*Journal of Hydrology*, Vol. 361, No. 1-2, 2008, pp 118-130

# A new indirect multi-step-ahead prediction model for a long-term hydrologic prediction

Chun-tian Cheng[1,*] Jing-Xin Xie[2,3] Kwok-Wing Chau[4] Mehdi Layeghifard[5]

*1Department of Civil Engineering, Dalian University of Technology, Dalian, 116024, P.R. China*
*2College of Mechanical and Electronic Engineering, Hebei Agricultural University, Baoding 071001, P.R. China*
*3 Département d' Informatique, Université du Québec à Montréal (UQÀM), Montréa l H2X 3Y7, QC, Canada*
*4Department of Civil and Structural Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
*5Département des Sciences Biologiques, Université du Québec à Montréal (UQÀM), Montréal H2X 3Y7, QC, Canada*

**Abstract:** A dependable long-term hydrologic prediction is essential to planning, designing and management activities of water resources. A three-stage indirect multi-step-ahead prediction model, which combines dynamic spline interpolation into multilayer adaptive time-delay neural network (ATNN), is proposed in this study for the long term hydrologic prediction. In the first two stages, a group of spline interpolation and dynamic extraction units are utilized to amplify the effect of observations in order to decrease the errors accumulation and propagation caused by the previous prediction. In the last step, variable time delays and weights are dynamically regulated by ATNN and the output of ATNN can be obtained as a multi-step-ahead prediction .We use two examples to illustrate the effectiveness of the proposed model. One example is the sunspots time series that is a well-known nonlinear and non-Gaussian benchmark time series and is often used to evaluate the effectiveness of nonlinear models. Another example is a case study of a long-term hydrologic prediction which uses the monthly discharges data from the Manwan Hydropower Plant in Yunnan Province of China. Application results show that the proposed method is feasible and effective.

**Keywords:** time-delay neural network, adaptive time-delay neural network, indirect multi-step-ahead prediction, spline interpolation

## 1. Introduction

A dependable long-term hydrologic prediction is essential to planning, designing and management activities of water resources (Lin et al., 2006; Sivakumar et al., 2001; Mimikou and Rao, 1983 ). During the past few decades, a great deal of research has been devoted to the formulation and development of approaches and models to improve the quality of hydrological prediction, including mechanistic models and black-box models(Karunasinghe and Liong, 2006; Chau, 2006; Cheng et al., 2006; Lin et al., 2006; Wu and Chau, 2006; Chau et al., 2005; Liong et al., 2005; Cheng et al., 2004; Arora, 2002; Islam and Sivakumar, 2002;Ismaiylov and Fedorov, 2001; Sivakumar et al., 2001; Irvine and Eberhart, 1992).Hydrological processes vary both spatially and temporally with a high nonlinearity in spatial and temporal scales(Parasuraman and Elshorbagy, 2007). The mechanistic models used to model such processes would require a large amount of high-quality data associated with astronomical, meteorological, natural geographical characteristics as well as human activity (Arora, 2002; Maier and Dandy, 1999; Milly, 1994), while the burden of data constrains the application of mechanistic models. In the other hand, the black-box models, that at first were only designed to identify the connection between inputs and outputs, are widely applied to forecast the long-term streamflow because of their requirement of little data and their simple formulation. The earlier methods include time series techniques and multiple linear regression methods (Smith, 1991; Irvine and Eberhartdt, 1992). As an alternative to the aforementioned mathematical models, ANNs, which map the input to

[*] Chun-Tian Cheng, Corresponding author. Professor, Department of Civil Engineering, Dalian University of Technology, Dalian, 116024, P.R. China. Tel: +86-411-84708768. Fax: +86-411-84674141 Email: ctcheng@dlut.edu.cn

output without the need to identify the physics a priori, have been widely applied to hydrology field (ASCE Task Committee, 2000; Luk et al., 2000; Maier and Dandy, 1999; Atiya et al., 1999). Some applications of ANNs in long-term hydrologic prediction can be found in the literature (Parasuraman and Elshorbagy, 2007; Karunasinghe and Liong, 2006; Kisi, 2004).

For many engineering applications, a series of forecasts with a long ahead time are required. In recent decades, multi-step-ahead (MS) techniques (Williams and Zipser,1995), which can predict time series values of many time-steps into the future and are classical model predictive algorithms, have been developed to achieve this goal . MS prediction can be divided into direct and indirect categories which have their own advantages and disadvantages. Direct MS prediction models employ all the observations as inputs, while the indirect models use the recursive method of single-step (SS) predictor. Theoretically, the former models provide more precise results in comparison to the later models. However, the direct prediction demands the model hold more flexible ability for each step prediction. Furthermore, it is not easy to develop a direct prediction model.This is why we focus on developing a new indirect multi-step-ahead prediction model in this research.

The difficulty of developing MS predictors is because of the lack of measurements in the prediction horizon that necessitates the recursive use of SS predictors for reaching the end-point in the horizon. Even small SS prediction errors at the beginning of the horizon accumulate and propagate, often resulting in poor prediction accuracy. The situation is even worse for complex systems which are characterized by poorly understandable, noisy, and often nonlinear dynamics (Parlos et al., 2000). Recently, the recurrent neural network was proven to be able to improve MS-based prediction and found to attain promising performance (Bone and Crucianu, 2002; Khotanzad et al., 1994). However, training of a recurrent neural network is usually very time consuming and a single recurrent neural network might lack in robustness (Ahmad and Zhang, 2002). Relatively, feedforward network is easy to implement with a low complexity regarding time and space. Time-delay neural network (TDNN) and adaptive time-delay neural network (ATNN) were proven to be able to improve the efficiency of the MS prediction. TDNN, introduced by Waibel (Waibel et al., 1989) who employed time delays on connections in feedforward networks, has been successfully applied in many areas (Haffner and Waibel, 1992; Luk et al., 2000; Ng and Cook, 1998; Shi et al., 2003; Tan and Cauwenberghe, 1999; Yamashita, 1997). An adaptive version of TDNN, called ATNN, which was originally proposed by Day (Day and Davenport, 1999) adapts its time-delay values and its weights to better accommodate to changing temporal patterns, and also to provide more flexibility for optimization tasks. It has also been successfully utilized in nonlinear system identification(Lin et al., 1995; Yazdizadeh and Khorasani, 2002; Yazdizadeh et al., 2000). In the case of single stage MS prediction, the main idea behind both TDNN and ATNN is time-delay technology which utilizes current and delayed (or past) observations of the measured system inputs and outputs as inputs to the network (Parlos et al., 2000). As a result of time-delay technology, error iteration can deteriorate prediction accuracy very quickly with increased steps ahead. Naturally, to improve the MS prediction, it is required to reduce the use of iterative forecast values and add the observed values. Luckily, the interpolation for discrete sequences(Mery et al., 1998; Schafer and Rabiner, 1973; Tarczynski et al., 1994; Unser, 1999), which is usually employed in signal processing, can be used for this purpose. In our study, the spline interpolation is employed to expend the measurement data space of the model inputs and to increase the effect from observations. Moreover, ATNN can provide more flexibility for optimization tasks.

In this paper, a three-stage MS prediction model, which combines dynamic spline interpolation into multilayer ATNN (SATNN), is proposed. In the first stage, the discrete time series, which has a uniform interval considered as original sampling frequency, is enlarged into many derivative sequences with various sampling frequencies by spline interpolating approximation. In the next stage, the input data set of ATNN variables for each prediction step is dynamically constructed through the integration of the derivative sequences mentioned above. In the last stage, parameters of the two previous stages, variable time delays, and weights are dynamically regulated by ATNN and therefore the output of ATNN can be obtained as a multi-step-ahead prediction. Using interpolation algorithm, some dynamic virtual data are inserted into the original sequences at a point far from the current spot. Therefore, the impact of the insertion of the prediction errors of the previous steps into the next step will be decreased and the reliability of this indirect multi-step-ahead prediction model will be improved. To illustrate the advantages of the proposed model, two examples are used. One example is the sunspots time series that is a well-known benchmark nonlinear and non-Gaussian time series and is often used to evaluate the effectiveness of nonlinear models(Zhang, 2003). Another is a case study of a long-term hydrologic prediction which uses the monthly discharges data from the Manwan Hydropower Plant in Yunnan Province, China.

## 2. Brief review on MS prediction, spline interpolation and ATNN

### 2.1 MS prediction

The recursive relation between inputs and outputs in MS prediction is defined as

$$\hat{x}_{t+p} = F(\hat{x}_{t+p-1}, \hat{x}_{t+p-2}, \ldots, x_t, \ldots, x_{t+p-s}) \tag{1}$$

Where p is the MS prediction horizon, s is the input dimension, and $\hat{x}_{t+p}$ is an estimate of the output at time-step $t+p$. From equation (1), $\hat{x}_{t+p}$ not only depends on the observation values but also on the previous predictions. The prediction accuracy deteriorated very quickly with increased $p$. An approach to improve the prediction accuracy is to enlarge the observation sample.

### 2.2 Cubic spline interpolation for discrete sequences(Kahaner et al., 1988)

The problem for cubic spline interpolation is described as we know a table of points [$x_i$, $y_i$] for $i$=0,1,…,$n$. And the function $y = f(x)$ estimates the value of a function for arbitrary $x$ in a set of points $a = x_0 < x_1 < x_2 < \ldots < x_n = b$. The function $s(x)$ is called a cubic spline on [a, b] if

1) $s(x)$ is defined on [a, b];
2) $s(x)$ and its first and second derivative, i.e., $s'(x)$ and $s''(x)$, are all continuous functions on [a, b];
3) There are points (knots of the spline $s(x)$) such that $a = x_0 < x_1 < x_2 < \ldots < x_n = b$ and $s(x)$ is a polynomial of degree <=3 on each subinterval [$x_{i-1}$, $x_i$]

The fundamental idea behind cubic spline interpolation is used to draw smooth curves through a number of points. A third degree polynomial $s_i(x)$ is determined by

$$S_i(x) = M_{i-1}\frac{(x_i - x)^3}{6h_i} + M_i\frac{(x - x_{i-1})^3}{6h_i} + \left(y_{i-1} - \frac{M_{i-1}}{6}h_i^2\right)\frac{x_i - x}{h_i} + \left(y_i - \frac{M_i}{6}h_i^2\right)\frac{x - x_{i-1}}{h_i} \quad (i = 1, 2, \cdots, n) \tag{2}$$

Where $M_i = S_i''(x_i)$, $h_i = x_i - x_{i-1}$. Using the four conditions of cubic splines (Pollock, 1999), we can draw the following equation.

$$h_i M_{i-1} + 2(h_i + h_{i+1})M_i + h_{i+1}M_{i+1} = 6\left(\frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}\right) \quad (3)$$
$$i = 1, 2, \cdots, n-1$$

These equations can be much simplified if divided by $h_i + h_{i+1}$. Let $\lambda_i = h_{i+1} \big/ (h_i + h_{i+1})$, $\mu_i = 1 - \lambda_i$ and $d_i = 6\left[\frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}\right] \Big/ (h_i + h_{i+1})$ . The equation (3) is translated into

$$\mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = d_i \qquad (i = 1, 2, \cdots, n-1) \quad (4)$$

Note that this system has n-2 rows and n columns, and is therefore under-determined. In order to generate a unique cubic spline, two other conditions must be imposed upon the system. There are various methods of the stipulation to be imposed upon the system. Natural spline is one of methods. Let the second derivative be equal to zero at the endpoints, i.e., $M_1 = M_n = 0$. This results in the spline extending as a line outside the endpoints. Other second derivatives are determined accordingly. Correspondingly, $s_i(x)$ can be obtained.

Using spline interpolation methods, we can increase sampling points between observations. Therefore, the estimated quality of $\hat{x}_{t+p}$ will be improved once these interpolated values are pushed into current and delayed (or past) observations of the measured system input and output in equation 1.

### 2.3 Dynamic ATNN structure

ATNN adapts its time-delay values as well as its weights during training to better accommodate to changing temporal patterns and to provide more flexibility for optimization tasks (Day and Davenport 1999; Lin, et al., 1995; Yazdizadeh, 2002). A dynamic neuron structure is proposed by Lin et. al. (1995) and shown in Figure 1.The input-output mapping is then governed by

$$y(t) = \sigma\left(\sum_{i=1}^{M} \omega_i x_i \left(t - \tau_i\right)\right) \quad (5)$$

where $\omega_i$'s are the neuron weights, $\tau_i$'s are the delays, and $\sigma(\cdot)$ is a nonlinear activation function. It has been shown that, even by taking the above simplified assumption, the resulting input-output map is still capable of representing the nonlinear system (Waibel et al., 1989). For the continuous time series, the time point $t$ is rational sampling point, while in our study it is observation time. It should be noted that the output of the neuron at time $t$, which depends on the previous values of the outputs, results in a dynamic behavior. This dynamics will be modified subsequently for representing the nonlinear system.

*INSERT Figure 1 NEAR HERE*

## 3. SATNN model architecture and algorithm

The basic idea behind the design of the model is to use higher temporal resolution (i.e., a higher sampling rate and higher frequencies) for the long-term history and to use lower temporal resolution for the short-term history (human brain uses a similar approach when combining the “detailed” certain-memory with the “general” uncertain-memory to predict future events). By this means, we get more essential information on the “detailed” and “general” history of the time series while we use a relatively small number of inputs in the forecasting system. With interpolation algorithm, some dynamic virtual data which can be called the “detailed” are inserted into the original sequences at the point far from the current spot. So the impact of previous prediction errors that would be iterated into the model for the next step prediction is decreased. Therefore, the reliability of this indirect multi-step-ahead prediction model will be improved when we make multi-step ahead prediction.

### 3.1 SATNN model architecture

*INSERT Figure 2 NEAR HERE*

SATNN adopts a three-stage architecture that its structure sketch is illustrated in Figure 2. In the first stage, Gs, as a generator, can produce several time series *Xijl* with proper sampling frequencies which are interpolated from the original series *X*. The spline interpolation is applied once over the whole data set and the sequences *Xijl* are obtained with different rates time-delay technology, in the other words, different interpolations are run each time to produce each *Xijl*. In the second stage, dynamic sequences X’ is obtained from the time series *Xijl*. And the procedure is governed by series $\{c_{t1}, c_{t2}, \ldots, c_{tq}\}$ as a result of controller C. Here we can call the series $\{c_{t1}, c_{t2}, \ldots, c_{tq}\}$ controlling signal, because each variable provides information about how to extract the proper parts from series *Xijl* . In the third stage, ATNN is used for prediction based on the newly obtained sequences.

### 3.2 Algorithm

In the first stage, given the time series $\{X \mid x_i, i = 1,2,\ldots,n\}$, the three-stage architecture is summarized in Figure2. In this stage, $G_s$ is a spline interpolation generator with parameter $q$ equal to time window of delayed input series of ATNN in the third stage, which is equivalent to the number of neural network input nodes. $q$ is obtained by the method named Maximum Entropy Method 1 (MEM1)(Jaynes, 1957 ). MEM1, that is widely used to decompose period characteristic of representative hydrologic series(Letie, 1995; Singh, 1997; Wang and Zhu, 2002), is employed to estimate period of nonlinear time series. Given this period the neural network input nodes can be determined.

$\{SI_1, SI_2, \ldots, SI_q\}$ is spline interpolation digital filter (Unser, 1999). Among them, $SI_1$ is a simple linear function generating the same data set as $\{X \mid x_j, j = 1,2,\ldots,n\}$ that is noted as $\{X_{ijl} \mid x_{ijl} = x_j ;= 1; j = 1,2,\ldots,n; l = 1\}$. These interpolation units are employed to interpolate the original series into the smoothed series $\{X_{ijl} \mid i = 1,2,\ldots,q;\ j = 1,2,\ldots,n;\ l = 1,2,\ldots,q;\ l \le i\}$ with various sampling frequencies $\{f_1, f_2, \ldots, f_q\}$ where $f_1$ is the sampling frequency of the original series. It is observed that these interpolation units play the roles that by inserting the smooth virtual data, the original series with frequency $f_1$ can change into various sequences $X_{ijl}$ with frequency $i \times f_1$. Figure 3 describes the process of this stage.

*INSERT Figure 3 NEAR HERE*

In this stage, given the input data sequence $X$, the spline function can be denoted by $S(k)$ where $k$ is the time order of the sequence. The sequences from spline-interpolation units are obtained as follow ~~as~~

$$
\begin{aligned}
&X_{1jl} = S(k_1), \quad k_1 = 1, 2, \ldots, n \\
&X_{2jl} = S(k_2), \quad k_2 = 1, 1\frac{1}{2}, 2, \ldots (n-1)\frac{1}{2}, n \\
&\vdots \\
&X_{qjl} = S(k_q), \quad k_q = 1, 1\frac{1}{q}, 1\frac{2}{q}, \ldots 1\frac{q-1}{q}, 2, \ldots, (n-1)\frac{1}{q}, (n-1)\frac{2}{q}, \ldots (n-1)\frac{q-1}{q}, n.
\end{aligned} \tag{6}
$$

In the second stage, for every current time point $t$, dynamic sequences $X'$ are obtained by series $\{c_{t1}, c_{t2}, \ldots, c_{tq}\}$ as a result of controller $C$. From the outputs of the interpolation units, $\{X_{ijl} \mid i = 1, 2, \ldots, q; j = 1, 2, \ldots, n; l = 1, 2, \ldots, q; l \le i\}$, a set of proportion data are extracted to form a new sequence $\{X'_t \mid x'_{ti}; i = 0, 1, \ldots, J-1\}$. A glide record is utilized in the whole course. The rule of extraction is as follows: firstly, all the data are the content of extraction; secondly, the record backward glides along the time direction; lastly, the beginning point is a certain original data $x_j$ and the consequent data are those behind $x_{j-1}$. All the steps are illustrated in Figure 4. The new sequence is listed by Eq.7.

*INSERT Figure 4 NEAR HERE*

$$
\begin{cases}
x'_{t-0} = x_{1t1} \\
x'_{t-1} = x_{2t1}, \quad x'_{t-2} = x_{2t2} \\
\vdots \\
\cdots \qquad , \ \ldots, \quad x'_{t-J+1} = x_{q(t-q+1)q}
\end{cases}
\qquad J = \frac{q(q+1)}{2}, \quad t \in [p+1, n], \tag{7}
$$

where $t$ is the current time and $J$ the length of the sequence. Considering ~~of~~ the training of the network in the next stage, the scale of $t$ is defined as $t \in [p+1, n]$. When $t$ differs, the $X'$ is a dynamic sequence. After the first two stages, the time order in the original series will never work. Instead, we will focus on the order in $X'$. Figure 5 describes the dynamic combination of sequences in this stage. Then the indirect multi-step-ahead prediction based on the former two stages can be induced as follows:

$$
\begin{cases}
\hat{x}'_{t+p} = F(\hat{x}'_{t+p-1}, \hat{x}'_{t+p-2}, \ldots, \hat{x}'_{t+1}, x'_t, \ldots, x'_{t+p-\tau}) & \tau > p, \tau = 0, 1, \ldots J-1 \\
\hat{x}'_{t+p} = F(\hat{x}'_{t+p-1}, \hat{x}'_{t+p-2}, \ldots, \hat{x}'_{t+p-\tau}) & \tau \le p, \tau = 0, 1, \ldots J-1
\end{cases} \tag{8}
$$

where the variable with a cap "^" denotes the prediction value.

*INSERT Figure 5 NEAR HERE*

In the third stage, the network consists of $L$ layers with $N^L$ neurons in the $lth$ layer. The bipolar sigmoid function is applied as the activation function. In order to compare our model with other models proposed in literature, we choose the same bipolar sigmoid activation function $f(x) = \frac{2}{1+e^{-x}} - 1$. This bipolar sigmoid will generally yield an output that approaches 1 or –1.

By using the spline interpolation, the typical neuron governing equations are developed as follows:

$$\begin{cases} net_j^l(t) = \sum_{i=1}^{N^{l-1}} w_{ji}^l o_i^{l-1}(t-\tau_{ji}^l) \\ o_j^l(t) = \sigma^l\left(net_j^l(t)\right) \end{cases} \qquad \tau_{ji}^l \in \{0,1,\ldots,\tau_{\max}\}, \quad \tau_{\max} = q-1 \tag{9}$$

The output of the $j$ th neuron in the $l$th layer at time $t$ is denoted by $o_j^l(t)$. The first equation depicts the governing algorithm of original typical multilayer adaptive time-delay, in which the weight and associated delay connecting the $j$th neuron in the $l$th layer to the $i$ th neuron in the ($l$-$1$)th layer are denoted by $w_{ji}^l$ and $\tau_{ji}^l$, respectively. It should be noted that $j$ varies from 1 to $N^l$, $i$ varies from 1 to $N^{l-1}$, and $\tau_{ji}^l$ varies form 0 to $\tau_{\max}$, which is defined subsequently as the maximum delay used to represent the desired input-output map (Yazdizadeh, 2002 ). Moreover, most variables mentioned above, such as $i, j, l, t, \tau$ and $q$, are all integers. In our model, $N^1 = q$ is the number of spline interpolation units, and is equivalent to the number of derivative sequences. Clearly, the input and output that involved in the involved with above equation are depicted as

$$y(t) = \sigma\left(\sum_{i=1}^{M} w_i x_i\left(t - \tau_i\right)\right), \quad \tau_i = 0,1,\ldots,J-1. \tag{10}$$

Moreover, in order to avoid the problem of overfitting, we use Leave-one-out cross-validation to obtain a minimum best support value.

## 4. Case studies

Two case studies are used to illustrate the effectiveness of SATNN's perdition. The first one is the sunspots prediction which is a classical example of a combination of periodic and chaotic phenomena and has been served as a benchmark in the statistics literature of time series. This example is used to explore the SATNN model for general MS prediction problem. The second involves the long-term forecast of monthly discharge of a real hydropower plant. The goal is to explore the algorithm efficiency to long-term hydrologic prediction.

### 4.1. Case study I: sunspots prediction

Data series used in this study is from the literature (Boné and Crucianu, 2002). For convenient comparison with other methods (Boné and Crucianu, 2002), the same data sets are used for calibration and validation, i.e., the sunspots average of years 1700 through 1979 is chosen to train and test model for multi-step-ahead forecasting. The training set and two testing sets are selected from this data. The training set is from years 1700 through 1920 and the test sets are from years 1921 through 1954 (Set1) and years 1955 through 1979(Set2) (shown in Figure6).

*INSERT Figure 6 NEAR HERE*

We must pay more attention in designing a proper structure for ATNN in our model. In most prediction applications, the goal is to train the network to achieve a balance between the ability of the network to respond correctly to prediction results, and its ability to spend reasonable time to get those results. Hence, a simplified network structure with three layers is employed into our model, which can effectively perform prediction. The performance that is resulted from proper neural-network architecture is mainly based on two methodologies. The first methodology is the Maximum Entropy Method 1, and the second is Statistical Methodology. Firstly, With the MEM1, we get 10 as the number of input nodes for this three layers ANN. Secondly, if the conventional methods fail to calculate the system dimension, we can minimize output error of a neural network as a function of the number of hidden neurons (Gershenfeld and Weigend, 1993). This number can estimate the system dimension(Emad et al., 1998). Then a Statistical Methodology, which uses Normalized Mean Square Error (NMSE) to calculate prediction error, is implemented to determine the number of hidden neurons. The average relationship between the number of hidden units and NMSE is shown in Figure 7. It is clear that only at the point 13 on hidden units axis, both NMSE and mean of NMSE (tested for all steps ahead) have the least values. Therefore, the number of hidden units is 13. Then the structure of STANN is 10-13-1.

*INSERT Figure 7 NEAR HERE*

Three neural networks, TDNN, ATNN and our model SATNN are implemented over the sets mentioned earlier. Moreover, we also select three models from work of Boné (Boné and Crucianu, 2002). The first is a neural network based on the error back-propagation through time algorithm (RN_BPTT). This method makes use of measures computed during gradient descent and its order of complexity is the same as for BPTT. The second is recurrent neural network based on the constructive back propagation through time (RN_CBPTT), which is heuristic, a connection is considered useful if it can have an important contribution to the computation of the gradient of the error with respect to the weights. And the last is a recurrent neural network based on the exploratory back-propagation through time (RN_ECBPTT), which is also a heuristic, a sort of breadth-first search. It explores the alternatives for the location and the delay associated with a new connection by adding that connection and performing a few iterations of the underlying learning algorithm. According to the literature, these models have an input neuron, a linear output neuron, a bias unit, and a recurrent hidden layer composed of neurons with the symmetric sigmoid as activation function. We performed 20 experiments for each architecture, by randomly initializing the weights in [-0.3, 0.3]. The results of the above parameters are mostly the same as those from the referenced literature.

In order to compare to other models, we also employ the normalized mean squared error (NMSE) which is the ratio between the MSE and the variance of the time series. Comparison among six algorithms for Set 1 is listed in Table 1. SATNN holds the holds all best result in each steps ahead prediction. For example, $NMSE_{1\text{-step-ahead}}=0.0505$, $NMSE_{2\text{-step-ahead}}=0.1283$, $NMSE_{3\text{-step-ahead}}=0.1457$, $NMSE_{4\text{-step-ahead}}=0.1457$, $NMSE_{5\text{-step-ahead}}=0.1478$, $NMSE_{6\text{-step-ahead}}=0.150$. Furthermore, the mean of them is also the best $NMSE_{mean1\text{-}6}=0.1280$. Figure 8 displays the comparison of 6-step-ahead forecasting between TDNN, ATNN and SATNN. It is observed that SATNN provides the best prediction value. The analysis of errors in 6-step-ahead prediction is illustrated can be drown in Figure 8, in which x represents observation value, y represents prediction value, $R$ implies the correlation coefficient, and $B$ implies the slop of the linear fit. From the figure we can observe $R_{\mathrm{TDNN}}=0.85571$, $R_{\mathrm{ATNN}}=0.96294$, $R_{\mathrm{SATNN}}=0.97594$, $B_{\mathrm{TDNN}}=0.59544$, $B_{\mathrm{ATNN}}=1.32698$, and $B_{\mathrm{SATNN}}=0.69952$. These results display the capacity of our model for multi-step-ahead prediction over other models.

*INSERT Table. 1 NEAR HERE*

*INSERT Figure 8 NEAR HERE*

*INSERT Figure 9 NEAR HERE*

Predictions for Set 2 are displayed in Table 2 and Figure 10. In them, we observe that SATNN has a similar performance to RN_ECBPTT where the errors of RN_ECBPTT at the step 1、2、4 are as least as $NMSE_{1\text{-step-ahead}}$=0.2507, $NMSE_{2\text{-step-ahead}}$=0.8982 and $NMSE_{4\text{-step-ahead}}$=1.2537, while at step 3、5、6, SATNN are as least as $NMSE_{3\text{-step-ahead}}$=1.1987, $NMSE_{5\text{-step-ahead}}$=1.3536, $NMSE_{6\text{-step-ahead}}$=1.3817, and $NMSE_{mean1\text{-}6}$=1.0988. The above results show that SATNN provides more accurate prediction for single variable multi-step-ahead forecasting than other models.

*INSERT Figure 10 NEAR HERE*

*INSERT Table. 2 NEAR HERE*

## 4.2. Case study II: long-term hydrologic prediction

The long-term hydrologic prediction of the Manwan Hydropower plant is implemented in this case study. The Manwan Hydropower plant is located on the middle reaches of the Lancang river in Yunnan Province of China and is the first completed large hydropower plant in the cascading hydropower development of the Lancang river. The Lancang River is a large river in Asia, which originates from the Qinghai-Tibet Plateau, penetrates Yunnan from northwest to the south and passes through the Laos, Burma, Thailand, Cambodia and Vietnam, and finally ingresses into the South China Sea. The river is about 4,500 km long and has a drainage area of 744,000 $km^2$. The Manwan Hydropower merges on the middle reaches of the Lancang River and at borders of Yunxian and Jingdong counties. The catchment area at the Manwan dam site is 114,500 $km^2$, the length above Manwan is 1,579 km, and the mean elevation is 4,000 km. The average yearly runoff is 1,230 cubic meters per second at the dam site. Rainfall provides most of the runoff and snow melt accounts for 10%. Nearly 70% of the annual rainfall occurs from June to September.

The monthly discharge from 1953 to 2003 can be obtained wholly. Constrained by the change of hydrologic conditions because of dam projects, the monthly discharge series from January 1988 to December 2003 (Figure 11) are selected. The data set from January 1988 to December 2002 is used for training whilst that from January to December 2003 is used for validation.

*INSERT Figure 11 NEAR HERE*

Three neural networks, TDNN, ATNN and SATNN are implemented over the sets (Figure 11). The 12-step-ahead forecasting is considered to satisfy the engineering. The NMSE are employed as the forecasting accuracy measures. Figure 12 gives comparison among them. Points of interval from Jan. to Jul. are similar, while the SATNN obtains more accurate values than other models in Aug., and also at point of September, and especially at the peak value of each year, and at end of multi-step-ahead.

*INSERT Figure 12 NEAR HERE*

Figure 13 illustrates the relationships between observations and predictions of three layers ANN model. SATNN predicts better than TDNN and ATNN with the correlation coefficient of 0.9395 and slope of the best fit lines of 1.0069. Whereas TDNN (the correlation coefficient and slope of the best fit lines are 0.9194 and 0.9168, respectively) and ATNN (the correlation coefficient and slope of the best fit lines are 0.9225 and 0.9432, respectively) predicted poorly. SATNN gives the best performance of NMSE for three test sets, which are $MNSE_{Set1}$=0.1357, $MNSE_{Set2}$=0.1285, and $MNSE_{Set3}$=0.1111. These results show that SATNN utilization of interpolation technology and ATNN helps it in effective tackling of the drawbacks of MS. Therefore, our model (SATNN) performs much better in prediction of time series in comparison to the TDNN and ATNN models (Table 3). Even if we use it to solve the problem of hydrologic long-term prediction, SATNN can give the effective performance.

*INSERT Figure 13 NEAR HERE*

## 5. Conclusion

Hydrological time series analysis and forecasting has been an active research area over the past few decades. So the need for a long ahead process in prediction is obvious. The objective of this study is to present a method for improving MS prediction model for hydrologic prediction with a single variable. A three-stage indirect multi-step-ahead prediction model, which combines dynamic spline interpolation into multilayer ATNN, is proposed for the long term hydrologic prediction. Using spline interpolation techniques and ATNN, observations samples are enlarged and simultaneously the errors accumulation and propagation caused by the previous prediction are decreased.

The results of the case studies show that SATNN model produces the best results in most situations in comparison to other models. Considering the fact that sunspots prediction is a benchmark in the statistics literature of time series, the application results demonstrate that SATNN model can be widely applied in other fields. For the second case study, the monthly discharge prediction with the 12-step-ahead was analyzed. The SATNN also performed better than other models. The two case studies show that SATNN is capable of capturing potential information and relationship in the time series, and provide better predictions. The SATNN should become a potential method for long-term hydrologic prediction in the future.

**Acknowledgments**

This research was supported by the National Natural Science Foundation of China (No.50679011), Central Research Grant of Hong Kong Polytechnic University (G-U162) and Hebei Province Technology Office Foundation of China (No. 072135125)

**References**

Ahmad, Z. and Zhang, J., 2002. Improving long range prediction for nonlinear process modeling through combining multiple neural network. Proceeding of the 2002 IEEE Interational Conference on Control Applications 966-971.

Arora, V.K., 2002. The use of the aridity index to assess climate change effect on annual runoff. Journal of Hydrology 265,164-77.

ASCE Task Committee,2000. Artificial neural networks in hydrology-II: Hydrological applications. Journal of Hydrologic Engineering, ASCE 5(2) , 124-137

Atiya, A.F., El-Shoura, S.M., Shaheen, S.I., & El-Sherif, M.S., 1999. A comparison between neural network forecasting techniques-case study: river flow forecasting. IEEE Transactions on Neural Networks 10 (2), 402-409.

Boné, R. and Crucianu, M., 2002. An evaluation of constructive algorithms for recurrent networks on multi-step-ahead prediction. ICONIP'02 Proceedings of the 9th International Conference on Neural Information Processing 2 547-551.

Boné, R. and Crucianu, M., 2002. Multi-step-ahead prediction with neural networks: a review, Aproches Connectionistes en Sciences Economiques et en Gestion. Boulogne sur Mer, France, pp. 97-109.

Chau, K.W., **2006.**Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. Journal of Hydrology 329( 3-4), 363-367.

Chau, K.W., Wu, C.L. and Li, Y.S., **2005.** Comparison of several flood forecasting models in Yangtze River. Journal of Hydrologic Engineering- ASCE 10(6), 485-491.

Cheng, C.T., Chau, K.W., Li, X.Y. and Li, G., 2004.Developing a web-based flood forecasting system for reservoirs with J2EE. Hydrological Sciences Journal 49(6), 973-986.

Cheng, C.T., Zhao, M.Y., Chau, K.W. and Wu, X.Y., 2006.Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. Journal of Hydrology 316(1-4),129-140.

Day, S.P. and Davenport, M.R., 1999. Continuous-time temporal back-propagation with adaptive time delays, Ohio State University.

Emad, E.W., Prokhorov, D.V. and Wunsch, D.C., 1998. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks and probabilistic neural networks. IEEE Transactions on Neural Networks 9(6), 1456-1469.
Gershenfeld, N. and Weigend, A., 1993. The future of time series: learning and understanding. Time Series Prediction: Forecasting the Future and Understanding the Past.
Haffner, P. and Waibel, A., 1992. Multi-state time delay neural net works for continuous speech recognition. Advances in Neural Information Processing Systems 4, 135-142.
Irvine, K.N., & Eberhardt, A.J.,1992. Multiplicative,season ARIMA models for Lake Erie and Lake Ontario water levels. Water Resources Bulletin 28 (2), 385-396.
Islam, M.N., & Sivakumar, B. ,2002. Characterization and prediction of runoff dynamics: a nonlinear dynamical view. Advances in Water Resources 25 (2), 179-190.
Ismaiylov, G.K., & Fedorov, V.M. ,2001. Analysis of long-term variations in the Volga annual runoff. Water Resources 28 (5), 469-477.
Jaynes, E.T., 1957 Information theory and statistical mechanics Physical Review 106, 620-630.
Kahaner, D., Moler, C. and Nash, S., 1988. Numerical methods and software. Prentice Hall, Upper Saddle River,NJ.
Karunasinghe,Dulakshi S.K., Liong ,Shie-Yui, 2006. Chaotic time series prediction with a global model: Artificial neural network. Journal of Hydrology 323(1-4), 92-105
Khotanzad, A., Abaye, A. and Maratukulam, D., 1994. An adaptive recurrent neural network system for multi-step-ahead hourly prediction of power system loads. IEEE International Conference on Neural Networks and IEEE World Congress on Computational Intelligence, 5: 3393-3397.
Kisi, O., 2004.River flow modeling using artificial neural networks.Journal of Hydrologic Engineering 9 (1),60-63.
Letie, S.M., 1995. Maximum entropy spectral analysis of the Duero Basin International Journal of Climatology, 15(4), 463-472.
Lin, D.T., Dayhoff, J.E. and Ligomenides, P.A., 1995. Trajectory Production with the adaptive time-delay neural network. Neural Networks 8(3), 447-461.
Lin, J.Y., Cheng, C.T. and Chau, K.W., 2006.Using support vector machines for long-term discharge prediction. Hydrological Sciences Journal 51(4), 599-612.
Liong, S.Y., Phoon, K.K., Pasha, M.F.K., 2005.Efficient implementation of inverse approach for forecasting hydrological time series using micro GA. Journal Of Hydroinformatics 7 (3), 151-163
Luk, K.C., Ball, J.E. and Sharma, A., 2000. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. Journal of Hydrology 227(1), 56-65.
Maier, H., & Dandy, G.C. , 1999. Neural networks for the prediction and forecasting of water resources variables: A review of modeling issues and applications. Environmental Modelling Software 15, 101-124.
Mery, H., Moeneclaey, M. and Fechtel, S., 1998. Digital communication receivers: synchronization, channel estimation, and signal processing. John Wiley & Sons, NewYork.
Milly, P.C.D. ,1994. Climate, soil water storage, and the average annual water balance. Water Resoures Research 30, 2143-2156.
Mimikou, M., & Rao, A.R.,1983. Regional monthly rainfall-runoff model. Journal of Water Resources Planning and Management 109 (1), 75-93
Ng, G.W. and Cook, P.A., , , . 1998. Real-time control of systems with unknown and varying time-delays using neural networks. Engineering Applications of Artificial Intelligence 11(3), 401-409.
Parasuraman, K., Elshorbagy A., 2007.Cluster-based hydrologic prediction using genetic algorithm-trained neural networks.Journal of Hydrologic Engineering 12 (1), 52-62.
Parlos, A.G., Rais, O.T. and Atiya, A.F., 2000. Multi-step-ahead prediction using dynamic recurrent neural networks. Neural Networks 13(5), 765-786.
Pollock, D.S.G., 1999. A handbook of time series analysis-signal processing and dynamics. Academic Press.
Schafer, R.W. and Rabiner, L.R., 1973. A digital signal processing approach to interpolation. IEEE International Conference on Neural Networks and IEEE World Congress on Computational Intelligence, 61(6): 692-702.
Shi, D., Zhang, H.J. and Yang, L.M., 2003. Time-delay neural network for the prediction of carbonation tower's temperature. IEEE Transactions on Instrumentation and Measurement 52(4), 1125-1128.
Singh, V.P., 1997. The use of entropy in hydrology and water resource. Hydrological Processes 11, 587-626.
Sivakumar, B., Berndtsson, R., & Persson, M. ,2001. Monthly runoff prediction using phase space reconstruction. Hydrological Sciences Journal 46 (3), 377-387.
Smith, J.A. ,1991. Long-range streamflow forecasting using non-parametric regression. Water Resources Bulletin 27 (1), 39-46.
Tan, Y. and Cauwenberghe, A.V., 1999. Neural-network-based d-step-ahead predictors for nonlinear systems with time delay. Engineering Applications of Artificial Intelligence 12(1), 21-35.

Tarczynski, A., Kozinski, W. and Cain, G.D., 1994. Sampling rate conversion using fractional-sample delay. ICASSP: 285-288.

Unser, M., 1999. Splines: a perfect fit for signal and image processing. IEEE Signal Processing Magazine 16(6), 22-38.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K. and Lang, K.J., 1989. Phoneme recognition using time-delay neural networks. IEEE Transactions on Acoustics,Speech, Signal Procesings 37(3),328-339.

Wang, D. and Zhu, Y.S., 2002. Research on Cryptic Period of Hydrologic Time Series Based on MEM1 Spectral Analysis. Chinese Journal of Hydrology 22(2), 19-23.

Williams, R. and Zipser, D., 1995. Gradient-based learning algorithms for recurrent networks and their computational complexity. Backpropagation:Theory, Architecture, and Applications (LEA)

Wu, C.L. and Chau, K.W., 2006.A flood forecasting neural network model with genetic algorithm. International Journal of Environment and Pollution 28(3-4), 261-273.

Yamashita, Y., 1997. Time delay neural networks for the classification of flow regimes. Computers and Chemical Engineering 21, 6367-6371.

Yazdizadeh, A. and Khorasani, K., 2002. Adaptive time delay neural network structures for nonlinear system identification. Neurocomputing 47, 207-240.

Yazdizadeh, A., Khorasani, K. and Patel, R.V., 2000. Identification of a two-link flexible manipulator using adaptive time delay neural networks. IEEE Transactions on Systems, Man and Cybernetics, Part B, 30(1),165-172.

Zhang, P.G., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

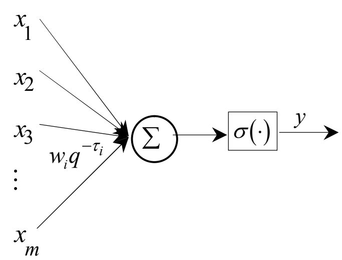


**Figure 1** Dynamic neuron in ATNN. $q^{-\tau}$, the shift operator. $\sigma(\cdot)$, activation function.

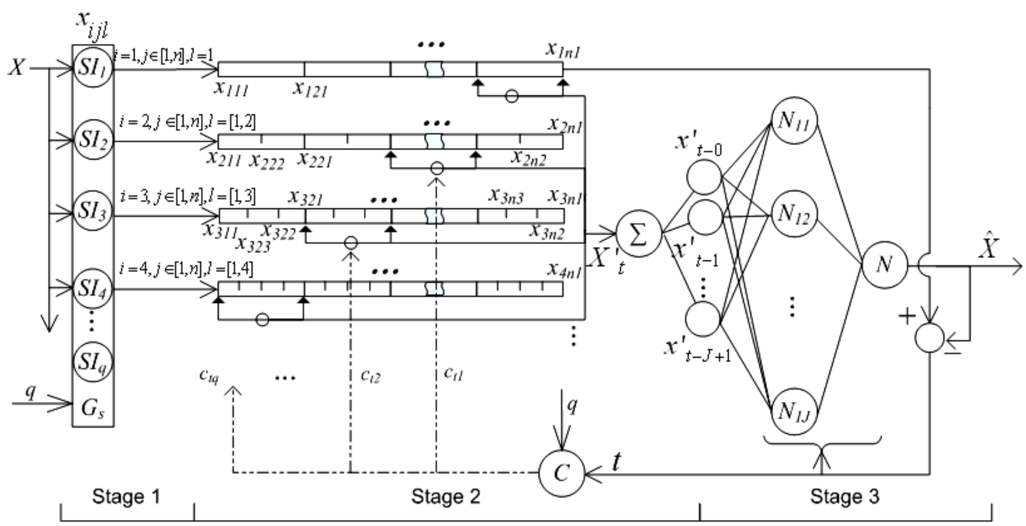


**Figure 2** The three-stage architecture of SATNN.

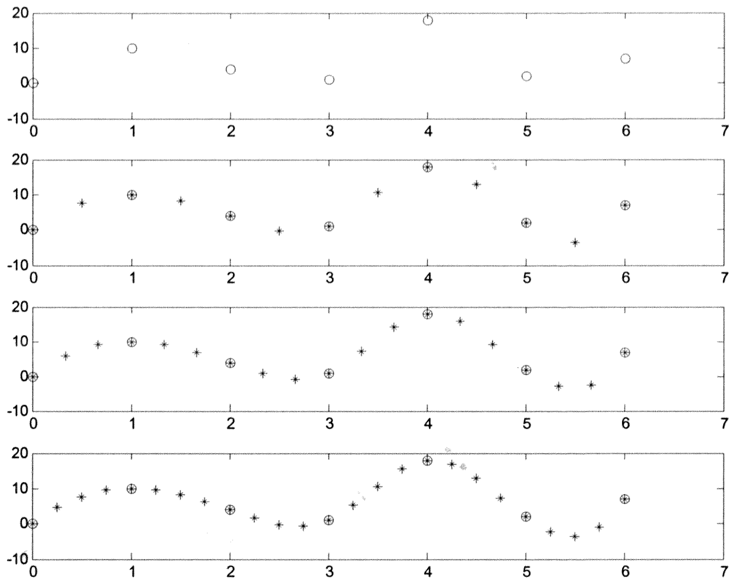

**Figure 3** Sketch of sequences production by spline-interpolation units in the first stage

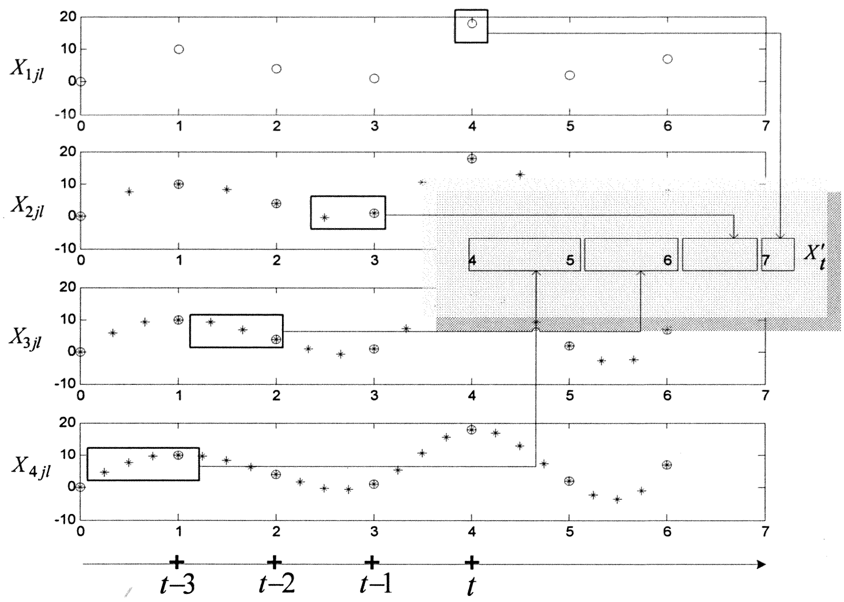


**Figure 4** Dynamic sequences production in the second stage

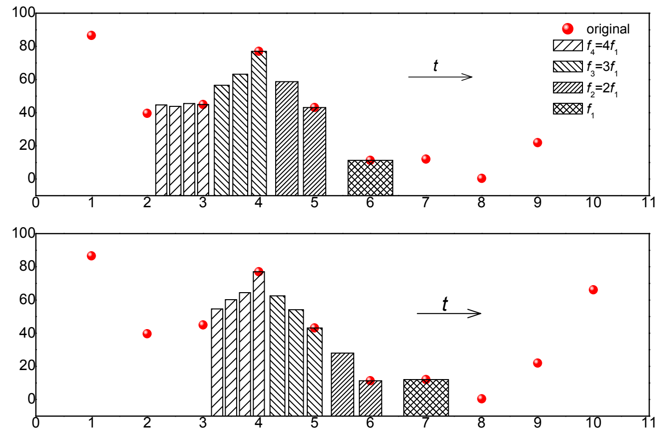


**Figure 5** Dynamic sequences from the second stage

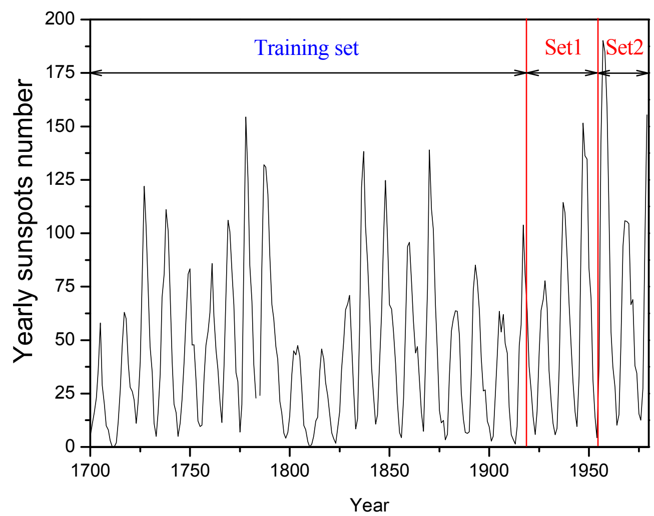


**Figure 6** Training set and two test sets of yearly sunspots number

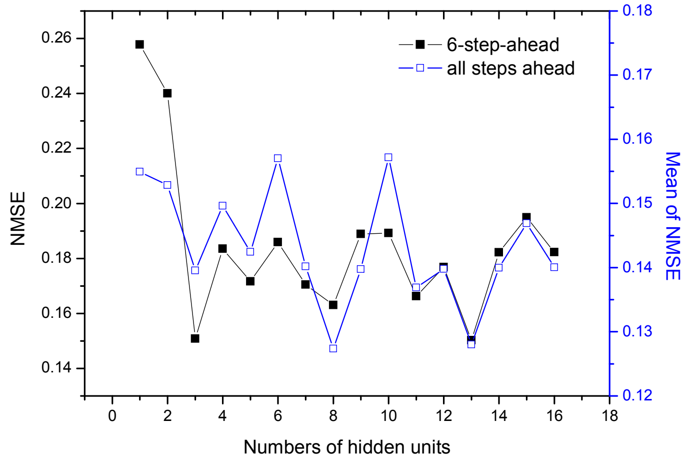


**Figure 7** Relationship between the value of hidden units and NMSE. The least NMSE can be obtained when the number of hidden units is 13.

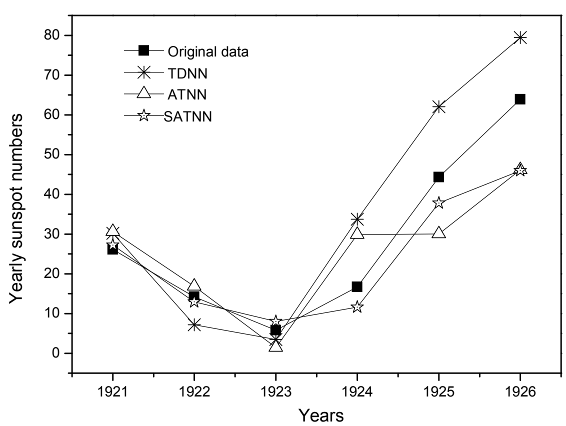


**Figure 8** Best prediction results of Set1 for 6-step-ahead

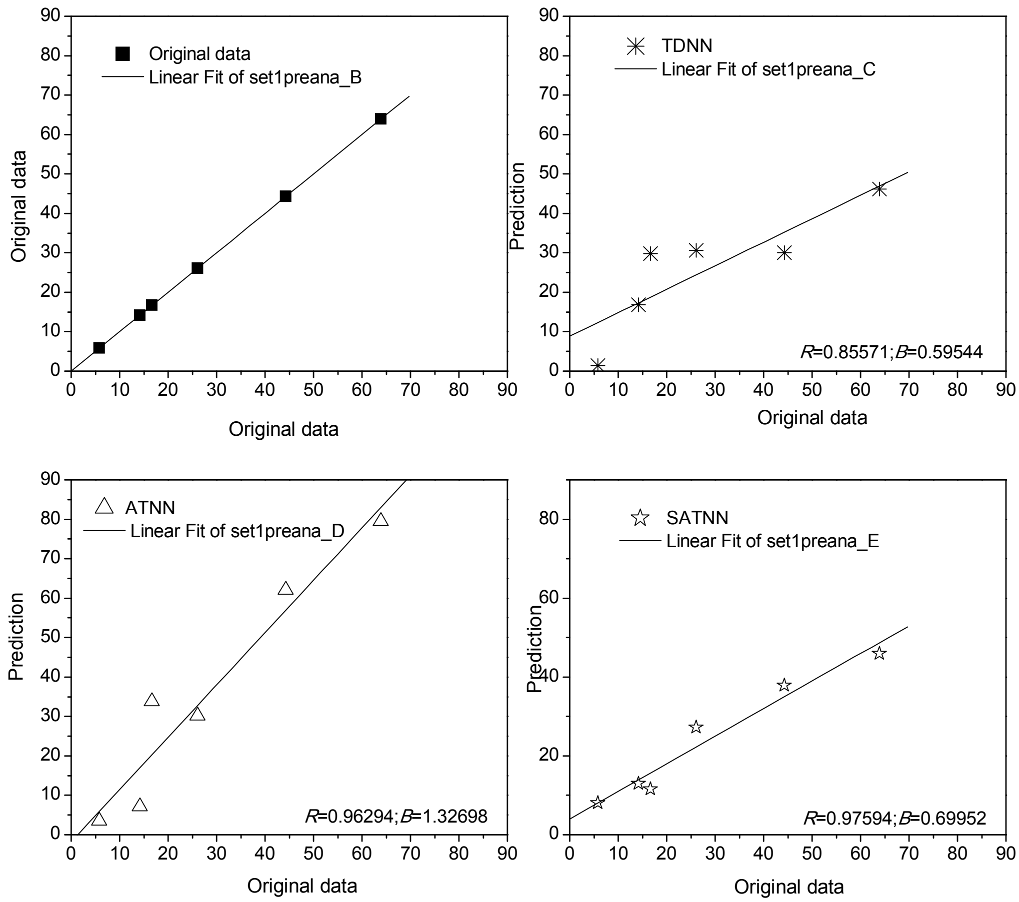


**Figure 9** Error analysis of prediction for Set1

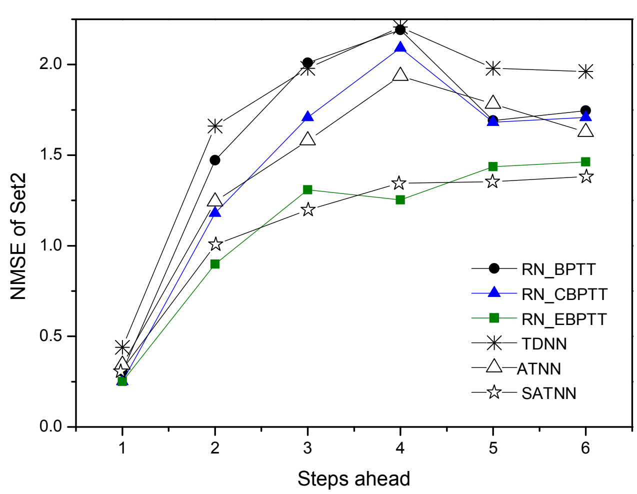

**Figure 10** Prediction errors of all steps for Set2

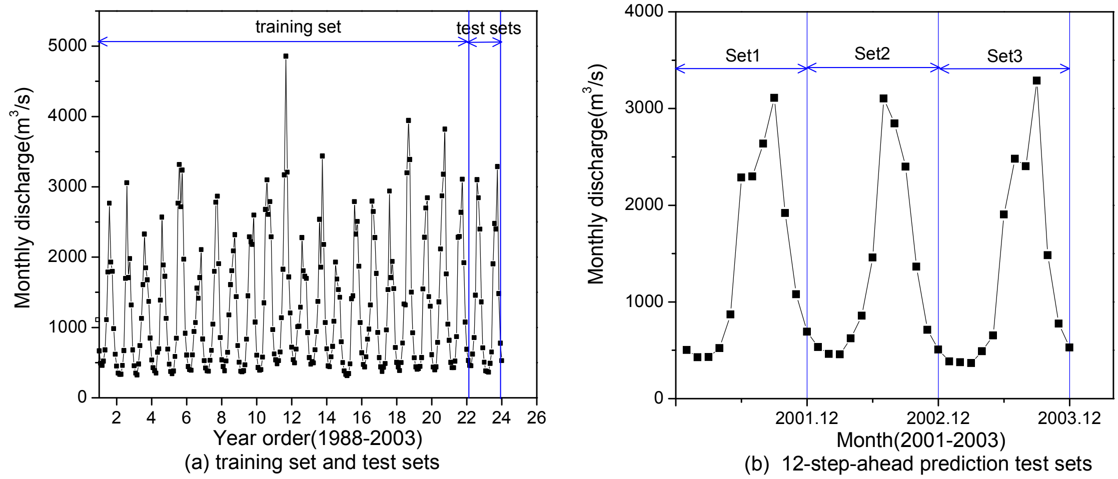


**Figure 11** The training and test sets for 12-step-ahead prediction which are selected from the monthly runoff observation. (b) is the enlarged chart of test sets in (a).

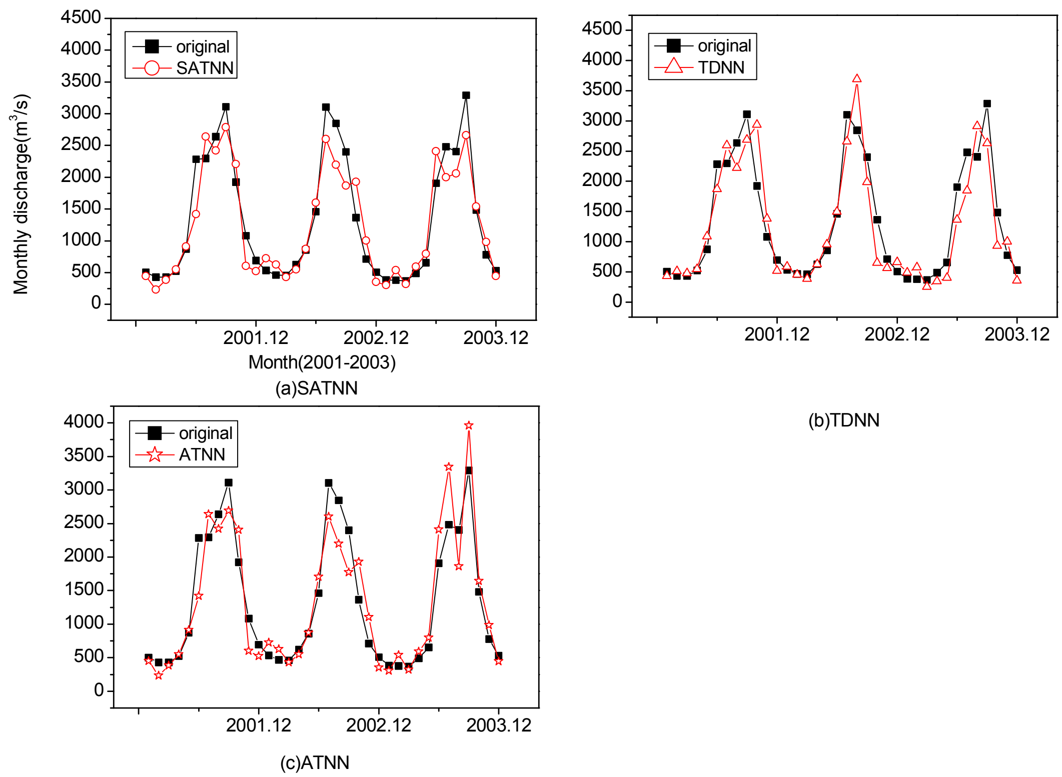


**Figure 12** Prediction comparison among ATNN, TDNN and our model SATNN from January to December 2001~2003.

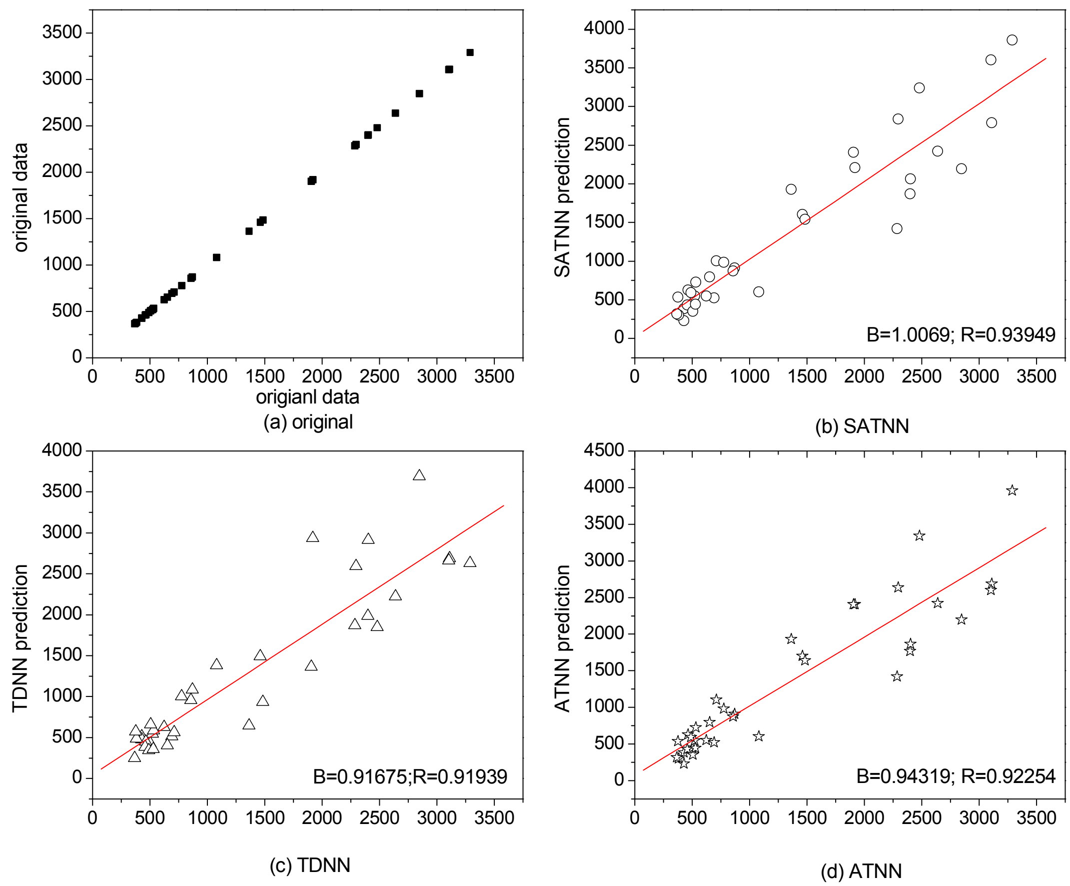


**Figure 13** Prediction error analysis of three ANN models and two main parameters of evaluation (correlation coefficient and slope).

Table. 1 Comparison among six algorithms for Set1

| Steps | RN_BPTT* | RN_CBPTT* | RN_EBPTT* | TDNN | ATNN | SATNN |
|---|---|---|---|---|---|---|
| 1 | 0.0605 | 0.0524 | 0.0519 | 0.0554 | 0.0522 | ***0.0505*** |
| 2 | 0.5015 | 0.4063 | 0.2677 | 0.4863 | 0.3063 | ***0.1283*** |
| 3 | 0.5354 | 0.4668 | 0.3805 | 0.5166 | 0.4068 | ***0.1457*** |
| 4 | 0.5273 | 0.5015 | 0.4322 | 0.5115 | 0.4315 | ***0.1457*** |
| 5 | 0.5096 | 0.4926 | 0.4491 | 0.5126 | 0.4726 | ***0.1478*** |
| 6 | 0.4757 | 0.4668 | 0.3628 | 0.5081 | 0.4608 | ***0.1501*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| $mean_{1-6}$ | 0.4350 | 0.3977 | 0.3240 | 0.4318 | 0.3550 | ***0.1280*** |

Note:* (Boné and Crucianu, 2002)

**Table 2** The comparison among several predictions for Set2 with six algorithms

| Steps | $MNSE_{RN\_BPTT}$ | $MNSE_{RN\_CBPTT}$ | $MNSE_{RN\_EBPTT}$ | $MNSE_{TDNN}$ | $MNSE_{ATNN}$ | $MNSE_{SATNN}$ |
|---|---|---|---|---|---|---|
| 1 | 0.3061 | 0.2507 | ***0.2507*** | 0.4396 | 0.3423 | 0.3061 |
| 2 | 1.4720 | 1.1807 | ***0.8982*** | 1.6612 | 1.2445 | 1.0077 |
| 3 | 2.0096 | 1.7087 | 1.3083 | 1.9799 | 1.5816 | ***1.1987*** |
| 4 | 2.1917 | 2.0915 | ***1.2537*** | 2.2078 | 1.9372 | 1.3448 |
| 5 | 1.6910 | 1.6814 | 1.4358 | 1.9799 | 1.7822 | ***1.3536*** |
| 6 | 1.7456 | 1.7087 | 1.4631 | 1.9614 | 1.6273 | ***1.3817*** |
| $mean_{1-6}$ | 1.5693 | 1.4369 | 1.1016 | 1.7049 | 1.4192 | ***1.0988*** |

**Table 3** The comparison of prediction error analysis among three algorithms for Set1, Set 2 and Set3

| | $MNSE_{Set1}$ | $MNSE_{Set2}$ | $MNSE_{Set3}$ |
|---|---|---|---|
| SATNN | ***0.1357*** | ***0.1285*** | ***0.1111*** |
| TDNN | 0.1567 | 0.1448 | 0.1534 |
| ATNN | 0.1403 | 0.1484 | 0.1499 |