Flood forecasting using an improved NARX network based on wavelet analysis coupled with uncertainty analysis by Monte Carlo simulations: a case study of Taihu Basin, China

Feiqing Jianga, Zengchuan Donga,*, Zeng’an Wangb, Yiqing Zhuc, Moyang Liud, Yun Luoa and Tianyan Zhanga

a Department of Hydrology and Water Resources, Hohai University, No. 1 Xikang Road, Nanjing, Jiangsu, China
b Jiangsu Expressway Company Limited, No. 6 Xianlin Avenue, Nanjing, Jiangsu, China
c Liyang Jiangnan Engineering Testing Company Limited, No. 1098 Pingling West Road, Kunlun Street, Liyang, Changzhou, Jiangsu, China
d Fenner School of Environment & Society, The Australian National University, Canberra, ACT 0200, Australia

*Corresponding author. E-mail: zcdong@hhu.edu.cn

ABSTRACT

Reliable flood forecasting can provide a scientific basis for flood risk assessment and water resources management, and the Taihu water level forecasting with high precision is essential for flood control in the Taihu Basin. To increase the prediction accuracy, a coupling model (DWT-iNARX) is established by combining the discrete wavelet transformation (DWT) with improved nonlinear autoregressive with exogenous inputs network (iNARX), for predicting the daily Taihu water level during the flood season under different forecast periods. And the DWT-iNARX model is compared with the back-propagation neural network (BP) and iNARX models to assess its capability in prediction. Meanwhile, we propose an uncertainty analysis method based on Monte Carlo simulations (MCS) for quantifying model uncertainty and performing probabilistic water level forecast. The results show that three models achieve good simulation results with higher accuracy when the forecast period is short, such as 1–3 days. In overall performance, iNARX and DWT-iNARX models show superiority in comparison with the BP model, while the DWT-iNARX model yields the best performance among all the other models. The research results can provide a certain reference for the water level forecast of the Taihu Lake.

Key words: discrete wavelet transformation, improved NARX network, Monte Carlo simulations, probabilistic forecast, Taihu Lake, water level prediction

HIGHLIGHTS

• This study investigates a new data-mining-based model, which incorporates the discrete wavelet transformation and improved nonlinear autoregressive with exogenous inputs network, for flood forecasting in different forecast periods.
• This study proposes an uncertainty analysis method framework based on Monte Carlo simulations for quantifying model uncertainty and performing probabilistic water level forecast.

GRAPHICAL ABSTRACT

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).
INTRODUCTION

Flood is one of the most frequent natural calamities, resulting in loss of life and damage to personal property and critical public health infrastructure. More than 2 billion people were affected by floods from 1998 to 2017 worldwide (Rehman et al. 2019), and those who live in floodplains, without efficient flood forecasting systems, are most vulnerable to floods. The Taihu Basin is a typical plain river network area in China, with a developed economy and dense population (Luo et al. 2019). It is characterized by a flat terrain, extensive river networks, and numerous polder areas, which are vulnerable to flooding (Zhai et al. 2020). In 2016, direct economic losses due to flood disasters in the Taihu Basin were $1.06 billion, accounting for 0.12% of its Gross Domestic Product (Wang 2018). Meanwhile, in 1991 and 1999, the direct economic losses due to flood disasters accounted for 6.7% and 1.6% of the GDP in the corresponding year (Xu 2009; Wang et al. 2018). Hence, there is an urgent need to pay attention to flood control in the Taihu Basin.

The Taihu water level is the critical indicator of flood control and dispatching in the Taihu Basin, which determines the operation modes of the Wangyuhe water conservancy project, Taipuhe water conservancy project, and other water conservancy projects in the basin. Therefore, a good forecast of the Taihu water level is essential for flood risk assessment and water resources management in the basin. The prediction of the water level in the plain river network area as a result of rainfall process, underlying surfaces variation, and hydraulic engineering dispatching events is an important water resources management in the basin. The prediction of the water level in the plain river network area as a result of rain-vancy projects in the basin. Therefore, a good forecast of the Taihu water level is essential for operation modes of the Wangyuhe water conservancy project, Taipuhe water conservancy project, and other water conservancy projects in the basin. The physical-based model describes the physical mechanisms of the hydrodynamic process and is considered to be one of the most complex models, due to a large number of variables involved in the modeling process, huge temporal and spatial variability of basin characteristics, and difficulties in discretizing and solving partial differential equations caused by the intricacies of river networks (Teng et al. 2017). Actually, the complex hydrological regime and engineering conditions in the Taihu Basin, coupled with the difficulty of obtaining the real-time water, precipitation, and engineering information, make Taihu water level forecasting using the physical-based simulation extremely complicated. The data-driven model considers the hydrodynamic system as a black-box and tries to establish a relationship between historical inputs, such as rainfall and discharge, and outputs. It is the most intuitive and straightforward approach to forecast the Taihu water level.

Among the black-box models, artificial neural networks (ANNs) have been proved to have excellent nonlinear and indeterminate mapping capability, with advantages of no prior knowledge required and using less assumption in its applications (Hsu et al. 1995), which has been successfully applied in many hydrological studies (Anmala et al. 2000; Abrahart et al. 2012; Arsenault et al. 2015). The back-propagation neural network (BP) is the most widely used neural network model (Ju et al. 2009; Chen et al. 2010; Liu et al. 2017), which is a static network as it allows only one-way information flow from the input layer to the output layer, and is incapable of examining the temporal dimension of data. Nevertheless, this is vital for a hydrological system since its current response is very reliant on its preceding states. The dynamic recurrent neural network (RNN) has the capability to learn from the preceding conditions of the system as it facilitates time delay units through feedback connections (Shoaib et al. 2016). Nonlinear autoregressive with the exogenous inputs network (NARX) is a special type of dynamic RNN that describes the modeled process based on lagged input–output variables and prediction errors. It has been widely used in the prediction of time series (Chen et al. 2015; Lee & Tuan Resdi 2016; Guzman et al. 2017; Di Nunno & Granata 2020) as it has powerful performance and can approximate almost all nonlinear functions. The embedded memory of the NARX network provides a shorter path to optimize the propagation of information and back-propagation of error signals, which reduces the long-term dependency of the model (Guzman et al. 2017). Compared with traditional RNNs, it can provide optimal predictions without calculating losses (Siegelmann et al. 1997).

In recent years, wavelet transformation (WT) has been found to be effective in simultaneously elucidating both temporal and spectral information within the signal. Introduced by Grossmann & Morlet (1984), WT is developed on the basis of Fourier analysis, which has a serious drawback: a signal's Fourier transform into the frequency domain results in the loss of time information, such that it becomes impossible to tell when a particular event took place. In contrast, WT allows the use of long time intervals when more precise low-frequency information is needed, and shorter regions when high-frequency information is of interest (Nourani et al. 2014). It also has flexibility in choosing the mother wavelet (i.e., the transform function) according to the characteristics of time series (Seo et al. 2015). Accordingly, the time series can be decomposed into its subcomponents using wavelet transformation and these multi-scale signals are used as input data for the ANN model, the
result of which is a hybrid wavelet-ANN model. A majority of studies have been dedicated to applying the hybrid wavelet-ANN model to predict hydrological elements such as the water level (Seo et al. 2015; Loh et al. 2019), runoff (Shoaib et al. 2016, 2018), and precipitation (Nourani et al. 2009a). Nourani et al. (2014) pointed out that the neural network model coupled with wavelet analysis has dominant advantages in time series de-noising, multi-resolution analysis, and hydrological forecasting (Nourani et al. 2009b; Remesan et al. 2009; Nourani & Parhizkar 2013). The hybrid wavelet-ANN model has proven to be superior compared with the single model as it may present more probable forecasting rather than a single pattern input. However, the existing research focused on the coupling of wavelet analysis and feedforward neural network (FNN), while the coupling of wavelet analysis and NARX network has not been exhausted.

Most previous studies performed linear mixture regression for Taihu water level prediction (Lu et al. 2000; Wang et al. 2001; Wang & Dong 2003; Hu 2008). Practically, there is a complex and highly nonlinear relationship between the Taihu water level, and the previous Taihu water levels, rainfall in the basin, and the operation mode of flood control projects. To fill this knowledge gap, in this study, considering the influencing factors of Taihu water level, taking advantage of the NARX network in time series prediction, we combined the discrete wavelet transformation (DWT) with the improved NARX network (iNARX) to establish a coupling model (DWT-iNARX), for Taihu water level prediction. The main objectives of this study are (1) to forecast daily Taihu water level during the flood season under the forecast period of 1–7 days using the DWT-iNARX model, (2) to compare the DWT-iNARX model with BP and iNARX models to assess its capability in predicting water levels, and (3) to present an uncertainty analysis framework for quantifying model uncertainty and performing probabilistic water level forecast.

STUDY AREA AND DATA

Study area

As shown in Figure 1, the Taihu Basin is located downstream of the Yangtze River Delta, with a total drainage area of 36,895 km². The Taihu Lake is located in the middle of the basin, whose water surface area is about 2,336.8 km² under normal water level. Based on geographic and hydrological features, the Taihu Basin is usually divided into eight hydrological subregions: Hu Xi Region, Zhe Xi Region, Wu Cheng Xi Yu Region, Yang Cheng Dian Mao Region, Lake Region, Hang Jia Hu Region, Pudong Region, and Puxi Region. There is almost no water exchange between Pudong-Puxi regions and Taihu Lake, due to the Qingsong polder dike engineering in Shanghai.

The Taihu Basin lies in the subtropical monsoon climate zone, which is mild and humid, with an annual average rainfall of 1,189 mm. The distribution of the rainfall is uneven during the year, with the most rainfall occurring in the flood season (from May 1 to September 30), accounting for 75% of the annual rainfall. The Taihu water level generally drops at a rate of 2–3 cm/d when there is no rain, rises at a rate of 1–10 cm/d when there is significant rainfall, and rises at a rate of 20–30 cm/d when there is heavy rain. Due to the flat terrain, backwater effect of the tidal downstream and various water conservancy projects preventing water drainage, the Taihu water level often maintains above 3.5 m in the flood season, posing a serious flood threat to surrounding areas and low-lying flat downstream areas.

Data preparation

The data used in this study included:

(1) Daily Taihu water level data, daily water level data of the Ganlu, Fengqiao and Pingwang stations from 1990 to 2017.
(2) Daily discharge data of the Wangting Hydro-junction and the Taipu Gate from 1990 to 2017.
(3) Daily precipitation data of 65 hydrological stations evenly distributed in the Taihu Basin from 1990 to 2017.

The locations of each station are shown in Figure 1. The area-mean rainfall of six hydrological subregions was calculated using the Thiessen polygon method. All data were obtained from the Taihu Basin Authority of Ministry of Water Resources and had been reviewed to guarantee the completeness, accuracy, and consistency. The hydrological observation data of the flood season from 1990 to 2017 with consistent human activities were used for calibration models, of which the data from 1990 to 2011 were used to train and test the forecast model, while the data from 2012 to 2017 were used to predict the Taihu water level and validate the model.
The input variables were scaled to accelerate the convergence speed and make the training process faster. The following equation was used to normalize the input data:

\[ B_t = \frac{B_t - B_{\text{min}}}{B_{\text{max}} - B_{\text{min}}} \]  

where \( B_t \) and \( B_i \) are the observed and normalized values of the input variable at time \( t \), respectively; \( B_{\text{max}} \) and \( B_{\text{min}} \) are the maximum and minimum values of the input variable, respectively.

### METHODS

**iNARX model**

**Model structure**

NARX models are recurrent dynamic architectures with several hidden layers and are inspired by discrete-time nonlinear models (Leontaritis & Billings 1985), the common definition for which is given by:

\[ y(t) = f[y(t - 1), y(t - 2), \ldots, y(t - n), u(t - 1), u(t - 2), \ldots, u(t - n)] + e_t \]  

where \( u(t) \in \mathbb{R} \) and \( y(t) \in \mathbb{R} \) represent inputs and outputs of the network at a discrete time step \( t \), respectively; \( n \) is the time delay of input and output variables; and \( e_t \) is the model error between the target and prediction. The function \( f() \) is a nonlinear function that should be approximated. When this is done by a multilayer perceptron, the resulting topology is called a NARX recurrent neural network.
Since different input time series have the same time delay in the conventional NARX model, we improved the NARX model in our proposed method so that different input sequences had different time delays, which could also be different from the time delay of output sequences, thus the iNARX model with $k$ input variables and one output variable is obtained as:

$$y(t) = f_o \left\{ b_o + \sum_{h=1}^{H} w_{ho} \cdot f_h \left[ b_h + \sum_{i=1}^{d_{o1}} w_{ih} u_1(t - i_1) + \sum_{i=2}^{d_{o2}} w_{ih} u_2(t - i_2) \right. \right.$$ 
$$+ \cdots + \sum_{i=1}^{d_{oH}} w_{ih} u_h(t - i_h) \left. \right] + \sum_{j=1}^{d_{oj}} w_{jh} y(t - j) \right\}$$

where $H$ is the number of hidden neurons; $w_{ih}$ is the connection weight between the input neuron and hidden neuron; $w_{jh}$ is the connection weight between the hidden neuron and output feedback neuron; $w_{ho}$ is the connection weight between the hidden neuron and predicted output; $b_h$ and $b_o$ are the biases of the hidden neuron and predicted output, respectively; $f_h()$ is the hidden layer activation function; and $f_o()$ is the output layer activation function. The structure of the iNARX network is depicted in Figure 2.

Compared with FNN, the iNARX network can feedback outputs during the training procedure, so it contains information about the long sequence and has long-term memory, which can simulate the long-term dynamics of time series. Meanwhile, when an iNARX network is unfolded in time, the output delays will appear as jump-ahead connections in the unfolded network. Intuitively, these jump-ahead connections provide a shorter path for propagating gradient information, thus reducing the sensitivity of the network to long-term dependencies (Lin et al. 1996). Consequently, the iNARX network converges much faster and generalizes better than other networks.

**Figure 2** | Structure of the iNARX network ($z^{-1}$ indicates unit time delay).
Training algorithm

The Bayesian Regularization (BR) algorithm provides an efficient generalization to reduce the negative effects of large weights by modifying the training function of the neural network based on Bayesian inference techniques (Foresee & Hagan 1997), which has a small network structure, high learning efficiency, and good generalization ability (Hirschen & Schäfer 2006). The network performance function is expressed as:

\[ F(w) = \beta \cdot \frac{1}{K \cdot g} \sum_{j=1}^{k} (\bar{y}_j - y_j)^2 + \alpha \cdot \frac{1}{K \cdot g} ||Pf||^2 = \beta \cdot E_D + \alpha \cdot E_w \tag{4} \]

where \( k \) is the number of output nodes; \( g \) is the total number of training values; \( \bar{y}_j \) is the expected output value of the training value at node \( f \); \( y_j \) is the network actual output value of the training value at node \( f \); \( E_D \) is the sum of squared errors; \( E_w \) is the sum of the square of the network weights; and \( \beta \) and \( \alpha \) are objective function parameters.

We supposed that the network weights were random variables and the probability density function (PDF) of the weights could be computed according to the Bayes rules:

\[ P(w|D, \alpha, \beta, N) = \frac{P(D|w, \beta, N) \cdot P(w|\alpha, N)}{P(D|\alpha, \beta, N)} \tag{5} \]

where \( w \) is the vector of network weights; \( N \) is the iNARX network we used; and \( D \) is the training target data sample. \( P(w|\alpha, N) \) is the prior distribution of weights; \( P(D|w, \beta, N) \) is the likelihood function which is the probability of occurrence, giving the network weights; \( P(D|\alpha, \beta, N) \) is a normalization factor, which guarantees that the total probability is 1.

Assuming that the noise in the training data and the prior distribution of weights are Gaussian, the probabilities can be written as:

\[ P(D|w, \beta, N) = \frac{1}{Z_D(\beta)} \exp (-\beta E_D) \tag{6} \]

\[ P(w|\alpha, N) = \frac{1}{Z_w(\alpha)} \exp (-\alpha E_w) \tag{7} \]

where \( Z_D(\beta) = (\pi/\beta)^{m/2} \) and \( Z_w(\alpha) = (\pi/\alpha)^{M/2} \), with \( m \) being the total number of network weights and \( M \) the total number of parameters in the network. Then the Gaussian form for the posterior probability is:

\[ P(w|D, \alpha, \beta, N) = \frac{1}{Z_w(\alpha) Z_D(\beta)} \exp \left( -\left( \beta E_D + \alpha E_w \right) \right) = \frac{1}{Z_F(\alpha, \beta)} \exp (-F(w)) \tag{8} \]

In this Bayesian framework, the optimal weights should maximize the posterior probability \( P(w|D, \alpha, \beta, N) \). Maximizing the posterior probability of \( w \) is equivalent to minimizing the regularized objective function \( F(w) \).

Then Bayes’ rule is applied to optimize the objective function parameters \( \alpha \) and \( \beta \).

\[ P(\alpha, \beta|D, N) = \frac{P(D|\alpha, \beta, N) \cdot P(\alpha, \beta|N)}{P(D|N)} \tag{9} \]

If the prior density \( P(\alpha, \beta|N) \) is uniform for the regularization parameters \( \alpha \) and \( \beta \), then maximizing the posterior is achieved by maximizing the likelihood function \( P(D|\alpha, \beta, N) \), which is the normalization factor for Equation (5). We can
solve Equation (5) for the normalization factor:

\[
P(D | \alpha, \beta, N) = \frac{P(D | w, N) \cdot P(w | \alpha, N)}{P(w | D, \alpha, \beta, N)} = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \frac{1}{Z_w(\alpha)} \exp(-\alpha E_w)
\]

where the constants \( Z_D(\beta) \) and \( Z_w(\alpha) \) are known from Equations (6) and (7), and \( Z_F(\alpha, \beta) \) is the only unknown part, which can be estimated by Taylor series expansion. Since the objective function has the shape of a quadratic in a small area surrounding a minimum point, \( F(w) \) can be expanded around the minimum point of the posterior density \( w^{MP} \), where the gradient is zero. Solving for the normalizing constant yields:

\[
Z_F = (2\pi)^{M/2} (\text{det}(H^{MP})^{-1})^{1/2} \exp(-F(w^{MP}))
\]

where \( H = \beta^2 E_D + \alpha^2 E_w \) is the Hessian matrix of the objective function. Placing the result into Equation (10), the optimal values for \( \alpha \) and \( \beta \) at the minimum point can be solved. We do this by taking the derivative with respect to each log of Equation (10) and set them equal to zero. This yields:

\[
\alpha^{MP} = \frac{\gamma}{2E_w(w^{MP})}, \quad \beta^{MP} = \frac{m - \gamma}{2E_D(w^{MP})}
\]

where \( \gamma = M - 2\alpha^{MP}r(H^{MP})^{-1} \) is the effective number of parameters, which is a measure of how many parameters in the network are effectively used in reducing the error function.

The Bayesian optimization of the regularization parameters requires the computation of the Hessian matrix of \( F(w) \) at the minimum point \( w^{MP} \). The Gauss–Newton approximation to the Hessian matrix is applied to optimize regularization. Steps are as follows:

1. Initialize \( \alpha, \beta \), and the weights. After the first training step, the objective function parameters will recover from the initial setting.
2. Take one step of the Levenberg–Marquardt algorithm to minimize the objective function \( F(w) = \beta E_D + \alpha E_w \).
3. Compute the effective number of parameters \( \gamma = M - 2\alpha^{MP}r(H^{MP})^{-1} \) using the Gauss–Newton approximation to the Hessian available in the Levenberg–Marquardt training algorithm: \( H = \nabla^2 F(w) = 2\beta^2 J + 2\alpha J \), where \( J \) is the Jacobian matrix of the training set errors.
4. Compute new estimates for the objective function parameters \( \alpha = (\gamma/2E_w(w)) \) and \( \beta = (m - \gamma)/(2E_D(w)) \).
5. Iterate steps 2 through 4 until convergence.

DWT

There are two types of WTs: continuous wavelet transformation (CWT) and DWT. DWT is generally used in hydrological series, which can be thought of as the dyadic sampling of CWT. The discretely scaled and translated wavelet can be written as:

\[
\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)
\]

where \( k \) is a location index; \( j \) is referred to as the decomposition level; and \( \psi(t) \) is the wavelet function or mother wavelet, of which the discrete wavelet transform can be written as:

\[
W_{\psi}(j, k) = 2^{-j/2} \int_{-\infty}^{\infty} \bar{\psi}(2^{-j}t - k) f(t) dt, \quad j = 0, 1, \ldots, k \in \mathbb{Z}
\]

where \( \bar{\psi}(t) \) is the complex conjugate functions of \( \psi(t) \).
For a discrete time series $x(n)$, the dyadic wavelet transform becomes:

$$W_c(j, k) = 2^{-j/2} \sum_{n=0}^{N-1} \psi(2^{-j}n - k)x_n$$  \hspace{1cm} (15)$$

where $W_c(j, k)$ is the wavelet coefficient for the discrete wavelet of scale $a = 2^j$ and location $b = 2^j k$

$$x_n = T(t) + \sum_{j=1}^{J} \sum_{k=0}^{2^j - 1} W_c(j, k)2^{-j/2}\psi(2^{-j}n - k)$$  \hspace{1cm} (16)$$

which can also be written as:

$$x_n = T(t) + \sum_{j=1}^{J} W_j(t)$$  \hspace{1cm} (17)$$

where $T(t)$ is the residual term, called the approximation at level $J$, which represents the background information of data, and the wavelet coefficients $W_j(t)$ are details at levels $j = 1, 2, \ldots, J$, providing the detail signals, which can capture the small features of interpretational value in the data.

In this study, we considered the dyadic wavelet transform for discrete time series and used the Mallat algorithm (Mallat 1989) to do discrete wavelets decomposition and reconstruction. Multiple-level decomposition was performed to get high-scale, low-frequency components (approximations) and low-scale, high-frequency components (details) of original time series. Because of the simplicity of approximations and details after the decomposition, some interesting characteristics such as period, hidden period, dependence, and jump can be diagnosed easily through wavelet components.

**DWT-iNARX model**

The DWT-iNARX model is the coupling model of iNARX and DWT. Since the iNARX model predicts the current value of the time series from its past values, so the predicted time series is both the input and output of the model, to be specific, past values of the sequence are the input and the current value is the output. DWT is employed to decompose the time series into approximation and detail components, and to reconstruct the signal over the predicted frequency components using the wavelet decomposition structure. The decomposed time series are used as both inputs and outputs to the DWT-iNARX model. The DWT-iNARX model consists of the following three-step algorithm:

1. The wavelet decomposition is employed to pre-process the predicted time series. First, the decomposition level is determined and the appropriate wavelet type is selected according to the characteristics of sequences. Then, multilevel DWT is used to decompose the original signal into details ($D_1, D_2, \ldots, D_k$) and approximation ($A_k$), where $k$ is the decomposition level.
2. The iNARX model is trained and tested using the details and approximation as input. In this step, the strategies for modeling the iNARX model can be used, including the selection of time delay for input and output variables, the number of hidden nodes, learning algorithms, etc. The details and approximation of the final time series can be predicted in this step.
3. The wavelet reconstruction is employed to construct the signal, which is the final prediction result, over the predicted details and approximation using the wavelet decomposition structure in Step 1.

**Evaluation of performance**

In order to evaluate the accuracy of the constructed model, four statistical parameters were calculated. These parameters are qualification rate (QR), Nash–Sutcliffe efficiency coefficient (NSE), root mean square error (RMSE), and mean relative error (MRE). These parameters were calculated by using the model predictions and the corresponding measured data. These methods are expressed as:

$$QR = \frac{m}{n} \times 100\%$$  \hspace{1cm} (18)$$
Uncertainty analysis

This study examined a methodology for the probabilistic water level forecast to cope with uncertainties by combining the DWT-iNARX model and Monte Carlo simulations (MCS). The prediction interval is comprised of upper and lower bounds that bracket a future unknown value with a confidence level.

The model uncertainty is defined by the sum of three components: (1) inherent variability due to random fluctuations, (2) model structure uncertainty (Pappenberger et al. 2006; Sharma & Regonda 2021), and (3) input data uncertainty (Mohanty et al. 2020). In our study, model structure uncertainty was reduced through optimization of model structure and parameters.

We mainly considered the uncertainty of the area-mean rainfall in the input data, which could not be estimated with absolute certainty by the error type and magnitude of rain gauges, and their density and pattern of space distribution. Uncertainty and inherent variability of the input data are considered through the definition of probability distributions. The expected value of a function $g$ of the uncertain variables of $x$ with a PDF of $f_X(x)$ is characterized by the following equation if $x$ is continuous:

$$E[g(x)] = \int_{x \in X} g(x)f_X(x)dx$$

where $x$ represents the area-mean rainfall and $g(x)$ represents water level predictions based on DWT-iNARX modeling. A Monte Carlo estimate of $E[g(x)]$ can be obtained by taking an $n$ sample of $X,(x_1, x_2, \ldots, x_n)$ through the following equation:

$$\bar{g}_n(X) = \frac{1}{n} \sum_{i=1}^{n} g(x_i)$$

If $E[g(x)]$ exists, then the weak law of large numbers indicates that as $n$ gets large, the probability that $\bar{g}_n(X)$ deviates from $E[g(x)]$ becomes smaller. In fact, $\bar{g}_n(X)$ is an unbiased estimate for $E[g(x)]$.

The prediction intervals are estimated by taking a very similar approach: randomly drawing an $n$ sample $(x_1, x_2, \ldots, x_n)$ from the hypothetical distribution of the uncertain input variables (i.e., the area-mean rainfall), subsequently estimating the water levels $g(x_i)$ for each $x_i$ through DWT-iNARX modeling, and finally employing the resulting statistical populations of $g(x_i)$ to estimate the prediction intervals for each water level. The following procedures are performed.

First, the MCS is performed to simulate $N$ samples for each input area-mean rainfall. The normal distributions are assumed for the area-mean rainfall and different variances represent the different levels of uncertainty. In the random sampling process of MCS, the Latin Hypercube Sampling (LHS) is used to separately draw an $N$ sample from the hypothetical distribution of each of the uncertain input area-mean rainfall. LHS is more efficient than the simple random sampling when the sample size is less than a few thousand. In the LHS process, the range of each variable is divided into $N$ nonoverlapping intervals of equal probability of $1/N$, where $N$ is the sample size. Subsequently, a value is selected randomly from each interval according to the probability density of the interval. Here, $N$ is set equal to 1,000. And then Taihu water levels are predicted using the DWT-iNARX model which takes the simulated area-mean rainfall samples as input. Thus, we get $N$ predicted water levels at each.
RESULTS AND DISCUSSION

Selection of input variables and time delays

The rainfall of the Taihu Basin, water levels at key flood control stations, and discharges of the key flood control projects are the main factors affecting the water level of the Taihu Lake. In this study, based on the physical genesis and priori knowledge of characteristics of potential input variables, the following four types of predictors, i.e., input variables, were selected: (1) early water levels of the Taihu Lake; (2) area-mean rainfalls of six hydrological regions (i.e., Lake Region, Hu Xi Region, Zhe Xi Region, Hang Jia Hu Region, Wu Cheng Xi Yu Region, and Yang Cheng Dian Mao Region) in the Taihu Basin; (3) water levels at key flood control stations (i.e., Ganlu Station, Fengqiao Station, and Pingwang Station); and (4) discharges of key flood control projects (i.e., the Wangting Hydro-junction and the Taipu Gate). These input variables are listed in Table 1. The water level of the Taihu Lake is predicted, which is the output variable.
The time delays of input time series were determined by physical genesis combined with statistical analysis. The time delay of the rainfall sequence in each subregion was determined by the concentration time, which is defined as the time needed for water to flow from the most remote point in each subregion to the Taihu Lake. Based on the distance of each subregion from the Taihu Lake, topography, geology, and land use within each subregion, and layout of hydraulic engineering, the confluence time of each subregion is calculated as: 2 days in the Lake Region, 5 days in the Hu Xi Region, 7 days in the Zhe Xi Region, 7 days in the Hang Jia Hu Region, 5 days in the Wu Cheng Xi Yu Region, and 5 days in the Yang Cheng Dian Mao Region (Wang et al. 2001), which are regarded as time delays of rainfall sequences.

The time delays of other input sequences were determined by the autocorrelation and cross-correlation coefficients. The autocorrelation of Taihu water level ($Z_{Taihu}$) and its cross-correlation with the water levels at key flood control stations ($Z_{Ganlu}$, $Z_{Fengqiao}$, and $Z_{Pingwang}$) and discharges of key flood control projects ($Q_{Wangting}$ and $Q_{Taihu}$) are displayed in Figure 4. From Figure 4(a), the first to seventh order autocorrelation coefficients of $Z_{Taihu}$ were all greater than 0.9, indicating an obvious correlation between the current Taihu water level and its historical values in the previous 7 days. As shown in Figure 4(b)–4(d), the cross-correlation coefficients between $Z_{Taihu}$ and $Z_{Ganlu}$, $Z_{Fengqiao}$ and $Z_{Pingwang}$ did not change much with the increase of the order, but they all had a strong correlation. From Figure 4(e) and 4(f), the correlation coefficients between $Z_{Taihu}$ and $Q_{Wangting}$ and $Q_{Taipu}$ did not change significantly as the order increases, but there was a certain correlation between them. Considering all the above, the time delay of each input time series was determined as shown in Table 1.

### Optimization of model structure

The methods for multi-step ahead (also called long-term) forecasting of hydrological time series include recursive strategy, direct strategy, multi-input multi-output strategy (Taieb et al. 2012), etc. This study used the direct strategy for single-step prediction, that is, different neural network models were established for different forecast periods. Three types of models, which were BP, iNARX, and DWT-iNARX models, were built and the prediction periods were 1–7 days. Consequently, a total of 21 prediction models had been established to simulate the Taihu water level in different forecast periods.

Hornik et al. (1989) introduced that multilayer neural networks with as few as one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficient hidden units are available. Therefore, this study chose a three-layer network with one input layer, one hidden layer, and one output layer. Based on some rule-of-thumb methods, through trial and error, we obtained that networks with 26 neurons in the hidden layer generally produced the best results. The transfer functions of the hidden and output layer were ‘tansig’ and ‘purelin’, respectively. The network was trained for 10,000 epochs to minimize the mean squared error using the BR algorithm.

During the establishment of the DWT-iNARX model, the time series of Taihu water level should be decomposed first. The selection of suitable wavelet function and decomposition level should be considered when using DWT for sequence decomposition. Commonly used wavelet functions include the simple wavelet functions Haar wavelet, the db4 and db8

### Table 1 | Input time series and their time delay

| ID | Input time series                        | Time delay (day) |
|----|------------------------------------------|------------------|
| 1  | The early water level of Taihu Lake      | 7                |
| 2  | The area-mean rainfall of Lake Region    | 2                |
| 3  | The area-mean rainfall of Hu Xi Region   | 5                |
| 4  | The area-mean rainfall of Zhe Xi Region  | 7                |
| 5  | The area-mean rainfall of Hang Jia Hu Region | 7            |
| 6  | The area-mean rainfall of Wu Cheng Xi Yu Region | 5          |
| 7  | The area-mean rainfall of Yang Cheng Dian Mao Region | 5        |
| 8  | The water level at Ganlu Station         | 7                |
| 9  | The water level at Fengqiao Station      | 7                |
| 10 | The water level at Pingwang Station      | 7                |
| 11 | The discharge of Wangting Hydro-junction | 7                |
| 12 | The discharge of Taipu Gate              | 7                |

The time delays of input time series were determined by physical genesis combined with statistical analysis. The time delay of the rainfall sequence in each subregion was determined by the concentration time, which is defined as the time needed for water to flow from the most remote point in each subregion to the Taihu Lake. Based on the distance of each subregion from the Taihu Lake, topography, geology, and land use within each subregion, and layout of hydraulic engineering, the confluence time of each subregion is calculated as: 2 days in the Lake Region, 5 days in the Hu Xi Region, 7 days in the Zhe Xi Region, 7 days in the Hang Jia Hu Region, 5 days in the Wu Cheng Xi Yu Region, and 5 days in the Yang Cheng Dian Mao Region (Wang et al. 2001), which are regarded as time delays of rainfall sequences.

The time delays of other input sequences were determined by the autocorrelation and cross-correlation coefficients. The autocorrelation of Taihu water level ($Z_{Taihu}$) and its cross-correlation with the water levels at key flood control stations ($Z_{Ganlu}$, $Z_{Fengqiao}$, and $Z_{Pingwang}$) and discharges of key flood control projects ($Q_{Wangting}$ and $Q_{Taihu}$) are displayed in Figure 4. From Figure 4(a), the first to seventh order autocorrelation coefficients of $Z_{Taihu}$ were all greater than 0.9, indicating an obvious correlation between the current Taihu water level and its historical values in the previous 7 days. As shown in Figure 4(b)–4(d), the cross-correlation coefficients between $Z_{Taihu}$ and $Z_{Ganlu}$, $Z_{Fengqiao}$ and $Z_{Pingwang}$ did not change much with the increase of the order, but they all had a strong correlation. From Figure 4(e) and 4(f), the correlation coefficients between $Z_{Taihu}$ and $Q_{Wangting}$ and $Q_{Taipu}$ did not change significantly as the order increases, but there was a certain correlation between them. Considering all the above, the time delay of each input time series was determined as shown in Table 1.

### Optimization of model structure

The methods for multi-step ahead (also called long-term) forecasting of hydrological time series include recursive strategy, direct strategy, multi-input multi-output strategy (Taieb et al. 2012), etc. This study used the direct strategy for single-step prediction, that is, different neural network models were established for different forecast periods. Three types of models, which were BP, iNARX, and DWT-iNARX models, were built and the prediction periods were 1–7 days. Consequently, a total of 21 prediction models had been established to simulate the Taihu water level in different forecast periods.

Hornik et al. (1989) introduced that multilayer neural networks with as few as one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficient hidden units are available. Therefore, this study chose a three-layer network with one input layer, one hidden layer, and one output layer. Based on some rule-of-thumb methods, through trial and error, we obtained that networks with 26 neurons in the hidden layer generally produced the best results. The transfer functions of the hidden and output layer were ‘tansig’ and ‘purelin’, respectively. The network was trained for 10,000 epochs to minimize the mean squared error using the BR algorithm.

During the establishment of the DWT-iNARX model, the time series of Taihu water level should be decomposed first. The selection of suitable wavelet function and decomposition level should be considered when using DWT for sequence decomposition. Commonly used wavelet functions include the simple wavelet functions Haar wavelet, the db4 and db8

### Table 1 | Input time series and their time delay

| ID | Input time series                        | Time delay (day) |
|----|------------------------------------------|------------------|
| 1  | The early water level of Taihu Lake      | 7                |
| 2  | The area-mean rainfall of Lake Region    | 2                |
| 3  | The area-mean rainfall of Hu Xi Region   | 5                |
| 4  | The area-mean rainfall of Zhe Xi Region  | 7                |
| 5  | The area-mean rainfall of Hang Jia Hu Region | 7            |
| 6  | The area-mean rainfall of Wu Cheng Xi Yu Region | 5          |
| 7  | The area-mean rainfall of Yang Cheng Dian Mao Region | 5        |
| 8  | The water level at Ganlu Station         | 7                |
| 9  | The water level at Fengqiao Station      | 7                |
| 10 | The water level at Pingwang Station      | 7                |
| 11 | The discharge of Wangting Hydro-junction | 7                |
| 12 | The discharge of Taipu Gate              | 7                |
wavelet functions of most commonly wavelet family Daubeches, sym4 and sym8 wavelet functions with the sharp peak of the Symlet wavelet family and discrete approximation of the Meyer wavelet function. The research of Shoaib et al. (2016) showed that the db8 wavelet function had the best results for hydrological time series, because it had reasonable support width and also contained good time–frequency localization property, which allowed the db8 wavelet function to capture both the

Figure 4 | Autocorrelation and cross-correlation coefficients. (a) Autocorrelation of $Z_{Taihu}$, (b) Cross-correlation of $Z_{Taihu}$ and $Z_{Ganlu}$, (c) Cross-correlation of $Z_{Taihu}$ and $Z_{Fengjiao}$, (d) Cross-correlation of $Z_{Taihu}$ and $Z_{Fengxian}$, (e) Cross-correlation of $Z_{Taihu}$ and $Q_{Wangting}$, and (f) Cross-correlation of $Z_{Taihu}$ and $Q_{Taipu}$.
underlying trend as well as the short-term variability in the time series data. Consequently, the db8 wavelet function was considered for this study.

The suitable decomposition level is selected using the following formula (Nourani et al. 2011):

\[ L = \text{int}\left(\log(N)\right) \]

where \( L \) is the level and \( N \) is the total data points. In this study, the decomposition level was calculated as 3 according to the length of time series.

Level 3 decomposition of the Taihu water level series by db8 wavelet is presented in Figure 5. Figure 5(a) shows the approximation subsignal at Level 3, while Figure 5(b) and 5(c) shows detail subsignals at Levels 1, 2, and 3, respectively.

Predicting water levels

The BP, iNARX, and DWT-iNARX models were validated by comparing the simulation results with 6 years of data that were not previously included in the training set. The performances of three constructed models were assessed by QR, NSE, RMSE, and MRE. The results in terms of four evaluation indexes for the training and prediction data sets in different forecast periods are shown in Table 2 and Figure 6.

It could be seen from Table 2 and Figure 6 that NSEs of iNARX and DWT-iNARX models were higher than 0.9, and their QRs were not less than 80% in all forecast periods. NSE of the BP model was greater than 0.85, and its QR was more than 70% in different forecast periods. In addition, as the forecast period increased, the prediction accuracy of three models decreased in varying degrees. The accuracy of the BP model dropped the most, while the iNARX and DWT-iNARX models were relatively stable, as the prolongation of the forecast period failed to reduce their accuracy rapidly. When the forecast period was short, such as 1–3 days, three models achieved good simulation results with higher accuracy for both training and prediction data sets. This indicated that the neural network model had certain advantages in forecasting hydrological time series because of its universal approximation capability and flexible structure that allowed us to capture complex nonlinear behaviors. When the forecast period was long, such as 6 or 7 days, the performance of the DWT-iNARX model was obviously superior to the other models, showing that it is robust and accurate.

The gap of each evaluation index between training and prediction periods became larger as the forecast period increased, among which the gap of the BP model was the largest. In the training period, evaluation indexes of the iNARX model were not significantly different from those of the BP model, and even the QR of the BP model was sometimes better than the iNARX model, but in the prediction period, the iNARX model was significantly better than the BP model. The reason was that compared with the traditional BP model, the iNARX model had feedback connections between layers of neurons, which could effectively prevent overfitting, and led to a good performance in the prediction period. Moreover, the gap of each evaluation index between training and prediction periods for the DWT-iNARX model was the smallest, indicating that the DWT-iNARX model had the best generalization ability.

Taking the year 2016 as an example, Figure 7 shows the observed Taihu water level and its predicted values of the three models under the forecast periods of 1–7 days during the flood season.

In 2016, affected by the strong El Niño event, spring floods, plum floods, and autumn floods occurred successively in the Taihu Basin. The large amount of rainfall, high water level of the Taihu Lake, and long duration of flood were rare in history. The Taihu water level was 3.51 m on 1 May 2016, which was the highest water level entering the flood season in the same period in history. On the day of entering the plum, the Taihu water level rose to 3.77 m, which was the second highest in history. After entering the plum, the Taihu water level rose rapidly due to heavy rain. On July 3, the Taihu water level reached and exceeded 4.65 m, and a flood exceeding the designed level occurred in the Taihu Basin. On July 6, the Taihu water level rose to 4.80 m, and an extraordinary flood occurred in the basin. On July 8, the Taihu Lake reached a peak water level of 4.87 m, second only to the highest level recorded in 1999.

From Figure 7, when the observed water level changed greatly and extreme values appeared, the water level hydrograph predicted by the BP model showed severe oscillations, and the hydrograph predicted by the iNARX model showed a jagged shape. Additionally, the DWT-iNARX model simulated the rising trend of water level well, and the hydrograph predicted by it was relatively smoother and closer to the observed water level under the same forecast period. Rainfall in the basin during the flood season, upstream flooding into the lake, dispatch of pumps and sluices in the lake region, and so on would all cause large fluctuations in the Taihu water level, but the forecast factors selected in this study were limited for convenience and efficiency. Thus, the forecast accuracy would be lower when
Figure 5 | Approximation and detail subsignals of Taihu water level series using db8 wavelet (Level 3). (a) Approximation subsignal at Level 3, (b) detail subsignal at Level 1, (c) detail subsignal at Level 2, and (d) detail subsignal at Level 3.
The water level changed greatly. The reason the iNARX model did not oscillate as sharply as the BP model was that the iNARX model had long-term memory capabilities, which could capture the details on the water level's fluctuations. The DWT-iNARX model performed best, proving that the use of wavelet transformation would capture useful information on various resolution levels and significantly improve the ability of a forecasting model.

As shown in Figure 7, with the extension of the forecast period, the deviation between the actual water level process and predicted hydrograph of all three models became larger, among which the BP model was the largest, and the DWT-iNARX model was the smallest. Especially when the forecast period was long, such as 6 or 7 days, the performance of the BP model was obviously inferior to the other two models. This was consistent with the results shown in Table 2 and Figure 6.

Figure 8 shows the scatter plot of the observed and forecast daily Taihu water level for three models in different forecast periods during the flood season from 2012 to 2017. With the increase of forecast period, the points of BP and iNARX models became more scattered compared with the DWT-iNARX model, of which the BP model was more obvious.

Figure 9 shows the cumulative probability distribution curve of the absolute error for three models in different forecast periods for the prediction data set. It can be clearly seen from Figure 9 that in the same forecast period, the prediction error of the BP model was the largest among the three models, and the performance of the iNARX model was slightly inferior to that of the DWT-iNARX model. As the forecast period increased, the gap between the prediction error of the BP model and that of the other two models also increased.

In overall performance, iNARX and DWT-iNARX models showed superiority in comparison with the BP model, while the DWT-iNARX model yielded the best performance among all other models. The DWT-iNARX model used various scale and frequency components of the wavelet decomposition as predictors, which can reflect more detailed information on hydrological sequences, so its accuracy was the highest in the same forecast period. It proved that the use of wavelet transformation would capture useful information on various resolution levels and significantly improve the ability of a forecasting model.

Uncertainty analysis
Uncertainty analysis allows decision makers to consider multiple scenarios and solutions for various conditions and avoid selecting highly risk-associated actions under uncertain conditions (Merwade et al. 2008). In this section, we computed prediction intervals of the water level to cope with the uncertainty of the area-mean rainfall in the input data. Results are shown in Figure 10. Water resources managers can identify how the uncertainty in input variables propagates through the model and

| Model     | Forecast period (day) | Training (flood season from 1990 to 2011) | Prediction (flood season from 2012 to 2017) |
|-----------|-----------------------|-----------------------------------------|---------------------------------------------|
|           | QR (%)                | NSE                       | RMSE             | MRE (%)                      | QR (%)                | NSE                       | RMSE             | MRE (%)                      |
| BP        | 1                     | 99.91                     | 0.999            | 0.0135                       | 0.26                   | 99.67                     | 0.989            | 0.0352                       | 0.42                   |
|           | 2                     | 99.20                     | 0.995            | 0.0259                       | 0.51                   | 97.49                     | 0.988            | 0.0368                       | 0.68                   |
|           | 3                     | 97.62                     | 0.989            | 0.0386                       | 0.76                   | 93.79                     | 0.975            | 0.0542                       | 1.07                   |
|           | 4                     | 94.50                     | 0.977            | 0.0565                       | 1.12                   | 88.78                     | 0.946            | 0.0791                       | 1.54                   |
|           | 5                     | 90.64                     | 0.962            | 0.0727                       | 1.29                   | 83.66                     | 0.929            | 0.0909                       | 1.88                   |
|           | 6                     | 85.32                     | 0.943            | 0.0887                       | 1.72                   | 78.98                     | 0.905            | 0.1048                       | 2.10                   |
|           | 7                     | 84.82                     | 0.939            | 0.0919                       | 1.70                   | 72.66                     | 0.851            | 0.1317                       | 2.56                   |
| iNARX     | 1                     | 99.97                     | 0.999            | 0.0124                       | 0.28                   | 99.56                     | 0.996            | 0.0204                       | 0.36                   |
|           | 2                     | 99.55                     | 0.997            | 0.0208                       | 0.40                   | 98.04                     | 0.987            | 0.0385                       | 0.81                   |
|           | 3                     | 98.13                     | 0.991            | 0.0354                       | 0.69                   | 94.44                     | 0.973            | 0.0555                       | 1.19                   |
|           | 4                     | 95.72                     | 0.983            | 0.0487                       | 0.93                   | 90.85                     | 0.963            | 0.0651                       | 1.23                   |
|           | 5                     | 90.52                     | 0.968            | 0.0668                       | 1.28                   | 87.25                     | 0.940            | 0.0836                       | 1.59                   |
|           | 6                     | 86.30                     | 0.943            | 0.0890                       | 1.60                   | 82.68                     | 0.913            | 0.1008                       | 1.88                   |
|           | 7                     | 84.55                     | 0.940            | 0.0912                       | 1.67                   | 81.37                     | 0.909            | 0.1026                       | 1.92                   |
| DWT-iNARX | 1                     | 100.00                    | 0.999            | 0.0080                       | 0.18                   | 99.89                     | 0.999            | 0.0115                       | 0.19                   |
|           | 2                     | 99.91                     | 0.999            | 0.0118                       | 0.23                   | 98.37                     | 0.992            | 0.0297                       | 0.56                   |
|           | 3                     | 98.46                     | 0.994            | 0.0292                       | 0.55                   | 94.77                     | 0.975            | 0.0537                       | 0.94                   |
|           | 4                     | 96.05                     | 0.980            | 0.0522                       | 1.01                   | 91.39                     | 0.963            | 0.0660                       | 1.23                   |
|           | 5                     | 93.49                     | 0.972            | 0.0619                       | 1.25                   | 90.41                     | 0.956            | 0.0718                       | 1.35                   |
|           | 6                     | 92.19                     | 0.968            | 0.0665                       | 1.31                   | 86.71                     | 0.940            | 0.0837                       | 1.58                   |
|           | 7                     | 89.33                     | 0.956            | 0.0783                       | 1.39                   | 82.14                     | 0.917            | 0.0980                       | 1.85                   |
Figure 6 | Four evaluation indexes of three models for training and prediction sets in different forecast periods. (a) QR of three models for training set, (b) NSE of three models for training set, (c) RMSE of three models for training set, (d) MRE of three models for training set, (e) QR of three models for prediction set, (f) NSE of three models for prediction set, (g) RMSE of three models for prediction set, and (h) MRE of three models for prediction set.
Figure 7 | Observed Taihu water level and its predicted values of three models with leading time of (a) 1 day, (b) 2 days, (c) 3 days, (d) 4 days, (e) 5 days, (f) 6 days, and (g) 7 days during the flood season of 2016.
Figure 8 | Scatter plot of observed and forecast daily Taihu water level for three models with leading time of (a) 1 day, (b) 2 days, (c) 3 days, (d) 4 days, (e) 5 days, (f) 6 days, and (g) 7 days during the flood season from 2012 to 2017.
Figure 9 | Cumulative probability distribution curve of absolute error for three models with leading time of (a) 1 day, (b) 2 days, (c) 3 days, (d) 4 days, (e) 5 days, (f) 6 days, and (g) 7 days for prediction data set.
influences the resulting water level predictions from Figure 10. It can be seen from Figure 10 that the prediction interval covered the observed water level, indicating that the probabilistic daily water level prediction is reliable. Wide PIs are an indication of the presence of a high level of uncertainty in the prediction of the water level with a high chance of confronting an unexpected condition. Under the same variance of uncertain input variable distributions, as the forecast period increased, the width of the PIs became larger, which illustrated that the forecast period and the uncertainty of forecast results increased consistently. Near the extreme water level, the PI was wider than usual, indicating that the forecast uncertainty for extreme values was relatively larger.
The innovative work of this study is as follows:

1. This study proposed a new data-mining-based method for the water level prediction in different forecast periods. The proposed method, called DWT-iNARX, incorporating DWT and iNARX networks, is capable of taking advantage of both temporal and spectral information within the signal, and simulating the long-term dynamics of time series, with fast convergence speed and good generalization ability. Research results validated the effectiveness of the proposed approach and showed a considerable improvement in the prediction precision. This method was simple, affordable, and practical, and it required less information about real-time rain, water and project operation compared with the simulation of hydrodynamic models. It could provide the accurate forecast of Taihu water level for flood control in the Taihu Basin under the circumstance of less real-time hydrological information.

2. Furthermore, this study investigated a methodology of combining DWT-iNARX and MCSs to quantify model uncertainty and provide the probabilistic water level forecast. The probabilistic forecast offered great potential for water level predictions, and facilitated communications between modelers and urban decision makers. Besides this, the uncertainty analysis method proposed in this study was independent of deterministic hydrological forecasting models, and could be integrated with any deterministic hydrological forecasting model without any assumptions, which provided a methodological framework for various probabilistic hydrological forecasting systems.

**CONCLUSIONS**

Floods are made more likely by more extreme weather patterns caused by long-term global climate change. The Taihu Basin is a typical plain river network area, with extensive rivers, lakes, and polder areas, and is vulnerable to flooding. Within this context, the Taihu water level forecast is of great significance to the flood risk assessment, flood control dispatching, and water resources management in the Taihu Basin. In this study, a coupling model of DWT and the iNARX network was established to predict the daily Taihu water level under different forecast periods, which was compared with BP and iNARX models to verify its effectiveness and applicability. Meanwhile, an uncertainty analysis method was proposed based on MCS for quantifying model uncertainty and performing probabilistic water level forecast. The conclusions were as follows:

1. Three models achieved good simulation results with higher accuracy when the forecast period was short, such as 1–3 days, indicating that the neural network model had certain advantages in forecasting hydrological time series because of its universal approximation capability and flexible structure that allowed us to capture complex nonlinear behaviors.

2. When the forecast period was long, such as 6 or 7 days, the iNARX and DWT-iNARX models showed obvious superiority in comparison with the BP model. This reflected the superiority of the structure of the iNARX network, which had strong generalization ability and long-term memory capability.

3. When the observed water level changed greatly and extreme values appeared, the water level hydrograph predicted by the BP model showed severe oscillations, the hydrograph predicted by the iNARX model showed a jagged shape, and the hydrograph predicted by the DWT-iNARX model was relatively smoother and closer to the observed water level. This proved that the use of wavelet transformation would capture useful information on various resolution levels and significantly improve the ability of a forecasting model.

4. In overall performance, iNARX and DWT-iNARX models showed superiority in comparison with the BP model, while the DWT-iNARX model yielded the best performance among all other models, indicating that the proposed method could not only improve the prediction accuracy but also prolong the forecast period.

5. The forecast period and uncertainty of forecast results increased consistently, and the forecast uncertainty for extreme values was relatively larger.

The results of this study improve the prediction accuracy of the Taihu water level and prolong the forecast period, which can improve the reliability of decision-making basis for flood control dispatching in the Taihu Basin, and provide references for efficient use of water resources. The probabilistic water level forecast facilitated communications between modelers and decision makers. Moreover, the proposed methodology, which is a black-box model, is the most intuitive and straightforward approach to retrieve useful flood information from observations. It is especially suitable for areas where the physical-based simulation is difficult due to complex hydrodynamic processes. It uses snapshots of the past and cannot directly predict responses to future or changed conditions. This is the limitation of this model compared with the hydrodynamic models. However, if enough data meeting actual requirements can be obtained, then the proposed method is robust and accurate.
FUNDING
This research was funded by the National Key Research and Development Program of China, grant number 2018YFC1508200.

DATA AVAILABILITY STATEMENT
All relevant data are included in the paper or its Supplementary Information.

REFERENCES
Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E. & Wilby, R. L. 2012 Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Progress in Physical Geography 36 (4), 480–513.

Anmala, J., Zhang, B. & Govindaraju, R. S. 2000 Comparison of ANNs and empirical approaches for predicting watershed runoff. Journal of Water Resources Planning & Management 126 (3), 156–166.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F. O. & Martel, J. L. 2015 A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. Journal of Hydrology 529, 754–767.

Chan, R. W., Yuen, J. K., Lee, E. W. & Arashpouir, M. 2015 Application of Nonlinear-Autoregressive-Exogenous model to predict the hysteretic behaviour of passive control systems. Engineering Structures 85, 1–10.

Chen, C.-S., Chen, B. P.-T., Chou, F. N.-F. & Yang, C.-C. 2010 Development and application of a decision group Back-Propagation Neural Network for flood forecasting. Journal of Hydrology 385 (1–4), 173–182.

Di Nunno, F. & Granata, F. 2020 Groundwater level prediction in Apulia region (Southern Italy) using NARX neural network. Environmental Research 190, 110062.

Foresce, F. D. & Hagan, M. T. 1997 Gauss-Newton approximation to Bayesian learning. In: Proceedings of International Conference on Neural Networks (ICNN’97). IEEE, pp. 1930–1935.

Grossmann, A. & Morlet, J. 1984 Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM Journal on Mathematical Analysis 15 (4), 723–736.

Guzman, S. M., Praz, J. O. & Tagert, M. L. M. 2017 The use of NARX neural networks to forecast daily groundwater levels. Water Resources Management 31 (5), 1591–1603.

Hirschen, K. & Schäfer, M. 2006 Bayesian regularization neural networks for optimizing fluid flow processes. Computer Methods in Applied Mechanics and Engineering 195 (7–8), 481–500.

Hornik, K., Stinchcombe, M. & White, H. 1989 Multilayer feedforward networks are universal approximators. Neural Networks 2 (5), 359–366.

Hsu, K. L., Gupta, H. V. & Sorooshian, S. 1995 Artificial neural network modeling of the rainfall-runoff process. Water Resources Research 31 (10), 2517–2530.

Hu, W. 2008 Application of statistical correlation analysis in water level forecast for Taihu River Basin. Journal of Economics of Water Resources 08 (03), 55–8 + 78.

Ju, Q., Yu, Z., Hao, Z., Ou, G., Zhao, J. & Liu, D. 2009 Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. Neurocomputing 72 (13–15), 2873–2883.

Lee, W.-K. & Tuan Resdi, T. A. 2016 Simultaneous hydrological prediction at multiple gauging stations using the NARX network for Kemaman catchment, Terengganu, Malaysia. Hydrological Sciences Journal 61 (16), 2930–2945.

Leontaritis, I. & Billings, S. A. 1985 Input-output parametric models for non-linear systems part I: deterministic non-linear systems. International Journal of Control 41 (2), 303–328.

Lin, T., Horne, B. G., Tino, P. & Giles, C. L. 1996 Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on Neural Networks 7 (6), 1329–1338.

Liu, F., Xu, F. & Yang, S. 2017 A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network. In: 2017 IEEE Third International Conference on Multimedia Big Data (BigMM). Ieee, pp. 58–61.

Loh, E. C., binti Ismail, S., Mustafa, A. & Shaharudin, S. M. 2019 Application of discrete wavelet transformation with artificial neural network in Kelantan water level prediction. Journal of Advanced Research in Dynamical and Control Systems 11 (12), 82–88.

Lu, T., Fan, B. & Zhu, L. 2000 Relation between precipitation around Taihu in the flood season and Taihu water level. Scientia Meteorologica Sinica 00 (01), 51–56.

Luo, J., Dong, Z., Guan, X. & Liu, Y. 2019 Flood risk analysis of different climatic phenomena during flood season based on copula-based Bayesian network method: a case study of Taihu Basin, China. Water 11 (8), 1534.

Mallat, S. G. 1989 A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7), 674–693.

Merwade, V., Olivera, F., Arabi, M. & Edleman, S. 2008 Uncertainty in flood inundation mapping: current issues and future directions. Journal of Hydrologic Engineering 13 (7), 608–620.
Mohanty, M. P., Nithya, S., Nair, A. S., Indu, J., Ghosh, S., Bhatt, C. M., Rao, G. S. & Karmakar, S. 2020 Sensitivity of various topographic data in flood management: implications on inundation mapping over large data-scarce regions. Journal of Hydrology, 590, 125523.

Nandalal, K. D. W. 2009 Use of a hydrodynamic model to forecast floods of Kalu River in Sri Lanka. Journal of Flood Risk Management, 2 (3), 151-158.

Nourani, V. & Parhizkar, M. 2013 Conjunction of SOM-based feature extraction method and hybrid wavelet-ANN approach for rainfall-runoff modeling. Journal of Hydroinformatics, 15 (3), 829-848.

Nourani, V., Alami, M. T. & Aminfar, M. H. 2009a A combined neural-wavelet model for prediction of Ligyanchai watershed precipitation. Engineering Applications of Artificial Intelligence, 22 (5), 466-472.

Nourani, V., Komasi, M. & Mano, A. 2009b A multivariate ANN-wavelet approach for rainfall-runoff modeling. Water Resources Management, 23 (14), 2877-2894.

Nourani, V., Kisi, Ö. & Komasi, M. 2011 Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. Journal of Hydrology, 402 (1-2), 41-59.

Nourani, V., Baghanam, A. H., Adamowski, J. & Kisi, O. 2014 Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. Journal of Hydrology, 514, 358-377.

Pappenberger, F., Matgen, P., Beven, K. J., Henry, J.-B., Pfister, L. & De, P. F. 2006 Influence of uncertain boundary conditions and model structure on flood inundation predictions. Advances in Water Resources, 29 (10), 1430-1449.

Rehman, J., Sohaib, O., Asif, M. & Pradhan, B. 2019 Applying systems thinking to flood disaster management for a sustainable development. International Journal of Disaster Risk Reduction, 36, e101101.

Remesan, R., Shamim, M. A., Han, D. & Mathew, J. 2009 Runoff prediction using an integrated hybrid modelling scheme. Journal of Hydrology, 372 (1), 48-60.

Seo, D., Sigdel, R., Kwon, K. H. & Lee, Y. S. 2010 3-D hydrodynamic modeling of Yongdam Lake, Korea using EFDC. Desalination and Water Treatment, 19 (1-3), 42-48.

Seo, Y., Kim, S., Kisi, O. & Singh, V. P. 2015 Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. Journal of Hydrology, 520, 224-243.

Sharma, V. C. & Regonda, S. K. 2021 Two-dimensional flood inundation modeling in the Godavari River Basin, India-insights on model output uncertainty. Water, 15 (2), 191.

Shoaib, M., Shamseldin, A. Y., Melville, B. W. & Khan, M. M. 2016 A comparison between wavelet based static and dynamic neural network approaches for runoff prediction. Journal of Hydrology, 535, 211-225.

Shoaib, M., Shamseldin, A. Y., Khan, S., Khan, M. M., Khan, Z. M., Sultan, T. & Melville, B. W. 2018 A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. Water Resources Management, 32 (1), 83-103.

Siddique-E-Akbor, A. H. M., Hossain, F., Lee, H. & Shum, C. K. 2011 Inter-comparison study of water level estimates derived from hydrodynamic-hydrologic model and satellite altimetry for a complex deltaic environment. Remote Sensing of Environment, 115 (6), 1522-1531.

Siegelmann, H. T., Horne, B. G. & Giles, C. L. 1997 Computational capabilities of recurrent NARX neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 27 (2), 208-215.

Taieb, S. B., Bontempi, G., Aliya, A. F. & Sorjamaa, A. 2012 A review and comparison of strategies for multi-step ahead time series forecasting based on the NNS forecasting competition. Expert Systems with Applications, 39 (8), 7067-7083.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D. & Kim, S. 2017 Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. Environmental Modelling & Software, 90, 201-216.

Timbadiya, P. V., Patel, P. L. & Porey, P. D. 2015 A 1D-2D coupled hydrodynamic model for river flood prediction in a coastal urban floodplain. Journal of Hydrologic Engineering, 20 (2), 05014017.1-18.

Wang, T. 2018 Analysis of 2016 major floods in Taihu Basin. China Flood & Drought Management, 28 (6), 60-62.

Wang, J. & Dong, Z. 2003 On the forecasting of Pingwang water level and Mishidu tide level hydrograph in Taihu Basin. Journal of Lake Sciences, 03 (03), 229-235.

Wang, J., Dong, Z. & Wu, H. 2001 Global optimum approximation of feed forward artificial neural network for Taihu flood forecast. Journal of Hohai University (Natural Sciences), 01 (04), 84-90.

Wang, L., Hu, Q., Wang, Y., Lin, H., Cui, T. & Zhu, X. 2018 Comparison of the two flood events occurred in 2016 and 1991 in the Taihu Lake Basin. Journal of Hohai University (Natural Sciences), 046 (006), 471-478.

Xu, J. 2009 Analysis on the losing and cause of flood disaster in Taihu drainage area. Shanxi Architecture, 35 (03), 359-360.

Young, C.-C., Liu, W.-C. & Hsieh, W.-L. 2015 Predicting the water level fluctuation in an alpine lake using physically based, artificial neural network, and time series forecasting models. Mathematical Problems in Engineering, 2015, 708204.

Zeng, Z., Liang, J., Yang, M., Zeng, Z., Lang, Y., Lei, X., Cai, S., Yu, Y. & Li, H. 2018 A hydrodynamic model coupled multiple scenarios for plain river network and its application. MATEC Web of Conferences, 246, 01016.

Zhai, Y., Wang, C., Chen, G., Wang, C., Li, X. & Liu, Y. 2020 Field-based analysis of runoff generation processes in humid lowlands of the Taihu Basin, China. Water, 12 (4), 1216.

Received 18 January 2021; accepted in revised form 20 April 2021. Available online 1 June 2021.