Hydrol. Earth Syst. Sci. Discuss., 10, 9239–9269, 2013
www.hydrol-earth-syst-sci-discuss.net/10/9239/2013/
doi:10.5194/hessd-10-9239-2013

Hydrology and Earth System Sciences Discussions
Open Access

Multi-step-ahead predictor design for effective long-term forecast of hydrological signals using a novel wavelet-NN hybrid model

J.-S. Yang[1,2], S.-P. Yu[1,2], and G.-M. Liu[1]

[1]State Key Laboratory of Soil and Sustainable Agriculture, Nanjing Institute of Soil Science, Chinese Academy of Sciences, 71 East Beijing Road, Nanjing 210008, China
[2]Dongtai Institute of Tidal Flat, Nanjing Branch of the Chinese Academy of Sciences, 8 Beihai Road, Dongtai 224200, China

Received: 23 June 2013 – Accepted: 24 June 2013 – Published: 15 July 2013

Correspondence to: S.-P. Yu (spyu@issas.ac.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

In order to increase the accuracy of serial-propagated long-range multi-step-ahead (MSA) prediction, which has high practical value but also great difficulty to conduct because of huge error accumulation, a novel wavelet-NN hybrid model CDW-NN, combining continuous and discrete wavelet transforms (CWT and DWT) and neural networks (NN), is designed as the MSA predictor for effective long-term forecast of hydrological signals. By the application of 12 types of hybrid and pure models in estuarine 1096 day river stage series forecasting, different forecast performances and the superiorities of CDW-NN model with corresponding driving mechanisms are discussed, and one type of CDW-NN model (CDW-NF), which uses Neuro-Fuzzy as the forecast submodel, has been proven to be the most effective MSA predictor for the accuracy enhancement in the overall 1096 days long-term forecast. The special superiority of CDW-NF model lies in the CWT based methodology, which determines the 15 and 28 day prior data series as model inputs by revealing the significant short-time periodicities involved in estuarine river stage signals. Comparing conventional single-step-ahead based long-term forecast models, the CWT based hybrid models broaden the prediction range in each forecast step from 1 day to 15 days, thus reduce the overall forecasting iteration steps from 1096 steps to 74 steps and finally creates significant decrease of error accumulations. In addition, combination of the advantages of DWT method and Neuro-Fuzzy system also very benefit filtering the noisy dynamics for model inputs and enhancing the simulation and forecast ability of the complex hydro-system.

1 Introduction

Hydrological signal forecasts, especially a long-term forecast, are important for the study and guidance of water resource management. Nevertheless, hydrological signals are highly complex nonlinear systems and have severe variations in time and space, which make accurate forecasts difficult. Generally, the hydrological time series are predicted with models based on physical considerations or other numerical theories, such as LR (linear regressive) analysis methods (Salas et al., 1980) based on stochastic theory; grey models (Deng, 1992) based on grey information theory; chaos models (Jayawardena and Lai, 1994; Islam and Sivakumar, 2002) based on a local similarity of signals; fuzzy prediction models (Jang, 1993; Jang et al., 1997; Chen, 2005) based on fuzzy theory; TAR (threshold auto-regression), BL (bilinear time series), and SVM (Support Vector Machine) models (Tong, 1990; Jin and Ding, 2002; Liong and Sivapragasm, 2002; Zou et al., 2010) based on nonlinear time series analysis; ANN (artificial neural networks) models (Raman and Sunlikumar, 1995; Yu et al., 2008; Yang et al., 2009) based on black-box theory; and NNB (nearest neighbour bootstrapping) regressive models (Wang et al., 2001) based on nonparametric prediction theory. However, these models are generally not successful enough in producing accurate predictions due to some inaccurate initial conditions, parameterisation schemes of sub-scale phenomena, and limited spatial resolution (Olson et al., 1995).

Many hybrid models have been proposed as predictors to improve the accuracy of hydrological time series forecasts, such as the wavelet-ANN model (Anctil and Tape, 2004), the periodic ANN (PANN) model (Wang et al., 2006), the chaotic-ANN model (Karunasinghe and Liong, 2006), the wavelet-based grey model (Chou, 2007), the wavelet-based NF (Neuro-Fuzzy) model (Partal and Kisi, 2007; Engin et al., 2007; El-Shafie et al., 2007), the non-supervised ANN-EA (evolutionary algorithms) model (Cao and Park, 2007; Chang et al., 2007), the fuzzy-SVM model (Hua et al., 2008), the wavelet-based multi-layer perceptron model (Kisi, 2008), and the wavelet-regression (WR) model (Kisi, 2011). These hybrid models have shown different advantages for accurate predictions due to their capabilities of utilising present information effectively. Among these hybrid models, the neural network (NN) models, such as NF (Neuro-Fuzzy) and ANN, are the most popularly utilised sub-models for signal forecast due to their capabilities of effectively learning complex and nonlinear relationships (Maier et al., 2010). The ANN model has been popularly used in hydrological signal forecasts in recent years by a number of researchers (French et al., 1992; Jain et al., 1999; ASCE,

HESSD
10, 9239–9269, 2013

2000; Cigizoglu, 2005; Marzano et al., 2006; Zou et al., 2010). The NF model has been successfully used in the hydrological sciences in recent years (Nayak et al., 2004; Kisi, 2005; Chang and Chang, 2006). In addition, the wavelet transform is a strong mathematical tool that provides a good local representation of the signal in both the time and frequency domains, and it has become a useful method for analysing variations, periodicities, and trends in time series (Daubechies, 1994; Torrence and Compo, 1998; Coulibaly and Burn, 2004; Partal and Kucuk, 2006). Among the various types of wavelet transforms, the discrete wavelet transform (DWT) as the data preprocessing method in a hybrid model is popularly used to decompose the original signal input due to its capabilities of effectively classifying a hydro-meteorological time series into distinct time and frequency domains (Smith et al., 1998; Kim and Valdes, 2003; Labat, 2005).

Because of the common Markovian property (Bolch et al., 2006) embedded in hydro-meteorological time series, most recent pure and hybrid models use data series at different previous time points as model inputs to forecast the original data series at the current time point. For daily time series, the data series from one day prior to a few days prior are usually used as model inputs, namely using data series S_{t-1}, S_{t-2}, ... as inputs, to forecast S_t. The data series at one day prior is always selected as one of the inputs because of the usually high lag-1 autocorrelation (Kisi, 2008, 2011; Zhou et al., 2008). This selection principle denotes a type of popular used single-step-ahead (SSA) prediction (Parlos et al., 2000), in which each single forecasting step of the Markovian property-based model can only predict the next one-day datum (Fig. 1a). However, SSA prediction may not provide enough information, especially in the situation in which it is desirable to understand the behaviour of multiple steps in the future, such as signal processing and time series prediction. Given this issue, the serial-propagated multi-step-ahead (MSA) prediction (Fig. 1b), which attempts to make predictions several time steps into the future without the availability of output measurements, has attracted an increasing number of scientific studies (Su et al., 1992; Schenker and Agarwal, 1995; Coulibaly et al., 2000; Gao et al., 2002; Chang et al., 2007, 2012; Yong et al., 2010). However, MSA predictors are difficult to develop because the lack of measurements in the prediction horizon necessitates the recursive use of SSA predictors to reach the end-point on the horizon, especially difficult for a long-range MSA predictor design. Even small SSA prediction errors at the beginning of the horizon accumulate and propagate, often resulting in a poor prediction accuracy. Over the last twenty years, MSA predictor design to increase the MSA prediction accuracy has received much attention, and different types of neural networks have been used successfully for some short-range MSA predictions (Su et al., 1992; Parlos, et al., 2000; Chang et al., 2007).

As mentioned above, the crucial and most difficult point in a long-term forecast of hydrological signal is the development of effective models to reduce the error accumulation and increase the accuracy of the long-range serial-propagated MSA prediction. In view of this, the present study designed a novel hybrid model CDW-NN, combining continuous and discrete wavelet transforms and neural networks, as MSA predictor for effective long-term forecast of hydrological signals by broadening the prediction range in each forecast step and reducing the total iteration steps in the long-term forecasting process. In the remainder of this paper, the long-term forecast methodologies of the MSA predictor CDW-NN are presented. In the next section, the details of daily river stage data series in different hydro-stations in Yangtze River Estuary, China are presented and CDW-NN hybrid models are applied to the long-term forecasts of different river stage signals. The results are discussed by comparing with the performances of other pure and hybrid models in the subsequent section, and finally, conclusions are drawn.

2 Methodologies

2.1 Continuous wavelet transform (CWT) and discrete wavelet transform (DWT)

Wavelet transform is a mathematical tool that allows the decomposition of the signal $f(t)$ in terms of elementary contributions called wavelets (Sadowskey, 1996; Labat et al., 2005). For the time series $f(t) \in L^2(R)$ or finite energy signal, the continuous

HESSD
10, 9239–9269, 2013

wavelet transform (CWT) of the signal $f(t)$ with the analysing wavelet ϕ is the convolution of $f(t)$ with a set of dilated and translated wavelets:

$$W_f(a,b) = \langle f(t), \varphi_{a,b}(t) \rangle = \sqrt{\frac{\delta t}{a}} \int_R f(t) \overline{\varphi}\left(\frac{t-b}{a}\right) \mathrm{d}t, a, b \in R, a > 0 \tag{1}$$

where $\overline{\phi}(t)$ is the complex conjugate function of $\phi(t)$, a is the dilation (scale or frequency) parameter, b is the translation (position or time) parameter, R is the domain of real numbers, and δt is the time interval of the data series. In this paper, the time interval of the data series equals 1.0 day, and the popularly used Morlet wavelet is selected as ϕ (Mallat, 1989; Daubechies, 1994; Torrence and Compo, 1998). The Morlet wavelet, which is a complex wavelet consisting of a plane wave modulated by a Gaussian function, is defined by:

$$\phi(t) = \pi^{-1/4} e^{i\omega_0 t} e^{(-t^2/2)} \tag{2}$$

where ω_0 is the non-dimensional frequency (usually taken to be 6 to satisfy the admissibility condition) (Farge, 1992).

The global wavelet power spectrum is defined as the power density at different time scale a, which is calculated by:

$$E_a = \frac{1}{N} \sum_{b=1}^{N} |W_f(a,b)|^2 \tag{3}$$

where N is the length of the data. The signal's periodicity can be indicated at the time scale at which the wave crest of wavelet power spectrum is observed. The significance of the global wavelet power spectrum is tested using a white or red noise model by compared with the theoretical global wavelet power spectrum (P). P is given as (Torrence and Compo, 1998):

$$P = \sigma^2 P_a \frac{x_v^2(p)}{v} \tag{4}$$

where σ^2 is the variance of data series, $x_v^2(p)$ is the inverse of chi-square cumulative distribution with v degrees of freedom at the requested confidence level $1-p$, and p is the distribution fraction. For the lag-1 autocorrelation, $r(1) < 0.1P_a$ is the white noise spectrum, and for $r(1) > 0.1, P_a$ is the red noise spectrum. For the Morlet wavelet, P_a is given as Eq. (5), and v is given as Eq. (6). In this study, a significance level of 0.005 was selected, e.g., $\chi_2^2(99.5\,\%) = 10.597$.

$$P_a = \frac{1 - r(1)^2}{1 + r(1)^2 - 2r(1)\cos\left(\frac{2\pi\delta t}{1.033a}\right)} \tag{5}$$

$$v = 2\sqrt{1 + \left(\frac{N\delta t}{2.32a}\right)^2} \tag{6}$$

The continuous wavelet (Eq. 1) is often discrete in real applications. When $a = a_0^j$, $b = kb_0a_0^j$, $a_0 > 1$, $b_0 \in R$, and k and j are integer numbers, the Discrete Wavelet Transform (DWT) of $f(t)$ can be written as:

$$W_f(j,k) = \frac{1}{\sqrt{a_0^j}} \int_R f(t)\overline{\phi}\left(a_0^{-j}t - kb_0\right) \mathrm{d}t \tag{7}$$

Based on the commonly used Mallat algorithm for calculating discrete wavelet coefficients, the most common and simplest choice for the parameters a_0 and b_0 is 2 and 1 time steps, respectively, and the Daubechies wavelet, which has no explicit mathematical expressions and can be calculated only numerically, is commonly used in the DWT (Mallat, 1989; Daubechies, 1994; Partal and Kucuk, 2006; Kisi, 2011). For a discrete time series $f(t)$ occurring at different times t (e.g., integer time steps are used herein), the DWT can be defined as:

$$W_f(j,k) = \frac{1}{\sqrt{2^j}} \sum_{t=0}^{N-1} f(t)\overline{\phi}(2^{-j}t - k) \tag{8}$$

HESSD
10, 9239–9269, 2013

where N is the number of discrete time steps and $W_f(j, k)$ is the wavelet coefficient for the discrete wavelet of scale $a = 2^j$ and time $b = 2^j k$.

2.2 Neuro-Fuzzy (NF) and BP-ANN

The popular neural network (NN) model Neuro-Fuzzy (NF), based on the Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang et al., 1997; Partal and Kisi, 2007), is utilised as the sub-model for the different hydro-meteorological signals forecast in this paper. ANFIS, first introduced by Jang (1993), is a universal approximator and, as such, is capable of approximating any real continuous function on a compact set to any degree of accuracy. The ANFIS is functionally equivalent to the Sugeno first-order fuzzy model (Jang et al., 1997; Drake, 2000), and its typical architecture with five learning layers is shown in Fig. 2. As mentioned by Jang et al. (1993), Partal and Kisi (2007), and Engin et al. (2007), two types of bell-shaped functions are generally used as transfer function in Layer 1 of the ANFIS architecture. The significant advantage of the NF model depends on the hybrid learning algorithm in ANFIS, which combines gradient descent, back-propagation, and the least-squares method and can rapidly train and adapt the ANFIS. Each learning epoch of the ANFIS is composed of a forward pass and a backward pass, and more information for Neuro-Fuzzy and ANFIS can be found in Jang's papers (Jang, 1993; Jang et al., 1997).

Another popular NN model, BP-ANN (back-propagation artificial neural networks), is utilised in our case to compare the forecast performance with the NF model. Based on the back-propagation algorithm, a common three-layer feed-forward type of BP-ANN is considered, the Levenberg–Marquardt methodology, which is more powerful than conventional gradient descent techniques (Hagan and Menhaj, 1994; Kisi, 2011), to adjust the weights of the ANN model, and the Tangent Sigmoid and linear activation functions are used for the hidden and output node(s), respectively. Because there is no theory yet to determine how many hidden layer nodes in the BP-ANN are needed to approximate any given function, the hidden layer node number in BP-ANN is commonly determined by the trial and error approach.

2.3 Architecture of the long-term forecasting based on the MSA predictor CDW-NN

In order to reduce the error accumulation and increase the accuracy of the long-range serial-propagated MSA prediction, the present study designed a novel hybrid model CDW-NN, combining CWT, DWT, and NN, as MSA predictor for effective long-term forecast of hydrological signals. The architecture of CDW-NN hybrid model was shown in Fig. 3. Firstly, for the original given daily data series $x(1) \sim x(t)$, the CWT method was utilized to reveal its short-term periodicities, i.e. periods at $a1 \sim ai$ days in Fig. 3 ($a1 < a2 < \ldots < ai$). Meanwhile, decomposition of the original signal DWT was conducted to get new data series $\mathrm{TD}(1) \sim \mathrm{TD}(t)$, which was constructed by selecting and combining optimal DWT decomposition components. Then, by combining the CWT and DWT results, the new TD series at $a1 \sim ai$ days ahead ($\mathrm{TD}(t - a1) \sim \mathrm{TD}(t - ai)$) were selected as NN model inputs for model training to forecast the datum $x(t)$. According to the serial-propagated prediction principle, using $\mathrm{TD}(t - a1 + 1) \sim \mathrm{TD}(t - ai + 1)$ as model inputs can predict the first future day datum $y(t + 1)$, and using $\mathrm{TD}(t) \sim \mathrm{TD}(t - ai + a1)$ as inputs can predict $y(t + a1)$. Here, the first batch of outputs $y(t + 1) \sim y(t + a1)$ are predicted from the first forecasting step, and then can be used as new observations for the second step DWT decomposing and NN forecasting to predict the second batch of outputs $y(t + a1 + 1) \sim y(t + 2a1)$. Just as the one day prediction from each SSA forecast involved in conventional MSA forecast, the day number of predictions in the output of each CDW-NN forecasting step is $a1$. So, after about $n/a1$ steps of forecasting process, the final long-term prediction series $y(t + 1) \sim y(t + n)$ can be obtained.

HESSD
10, 9239–9269, 2013

3 Application

3.1 Studied area and data

The hightide level data at two time points each day during 13 yr (4748 days) covering 1998–2010, supported by the Water Resources Department of Jiangsu Province, were observed and collected from estuarine Santiao Port Hydrologic Station (31.721° N, 121.698° E) and Qinglong Port Hydrologic Station (31.862° N, 121.239° E), which respectively locate about 19.0 km and 70.0 km upstream from Chinese Yangtze River entrance into East China Sea (Fig. 4). The daily river stage data series at each station was obtained based on the average value of hightide levels at two time points each day. The first 10 yr of river stage data (3652 days) was used for training and establishing hybrid models, i.e. $t = 3652$ in Fig. 3. The remaining 3 years of river stage data (1096 days) was used for testing the long-term forecasting performance of hybrid models, i.e. $n = 1096$ in Fig. 3.

3.2 Short-term periodic features of estuarine daily river stage series

Morlet wavelet transform coefficients of the training data series at relatively fine time scales (from one day to fifty days scales) were calculated by MATLAB language. The real parts of Morlet wavelet transform coefficients at Santiao Station (Fig. 2a) and Qinglong Station (Fig. 2c) clearly indicated the distribution conditions of the river stage signals for different time scales. The solid isograms indicate positive wavelet coefficients and a relatively high river stage period, and the dashed isograms indicate negative coefficients and a relatively low period. Further calculating the global wavelet power spectrums and the corresponding theoretical power spectrums using white noise model, the significances of wavelet power densities of daily river stage series at different time scales at Santiao and Qinglong Stations were calculated and shown in Fig. 2b and d.

Results showed that the Morlet wavelet transform coefficients of daily river stage series at Santiao Station generated obvious two kinds of quasi periodic oscillations

HESSD
10, 9239–9269, 2013

(QPOs), namely at 12 day and 23 day time scales, and both of their global wavelet power spectrums were prominent at the 99.5 % confidence level. The QPO of estuarine daily river stage at a fine time scale was often nested in a broad time scale. At the 12 day time scale, the average changing periodicity (T) of river stage time series was 15 days obtained by calculating and averaging the day numbers of each two neighboring high and low river stage periods. At the 23 day scale, the average T was 28 years. At Qinglong Station, the Morlet wavelet transform coefficients of daily river stage series generated obvious two QPOs at 12 day and 22 day time scales, of which corresponding T were 15 days and 28 days same as that at Santiao Station, and both of their global wavelet power spectrums were prominent at the 99.5 % confidence level. Based on the prominent short-term periodic features of estuarine daily river stage time series, estuarine daily river stage at 15 day prior and 28 day prior were determined to simulate and forecast the river stage at the current day, i.e. $a1$ and $a2$ in Fig. 3 equaled 15 and 28 respectively.

3.3 Decomposition of daily river stage time series and optimal DWT components combination

The decomposition process of DWT consists of a number of filtering steps following the Mallat algorithm. The original signal of training data series is first decomposed into an approximation (A_1) and details (D_1), and the A_1 is then broken down into many lower-resolution components (A_i and D_i). The details are the low-scale high-frequency components of the signal, while the approximations are the high-scale low-frequency components. The higher scales consist of the extended version of a wavelet, and the corresponding coefficients refer to the slowly changing coarse features of low-frequency components. The lower scales present the condensed wavelet and follow the rapidly changing details (high frequency components) of the signal (Mallat, 1989). In our case, ten decomposition components (A_{10}, D_1–D_{10}) of the original daily river stage signal (1998–2007) in ten resolution levels were calculated by MATLAB language, and were used to analyze the optimum input factors in our hybrid model.

The correlation coefficients (R) between each discrete wavelet component (A_{10}, $D_1 - D_{10}$) at 15 day and 28 day prior ($t-15$ and $t-28$) and original time series at time $t(S_t)$ were computed and presented in Table 1. By single factor and double factors analysis, at both Santiao and Qinglong Stations the D_3, D_4 and D_8 components at 15 day and 28 day prior showed prominently higher R with S_t than the other DWT components, especially significant in double factors analysis. Instead of using each DWT component individually as the model input, employment of the added suitable DWT components is more useful and can highly increase the forecast performance. Based on the revealed dominant DWT components of different hydrological series, the new series (TD) obtained by adding D_3, D_4 and D_8 at 15 day and 28 day prior were selected as two NN model inputs for the daily river stage forecast at both Santiao and Qinglong Stations. Comparing each DWT component with the original series (S), the new series (TD) showed significantly higher correlations at both 15 day and 28 day delay time nodes with S_t.

3.4 CDW-NN model training and a long-term forecasting of daily river stage signal

According to the above CWT and DWT results, two new daily series TD(1) ~ TD($t-28$) and TD(14) ~ TD($t-15$) extracted from the training data series were used as NN model inputs to simulate and forecast the original series $x(29) \sim x(t)$. Program codes were written in MATLAB language for training the Neuro-Fuzzy and BP-ANN submodels and determine their optimal model structures. At Santiao Station, by many trials the optimal CDW-ANN hybrid model structure was determined as CDW-ANN(2-3-1), which denotes two input layer nodes, three hidden layer nodes and one output layer node in BP-ANN submodel. At Qinglong Station the optimal CDW-ANN(2-4-1) model was obtained. When the training programs of NN submodels were done, two TD series TD($t-27$) ~ TD($t-13$) and TD($t-14$) ~ TD(t) were used as model inputs to forecast the future 15 days river stage series $y(t+1) \sim y(t+15)$, i.e. the first forecasting step in Fig. 3. The predicted 15 day river stage series was then treated by DWT and used as new inputs to forecast the next 15 days river stage in the next forecasting step. Because each single step forecasting using CDW-NN model obtained 15 days predictions, the total 1096 days future river stage values from 2008–2010 were predicted after 74 steps of forecasting.

The root mean square errors (RMSE), mean absolute errors (MAE) and correlation coefficient (R) statistics were used to evaluate the model performance of simulation and prediction. The RMSE and MAE are defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y i_{\text{observed}} - Y i_{\text{estimate}})^2} \tag{9}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y i_{\text{observed}} - Y i_{\text{estimate}}| \tag{10}$$

where n is the number of data sets, and $Y i$ is the daily river stage.

The forecast performances of CDW-NF and CDW-ANN models at Santiao and Qinglong Stations were shown in Fig. 4. Results showed that the CDW-NF models performed significantly better correlations between observed and predicted river stage data during 2008–2010 with the higher R of 0.533 and 0.283 at Santiao and Qinglong Stations respectively, while CDW-ANN with the lower R of −0.142 and 0.172 at Santiao and Qinglong Stations respectively. And the CDW-ANN hybrid model showed better forecast performances during the first year than that in the last two years.

4 Discussion

4.1 Forecast performances comparison between CDW-NN models and the other 10 types of hybrid and pure models

Similar to the establishing process of CDW-NN models, the hybrid CDW-LR model was established by combining CWT, DWT and a linear regression (LR) model. Using the

HESSD
10, 9239–9269, 2013

training data series, simulating equations of the hybrid models (CDW-LR) at Santiao Station and Qinglong Station were obtained and shown as Eqs. (11) and (12) respectively. Without the DWT treatment, three other CWT based (CW-) hybrid models were established by using the original daily river stage series at 15 day and 28 day previous (S_{t-15} and S_{t-28}) as input factors to simulate S_t. By model training, the optimal CW-ANN structures for Santiao and Qinglong Stations were determined as CW-ANN(2-3-1) and CW-ANN(2-1-1) respectively, and the CW-LR models for Santiao and Qinglong Stations were obtained and showed as Eqs. (13) and (14) respectively.

$$S_t = 0.454\mathrm{TD}_{t-28} + 0.533\mathrm{TD}_{t-15} + 3.966 \quad (11)$$

$$S_t = 0.496\mathrm{TD}_{t-28} + 0.487\mathrm{TD}_{t-15} + 4.004 \quad (12)$$

$$S_t = 0.428S_{t-28} + 0.493S_{t-15} + 0.315 \quad (13)$$

$$S_t = 0.473S_{t-28} + 0.460S_{t-15} + 0.266 \quad (14)$$

As mentioned in Fig. 1, many conventional studies on MSA predictor design focused on the methodology of recursive use of SSA predictors, e.g. generally using S_{t-1} and S_{t-2} as model inputs to forecast S_t. In view of this, 6 types of conventional SSA based long-term forecast hybrid and pure models were established for comparing with the CWT based hybrid models. Among the 6 types of models, DW-R, DW-ANN and DW-NF hybrid models utilized the optimum decomposition components combinations (TD_{t-1} and TD_{t-2}), determined by DWT, as model inputs to forecast S_t. With respect to the pure LR, BP-ANN and Neuro-Fuzzy models, the original daily river stage series at 1 day prior and 2 day prior (S_{t-1} and S_{t-2}) were used as model inputs to forecast S_t. By model training, the optimal DW-ANN structures for daily river stage forecasts at Santiao and Qinglong Stations were determined as DW-ANN(2-5-1) and DW-ANN(2-7-1) respectively, and the optimal structures of pure BP-ANN models for daily river stage forecasts at Santiao and Qinglong Stations were determined as BP-ANN(2-3-1) and BP-ANN(2-5-1) respectively. The DW-LR models for Santiao and Qinglong Stations were obtained and shown as Eqs. (15) and (16) respectively, and the pure LR models for Santiao and Qinglong Stations were obtained and shown as Eqs. (17) and (18)

HESSD
10, 9239–9269, 2013

respectively.

$$S_t = -0.973\mathrm{TD}_{t-2} + 1.818\mathrm{TD}_{t-1} + 3.966 \quad (15)$$

$$S_t = -0.971\mathrm{TD}_{t-2} + 1.826\mathrm{TD}_{t-1} + 4.005 \quad (16)$$

$$S_t = -0.316S_{t-2} + 1.124S_{t-1} + 0.761 \quad (17)$$

$$S_t = -0.397S_{t-2} + 1.234S_{t-1} + 0.649 \quad (18)$$

As shown in Table 2, the training and forecasting performances of 12 types of hybrid and pure models were compared with each other in respect to RMSE, MAE and R statistics. Due to the high lag-1 and lag-2 autocorrelations in hydrological time series, in the training periods the 6 types of conventional SSA based long-term forecast models performed better than the 6 types of CWT based hybrid models with higher R and smaller RMSE and MAE. Nevertheless, in the test periods, without the observation data as model inputs in each forecasting step, the CDW-NF hybrid model showed significant performances among all the 12 models, especially prominent at Santiao Station. In addition, due to the prominent ability of decomposition approach based on DWT in filtering weak correlated details from original signal, each DWT based hybrid model performed better than its corresponding model without DWT both in training and test periods.

4.2 Driving mechanism of advantages of the CDW-NN models on long-range MSA predictions

Prediction performance details in respect to R for the 12 types of hybrid and pure models at different forecasting steps during the overall 1096 day river stages forecasting were calculated and shown in Fig. 7. According to the serial-propagated MSA prediction theory, the error accumulation increases with the iteration steps increase in a long-term forecast. Therefore, the prediction performances of all 12 types of models had overall decreasing trends with the increasing of predicted data length to 1096 days. Nevertheless, during approximately the first 200 days ~ 600 days river stage forecasting all the 6 types of CWT based hybrid models performed better than the other 6 types of SSA based models. In particular, at Santiao Station the CDW-NF model showed significantly better performance than the other models covering the overall 3 yr (1096 days) river stage forecasting steps. With respect to the approximately last 500 days river stage forecasting at Qinglong Station, the CDW-NF model shared better performances with the 3 types of SSA based DW-hybrid models than the other models. The main explanations are that the CWT based models reduce the overall forecasting iteration steps to 74 steps by using the 15 day prior data series as the first model input, while the conventional SSA based models needs 1096 steps by using 1 day prior data series as the first model input. The prominent decrease of forecasting steps consequently brings significant reduction of error accumulation in the long-range MSA prediction. In addition, the combination of the advantages of the DWT method and Neuro-Fuzzy system also benefits weakening the noisy dynamics for the model inputs and enhancing the simulation and forecast ability of the complex hydro-system.

In view of the above discussion, the CDW-NF hybrid model is proven as an effective MAS predictor for long-term forecast of estuarine daily river stage signals. In addition, the CDW-ANN hybrid model can be taken as the second selection for shot-term and mid-term forecasts of estuarine daily river stage because of its high performance during the first year forecast process. It should be noted that the methodology of accurate MSA predictor design by reducing iteration steps and error accumulation is innovated in this study by revealing the short-time periodic features of estuarine river stage dynamics, which is mainly caused by the half-month periodicity involved in astronomical tidal fluctuation in river estuary. With respect to other kinds of hydro-meteorological signals, which have non-significant short-term periodic features, other kinds of algorithms and models for reducing error accumulation in long-term forecasting steps need further study in future research.

HESSD
10, 9239–9269, 2013

5 Conclusions

Studies on the long-term forecast of hydrological signals have high practical value, but the accurate long-range MSA predictor design is very difficult, especially in conventional SSA based MSA predictions, because of the huge error accumulation in serial-propagated long-term forecast. In this study, we designed a novel hybrid model CDW-NN, combining continuous and discrete wavelet transforms and neural networks, as MSA predictor for effective long-term forecast of hydrological signals. By the application of CDW-NN hybrid models and the other 10 types of hybrid and pure models in estuarine daily river stage series long-term forecasting, the 1096 days estuarine river stage data were forecasted, and the superiorities of CDW-NN models with corresponding driving mechanisms were proven as follows:

1. Comparing conventional SSA based models, the CWT based hybrid models broadened the prediction range in each forecast step from conventional 1 day to now 15 days and reduced the overall forecasting iteration steps from conventional 1096 steps to now 74 steps by using the 15 and 28 day prior data series as model inputs, which was determined by revealing signal's significant short-time periodicities from CWT. This prominent reduction of forecast steps has created significant decrease of error accumulations and increase of long-term forecast performances in the CWT based hybrid models.

2. Among the CWT based models, one type of CDW-NN model (CDW-NF) has been proven to be the most effective MSA predictor for the accuracy enhancement in the overall 1096 days long-term forecast of estuarine hydro-signal. The other type of CDW-NN model (CDW-ANN) has been proven to be the second selection for shot-term and mid-term forecasts of estuarine hydro-signal. The main explanation is the combination of the advantages of the CWT and DWT methods and Neuro-Fuzzy system in reducing the error accumulation, filtering weak correlated details from original signal and weakening the noisy dynamics for the model inputs, and enhancing the simulation and forecast ability of the complex hydro-system.

HESSD
10, 9239–9269, 2013

3. It should be noted that because the successful application of the novel CDW-NF model in hydro-signal long-term forecast largely depends on the significant short-term periodicities involved in estuarine hydro-signals, some other innovative algorithms and models still need to be further studied in future research for other kinds of hydro-meteorological signals without significant short-term periodic features.

Acknowledgements. This study is financially supported by the National Natural Science Foundation of China (No. 41101518), the National Basic Research Program of China (No. 2010CB429001), the Natural Science Foundation of Jiangsu Province of China (No. BK2010610 and No. BK2011425), and the National Non-Profit Research Program of China (No. 200903001 and No. 2011418020). Cordial thanks should be extended to the editors and reviewers for their invaluable comments, which greatly improved the quality of this paper.

References

Anctil, F. and Tape, D. G.: An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition, J. Environ. Eng. Sci., 3, S121–S128, 2004.

ASCE Task Committee: Artificial neural network in hydrology, J. Hydrol. Eng. ASCE , 5, 124–144, 2000.

Bolch, G., Greiner, S., de Meer, H., and Trivedi, K. S.: Queueing Networks and Markov Chains, John Wiley, New York, 2006.

Cao, H. Q. and Park, H. D.: Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms, Water Res., 41, 2247–2255, 2007.

Chang, F. J. and Chang, Y. T.: Adaptive neuro-fuzzy inference system for prediction of water level in reservoir, Adv. Water Res. , 29, 1–10, 2006.

Chang, F. J., Chiang, Y. M., and Chang, L. C.: Multi-step-ahead neural networks for flood forecasting, Hydrolog. Sci. J., 52, 114–130, 2007.

Chang, L. C., Chen, P. A., and Chang, F. J.: A Reinforced Two-Step-Ahead Weight Adjustment Technique for On-Line Training of Recurrent Neural Networks, IEEE T. Neural Networ. Learn. Syst. , 23, 1269–1278, 2012.

Chen, S. Y.: Theories and methods of variable fuzzy sets in water resources and flood control system, Dalian University of Technology Press, Dalian, 2005 (in Chinese).
Chou, C. M.: Applying multi-resolution analysis to differential hydrological grey models with dual series., J. Hydrol. (Amst) 332, 174–186, 2007.
Cigizoglu, H. K.: Application of the generalized regression neural networks to intermittent flow forecasting and estimation., J. Hydrol. Eng. ASCE , 10, 336–341, 2005.
Coulibaly, P. and Burn, H. D.: Wavelet analysis of variability in annual Canadian streamflows, Water Resour. Res., 40, W03105, doi:10.1029/2003WR002667, 2004.
Coulibaly, P., Anctil, F., and Bobee, B.: Neural network-based long-term hydropower forecasting system, Comput.-Aided Civ. Inf., 15, 355–364, 2000.
Daubechies, I.: Ten lectures on wavelets, CBMS, SIAM , 61, 194–202, 1994.
Deng, J. L.: Grey prediction and grey decision, Hua Zhong Li Gong University Press, Wuhan, 1992 (in Chinese).
Drake, J. T.: Communications Phase Synchronization Using the Adaptive Network Fuzzy Inference System, PhD Thesis, New Mexico State University, Las Cruces, New Mexico, 2000.
El-Shafie, A., Taha, M. R., and Noureldin, A.: A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam, Water Resour. Manag., 21, 533–556, 2007.
Engin, A., Davut, H., and Asaf, V.: An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition, Expert Syst. Appl., 33, 582–289, 2007.
Farge, M.: Wavelet transforms and their applications to turbulence, Annu. Rev. Fluid Mech., 24, 395–457, 1992.
French, M. N., Krajewski, W. F., and Cuykendal, R. R.: Rainfall forecasting in space and time using a neural network, J. Hydrol., 137, 1–37, 1992.
Gao, X. Z., Ovaska, S. J., and Vasilakos, A. V.: Temporal difference method-based multi-step-ahead prediction of long term deep fading in mobile networks, Comput. Commun., 25, 1477–1486, 2002.
Hagan, M. T. and Menhaj, M. B.: Training feed forward networks with the Marquardt algorithm, IEEE Trans. Neural Netw. , 6, 861–867, 1994.
Hua, B., Xiong, W., and Cheng, H.: Application of fuzzy support vector machine to runoff forecast, Eng. J. Wuhan Univ. , 41, 5–8, 2008 (in Chinese).
Islam, M. N. and Sivakumar, B.: Characterization and prediction of runoff dynamics: a nonlinear dynamical view, Adv. Water Resour., 25, 179–190, 2002.

Jain, S. K., Das, D., and Srivastava, D. K.: Application of ANN for reservoir inflow prediction and operation, J. Water Resour. Plan. Manag. , ASCE 125, 263–271, 1999.
Jang, J. S. R.: ANFIS: adaptive-network-based fuzzy inference system, IEEE T. Syst. Man. Cy. , 23, 665–685, 1993.
Jang, J. S. R., Sun, C. T., and Mizutani, E.: Neuro-Fuzzy and Soft Computing: a Computational Approach to Learning and Machine Intelligence, Prentice Hall, Upper Saddle River, New Jersey, 1997.
Jayawardena, A. W. and Lai, F.: Analysis and prediction of chaos in rainfall and streamflow time series, J. Hydrol., 153, 23–52, 1994.
Karunasinghe, D. S. K. and Liong, S. Y.: Chaotic time series prediction with a global model: artificial neural network, J. Hydrol., 153, 23–52, 2006.
Kim, T. W. and Valdes, J. B.: Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks, J. Hydrol. Eng. ASCE , 6, 319–328, 2003.
Kisi, O.: Suspended sediment estimation using neuro-fuzzy and neural network approaches, Hydrol. Sci. J. , 50, 683–696, 2005.
Kisi, O.: Stream flow forecasting using neuro-wavelet technique, Hydrol. Process, 22, 4142–4152, 2008.
Kisi, O.: Wavelet Regression model as an alternative to neural networks for river stage forecasting, Water Resour. Manag., 25, 579–600, 2011.
Labat, D.: Recent advances in wavelet analyses: Part 1, A review of concepts., J. Hydrol. 314, 275–288, 2005.
Labat, D., Ronchail, J., and Guyot, J. L.: Recent advances in wavelet analyses: Part 2, Amazon, Parana, Orinoco and Congo discharges time scale variability, J. Hydrol., 314, 289–311, 2005.
Liong, S. Y. and Sivapragasm, C.: Flood stage forecasting with SV M, J. Am. Water Resour. Assoc., 38, 173–186, 2002. Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. P.: Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions, Environ. Modell. Softw., 25, 891–909, 2010.
Mallat, S. G.: A theory for multi resolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. , 11, 674–693, 1989.
Marzano, F. S., Fionda, E., and Ciotti, P.: Neural-network approach to ground-based passive microwave estimation of precipitation intensity and extinction, J. Hydrol., 328, 121–131, 2006.

HESSD
10, 9239–9269, 2013

Effective long-term forecast of hydrological signals

J.-S. Yang et al.

Nayak, P. C., Sudheer, K. P., Rangan, D. M., and Ramasastri, K. S.: A neuro-fuzzy computing technique for modeling hydrological time series, J. Hydrol., 291, 52–66, 2004.
Olson, D. A., Junker, N. W., and Korty: Evaluation 33 years of quantitative precipitation forecasting at NMC, Weather Forecast., 10, 498–511, 1995.
Parlos, A. G., Rais, O. T., and Atiya, A. F.: Multi-step-ahead prediction using dynamic recurrent neural networks, Neural Networks, 13, 765–786, 2000.
Partal, T. and Kisi, O.: Wavelet and neuro-fuzzy conjunction model for precipitation forecasting, J. Hydrol., 342, 199–212, 2007.
Partal, T. and Kucuk, M.: Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey), Phys. Chem. Earth, 31, 1189–1200, 2006.
Raman, H. and Sunilkumar, N.: Multivariate modeling of water resource time series using artificial neural network, Hydrol. Sci. J., 40, 145–163, 1995.
Sadowskey, J.: Investigation of signal characteristics using the continuous wavelet transform, J. Hopkins APL Tech. D., 17, 258–269, 1996.
Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L.: Applied modeling of hydrologic time series, Water Resouces Pub., Michigan, 105–108, 1980.
Schenker, B. and Agarwal, M.: Long-range prediction for poorly-known systems, Int. J. Control., 62, 227–238, 1995.
Smith, L. C., Turcotte, D. L., and Isacks, B.: Stream flow characterization and feature detection using a discrete wavelet transform, Hydrol. Process, 12, 233–249, 1998.
Su, H. T., McAvoy, T. J., and Werbos, P.: Long-term prediction of chemical processes using recurrent neural networks: a parallel training approach, Indust. Appl. Chem. Engine. Res. , 31, 1338–1352, 1992.
Tong, H.: Non-linear time series: a dynamical system approach, Oxford University Press, Oxford, 1990.
Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, B. Am. Meteorol. Soc., 79, 61–78, 1998.
Wang, W. S., Xiang, H. L., and Ding, J.: Predication of hydrology and water resources with nearest neighbor bootstrapping regressive model, Water Resour. Pow., 19, 8–14, 2001.
Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K., and Ma, J.: Forecasting daily streamflow using hybridANN models, J. Hydrol., (Amst) 324, 383–399, 2006.

HESSD
10, 9239–9269, 2013

Effective long-term forecast of hydrological signals

J.-S. Yang et al.

Yang, L. N., Peng, L., Zhang, L. M., Zhang, L. L., and Yang, S. S.: A prediction model for population occurrence of paddy stem borer (*Scirpophaga incertulas*), based on back propagation artificial neural network and principal components analysis, Comput. Electron. Agr. , 68, 200–206, 2009.

Yong, S., Li, Y. B., Wang, Q., and Li, C. H.: Multi-steps prediction of chaotic time series based on echo state network, BIC-TA , 2, 669–672, 2010.

Yu, S. P., Yang, J. S., Liu, G. M., and Zou, P.: Simulation and prediction of soil salt dynamics in the Yangtze River Estuary with BP artificial neural network, Soil , 40, 976–979, 2008 (in Chinese).

Zhou, H. C., Peng, Y., and Liang, G. H.: The research of monthly discharge predictor-corrector model based on wavelet decomposition, Water Resour. Manag., 22, 217–227, 2008.

Zou, P., Yang, J. S., Fu, J. R., Liu, G. M., and Li, D. S.: Artificial neural network and time series models for predicting soil salt and water content, Agr. Water Mange., 97, 2009–2019, 2010.

HESSD

10, 9239–9269, 2013

Effective long-term forecast of hydrological signals

J.-S. Yang et al.

Table 1. Correlation coefficients between each discrete wavelet component series at 15 day and 28 day prior ($t-15$ and $t-28$) and original river stage data series at the current day (S_t).

Discrete wavelet components in ten resolution levels	Correlation coefficients (R) with S_t in Santiao Station			Correlation coefficients (R) with S_t in Qinglong Station		
	Single factor analysis at $t-15$	Single factor analysis at $t-28$	Double factors analysis at $t-15$ and $t-28$	Single factor analysis at $t-15$	Single factor analysis at $t-28$	Double factors analysis at $t-15$ and $t-28$
A_{10}	0.068	0.071	0.071	0.108	0.111	0.111
D_1	0.004	−0.017	0.017	0.007	−0.013	0.015
D_2	0.059	0.049	0.077	0.017	0.022	0.027
$\boldsymbol{D_3}$	0.656	0.556	0.708	0.632	0.547	0.684
$\boldsymbol{D_4}$	0.259	0.379	0.477	0.227	0.383	0.470
D_5	−0.095	0.013	0.097	−0.091	0.011	0.093
D_6	0.053	−0.020	0.082	0.054	−0.014	0.080
D_7	0.125	0.087	0.137	0.127	0.089	0.138
$\boldsymbol{D_8}$	0.320	0.292	0.333	0.367	0.334	0.382
D_9	0.078	0.074	0.082	0.089	0.077	0.116
D_{10}	0.081	0.080	0.084	0.016	0.015	0.026
$S = A_{10} + \sum D_i$	0.741	0.714	0.820	0.745	0.747	0.821
TD = $D_3 + D_4 + D_8$	0.759	0.734	0.825	0.746	0.751	0.836

HESSD
10, 9239–9269, 2013

Table 2. Comparison among performances of 12 types of river stage forecasting models in respect to root mean square errors (RMSE), mean absolute errors (MAE) and correlation coefficients (*R*) in training and test periods.

	Santiao Station						Qinglong Station					
	Training period			Test period			Training period			Test period		
	RMSE	MAE	*R*	RMSE	MAE	*R*	RMSE	MAE	*R*	RMSE	MAE	*R*
CDW-NF	0.314	0.250	0.838	0.476	0.381	0.533	0.298	0.234	0.849	0.486	0.396	0.283
CDW-ANN	0.318	0.254	0.834	0.774	0.635	−0.142	0.309	0.244	0.837	0.527	0.430	0.172
CDW-LR	0.325	0.261	0.825	0.562	0.463	0.110	0.309	0.244	0.836	0.499	0.410	0.119
CW-NF	0.316	0.250	0.835	0.622	0.494	0.183	0.309	0.243	0.836	0.602	0.494	−0.005
CW-ANN	0.322	0.255	0.829	0.702	0.568	0.019	0.313	0.247	0.832	0.505	0.409	0.043
CW-LR	0.329	0.262	0.820	0.565	0.463	0.073	0.322	0.255	0.821	0.487	0.397	0.105
DW-NF	0.211	0.166	0.931	0.582	0.480	0.172	0.211	0.169	0.927	0.471	0.384	0.285
DW-ANN	0.214	0.168	0.928	0.568	0.468	0.193	0.214	0.171	0.926	0.474	0.386	0.256
DW-LR	0.225	0.177	0.921	0.554	0.454	0.191	0.226	0.181	0.916	0.492	0.400	0.230
Neuro-Fuzzy	0.263	0.198	0.889	0.558	0.454	0.094	0.223	0.165	0.918	0.526	0.425	0.090
BP-ANN	0.271	0.203	0.883	0.557	0.451	0.128	0.226	0.168	0.916	0.497	0.402	0.094
LR	0.284	0.216	0.870	0.559	0.456	0.115	0.242	0.180	0.903	0.498	0.403	0.110

HESSD
10, 9239–9269, 2013

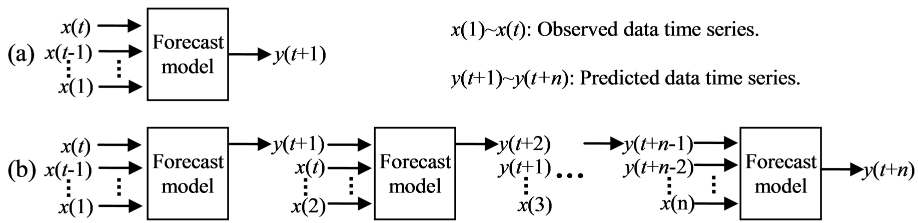

Fig. 1. Architecture of the conventional single-step-ahead (SSA) forecast **(a)** and the SSA based long-range serial-propagated multi-step-ahead (MSA) forecast **(b)**.[1]

[1] In this paper, Figs. 1–3 were created in MS Office 2003. Figure 4 was created in ArcGIS 9.3. Figures 5 and 6 were created in MATLAB-R2006b. Figure 7 was created in OriginPro v7.5.

HESSD
10, 9239–9269, 2013

Effective long-term forecast of hydrological signals

J.-S. Yang et al.

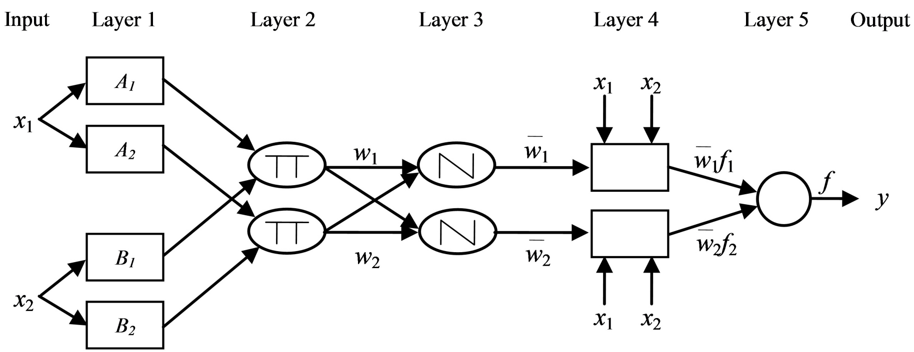

Fig. 2. Schematic diagram of a typical five-layer ANFIS structure.

HESSD
10, 9239–9269, 2013

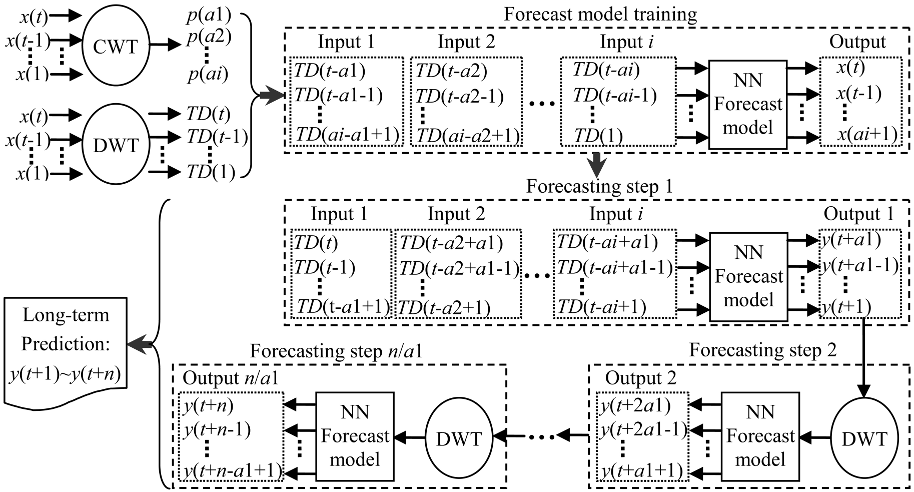

Fig. 3. Architecture of the novel CDW-NN hybrid model.

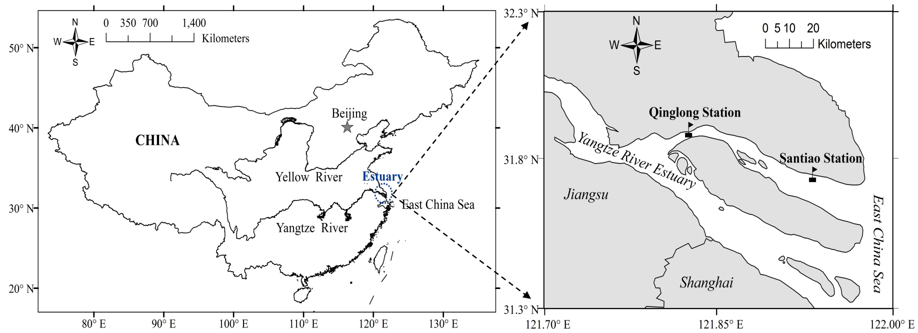

Fig. 4. Location map of the hydrological stations.

HESSD
10, 9239–9269, 2013
Effective long-term forecast of hydrological signals
J.-S. Yang et al.

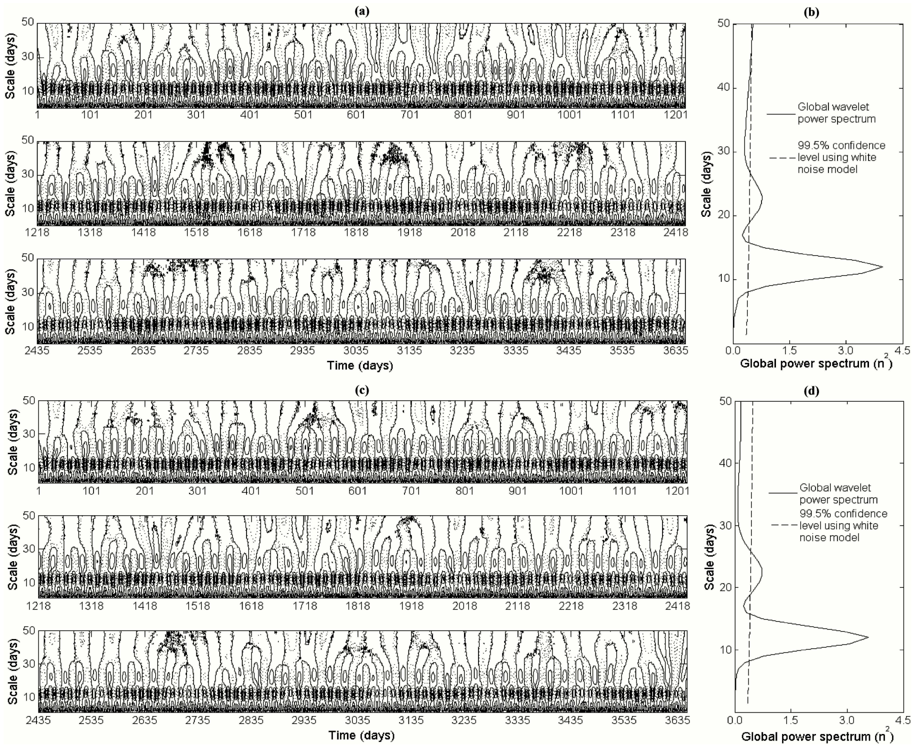

Fig. 5. Real parts of Morlet wavelet transform coefficients at Santiao Station **(a)** and Qinglong Station **(c)**, and their global wavelet power spectrums and corresponding confidence tests using white noise models **(b, d)**, based on 3652 daily river stage data during 1998–2007.

HESSD
10, 9239–9269, 2013

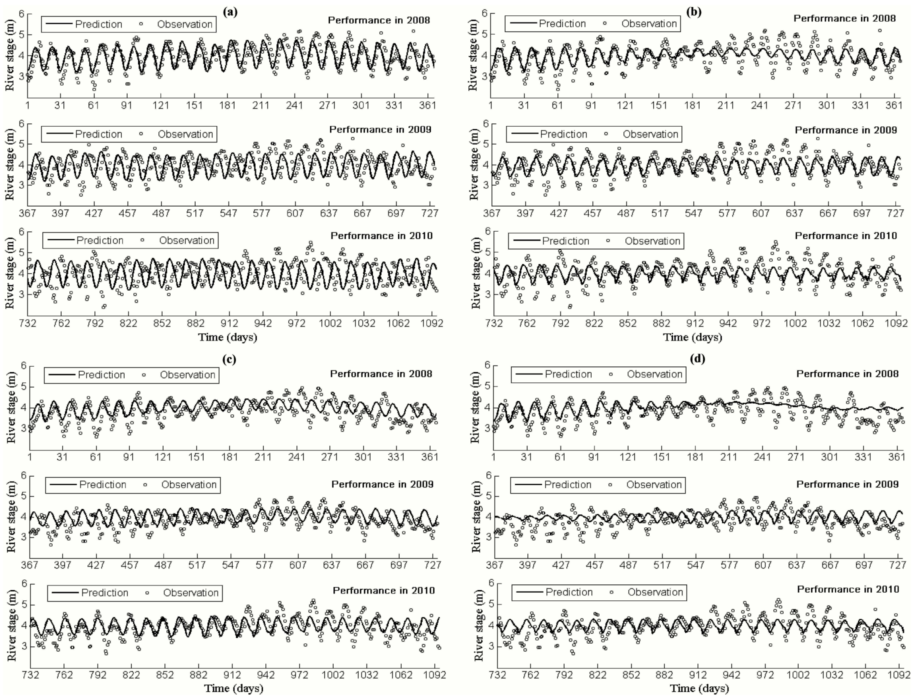

Fig. 6. Observed and predicted values of daily river stages from 2008 to 2010 using CDW-ANN and CDW-NF hybrid models at Santiao Station **(a, b)** and Qinglong Station **(c, d)**.

HESSD
10, 9239–9269, 2013

Effective long-term forecast of hydrological signals

J.-S. Yang et al.

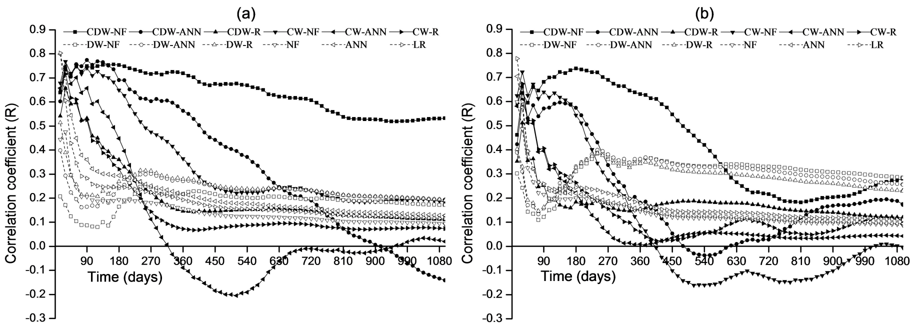

Fig. 7. Forecast performances in correlation coefficients (R) of 12 types of hybrid and pure models during 1096 days river stages forecasting at Santiao Station **(a)** and Qinglong Station **(b)**.

HESSD
10, 9239–9269, 2013