A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting
Mukesh K. Tiwari and Chandranath Chatterjee

ABSTRACT

A new hybrid model, the wavelet–bootstrap–ANN (WBANN), for daily discharge forecasting is proposed in this study. The study explores the potential of wavelet and bootstrapping techniques to develop an accurate and reliable ANN model. The performance of the WBANN model is also compared with three more models: traditional ANN, wavelet-based ANN (WANN) and bootstrap-based ANN (BANN). Input vectors are decomposed into discrete wavelet components (DWCs) using discrete wavelet transformation (DWT) and then appropriate DWCs sub-series are used as inputs to the ANN model to develop the WANN model. The BANN model is an ensemble of several ANNs built using bootstrap resamples of raw datasets, whereas the WBANN model is an ensemble of several ANNs built using bootstrap resamples of DWCs instead of raw datasets. The results showed that the hybrid models WBANN and WANN produced significantly better results than the traditional ANN and BANN, whereas the BANN model is found to be more reliable and consistent. The WBANN and WANN models simulated the peak discharges better than the ANN and BANN models, whereas the overall performance of WBANN, which uses the capabilities of both bootstrap and wavelet techniques, is found to be more accurate and reliable than the remaining three models.

Key words | bootstrap, discharge, forecasting, resampling, wavelet

INTRODUCTION

Daily discharge forecasting is desirable for water resource planning and management. A wide variety of rainfall–runoff models have been developed and applied for discharge forecasting. Most of the rainfall–runoff models have been developed either based on a mechanistic approach or on a systems theoretic approach. Hydrological prediction usually relies on incomplete and uncertain process descriptions that have been deduced from sparse and noisy datasets. Spatially distributed modeling is a typical example of the mechanistic approach to construct a model that explicitly accounts for as much of the small-scale physics and the natural heterogeneity as computationally possible (Loague & VanderKwaak 2004). The approach has been criticized for resulting in models that are overly complex, leading to problems of over-parameterization and equifinality (Beven 2006), which may manifest itself in large prediction uncertainty (Uhlenbrook et al. 1999), while in the system theoretic approach the concern is with the system operation, not the nature of the system by itself or the physical laws governing its operation.

Black box models in the form of artificial neural networks (ANNs) have gained momentum in the last few decades for river flow forecasting. The FFBP (Feed Forward Back Propagation) is the most popular ANN training method in water resources literature. The major advantage of the FFBP ANN is that it is less complex than other ANNs such as Radial Basis Function (RBF) and has the same nonlinear input–output mapping capability (Sudheer & Jain 2003; Coulibaly & Evora 2007). Some researchers have successfully applied fuzzy inference systems in river flow forecasting (Nayak et al. 2004, 2005a; Mukerji et al. (2009); Pramanik & Panda...
Jacquin & Shamseldin (2009) reviewed the existing studies dealing with the use of fuzzy inference systems in river flow forecasting. This review shows that fuzzy inference systems can be used as effective tools for river flow forecasting, even though their application is rather limited in comparison to the popularity of neural network models. They found that there are several unresolved issues requiring further attention before more clear guidelines for the application of fuzzy inference systems can be given. Wu & Chau (2006) observed that the hybrid GA-based ANN algorithm, under cautious treatment to avoid over-fitting, is able to produce better accuracy in performance, although at the expense of additional modeling parameters and longer computation time. ANNs are known to have several dozens of successful applications in river basin management and related problems (Solomatine & Ostfeld 2008). The ANNs are still widely applied and are a very popular tool compared to other data-driven techniques in river basin management. Therefore in this study we have used the FFBP ANN model as it has been increasingly used in rainfall–runoff modeling and river discharge forecasting due to their ability to map complex nonlinear rainfall–runoff relationships (Baratti et al. 2003; Cigizoglu 2003a, b; Jain & Srinivasulu 2004; Altunkaynak 2007; Demirel et al. 2009). Substantial literatures on ANN have been reported in ASCE (2000a, b).

The reliability of the hydrological predictions is affected by three major sources of uncertainties (Bates & Townley 1988): data uncertainty (quality and representativeness of data), model structure uncertainty (ability of the model to describe the catchment’s response) and parameter uncertainty (adequate values of model parameters). Han et al. (2007) studied the uncertainties involved in real-time prediction in using an ANN model. It was concluded that, for long-term predictions, the ANN showed superior performance but that was only probabilistic, depending on how the calibration and test events were arranged. Srivastav et al. (2007) proposed a method of uncertainty analysis for ANN hydrological models and showed that the ANN predictions contain a significant amount of uncertainty. Boucher et al. (2009) analyzed one-day-ahead hydrological ensemble forecasts obtained by stacked neural networks and found that ensemble forecasts outperform point forecasts. In order to overcome the limitations inherent in the conventional treatment of uncertainty in ANN model predictions, the recent trend has been to combine the outputs of several member bootstrap ANN models to reduce the uncertainty involved by controlling the generalization of the final predictive model and to produce more reliable and consistent predictions (Cannon & Whitfield 2002; Jeong & Kim 2005).

The bootstrap is a computational procedure that uses intensive resampling with replacement, in order to reduce uncertainty (Efron & Tibshirani 1993). In addition, it is the simplest approach since it does not require complex computations of derivatives and Hessian matrix inversion involved in linear methods or the Monte Carlo solutions of the integrals involved in the Bayesian approach (Dybowski & Roberts 2001). Applications of the bootstrap method range from estimating means, confidence intervals and parameter uncertainties to network design techniques (Lall & Sharma 1996; Sharma et al. 1997; Tasker & Dunne 1997). The bootstrap technique has also been used in artificial neural network model development. Abrahart (2005) employed the bootstrap technique to continuously sample the input space in the context of rainfall–runoff modeling and reported that it offered marginal improvement in terms of greater accuracies and better global generalizations. Antcil & Lauzon (2004) recommended the joint use of stop training or Bayesian regularization with either bagging or boosting for improvement and stability in modeling performance. Jeong & Kim (2005) used ensemble neural networks (ENN) using the bootstrap technique to simulate monthly rainfall–runoff. They concluded that ENN is less sensitive to the input variable selection and the number of hidden nodes than the single neural network (SNN). Jia & Culver (2006) used the bootstrap technique to estimate the generalization errors of neural networks with different structures and to construct the confidence intervals for synthetic flow prediction with a small data sample. Sharma & Tiwari (2009) applied the bootstrap technique for monthly runoff prediction and showed that the BANN model consistently outperformed the ANN models. Tiwari & Chatterjee (2010) applied the bootstrap technique for hourly flood forecasting and showed that the bootstrap technique is capable of quantifying uncertainty in flood forecasts and ensemble predictions are more stable and accurate.

A major criticism of ANN models is their limited ability to account for any physics of the hydrologic processes in a catchment (Aksoy et al. 2007; Koutsoyiannis 2007)).
discharge is widely perceived as being nonlinear and nonstationary (Rao et al. 2003; Wang et al. 2006). Nonstationarity in the data such as trends and seasonal variations influences the simulation of daily discharge greatly and often causes poor predictability in practical applications. The pronounced seasonality is driven by different underlying physical processes of streamflow generation for different periods. For instance, the low flows are mainly sustained by base flows, while high flows are affected by intensive rainfalls.

Wavelet transforms, which provide information in both time and frequency domains of the signal, give considerable information about the physical structure of the data, and wavelet analysis provides a time–frequency representation of a signal at many different periods in the time domain (Daubechies 1990). In spite of the suitable flexibility of ANN in modeling hydrologic time series such as runoff (Hsu et al. 1995; Zhang & Dong 2001) non-stationarity in the signal limits the use of ANN. In such a situation, ANNs will not produce good results with nonstationary data if pre-processing of the input and/or output data is not performed (Cannas et al. 2006). The wavelet transform technique decomposes the original time series data into sub-time series of different resolution levels, and these wavelet-transformed data improve the performance of a forecasting model by capturing different information on various resolution levels. The sub-time series obtained using wavelet decomposition of the signal of different resolution levels provides an interpretation of the series structure and extracts the useful information. This is the reason that this technique is largely applied to time series analysis of nonstationary signals (Nason & Von Sachs 1999). In the recent past, wavelet transforms have become a useful method for analyzing variations, periodicities and trends in time series (Xingang et al. 2005; Yueqing et al. 2004; Partal & Kucuk 2006). Recently, the WANN model has been successfully employed in some hydrology and water resource studies. For example, Wang & Ding (2003) proposed a wavelet network model with a combination of the wavelet transform and the ANN, and decomposed the original time series into periodic components by wavelet transform. Later, sub-time series were used as the inputs for ANNs and the resulting model was applied to forecast the original time series. This approach was used for monthly groundwater level and daily discharge forecasting. Kim & Valdes (2003) presented a hybrid neural network model combined with dyadic wavelet transforms to improve the forecasting of regional droughts. These researchers used the neural network model in two phases. First, neural networks were employed to forecast the signals decomposed by wavelet transform at various resolution levels and then the forecasted decomposed signals were reconstructed into the original time series. The researchers applied this model to the monthly and annual inflow and rainfall data and showed that the model significantly improved the neural network’s forecasting performance. Ancil & Tape (2004) used a wavelet–neural network model for one-day-ahead rainfall–runoff forecasting. The time series was decomposed by wavelet transform into three sub-series: short, intermediate and long wavelet periods. Then, multiple-layer ANN forecasting models were trained for each wavelet-decomposed sub-series, and later forecasted decomposed signals were reconstructed into the original time series. Partal & Cigizoglu (2008) predicted the suspended sediment load in rivers by a combined wavelet–ANN method. Measured data were decomposed into wavelet components via DWT, and the new wavelet series, consisting of the sum of selected wavelet components, was used as input for the ANN model. The wavelet–ANN model provided a good fit to observed data for the testing period. Zhou et al. (2008) developed a wavelet predictor–corrector model for prediction of monthly discharge time series and showed that the model has higher prediction accuracy than ARIMA and seasonal ARIMA. Hence a hybrid ANN–wavelet model which uses multi-scale signals as input data may present more reliable forecasting rather than a single pattern input. Adamowski (2008a, b) developed a new method of standalone short-term river flood forecasting based on wavelet and cross-wavelet analysis. He found that the proposed wavelet-based forecasting method can be used with great accuracy as a standalone forecasting method for short-term river flood forecasting. It was also shown that forecasting models based on wavelet and cross-wavelet constituent components for forecasting river floods were not accurate for longer lead time forecasting with the artificial neural network models providing more accurate results. Kisi (2008) investigated the accuracy of the neuro-wavelet (NW) technique for modeling monthly streamflows. The NW and ANN models were tested by applying them to various input combinations of monthly streamflow data from Gerdelli and Isakoy Stations in the Eastern Black Sea region of Turkey. The test results indicated that the DWT could
increase the accuracy of ANN models in modeling monthly streamflows. The comparison results showed that the NW model performed much better than the ANN, multi-linear regression (MLR) and autoregressive (AR) models. He recommended further studies using more data from different areas to strengthen these conclusions. Kisi (2009) proposed the application of a conjunction model (neurowavelet) for forecasting daily intermittent streamflow. The comparison results revealed that the suggested model could significantly increase the forecast accuracy of single ANN in forecasting daily intermittent streamflows. Partial (2009) employed the wavelet–neural network structure that combines wavelet transform and artificial neural networks to forecast the river flows of Turkey. He studied the performance comparison of generalized neural networks and radial basis neural networks with feed-forward back-propagation methods. Six different models were studied for forecasting of monthly river flows. It was seen that the wavelet and feed-forward back-propagation model was superior to the other models in terms of selected performance criteria. Satyajirao & Krishna (2009) proposed a wavelet–neural network (WNN) model, based on a combination of wavelet analysis and ANN, in order to improve the precision of daily streamflow and monthly groundwater level forecasts where there is a scarcity of other hydrological time series data. The results of daily streamflow and monthly groundwater level series modeling indicated that the performances of WNN models are more effective than the ANN models. The study also highlights the capability of WNN models in estimating low and high values in the hydrological time series data. Wang & Meng (2007) analyzed characteristics, abruptions, tendencies and causes of annual runoff variations in the Heihe River drainage basin by the wavelet–neural network model, and detected the basic variation characteristics and tendencies of the Heihe River surface hydrological system. Wang et al. (2009) developed a new hybrid model, the wavelet–network model (WNM), based on reasonable combination of wavelet analysis with ANN. The results from a given case study showed that the WNM has some significant advantages for short- and long-term prediction of runoff in hydrology. The WNM was compared with the threshold auto-regressive model (TAR). The accuracy of prediction using the WNM was better than the TAR. They recommended further researches on investigating the method for reasonable selection of resolution level, which is an important parameter in WNM. Partial & Cigizoglu (2009) predict the daily precipitation from meteorological data from Turkey using the wavelet–neural network method. The new approach in estimating the peak values showed a noticeably high positive effect on the performance evaluation criteria.

To the best of our knowledge, no study has been reported in the hydrologic literature that has used the combined strength of wavelet and bootstrap-based ANNs for hydrological modeling. The present study is the first application where ANN models are coupled with DWT and the bootstrap technique to utilize the individual strength of each approach. An attempt is made to forecast river discharge for 1 to 5 days using the novel approach that uses wavelet and bootstrap-based ANNs in a large river basin where flood forecasting and water resources planning are critical issues.

**METHODOLOGY**

**Artificial neural networks**

Artificial neural networks are information processing systems composed of simple processing elements (nodes) linked by weighted synaptic connections (Muller & Reinhardt 1991). Neural network models are developed by training the network to represent the relationships and processes that are inherent within the data. Being essentially nonlinear regression models, they perform an input–output mapping using a set of interconnected simple processing nodes or neurons. They reconstruct the complex nonlinear input/output relations by combining multiple simple functions, by analogy with the functioning of the human brain. This approach is fast and robust in noisy environments, flexible in the range of problems it can solve, and highly adaptive to newer environments. Owing to these established advantages, ANNs currently have numerous real-world applications, such as time series prediction, rule-based control and rainfall–runoff modeling (Jain et al. 1999). The multilayer feed-forward neural network consists of a set of sensory units that constitute the input layer, one or more hidden layers of computation nodes and an output layer of computation nodes. The input signal propagates through the network in a forward direction, layer by layer. These neural networks are commonly
referred to as multilayer perceptrons. A detailed explanation of different properties of ANN are beyond the scope of this paper. Interested readers are directed to refer to texts such as Bishop (1995) and Haykin (1999) for discussion on the general properties of ANN and Maier & Dandy (2000) for an overview of different applications of ANN in water resources.

The wavelet analysis

Wavelet analysis is a multiresolution analysis in the time and frequency domains and is the important derivative of the Fourier transform. The wavelet function \( \psi(t) \), called the mother wavelet, has finite energy and is mathematically defined as \( \int_{-\infty}^{\infty} \psi(t) dt = 0 \).

\[
\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right)
\]

(1)

where \( a \) and \( b \) are real numbers, \( \psi_{a,b}(t) = \) successive wavelet, \( a = \) scale or frequency factor and \( b = \) time factor. Thus, the wavelet transform is a function of two variables, \( a \) and \( b \). The parameter \( a \) can be interpreted as a dilation \( a > 1 \) or contraction \( a < 1 \) factor of the wavelet function \( \psi(t) \) corresponding to different scales of observation. The parameter \( b \) can be interpreted as a temporal translation or shift of the function \( \psi(t) \).

For the time series \( f(t) \in \mathbb{L}^2(\mathbb{R}) \) or finite energy signal (Rosso et al. 2004) the continuous wavelet transform of the time series \( f(t) \) is defined as

\[
W_f(a,b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi^\ast\left(\frac{t-b}{a}\right) dt
\]

(2)

where \( W_f(a,b) \) is the wavelet coefficient and \( \psi^\ast \) corresponds to the complex conjugate.

The wavelet transform searches for correlations between the signal and wavelet function. This calculation is done at different scales of \( a \) and locally around the time of \( b \). The result is a wavelet coefficient \( W_f(a,b) \) contour map known as a scalogram. However, computing the wavelet coefficients at every possible scale (resolution level) causes the generation of a large amount of data. If one chooses the scales and the positions based on the powers of two (dyadic scales and positions), then the analysis will be much more efficient as well as more accurate. This transform is called the discrete wavelet transform (DWT) and can be defined as (Mallat 1989)

\[
\psi_{m,n}(\frac{t-b}{a}) = a^{-m/2} \psi^\ast\left(\frac{t-nb_0d^n}{a_0^n}\right)
\]

(3)

where \( m \) and \( n \) are integers that control the wavelet dilation and translation, respectively, \( a_0 \) is a specified fined dilation step greater than 1 and \( b_0 \) is the location parameter which must be greater than zero. The most common and simplest choices for parameters are \( a_0 = 2 \) and \( b_0 = 1 \).

This power-of-two logarithmic scaling of the translations and dilations is known as a dyadic grid arrangement and is the simplest and most efficient case for practical purposes (Mallat 1989). For a discrete time series \( f(t) \), when it occurs at a different time \( t \) (i.e. here integer time steps are used), the discrete wavelet transform becomes

\[
W_f(m,n) = 2^{-m/2} \sum_{i=0}^{N-1} f(t) \psi^\ast(2^{-m}i-n)
\]

(4)

where \( W_f(m,n) \) is the wavelet coefficient for the discrete wavelet of scale \( a = 2^m \) and location \( b = 2^m n \). \( f(t) \) is a finite time series \( t = 0, 1, 2, \ldots, N-1 \), \( N \) is an integer power of 2 \( N = 2^M \) and \( n \) is the time translation parameter, which changes in the range \( 0 < n < 2^M - 1 \), where \( 1 < m < M \). At the largest wavelet scale (i.e. \( 2^m \) where \( m = M \)) only one wavelet is required to cover the time interval and only one coefficient is produced. At the next scale \( 2^{m-1} \), two wavelets cover the time interval, hence two coefficients are produced, and so on down to \( m = 1 \). At \( m = 1 \), the scale is \( 2^1 \), i.e. \( 2^{-1} \) or \( N/2 \) coefficients are required to describe the signal at this scale. The total number of wavelet coefficients for a discrete time series of length \( N = 2^M \) is then \( 1 + 2 + 4 + 8 + \ldots + 2M - 1 = N - 1 \).

In addition to this, a signal smoothed component, \( \overline{W} \), is left, which is the signal mean. Thus, a time series of length \( N \) is broken into \( N \) components, i.e. with zero redundancy. The inverse discrete transform is given by

\[
f(t) = \overline{W} + \sum_{m=1}^{M} \sum_{n=0}^{2^M-1} W_f(m,n)2^{-\frac{m}{2}} \psi^\ast(2^{-m}i-n)
\]

(5)

or in a simple format as

\[
f(t) = \overline{W}(t) + \sum_{m=1}^{M} W_m(t)
\]

(6)
where $\mathcal{W}(t)$ is called the approximation sub-signal at level $M$ and $W_m(t)$ are detailed sub-signals at levels $m = 1, 2, \ldots, M$.

DWT operates two sets of functions viewed as high-pass and low-pass filters (Figure 1). The original time series is passed through high-pass and low-pass filters and separated at different scales. The time series is decomposed into one comprising low frequencies or its trend (the approximation) and one comprising the high frequencies or the fast events (the detail). The wavelet coefficients, $W_m(t)$ ($m = 1, 2, \ldots, M$), provide the detail signals, which can capture small features of interpretational value in the data; the residual term, $\mathcal{W}(t)$, represents the background information (approximation) of data. Because of the simplicity of $W_1(t)$, $W_2(t)$; $\ldots$, $W_M(t)$, $\mathcal{W}(t)$, some interesting characteristics, such as period, hidden period, dependence and jump can be diagnosed easily through these discrete wavelet components (DWCs).

**Bootstrapped artificial neural networks (BANNs)**

The bootstrap is a data-driven simulation method that uses intensive resampling, with replacement, to reduce uncertainties (Efron 1979; Efron & Tibishirani, 1993). The technique is based on resampling with replacement of the available dataset and training an individual network on each resampled instance of the original dataset. The bootstrap method can be used to expand upon a single realization of a distribution or process to create a set of bootstrap samples that can provide a better understanding of the average and variability of the original unknown distribution or process.

Assume that the data consists of a random sample $T_n = \{ (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \}$ of size $n$ drawn from a population of unknown probability distribution $F$, where $f_t(x_i, y_i)$ is a realization drawn independently and identically distributed (i.i.d.) from $F$ and consists of a predictor vector $x_i$ and the corresponding output variable $y_i$. Let $\hat{F}$ be the empirical distribution function for $T_n$ with mass $1/n$ on $t_1, t_2, \ldots, t_n$; and let $T^*$ be a random sample of size $n$ taken from i.i.d. with replacement from $\hat{F}$. The set of $B$ bootstrap samples can be represented as $T_1^*, T_2^*, \ldots, T_B^*$, in which $B$ is the total number of bootstrap samples and ranges usually from 50–200 (Efron 1979). For each $T^*_b$, an ANN prediction model is constructed and the output is represented as $\hat{f}_{ANN}(x_i, w_b/T^*_b)$, built using all $n$ observations. The performance of the trained ANN model is evaluated using the observation pairs that are not included in a bootstrap sample and the average performance of these ANNs on their corresponding validation sets is used as an estimate of the generalization error of the ANN model developed on $T_n$. The generalization error of an ANN model can be estimated by its ‘$E_0$’ estimate (Twomey & Smith 1998):

$$E_0 = \frac{\sum_{b=1}^{B} \sum_{i \in A_b} (y_i - \hat{f}_{ANN}(x_i, w_b/T^*_b))^2}{\sum_{b=1}^{B} |A_b|}$$  (7)

The output of the ANN developed based on the bootstrap sample $T^*_b$ is represented as $\hat{f}_{ANN}(x, w_b/T^*_b)$, where $x$ is a particular input vector, $w_b$ is the weight vector, $A_b$ is the set of indices for the observation pairs not included in the bootstrap sample $T^*_b$ and $|A_b|$ is the number of observation pair indices in $A_b$. The BANN estimate $\hat{\theta}(x)$ is given by the average of the $B$ bootstrapped estimates:

$$\hat{\theta}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_{ANN}(x, w_b/T^*_b)$$  (8)

**Multiple linear regression (MLR)**

Multiple linear regression attempts to model the relationship between two or more independent variables and a dependent variable by fitting a linear equation to the data points. A multiple linear regression equation takes the following form:

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$  (9)

where $y$ is the dependent variable, $a$ is a constant and $b_1$ to $b_n$ are multipliers for $x_1$ to $x_n$ independent variables. Constant and multipliers are estimated through minimizing the sums of
squares of deviations between each data point and the regression line. MLR has been the traditional approach utilized in water resources hydrology for several decades since the last century. Some recent applications appear in Leclerc & Ouarda (2007) and Sahoo et al. (2009).

**Performance indices**

Dawson et al. (2007) discussed 20 performance measures generally used in hydrological forecasting. In this study, we have selected four performance measures to evaluate model performance. The Nash–Sutcliffe efficiency \( E \), root mean square error \( \text{RMSE} \), mean absolute error \( \text{MAE} \) and persistence index \( \text{PERS} \) performance measures are used to evaluate the accuracy of the developed models. The Nash–Sutcliffe efficiency \( E \), introduced by Nash & Sutcliffe (1970), is still one of the most widely used criteria for the assessment of model performance. \( E \) provides a measure of the ability of a model to predict values that are different from the mean. It records as a ratio the level of overall agreement between the observed and modelled means and variances, it has been strongly recommended (e.g. NERC 1975; ASCE 1993). \( \text{RMSE} \) and \( \text{MAE} \) provide different types of information about the predictive capabilities of the model. The \( \text{RMSE} \) measures the goodness-of-fit relevant to high flow values whereas the \( \text{MAE} \) is not weighted towards high(er) magnitude or low(er) magnitude events, but instead evaluates all deviations from the observed values, in both an equal manner and regardless of sign. \( \text{PERS} \) is the substitution of the last known value as the current prediction and represents a good benchmark against which other predictions can be measured (Cannas et al. 2006; Kitanidis & Bras 1980).

(i) Nash–Sutcliffe coefficient \( (E) \). It is expressed as

\[
E = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O}_i)^2}
\]  

where \( O_i \) and \( P_i \) are the observed and predicted flow, \( \bar{O}_i \) is the mean of the observed flow and \( n \) is the number of data points. The value of the Nash–Sutcliffe coefficient varies between \(-\infty \) to 1. The closer the value to 1, the better is the model performance.

(ii) Root mean square error \( (\text{RMSE}) \). It is expressed as

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - P_i)^2}
\]  

(iii) Mean absolute error \( (\text{MAE}) \). It is expressed as

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |O_i - P_i|
\]  

(iv) Persistence index \( (\text{PERS}) \). It is expressed as

\[
\text{PERS} = 1 - \frac{\text{SSE}}{\text{SSE}_\text{naive}}
\]

where

\[
\text{SSE} = \sum_{i=1}^{n} (O_i - P_i)^2
\]  

and

\[
\text{SSE}_\text{naive} = \sum_{i=1}^{n} (O_i - \bar{O}_i - L)^2
\]  

in which the SSE terms are the sum of square errors, \( O_i - L \) is the observed discharge at time \( i \) minus the lead time \( L \). Persistence consists of a comparison between the model under study and the naive model where a value of \( \text{PERS} \) smaller than or equal to 0 indicates that the model under study performs worse or no better than the easy-to-implement naive model. A \( \text{PERS} \) value of 1 is obtained when the model under study provides exact estimates of observed discharge.

**STUDY AREA AND DATA USED**

The Mahanadi river basin, which is the fourth largest river basin in India, was selected for this study. The Mahanadi River flows to the Bay of Bengal in east-central India draining an area of 141,589 km² and has a length of 851 km. It lies between east longitudes 80°30’ to 86°30’ and north latitudes 19°21’ to 23°35’. About 53% of the basin is in the state of Chhattisgarh, 46% is in the coastal state of Orissa and the remainder of the basin is in the states of Jharkhand and Maharastra. Numerous dams, irrigation projects and barrages...
are present in the Mahanadi river basin, the most prominent of which is the Hirakud reservoir. The middle reaches of the lower Mahanadi river basin, located in Orissa between 82°E 19’N and 86°E 22’N and encompassing a geographical area of 47 558.6 km², forms the study area (Figure 2). The main river reach extends from Hirakud Dam to Naraj, having a total length of 358.4 km. The main soil types found in the study area are red and yellow soils. The normal annual rainfall is 1458 mm and temperature in this region varies from 14°C to 40°C. The average monthly pan evaporation of the area varies from 2.4 mm to 14.6 mm. Most of the rainfall and river flow occur during the monsoon season, between June and September. In the delta region of the Mahanadi river basin almost every year flooding is a serious problem during monsoon seasons. Naraj, which is situated at the mouth of the delta, was selected for daily discharge forecasting.

The data used for this study consists of daily discharge data of seven upstream gauging stations during the period 2000–2006. Some of the statistical properties of the discharge data are presented in Table 1. The locations of different gauging stations are shown in Figure 2. Seven years of daily discharge data of seven discharge gauging stations for the monsoon period (17 July to 25 September) from years 2000–2006 (497 sets) are divided into three datasets. Daily discharge of years 2000–2004 (355 patterns) are taken for training, 2005 (71 patterns) for cross-validation and 2006 (71 patterns) for testing. The training dataset serves for model training and the testing dataset is used to evaluate the performances of the models. The cross-validation set helps to implement an early stopping approach in order to avoid overfitting of the training data.

### MODEL DEVELOPMENT

One of the most important steps in the ANN hydrologic model development process is the determination of significant input variables. The current study used a statistical approach suggested by Sudheer et al. (2002) to identify the appropriate input vectors. The method is based on the heuristic that the potential influencing variables corresponding to different time lags can be identified through statistical analysis of the data series that uses cross-correlation (CCF), autocorrelation (ACF) and partial autocorrelation functions (PACF) between the variables. This process is applied to select significant inputs from the discharge data of seven
discharge gauging stations for daily river flow forecasting. The total input vectors identified are presented in Table 2.

The identification of the optimal network geometry is one of the major tasks in developing an ANN model. While the number of input output nodes is problem-dependent, there is no direct and precise way of determining the optimal number of hidden nodes. The model architecture of an ANN is generally selected through a trial-and-error procedure as currently there is no established methodology for selecting the appropriate network architecture prior to training. The optimum numbers of hidden neurons are calculated using the generalization error. The generalization errors (Equation (7)) of various ANN structures are representative of the ANN performance (Jia & Culver 2006). The ANN structures are tested for 1–6 hidden neurons and the structure with 4 hidden neurons for which the generalization error is lowest is chosen as the optimal structure.

Four ANN models developed in this study are traditional ANN, wavelet-based ANN (WANN), bootstrap-based ANN (BANN) and bootstrap–wavelet ANN conjunction model (BWANN). At first, a multi-layer perceptron (MLP) feed-forward ANN model is developed. The ANN models are

Table 1 | Statistics of the datasets for daily river flow forecasting

<table>
<thead>
<tr>
<th>Year</th>
<th>Statistics</th>
<th>Khairmal</th>
<th>Hirakud Release</th>
<th>Salebhata</th>
<th>Kesinga</th>
<th>Kantamal</th>
<th>Tikarpara</th>
<th>Naraj</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Mean (m³/s)</td>
<td>1792.9</td>
<td>784.9</td>
<td>25.0</td>
<td>359.8</td>
<td>569.5</td>
<td>2034.8</td>
<td>2296.9</td>
</tr>
<tr>
<td></td>
<td>Standard deviation (m³/s)</td>
<td>897.7</td>
<td>701.6</td>
<td>19.2</td>
<td>367.0</td>
<td>645.9</td>
<td>834.5</td>
<td>641.6</td>
</tr>
<tr>
<td></td>
<td>Maximum (m³/s)</td>
<td>5720.0</td>
<td>4601.3</td>
<td>101.2</td>
<td>2452.0</td>
<td>3661.0</td>
<td>4774.8</td>
<td>5050.0</td>
</tr>
<tr>
<td></td>
<td>Minimum (m³/s)</td>
<td>509.7</td>
<td>132.9</td>
<td>4.3</td>
<td>113.5</td>
<td>105.0</td>
<td>600.0</td>
<td>1090.0</td>
</tr>
<tr>
<td>2001</td>
<td>Mean (m³/s)</td>
<td>6983.8</td>
<td>4057.8</td>
<td>269.2</td>
<td>1049.5</td>
<td>1627.8</td>
<td>6035.1</td>
<td>9932.2</td>
</tr>
<tr>
<td></td>
<td>Standard deviation (m³/s)</td>
<td>6383.9</td>
<td>4401.3</td>
<td>431.6</td>
<td>940.6</td>
<td>1187.3</td>
<td>5568.9</td>
<td>9031.4</td>
</tr>
<tr>
<td></td>
<td>Maximum (m³/s)</td>
<td>30,468.9</td>
<td>21,567.3</td>
<td>3225.0</td>
<td>5293.0</td>
<td>3661.0</td>
<td>4774.8</td>
<td>5050.0</td>
</tr>
<tr>
<td></td>
<td>Minimum (m³/s)</td>
<td>894.8</td>
<td>474.5</td>
<td>8.6</td>
<td>116.5</td>
<td>336.7</td>
<td>1369.6</td>
<td>1842.1</td>
</tr>
<tr>
<td>2002</td>
<td>Mean (m³/s)</td>
<td>2230.1</td>
<td>1187.6</td>
<td>141.1</td>
<td>253.0</td>
<td>470.0</td>
<td>2206.6</td>
<td>3111.9</td>
</tr>
<tr>
<td></td>
<td>Standard deviation (m³/s)</td>
<td>2801.1</td>
<td>1964.7</td>
<td>243.9</td>
<td>327.1</td>
<td>579.5</td>
<td>2412.0</td>
<td>3583.7</td>
</tr>
<tr>
<td></td>
<td>Maximum (m³/s)</td>
<td>14,724.8</td>
<td>10,329.4</td>
<td>1545.4</td>
<td>1990.9</td>
<td>2900.0</td>
<td>12,305.6</td>
<td>16,630.3</td>
</tr>
<tr>
<td></td>
<td>Minimum (m³/s)</td>
<td>305.8</td>
<td>43.2</td>
<td>1.8</td>
<td>10.0</td>
<td>20.2</td>
<td>436.8</td>
<td>215.1</td>
</tr>
<tr>
<td>2003</td>
<td>Mean (m³/s)</td>
<td>9038.2</td>
<td>4887.2</td>
<td>664.5</td>
<td>1256.2</td>
<td>1645.7</td>
<td>7533.6</td>
<td>11,701.5</td>
</tr>
<tr>
<td></td>
<td>Standard deviation (m³/s)</td>
<td>8322.0</td>
<td>5306.2</td>
<td>1154.0</td>
<td>1327.7</td>
<td>2061.4</td>
<td>6445.3</td>
<td>10,014.6</td>
</tr>
<tr>
<td></td>
<td>Maximum (m³/s)</td>
<td>34,150.1</td>
<td>24,267.2</td>
<td>7916.0</td>
<td>8908.4</td>
<td>12,915.1</td>
<td>25,062.0</td>
<td>35,253.2</td>
</tr>
<tr>
<td></td>
<td>Minimum (m³/s)</td>
<td>1257.3</td>
<td>471.8</td>
<td>4.5</td>
<td>308.6</td>
<td>385.2</td>
<td>794.4</td>
<td>2620.3</td>
</tr>
<tr>
<td>2004</td>
<td>Mean (m³/s)</td>
<td>4581.7</td>
<td>2613.9</td>
<td>190.5</td>
<td>725.0</td>
<td>946.6</td>
<td>4208.4</td>
<td>5513.7</td>
</tr>
<tr>
<td></td>
<td>Standard deviation (m³/s)</td>
<td>4620.5</td>
<td>3240.1</td>
<td>236.2</td>
<td>633.9</td>
<td>687.2</td>
<td>3844.7</td>
<td>5452.1</td>
</tr>
<tr>
<td></td>
<td>Maximum (m³/s)</td>
<td>20,331.5</td>
<td>14,952.0</td>
<td>1004.0</td>
<td>3240.6</td>
<td>4014.0</td>
<td>17,744.4</td>
<td>22,465.4</td>
</tr>
<tr>
<td></td>
<td>Minimum (m³/s)</td>
<td>1121.3</td>
<td>538.5</td>
<td>13.6</td>
<td>182.6</td>
<td>345.5</td>
<td>1139.1</td>
<td>1320.7</td>
</tr>
<tr>
<td>2005</td>
<td>Mean (m³/s)</td>
<td>5570.8</td>
<td>3323.0</td>
<td>150.8</td>
<td>665.1</td>
<td>1001.2</td>
<td>4950.3</td>
<td>8229.0</td>
</tr>
<tr>
<td></td>
<td>Standard deviation (m³/s)</td>
<td>5565.3</td>
<td>3712.7</td>
<td>284.9</td>
<td>1210.2</td>
<td>1783.1</td>
<td>4141.0</td>
<td>7006.9</td>
</tr>
<tr>
<td></td>
<td>Maximum (m³/s)</td>
<td>26,249.7</td>
<td>13,196.6</td>
<td>1924.0</td>
<td>8120.8</td>
<td>11,030.1</td>
<td>24,399.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum (m³/s)</td>
<td>849.5</td>
<td>393.9</td>
<td>1.0</td>
<td>113.4</td>
<td>131.8</td>
<td>1130.0</td>
<td>1274.9</td>
</tr>
<tr>
<td>2006</td>
<td>Mean (m³/s)</td>
<td>7608.3</td>
<td>3343.9</td>
<td>573.8</td>
<td>1274.3</td>
<td>2138.1</td>
<td>7127.6</td>
<td>11,979.8</td>
</tr>
<tr>
<td></td>
<td>Standard deviation (m³/s)</td>
<td>5745.6</td>
<td>3036.7</td>
<td>836.8</td>
<td>1307.5</td>
<td>2985.1</td>
<td>5449.5</td>
<td>8737.6</td>
</tr>
<tr>
<td></td>
<td>Maximum (m³/s)</td>
<td>27,099.2</td>
<td>11,870.3</td>
<td>4681.4</td>
<td>6483.1</td>
<td>15,276.2</td>
<td>29,000.0</td>
<td>33,979.2</td>
</tr>
<tr>
<td></td>
<td>Minimum (m³/s)</td>
<td>1659.4</td>
<td>240.7</td>
<td>9.3</td>
<td>167.9</td>
<td>305.6</td>
<td>923.8</td>
<td>696.7</td>
</tr>
</tbody>
</table>
developed using the most significant inputs which are first log-transformed and then linearly scaled to the range (0, 1) for ANN modeling (Campolo et al. 1999). The computational efficiency of the training process is an important consideration for ANN modeling. A computationally efficient second-order training method, the Levenberg–Marquardt method, is used to minimize the mean squared error between the forecast and observed river flows. In the next step, the DWC data are used as inputs to the ANN model to develop the WANN model. The DWT decomposes an original discharge time series into many components (i.e. DWCs) at different scales (or frequencies). Each component plays a distinct role in the original flow series. The low-frequency component generally reflects the identity (periodicity and trend) of the signal whereas the high-frequency component uncovers details (Kucuk & Agrafologlou 2006). The wavelet function is derived from the family of Daubechies wavelets (Nourani et al. 2009; Wu et al. 2009). To choose the number of decomposition level or DWCs the following formula is used to determine the decomposition levels (Aussem et al. 1998; Nourani et al. 2008):

$$L = \text{int} \left[ \log(N) \right]$$  \hspace{1cm} (16)

where $L$ and $N$ are decomposition level and number of time series data, respectively. This study uses $N = 497$, which produces $L = 3$ approximately. The WANN models are developed employing subseries DWCs obtained using DWT. For this purpose, firstly, the original time series is decomposed into three levels of DWCs by DWT. Three levels of decomposition (i.e., D1, D2 and D3) and approximation (A3) for discharge data of Khairmal and Hirakud releases are shown in Figure 3. The effective DWCs are determined using the correlation coefficients between each wavelet components and the observed discharge at Naraj. Correlation between the periodic component and the original discharge data reveals the effectiveness of the component. Table 3 shows that the significant periodic components are A3 and D3 for Khairmal, Hirakud release data and Kantamal; A3, D2 and D3 for Tikarpara; A3, D1, D2 and D3 for Naraj; and only approximation A3 for Salebhata and Kesinga. These constitute the new wavelet discharge series. The significant wavelet components of a particular gauging station are added and employed to constitute the new inputs of the WANN for daily discharge forecasting. The newly constructed time series is used as WANN model inputs. The BANN model is developed as an ensemble of several ANNs built using bootstrap resamples of raw datasets, whereas the WBANN model is developed as an ensemble of several ANNs built using bootstrap resamples of DWCs instead of raw datasets. In this way the WBANN model uses the capabilities of both bootstrap resampling and wavelet transformation technique. Similar to the BANN model, the WBANN model is also developed using 200 resamples to maintain consistency. Bootstrap.xla, an Excel add-in (Barreto & Howland 2006), is used to generate bootstrap resamples of raw datasets and DWCs for the BANN and WBANN model development, respectively. A simple flow-chart depicting the development of ANN, BANN, WANN and WBANN models is shown in Figure 4. The number of inputs for all four models is taken as the same to maintain consistency. Similar to ANN the WANN, BANN and WBANN model structures are tested for 1–6 hidden neurons and all three models with 4 hidden neurons for which the generalization errors are minimum is chosen as the optimal structure. Therefore, for all four models the number of hidden neurons is 4.

**DAILY DISCHARGE FORECASTING**

Figure 5–7 show the hydrograph and scatter plot of observed and predicted discharges using WBANN for 1-, 3-, and 5-day lead time forecasts for the testing period. The figures show that WBANN forecasts are in good agreement with the observed data. For 1-day lead time forecasts the observed and predicted values are in good agreement as the predicted
Figure 3 | The discrete wavelet components of the discharge series of (a) Khairmal and (b) Hirakud release data for the year 2006.
values show the general behavior of the observed discharge. For longer lead times, model predictions gradually deteriorate. This is because, for longer lead time, the river flow values contain less information than for the shorter lead times for modeling discharge values at the longer time horizon. The results of the WBANN model in terms of different performance indices for the testing period is shown in Table 4. The RMSE statistic is a measure of residual variance and is indicative of the model’s ability to predict high flows. Considering a high value of 33,979.2 m$^3$/s and a very low value of 696.7 m$^3$/s at the Naraj gauging station, the WBANN model with RMSE value of 4981.3 m$^3$/s performed satisfactorily up to 5-day lead forecasts.

Table 5 shows the performance of ANN, WANN and BANN models in terms of $E$, RMSE and MAE. It is clear from the Tables 4 and 5 that for 1–5-day lead time forecasts the WBANN and WANN models perform better than the traditional ANN and BANN. The time of concentration of the basin is about 3 days. Therefore, the performance of the model is not very good for longer lead time forecasts (i.e. 4- and 5-day lead). Performance of the WANN model for 4-day lead time and performance of the WBANN model for 4- and 5-day lead times are still significantly better than ANN and BANN, as these models simulate the antecedent information for lead time equal to the time of concentration by utilizing the capability of wavelet and bootstrap techniques. Figures 8–10 show the hydrographs and scatter plots of observed and predicted discharges using ANN, WANN, BANN and WBANN for 1-, 3-, and 5-day lead time forecasts for the testing period. It is obvious from the figures that for longer lead time (> 1 day) the traditional ANN and BANN models underestimate the observed data, whereas the WANN and WBANN models are consistent and do not significantly underestimate or overestimate the observed values. It is observed from the figures that for 1-, 3- and 5-day lead time forecasts the WANN and WBANN models forecast higher discharge values better than the ANN and BANN models. Moreover, performance of the WANN and WBANN models in predicting peak values for 1- and 5-day lead time forecasts is almost similar, even though the 5-day lead time forecast

---

**Table 3 | Correlation coefficients between the discrete wavelet components and the observed discharge data at Naraj**

<table>
<thead>
<tr>
<th>Discrete wavelet components</th>
<th>Khairmal</th>
<th>Hirakud Release</th>
<th>Salebhata</th>
<th>Kesinga</th>
<th>Kantamal</th>
<th>Tikarpura</th>
<th>Naraj</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3</td>
<td>0.91</td>
<td>0.87</td>
<td>0.75</td>
<td>0.60</td>
<td>0.66</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>D1</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>D2</td>
<td>0.06</td>
<td>0.05</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>D3</td>
<td>0.21</td>
<td>0.13</td>
<td>0.07</td>
<td>0.09</td>
<td>0.11</td>
<td>0.23</td>
<td>0.27</td>
</tr>
<tr>
<td>Original</td>
<td>0.88</td>
<td>0.81</td>
<td>0.52</td>
<td>0.39</td>
<td>0.50</td>
<td>0.93</td>
<td>1.00</td>
</tr>
</tbody>
</table>

---

**Figure 4 | Flowchart showing the development of ANN, BANN, WANN and WBANN models.**
The performance of the WANN model is significantly better compared to the WBANN model. The reason may be that the WBANN model is the ensemble of several (in this study 200) ANNs built using resamples of DWCs. There may be a chance that the high discharge values, which are very less, are not included in any of the 200 bootstrap resamples for 5-day lead time forecasts. The WANN model developed with newly constructed time series using DWT is found to be better compared to the simple ANN and BANN models for daily discharge forecasting. This shows the capabilities of DWT in

![Figure 5](https://example.com/figure5.png)

**Figure 5** | Hydrograph (a) and scatter plot (b) of observed and predicted discharge of testing dataset for Naraj using the WBANN model for 1-day lead time.

![Figure 6](https://example.com/figure6.png)

**Figure 6** | Hydrograph (a) and scatter plot (b) of observed and predicted discharge of testing dataset for Naraj using the WBANN model for 3-day lead time.

![Figure 7](https://example.com/figure7.png)

**Figure 7** | Hydrograph (a) and scatter plot (b) of observed and predicted discharge of testing dataset for Naraj using the WBANN model for 5-day lead time.

<table>
<thead>
<tr>
<th>Lead time (d)</th>
<th>E (%)</th>
<th>RMSE (m$^3$/s)</th>
<th>MAE (m$^3$/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93.5</td>
<td>2247.9</td>
<td>1628.5</td>
</tr>
<tr>
<td>2</td>
<td>89.0</td>
<td>2921.6</td>
<td>2312.4</td>
</tr>
<tr>
<td>3</td>
<td>85.4</td>
<td>3367.2</td>
<td>2569.8</td>
</tr>
<tr>
<td>4</td>
<td>69.8</td>
<td>4850.9</td>
<td>3747.5</td>
</tr>
<tr>
<td>5</td>
<td>68.2</td>
<td>4981.3</td>
<td>3832.1</td>
</tr>
</tbody>
</table>

**Table 4** | Performance indices for 1–5-day lead time forecasts for the WBANN model

The performance of the WANN model is significantly better compared to the WBANN model. The reason may be that the WBANN model is the ensemble of several (in this study 200) ANNs built using resamples of DWCs. There may be a chance that the high discharge values, which are very less, are not included in any of the 200 bootstrap resamples for 5-day lead time forecasts. The WANN model developed with newly constructed time series using DWT is found to be better compared to the simple ANN and BANN models for daily discharge forecasting. This shows the capabilities of DWT in
daily discharge forecasting. WANN forecasts are found to be more accurate due to the fact that the features (such as periodicity) of the subseries are obvious (Ning & Yunping 1998). It is to be clarified that, even though WANN for all the lead times and ANN for 3- and 5-day lead times perform better than BANN, the BANN and WBANN models are more reliable and consistent. This is because the ANN model is uncertain and is not reproducible. Even a change in the arrangement of training, cross-validation and testing datasets lead to a different performance of ANN and WANN models and it cannot be predicted. On the other hand, the BANN and WBANN models are consistent and more reliable even if the arrangement of datasets for training, cross-validation and testing changes. This is the reason why ensemble forecasting, by averaging all the 200 models, are taken instead of selecting a best model. The BANN model is the aggregation of several (200 in this study) ANN models, each developed on resampled original time series. As the BANN model is developed on different realizations of the actual dataset using the bootstrap method, the ensemble predictions of these competing models lead to comprehensive, unbiased and realistic predictions. For longer lead time, the performance of the WBANN model is significantly better than the ANN, BANN and WANN models. Overall performance of the WBANN model is found to be superior due to the reason that it is developed by generating 200 bootstrap resamples of DWCs instead of a raw dataset as for BANN. In this way the WBANN model uses the capabilities of both bootstrap and wavelet techniques by taking inputs which are generated by resampling the newly constructed time series using DWCs.

In order to assess the performance of models to predict discharge values for different lead time forecasts for different magnitude flows, a “partitioning analysis” (Jain & Srinivasulu 2004) is carried out by dividing the total discharge values into low-, medium- and high- magnitude discharges. Table 6 presents the partitioning of discharge values for testing periods based on the relative spread of the discharge from the mean. It has been reported that the coefficient of efficiency can be high (80 or 90%) even for poor models, and the best models do not produce values which, on first examination, are impressively higher (Legates & McCabe 1999;
The RMSE statistic indicates only the model’s ability to predict a value away from the mean (Hsu et al. 1995). Therefore, it is important to test the model using some other performance evaluation criteria such as threshold statistics (Nayak et al. 2005b). The threshold statistics (TS) not only give the performance index in terms of predicting discharges but also the distribution of the prediction errors.

The threshold statistic for a level of x% is a measure of consistency in forecasting errors from a particular model. The threshold statistics is represented as $TS_x$ and expressed as a percentage. This criterion can be expressed for different levels of absolute relative error ($ARE$) from the model. $ARE$ is given as

$$ARE_i = \frac{O_i - P_i}{O_i} \times 100$$  \hspace{1cm} (17)

$TS_x$ is computed for $x\%$ level as

$$TS_x = \frac{Y_x}{n} \times 100$$  \hspace{1cm} (18)

where $Y_x$ is the number of computed discharges (out of $n$ total computed) for which the absolute relative error is less than $x\%$ from the model.

Table 7 shows the distribution of forecast error predicted from different models across different error thresholds for 1–5-day lead time forecasts for low, medium and high discharge profiles. Even though there are only two high discharge values, which constitute only 2.8% of the total testing dataset (Table 6), it is observed that the performance of the WBANN and WANN models for high flow forecasts is better than for ANN and BANN. It may be due to the reason that wavelets are reducing the noise in the discharge time series, making it easier to predict the discharge whereas bootstrapping is reducing the variance. Noise in high discharge may be due

**Table 6 | Number of data points in low-, medium- and high-discharge categories**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of points</th>
<th>Percentage of the total data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low ($x &lt; \mu$)</td>
<td>38</td>
<td>53.5</td>
</tr>
<tr>
<td>Medium ($\mu \leq x \leq \mu + 2\sigma$)</td>
<td>31</td>
<td>43.7</td>
</tr>
<tr>
<td>High ($x &gt; \mu + 2\sigma$)</td>
<td>2</td>
<td>2.8</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>100</td>
</tr>
</tbody>
</table>

$\mu$ is the mean and $\sigma$ is the standard deviation.
to the reason that the discharge values are computed using the rating curves; however, since rating curves are developed with limited stage-discharge measurements and since measurements of high flows are rare, there could be significant errors in rating curves at high levels (Sahoo & Ray 2006). This error or noise in high discharge is well approximated using discrete wavelet decomposition. Performance of the WBANN model for medium flow forecasts is significantly better compared to the remaining three models as the WBANN model can forecast 93.5%, 80.6%, 74.2%, 67.5% and 64.5% of discharge values within 25% relative error for 1–5-day lead time forecast, respectively, whereas the WANN model can forecast 77.4%, 64.5%, 71.0%, 58.1% and 35.5% of discharge values within 25% relative error. All four models perform more or less similarly during low discharge. It is observed that medium discharges are simulated more satisfactorily compared to low and high discharges profiles. The reason may be that the low discharge data are influenced by the unsaturated condition of the basin and the high discharge values do not have proper representation in training the models, whereas medium discharge values are not influenced much by the unsaturated condition of the basin and have proper representation in training datasets.

In order to compare ANNs with an empirical model, we also developed a multiple linear regression (MLR) model for daily discharge forecasting. The discharge at

Table 7  Threshold statistics for all four models for testing period

<table>
<thead>
<tr>
<th></th>
<th>Low flow</th>
<th>Medium flow</th>
<th>High flow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TS (%)</td>
<td>1-d lead</td>
<td>2-d lead</td>
</tr>
<tr>
<td>WBANN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>18.4</td>
<td>5.3</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>26.3</td>
<td>7.9</td>
<td>0.0</td>
</tr>
<tr>
<td>20</td>
<td>50.0</td>
<td>28.4</td>
<td>21.1</td>
</tr>
<tr>
<td>25</td>
<td>65.8</td>
<td>41.1</td>
<td>26.3</td>
</tr>
<tr>
<td>50</td>
<td>86.8</td>
<td>71.1</td>
<td>68.4</td>
</tr>
<tr>
<td>BANN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10.5</td>
<td>13.2</td>
<td>5.3</td>
</tr>
<tr>
<td>10</td>
<td>21.1</td>
<td>13.2</td>
<td>26.3</td>
</tr>
<tr>
<td>20</td>
<td>55.3</td>
<td>34.2</td>
<td>39.5</td>
</tr>
<tr>
<td>25</td>
<td>63.2</td>
<td>54.7</td>
<td>44.7</td>
</tr>
<tr>
<td>50</td>
<td>92.1</td>
<td>78.9</td>
<td>68.4</td>
</tr>
<tr>
<td>WANN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.9</td>
<td>5.3</td>
<td>2.6</td>
</tr>
<tr>
<td>10</td>
<td>28.9</td>
<td>15.8</td>
<td>7.9</td>
</tr>
<tr>
<td>20</td>
<td>54.2</td>
<td>34.2</td>
<td>23.7</td>
</tr>
<tr>
<td>25</td>
<td>52.6</td>
<td>44.7</td>
<td>31.6</td>
</tr>
<tr>
<td>50</td>
<td>86.8</td>
<td>68.4</td>
<td>47.4</td>
</tr>
<tr>
<td>ANN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>18.4</td>
<td>7.9</td>
<td>7.9</td>
</tr>
<tr>
<td>10</td>
<td>23.7</td>
<td>23.7</td>
<td>18.4</td>
</tr>
<tr>
<td>20</td>
<td>63.2</td>
<td>44.7</td>
<td>36.8</td>
</tr>
<tr>
<td>25</td>
<td>71.1</td>
<td>47.4</td>
<td>57.9</td>
</tr>
</tbody>
</table>
| 50    | 90.7   | 74.9     | 78.9     | 57.9     | 55.3     | 100.0    | 93.5     | 93.5     | 87.1     | 61.3     | 100.0    | 100.0    | 50.0     | 0.0      | 0.0
Naraj station is selected as the dependent variable and the input variables as of ANNs are selected as independent variables. The best-fit model is estimated and selected based on the highest coefficient of determination ($R^2$). All of the MLR models are first trained (to determine the regression coefficients) using the data in the training set (2000–2005) and then tested using the testing dataset (2006). The SPSS software package (version 10, SPSS Inc., Chicago, IL) is used for regression calculations. The performance comparison of the ANN, WANN, BANN, WBANN and MLR models is carried out with a simple naive model in terms of persistence index ($PERS$). It is obvious from Table 8 that the MLR model is better compared to a simple naive model only for 1-day lead time forecasts. All four ANN models (i.e. ANN, WANN, BANN and WBANN) perform better compared to the MLR model for 1–5-day lead time forecasts. It is clear that the $PERS$ of the WANN and WBANN models are greater than 0 and therefore both the models are better compared to a simple naive model for 1–5-day lead time forecasts. Performance of ANN and BANN are better up to 4-day lead forecasts, and negative $PERS$ for 5-day lead time forecast indicate that it is better to implement a simple naive model instead of ANN and BANN. The performance of the WBANN model is better than all the remaining models for all the lead times forecast; these findings again show the strength and capability of WBANN modeling for daily discharge forecasting. Negative values of $PERS$ for different models show that for these lead times the models are not able to extract any linear or nonlinear relationship and it is better to implement a simple naive model. It is clear from these findings that WBANN model performance is very good up to 5-day lead time forecasts.

### Table 8

<table>
<thead>
<tr>
<th>Lead time (d)</th>
<th>ANN</th>
<th>WANN</th>
<th>BANN</th>
<th>WBANN</th>
<th>MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.40</td>
<td>0.58</td>
<td>0.50</td>
<td>0.63</td>
<td>0.3786</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.48</td>
<td>0.21</td>
<td>0.69</td>
<td>-0.0185</td>
</tr>
<tr>
<td>3</td>
<td>0.21</td>
<td>0.43</td>
<td>0.14</td>
<td>0.70</td>
<td>-0.2864</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.47</td>
<td>0.01</td>
<td>0.49</td>
<td>-0.4608</td>
</tr>
<tr>
<td>5</td>
<td>-0.02</td>
<td>0.17</td>
<td>-0.07</td>
<td>0.59</td>
<td>-0.3678</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

The aim of this paper is to introduce a conjunction model WBANN based on wavelet and bootstrap techniques for daily discharge forecasting. The decomposed time series of the observed data present different periodic components. Each of the wavelet components makes a distinct contribution to the original time series. The significant wavelet components are selected based on the correlation between the particular component series and the observed discharge at Naraj and a new series is composed by adding the significant components for each time series. These series are employed as inputs to constitute the WANN model for daily discharge forecasting. Use of wavelet components supports the recent studies about the physics involved in the ANNs (Jain et al. 2004; Sudheer & Jain 2004; Sudheer 2005). The significant inputs to the ANN model are selected using the cross-correlation statistics. The bootstrap resampling method is used to generate different realizations of the datasets to create a set of bootstrap samples that provide a better understanding of the average and variability of the original unknown distribution or process. Two hundred bootstrapped ANN models are generated to reduce the uncertainty involved in ANN predictions. The inclusion of the DWCs in the ANN input layer enabled consideration of the effective periodic components in the discharge forecasting whereas bootstrapping techniques are used to reduce the uncertainty inherent in the conventional ANN modeling. WBANN models are developed to incorporate the individual capabilities of wavelet and bootstrap techniques. The WBANN model is found to be superior to the traditional ANN, BANN and WANN in terms of the selected performance criteria for 1–5-day lead forecasts. The peak discharge forecasts obtained by the WANN model are noticeably closer to the observations compared to all three remaining models. ARE and TS statistics show that performance of the WBANN and WANN models for high flow forecasts are better than ANN and BANN, the performance of the WBANN model for medium flow forecasts is significantly better compared to the remaining three models, whereas all four models perform more or less similar during low discharge forecasts. This study shows that the WBANN method is quite an appropriate tool for...
modeling the discharge for short as well as long lead time forecasts.

REFERENCES


ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000b Artificial neural networks in hydrology II: hydrologic applications. J. Hydro. Engng. 5(2), 124–137.


First received 1 July 2009; accepted in revised form 1 February 2010. Available online 1 October 2010