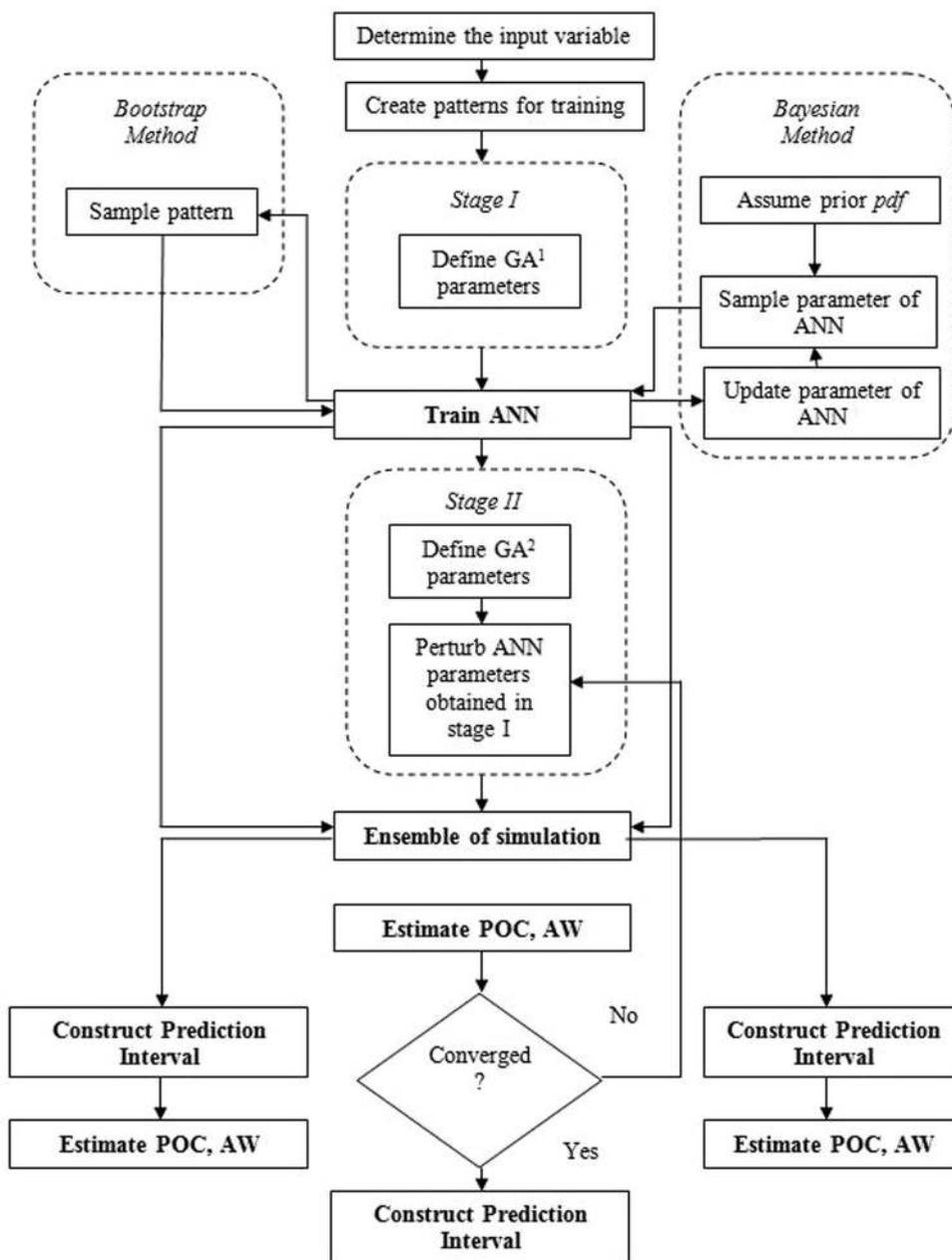
Bayesian approach

The traditional learning of ANN employs error minimization function which attempts to find a set of deterministically optimized weights. In contrast, the Bayesian learning involves training the neural network for the distribution of weights. This is carried out using Bayes' theorem which optimizes weights (i.e. posterior distribution) from the assumed prior distribution. The posterior distribution is then used to evaluate the predictive distribution of network outputs. According to Bayes' rule, the posterior probability distribution of parameters

of ANN model 'M' given the input–output pattern (X, Y) is,

$$p(\theta|X, Y, M) = \frac{p(X, Y|\theta, M)p(\theta|, M)}{p(X, Y|M)} \qquad (2)$$

where $p(X, Y|M)$ is a normalization factor which ensures the total probability is one. M denotes the model with specified connection weights for selected network architecture.

$p(X, Y|\theta, M)$ represents likelihood of the parameter θ, it is assumed that the model residuals follow Gaussian distribution and this can be written as,

$$p(X, Y | \theta, M) \propto \exp\left(-\frac{\beta |Y - f(X, \theta)|^2}{2} \right) \qquad (3)$$

$p(\theta|, M)$ is the prior probability distribution of parameter θ. This is assumed to follow Gaussian distribution and it is written as,

$$P(w) \propto \exp\left(-\frac{\alpha |w|^2}{2} \right) \qquad (4)$$

where α, β are called hyper parameters of distribution which follows Inverse-gamma distribution. These values are updated using Bayes's theorem given the input–output patterns. The model prediction is integration of posterior distribution of weight vectors given the data and is represented as

$$E[Y_{n+1}] = \int f(X_{n+1}, \theta) p(\theta|(X, Y)) d\theta \qquad (5)$$

Solving the above integral Eq. (5) analytically is computationally a challenging task. Therefore, it requires suitable sampling techniques to numerically solve. This study used Marcov Chain Monte Carlo (MCMC) algorithm to sample the parameters through initial and actual sampling phase (Neal 1996). During initial sampling phase, only the parameters of ANN were updated, however the hyper parameters were fixed at certain values. This prevents taking biased values of hyper parameter before ANN parameter reaches reasonable values. Once these starting values are fixed, actual sampling phase determines the values of hyper parameters. This progressively changes the shape of distribution and leads to posterior convergence of ANN parameters. In such way, many combinations of finally converged parameters from posterior distribution were stored and that were used to predict the variable of interest for the given input with quantified prediction interval.

Prediction interval (PI) method

Since neural network calibrates its parameters based on parallel computing, quantification of uncertainty along with calibration is a difficult task, plausibly due to the complexity in computations. Therefore the quantification of uncertainty is generally carried out after the model calibration. However in this method, the prediction interval of the ANN model outputs was constructed during calibration itself with a consideration of generating ensemble of predictions (Kasiviswanathan et al. 2013). A two stage optimization procedure is envisaged for constructing the prediction interval of ANN model outputs. During first stage of optimization, the optimal weights of an ANN were obtained. In the second stage, optimal variability of these

weights were identified that help generate ensembles with minimum residual variance for the ensemble mean, while ensuring a maximum of the measured data to fall within the estimated prediction interval, whose width also is minimized simultaneously. A genetic algorithm based optimization method was applied to generate the desired solution during stage I and II (Goldberg 1989). In which, the number of iteration was fixed as 1000.

In theory, if the width of prediction band is wider, it covers most of the observed values. However, in order to include more observed values in the prediction band, compromising on the width of the prediction band is not desirable. Since these measures are conflicting, a desired solution is to have maximum coverage with a narrow prediction band. Therefore, the second stage optimization is formulated as a multi-objective optimization problem that considers both these measures of uncertainty. The two uncertainty indices i.e. percentage of coverage (POC) and average width (AW) are defined as follows.

$$POC = \left(\frac{1}{n} \sum_{i=1}^{n} c_i \right) \times 100 \qquad (6)$$

$$AW = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{y}_i^U - \hat{y}_i^L \right] \qquad (7)$$

where n is the total number of patterns used for constructing the prediction interval; $\hat{y}_i^U$ and $\hat{y}_i^L$ are the upper and lower bound estimation of the ith pattern; $c_i = 1$ if the observed values of target fall in the prediction band $\left[\hat{y}_i^U, \quad \hat{y}_i^L \right]$, otherwise $c_i = 0$. In addition to that, the mean square error was used to preserve the shape of the hydrograph as well better convergence.

Suppose, if K individual networks derived, and ith pattern has a predicted value $\hat{y}_i^k$ obtained from the kth network ($k = 1, 2, \ldots K$) and y_o is the observed flow value. The ensemble average is considered to estimate the mean square error with observed values and it is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left[y_i^o - \frac{1}{K} \sum_{k=1}^{K} \hat{y}_i^k \right] \quad i = 1, 2, \ldots \ldots n \qquad (8)$$

Model development

It can be seen from Fig. 1 that the prediction interval method quantifies the uncertainty during calibration of ANN, where as other two method estimates after the calibration. The initial step was to determine the significant input variables for the ANN modeling (Sudheer et al. 2002). Generally some degree of a priori knowledge is required to determine the initial set of candidate inputs (e.g. Campolo et al. 1999; Thirumalaiah and Deo 2000).

However, the relationship between the variables is not clearly known a priori, and hence often an analytical technique, such as cross-correlation, is employed (*e.g.* Sajikumar and Thandaveswara 1999; Sudheer et al. 2002). The major disadvantage associated with using cross-correlation is that it is only able to detect linear dependence between two variables, while the modeled relationship may be highly nonlinear. Nonetheless, the cross-correlation methods represent the most popular analytical techniques for selecting appropriate inputs.

The following inputs were identified and patterns were created based on the methodology suggested by Sudheer et al. (2002): [R(t-9), R(t-8), R(t-7), Q(t-2), Q(t-1)], where R(t) represents the rainfall, Q(t) represents the runoff at time period 't'. The output of the network was considered as Q(t). It may be noted that the present study considers one step lead forecast for demonstrating the potential of methods used for quantifying the prediction interval. However, one can extend for the long lead time forecast in order to analyze the uncertainty in models. The short lead forecast should be accurate with less uncertainty and mainly used for operating the hydraulic structures under various flooding conditions. A total of 6525 patterns (input–output pairs) were created from the available data of three years for the study. Further, the data was split into training (5500 sets) and validation (1025 sets) data sets for the analysis. A single hidden layer was considered based on various research studies conducted on this basin (Nayak et al. 2005; Chetan and Sudheer 2006). As mentioned, the number of hidden neurons in the network, which is responsible for capturing the dynamic and complex relationship between various input and output variables, was identified by various trials. The trial and error procedure started with two hidden neurons initially, and the number of hidden neurons was increased up to 10 with a step size of 1 in each trial. The optimal number of hidden neurons was found to be 3 after trial and error. The ANN network structure used in this study is shown in Fig. 2. The following nomenclature is used to represent the links: for the connection between input nodes and hidden nodes, the link connecting I_1 (input 1) to H_1 (hidden node 1) is WI_1H_1, I_1 to H_2 is WI_1H_2, I_1 to H_3 is WI_1H_3 and so on; links connecting hidden nodes to the output node are designated as WH_1O_1 (H_1 to O_1), WH_2O_1 (H_2 to O_1), and WH_3O_1 (H_3 to O_1); the bias connection are represented as BH_1 (at H_1), BO_1 (at O_1).

A sigmoid function was used as the activation function in the hidden layer and linear function in output layer. As the sigmoid function was used in the model, the input–output variables have been scaled appropriately to fall within the function limits using the range of the data. The training of the ANN was carried out using genetic algorithm. For each set of hidden neurons, the network was
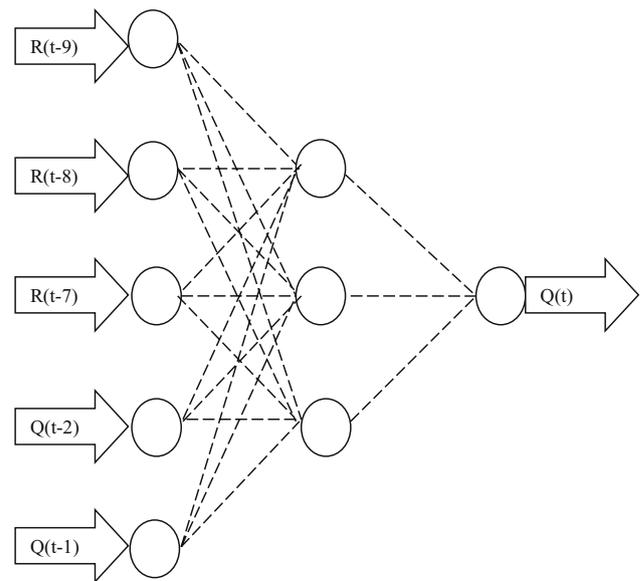

Fig. 2 The final ANN architecture identified for Kolar Basin

trained in batch mode (offline learning) to minimize the mean square error at the output layer. In order to check any over-fitting during training, a cross validation was performed by keeping track of the efficiency of the fitted model. The training was stopped when there was no significant improvement in the model efficiency, and the model was then tested for its generalization properties. The parsimonious structure that resulted in minimum error and maximum efficiency during training as well as validation was selected as the final form of the ANN model.

The total number of ensemble of simulation for all the methods was fixed as 100 as suggested by (Tiwari and Adamowski 2013) so as to maintain the uniformity while comparing the model performance and uncertainty. In the case of bootstrap method, 4500 patterns were randomly bootstrapped out of 5500 patterns available for training to create ensemble of simulation. A total of 100 sets of parameter were obtained through bootstrap sampling, which leads to 100 ensemble of simulation. The ensemble of simulation was then used to establish prediction interval around the observed value and hence the uncertainty was estimated using the measures POC and AW (Fig. 1). In the case of Bayesian approach, the parameters of ANN were optimized in the form of probability distribution function (*pdf*). The prior *pdf* of parameters was assumed to be uniform distribution and the samples of parameters were generated to train ANN model. The distribution was then updated using likelihood values of model under Bayes theorem. The Markov chain Monte Carlo algorithm was used to generate parameter samples repeatedly until it converges. Through this procedure, the ensemble of

simulation was made which leads to predictive uncertainty and which was quantified using POC and AW (Fig. 1).

In the case of prediction interval method, the first stage optimization helps in identifying optimal set of deterministic parameters using Genetic algorithm. During the second stage, the optimal weights obtained during Stage I were perturbed to generate ensemble of simulation. As mentioned, the multi-objective function is formulated with an objective of generating ensemble of simulation which has less uncertainty along with accurate prediction (Fig. 1).

Study area and data used

The presented method is demonstrated through a case study of Indian Basin for flood forecasting. Hourly rainfall and runoff data is collected during monsoon season (July, August, and September) for 3 years (1987–1989). Note that areal average values of rainfall data for three rain gauge stations have been used in the study. The basin has a complex topography which makes the non-linear response of hydrology processes, hence suitable for demonstrating the presented methods. Some of the previous studies have been reported for developing various modeling procedures using the data from this basin (Nayak et al. 2005; Chetan and Sudheer 2006; Srivastav et al. 2007; Kasiviswanathan et al. 2013).

The Kolar River is a tributary of the river Narmada that drains an area about 1350 km² before its confluence with Narmada near Neelkant. An index map of the watersheds is presented in Fig. 3. In the present study the catchment area up to the Satrana gauging site is considered, which constitutes an area of 903.87 km². The 75.3 km long river course lies between north latitude 22°09′–23°17′ and east longitude 77°01′–77°29′.

Results and discussion

Evaluation of model performance

The performance of the ANN model developed using different methods was evaluated with the statistical measures such as such as Correlation Coefficient (CC), Nash–Sutcliffe efficiency (NSE), Root mean square error (RMSE) and Mean biased error (MBE). The summary statistics of model performance is presented in Table 1. It may be noted that listed performance indices values were estimated for the ensemble mean generated from 100 simulations of respective methods.

It can be seen from Table 1 that the bootstrap method produced high amount of RMSE and MBE when compared to the other two methods. This indicates an inferior performance of bootstrap method. The reason for such poor performance may be attributed to random sampling of input–output patterns used for training without maintaining the similar statistical characteristics between the bootstrapped samples. The Bayesian and PI
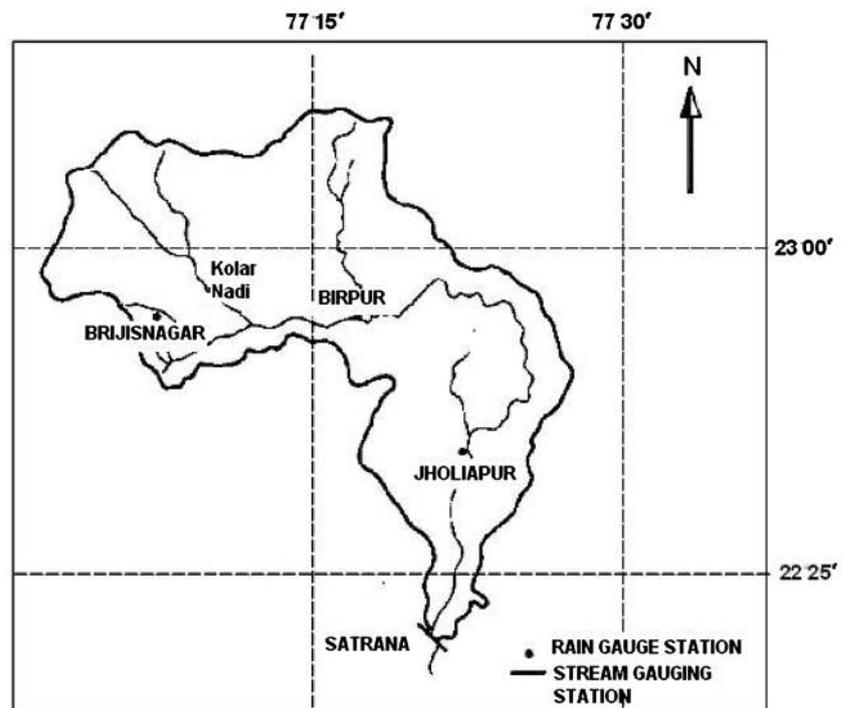
Fig. 3 Map of the study area (Kolar Basin)

Table 1 Comparison of model performance by different method

Category	Calibration					Validation			
Statistical indices	CC	NSE (%)	RMSE (m^3/s)	MBE (m^3/s)	CC	NSE (%)	RMSE (m^3/s)	MBE (m^3/s)
Bootstrap method	0.96	92.62	50.64	5.77	0.96	91.00	56.64	3.05
Bayesian method	0.99	97.04	32.02	0.98	0.99	98.27	24.78	0.79
Prediction interval method	0.99	97.13	31.55	0.33	0.99	98.64	23.04	−0.83

method produced a comparable performance during calibration and validation periods. Further, the values of CC and NSE during calibration and validation did not show any significant change, which supports the consistent performance of these methods. Consequently, the Bayesian method has an NSE of 97.04 % in calibration and 98.27 % in validation; the PI method has NSE of 97.13 % in calibration and 98.64 % in validation. The positive values of MBE in calibration and validation shows a consistent underestimation of model in the case of Bootstrap and Bayesian approach. However, the PI method produced a negative value of MBE in validation, which shows slight over estimation of model. Overall, the ensemble mean obtained through PI method closely matches the observed values with comparable performance than other two methods.

Parameter uncertainty

In this section, the parameter uncertainty estimated is compared between the methods. It is known that standard deviation is a statistical measure which can be used for assessing the variability or amount of uncertainty of the parameters. The standard deviation of each parameter of ANN model developed through different methods is presented in Table 2. It was found that the parameter uncertainty in Bootstrap method shows considerably high variability than other two methods. This might be due to model bias towards selection of patterns for training.

Further, it is clear that except few, most of the parameters in PI method have less values of standard deviation, which indicates minimum uncertainty as compared to other methods. Figure 4 illustrates the range of ANN parameters estimated using different methods. Overall, the results indicate that the bootstrap method has consistently high variability and other two methods have low variability (Fig. 4). The reason for less variability of parameter in PI method could be the objective function formulated in such way that tries to minimize the uncertainty through the best possible combination of weights and biases.

Table 2 Standard deviation of ANNs' 100 parameter sets obtained through different methods

Parameter	Bootstrap	Bayesian	PI
WI_1H_1	0.8434	0.0205	**0.0015**
WI_1H_2	0.7735	0.0549	**0.0013**
WI_1H_3	0.6531	0.0812	**0.0061**
WI_2H_1	0.8288	**0.0203**	0.0454
WI_2H_2	0.6895	0.0540	**0.0026**
WI_2H_3	0.8573	0.0432	**0.0279**
WI_3H_1	0.9242	**0.0218**	0.0423
WI_3H_2	0.6771	0.0568	**0.0012**
WI_3H_3	0.8016	0.1816	**0.0194**
WI_4H_1	1.0415	**0.1265**	0.1584
WI_4H_2	0.8423	0.2097	**0.0136**
WI_4H_3	0.8714	0.3908	**0.2870**
WI_5H_1	1.0852	**0.0972**	0.5259
WI_5H_2	0.9529	0.2443	**0.1280**
WI_5H_3	0.9542	0.4549	**0.2543**
BH_1	1.4880	0.0332	**0.0131**
BH_2	1.4265	0.0413	**0.1441**
BH_3	1.4175	0.2657	**0.2550**
WH_1O_1	0.6843	0.1073	**0.0715**
WH_2O_1	0.7063	0.0585	**0.0152**
WH_3O_1	0.7425	**0.1507**	0.4846
BO_1	0.5628	0.1578	**0.0117**

The nomenclature of the parameter is explained in the text

The parameter with less uncertainty is denoted as bold

Prediction uncertainty

The prediction uncertainty of each method estimated in terms of AW and POC is presented in Table 3. It may be noted that the presented results correspond to validation data only, however a consistent performance was obtained in calibration. In order to assess the uncertainty on different magnitude of flow, the flow values were statistically categorized as low, medium and high (Nayak et al. 2005). Out of 1025 patterns in validation, low flow values contain 843 patterns and 167, 15 patterns fall in medium, high flow, respectively. A better model will have less AW with more number of observed

Fig. 4 Parameter uncertainty
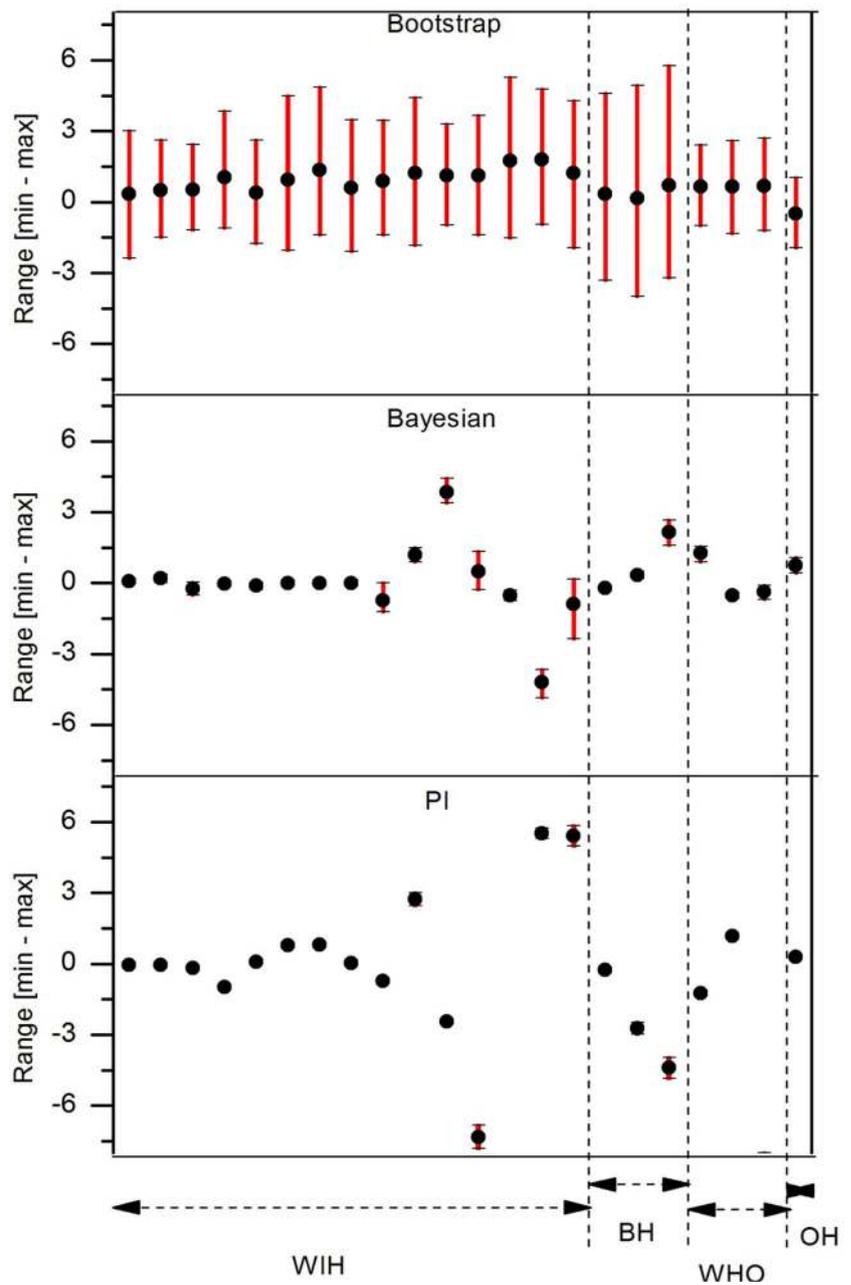quantified from the presented
methods

**Table 3** Comparison of predictive uncertainty by different method

Category	Complete flow Series		Low flow ($x < \mu$)		Medium flow ($\mu \leq x \leq + 2\sigma$)		High flow ($x > \mu + 2\sigma$)	
No of patterns	1025		843		167		15	
Statistical indices	AW (m^3/s)	POC (%)	AW (m^3/s)	POC (%)	AW (m^3/s)	POC (%)	AW (m^3/s)	POC (%)
Bootstrap method	75.95	98.14	38.06	99.88	198.63	95.81	839.56	20
Bayesian approach	3.42	46.58	0.44	49.46	6.78	33.53	133.49	26.67
Prediction interval	26.49	97.17	16.50	99.17	60.90	92.22	204.54	40

values falls over the prediction band (i.e. maximum POC). In other words, quantitatively an ideal prediction band will have POC of 100 % with AW approaches to low values. The prediction interval obtained through different method has varying magnitude of uncertainty in terms of POC and AW estimate (Fig. 5). In general, the Bootstrap method has high AW values across different flow domains as illustrated in Fig. 5, which indicates increased bias in model calibration resulted from sampling of input–output patterns. Consequently the ensemble simulation leads to high variability from the mean prediction. While comparing, different methods, the bootstrap method has high POC across different flow series. This is mainly due to increased width of prediction interval, which obviously will include more number of observed values. The minimum AW was found in ANN models that used Bayesian based model training. It was found that the Bayesian approach produced consistently less AW value across complete, low and medium flow series with 3.42, 0.44, and 6.78 m^3/s, respectively. The forgoing discussion clearly indicates that the Bayesian method is good in terms of less AW and bootstrap method is good with better estimate of maximum POC. However, the selection of particular method should not be biased by considering only either POC or AW as these measures conflict each others. In this regard, it was identified that PI method has better estimate of uncertainty with compromised values of AW and POC. Consequently PI method has acceptable values of POC such as 97.17, 99.17, 92.22 and 40 % in complete, low, medium and high flow, respectively. The AW values of PI in complete, low, medium and high flow were 26.49, 16.50, 60.90 and 204.54 m^3/s, respectively. Though the AW values were slightly higher than Bayesian method across different flow series, the method can be considered as a better estimate of uncertainty with improved confidence as it contains more number of observed values within the prediction interval (Fig. 5).

Uncertainty in peak flow prediction

Accurate estimation of peak flow helps in better decision making with improved reliability for developing the systems of flood warnings and flood protection measures.
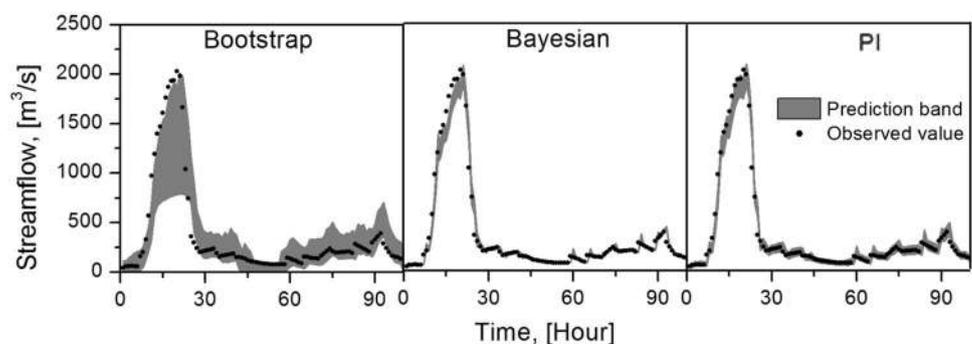
However in most of the cases, it was reported that the poor peak flow estimation is a general concern in ANN (Srivastav et al. 2007). One of the major reasons for such behavior is less number of peak flow patterns available for training, and that leads to more bias in predicting the values (mostly under prediction). The histogram (Fig. 6) shows the peak flow prediction of ensemble by selected methods. It is clear that none of the method predicts the peak reasonably, except PI method. The potential reason for such under prediction of other two methods could be the increased level of parameter uncertainty, where the model is unable to capture the peak responses of hydrograph. It is evident from Fig. 6 that the actual peak is 2029.18 m^3/s. However, most of the ensembles of bootstrap were able to predict the peak flow prediction between 1300 and 1900 m^3/s. The potential reason could be inferred that the sampled input–output patterns may not contain sufficient peak information for training. Similar findings were observed in Bayesian approach, where the peak flow prediction falls between 1700 and 1900 m^3/s. In this case, the reason for not predicting peak appropriately may be attributed to the assumption about prior probability distribution of ANN's parameter. However, the peak prediction of PI method was substantially good while comparing other two methods. This reinforces the discussion that performing uncertainty analysis under optimization framework leads to an improved result in all domains of flow with reduction in uncertainty. Still, understanding and modeling of hydrologic processes are far beyond the reality that needs considerable level of attention while quantifying the uncertainty.

Qualitative assessment of uncertainty

Based on the quantitative assessment of uncertainty across different methods, the qualitative classification is made and presented in Table 4. It is categorized as low, medium and high depends on the uncertainty evaluation method such as the formulation, computational burden, assumptions, validity of preserving probabilistic and statistical properties, convergence and accuracy in estimating peak flow. It
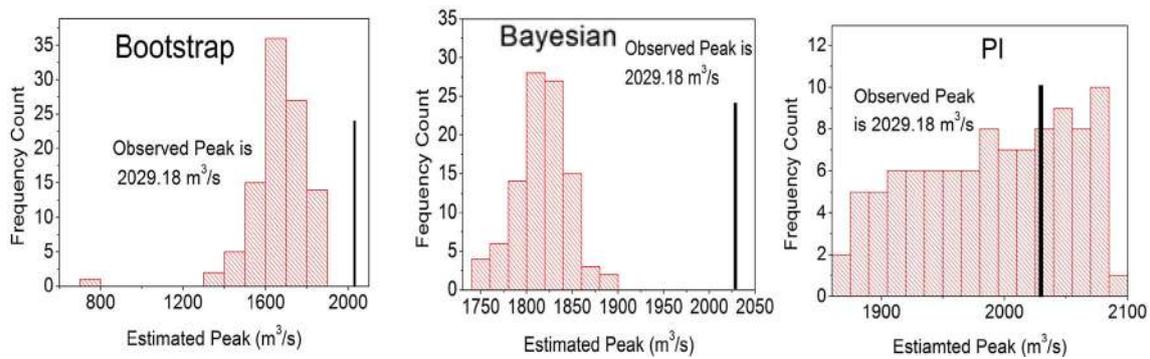
Fig. 5 Prediction interval quantified from the presented methods during model validation

Fig. 6 Histogram of peak flow prediction from the presented methods

Table 4 Qualitative analysis of methods

Method	Bootstrap	Bayesian	Prediction interval
Difficulty of implementation	Medium	High	Low
Computational efficiency	Low	Medium	High
Fulfilment of statistical and probabilistic assumption	Low	High	Medium
Parameter convergence	Low	Medium	High
Peak flow prediction	Low	Medium	High

is evident from Table 4 that the PI method relatively satisfies all the listed conditions besides quantitative estimate. This gives an indication of superiority of the PI method while quantifying the prediction uncertainty.

Summary and conclusions

The meaningful quantification of uncertainty in ANN models improves the reliability while making decisions. However, no clear evidence that show a specific method, which outperforms while estimating the uncertainty of hydrologic models compared to other methods. Hence, each method differs by its own complexity, principle and computational efficiency. This study presented three different uncertainty estimation method applied in ANN models. The whole modeling procedure is demonstrated through flow forecast model using rainfall—runoff data collected from Kolar basin, India. The quantitative and qualitative comparison was made between the methods in order to show their potential improvements and shortcomings. The parameter uncertainty resulted from PI method showed significant reduction. Consequently, the PI method produced improved model performance, with narrow prediction interval. In addition, the values of POC across different flow domains further enhance the confidence of the PI method in terms of uncertainty. Furthermore, the accurate estimation of peak flow is a general concern in ANN models, however the peak estimation of PI method encourages its potential application in flood forecasting.

References

Abrahart RJ, See LM, Dawson CW, Shamseldin AY, Wilby RL (2010) Nearly two decades of neural network hydrologic modeling. In: Sivakumar B, Berndtsson R (eds) Advances in data-based approaches for hydrologic modeling and forecasting. World Scientific Publishing, Hackensack, pp 267–346

Alvisi S, Franchini M (2011) Fuzzy neural networks for water level and discharge forecasting with uncertainty. Environ Model Softw 26(4):523–537

Boucher MA, Laliber\'e JP, Anctil F (2010) An experiment on the evolution of an ensemble of neural networks for streamflow forecasting. Hydrol Earth Syst Sci 14:603–612

Campolo M, Soldati A, Andreussi P (1999) Forecasting river flow rate during low flow periods using neural networks. Water Resour Res 35(11):3547–3552

Chetan M, Sudheer KP (2006) A hybrid linear-neural model for river flow forecasting. Water Resour Res 42:W04402. doi:10.1029/2005WR004072

Elshorbagy A, Corzo G, Srinivasulu S, Solomatine DP (2010a) Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: concepts and methodology. Hydrol Earth Syst Sci 14:1931–1941

Elshorbagy A, Corzo G, Srinivasulu S, Solomatine DP (2010b) Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology—Part 2: application. Hydrol Earth Syst Sci 14:1943–1961

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading

Han DT, Kwong, Li S (2007) Uncertainties in real-time flood forecasting with neural networks. Hydrol Process 21(2):223–228. doi:10.1002/hyp.6184

Hsu KI, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall-runoff process. Water Resour Res 31:2517–2530

Kasiviswanathan KS, Cibin R, Sudheer KP, Chaubey I (2013) Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. J Hydrol 499:275–288. doi:10.1016/j.jhydrol.2013.06.043

Nayak PC, Sudheer KP, Rangan DM, Ramasastri KS (2005) Short-term flood forecasting with a neurofuzzy model. Water Resour Res 41:W04004. doi:10.1029/2004WR003562

Neal RM (1996) Bayesian learning for neural networks, lecture notes in statistics no. 118. Springer, New York

Sajikumar N, Thandaveswara BS (1999) A non-linear rainfall-runoff model using an artificial neural network. J Hydrol 216:32–35

Schoups G, Vrugt JA (2010) A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resour Res. doi:10.1029/2009WR008933

Sharma N, Zakaullah Md, Tiwari H, Kumar D (2015) Runoff and sediment yield modeling using ANN and support vector machines: a case study from Nepal watershed. Model Earth Syst Environ 1:23. doi:10.1007/s40808-015-0027-0

Shrestha DL, Solomatine DP (2006) Machine learning approaches for estimation of prediction interval for the model output. Neural Netw 19:225–235

Srivastav RK, Sudheer KP, Chaubey I (2007) A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. Water Resour Res 43:W10407. doi:10.1029/2006WR005352

Sudheer KP, Gosain AK, Ramasastri KS (2002) A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrol. Process 16:1325–1330

Thirumalaiah K, Deo MC (2000) Hydrological forecasting using neural networks. J Hydrol Eng 5(2):180–189

Tiwari MK, Adamowski J (2013) Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour Res 49:6486–6507. doi:10.1002/wrcr.20517

Zhang X, Liang F, Srinivasan R, Van Liew M (2009) Estimating uncertainty of streamflow simulation using Bayesian neural networks. Water Resour Res 45:W02403. doi:10.1029/2008WR007030

Zhang X, Liang F, Yu B, Zong Z (2011) Explicitly integrating parameter, input, and structure uncertainties into Bayesian Neural Networks for probabilistic hydrologic forecasting. J Hydrol 409:696–709. doi:10.1016/j.jhydrol.2011.09.002