Journal of Hydrology 374 (2009) 294–306

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series

Wen-Chuan Wang [a,b], Kwok-Wing Chau [c,*], Chun-Tian Cheng [b], Lin Qiu [d]

[a] Faculty of Water Conservancy Engineering, North China Institute of Water Conservancy and Hydroelectric Power, Zhengzhou 450011, PR China
[b] Institute of Hydropower System and Hydroinformatics, Dalian University of Technology, Dalian 116024, PR China
[c] Department of Civil and Structural Engineering, Hong Kong Polytechnic University, Yuk Choi Road, Hung Hom, Kowloon, Hong Kong
[d] Institute of Environmental and Municipal Engineering, North China Institute of Water Conservancy and Hydroelectric Power, Zhengzhou 450011, PR China

ARTICLE INFO

Article history:
Received 13 August 2008
Received in revised form 21 January 2009
Accepted 16 June 2009

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Taha Ouarda, Associate Editor

Keywords:
Monthly discharge time series forecasting
ARMA
ANN
ANFIS
GP
SVM

SUMMARY

Developing a hydrological forecasting model based on past records is crucial to effective hydropower reservoir management and scheduling. Traditionally, time series analysis and modeling is used for building mathematical models to generate hydrologic records in hydrology and water resources. Artificial intelligence (AI), as a branch of computer science, is capable of analyzing long-series and large-scale hydrological data. In recent years, it is one of front issues to apply AI technology to the hydrological forecasting modeling. In this paper, autoregressive moving-average (ARMA) models, artificial neural networks (ANNs) approaches, adaptive neural-based fuzzy inference system (ANFIS) techniques, genetic programming (GP) models and support vector machine (SVM) method are examined using the long-term observations of monthly river flow discharges. The four quantitative standard statistical performance evaluation measures, the coefficient of correlation (R), Nash–Sutcliffe efficiency coefficient (E), root mean squared error (RMSE), mean absolute percentage error (MAPE), are employed to evaluate the performances of various models developed. Two case study river sites are also provided to illustrate their respective performances. The results indicate that the best performance can be obtained by ANFIS, GP and SVM, in terms of different evaluation criteria during the training and validation phases.

Introduction and review

The identification of suitable models for forecasting future monthly inflows to hydropower reservoirs is a significant precondition for effective reservoir management and scheduling. The results, especially in long-term prediction, are useful in many water resources applications such as environment protection, drought management, operation of water supply utilities, optimal reservoir operation involving multiple objectives of irrigation, hydropower generation, and sustainable development of water resources. As such, hydrologic time series forecasting has always been of particular interest in operational hydrology. It has received tremendous attention of researchers in last few decades and many models for hydrologic time series forecasting have been proposed to improve the hydrology forecasting.

These models can be broadly divided into three groups: regression based methods, time series models and AI-based methods. For autoregressive moving-average models (ARMA) proposed by Box and Jenkins (1970), it is assumed that the times series is stationary and follows the normal distribution. Since Carlson et al. (1970) presented significant developments in the form of ARMA models of the hydrologic times series, ARMA has been one of the most popular hydrologic times series models for reservoir design and optimization. ARMA is also applied to monthly hydrologic time series (Hipel and Mcleod, 1994). Extensive application and reviews of the several classes of such models proposed for the modeling of water resources time series were reported (Chen and Rao, 2002; Salas, 1993; Srikanthan and McMahon, 2001; Toth et al., 2000; Arena et al., 2006; Komornik et al., 2006).

In recent years, AI technique, being capable of analyzing long-series and large-scale data, has become increasingly popular in hydrology and water resources among researchers and practicing engineers. Since the 1990s, artificial neural networks (ANNs), based on the understanding of the brain and nervous systems, was gradually used in hydrological prediction. An extensive review of their use in the hydrological field is given by ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (ASCE, 2000a,b). Silverman and Dracup (2000) predicted the pattern of rainfall in California using an ANN model. Olsson et al. (2004) used the ANN technique to determine rainfall occurrence and rainfall intensity during rainy periods. Freiwan and Cigizoglu

* Corresponding author. Tel.: +852 2766 6014.
E-mail address: cekwchau@polyu.edu.hk (K.-W. Chau).

0022-1694/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.06.019

(2005) investigated the potential of feed-forward ANN technique in prediction of monthly precipitation amount. ANNs have been shown to give useful results in many fields of hydrology and water resources research (Alvisi et al., 2006; Campolo et al., 2003; Chau, 2006; Chang et al., 2002; Muttil and Chau, 2006; Sudheer et al., 2002; Sudheer and Jain, 2004; Wang et al., 2008).

The adaptive neural-based fuzzy inference system (ANFIS) model and its principles, first developed by Jang (1993), have been applied to study many problems and also in hydrology field as well. Chang and Chang (2001) studied the integration of a neural network and fuzzy arithmetic for real-time stream flow forecasting and reported that ANFIS helps to ensure more efficient reservoir operation than the classical models based on rule curve. Loukas (2001) developed an adaptive neuro-fuzzy inference system (ANFIS) for obtaining sufficient quantitative structure–activity relationships (QSAR) with high accuracy. Bazartseren et al. (2003) used neuro-fuzzy and neural network models for short-term water level prediction. Nayak et al. (2004) evaluated the potential of neuro-fuzzy technique in forecasting river flow time series. Nayak et al. (2005) used a neuro-fuzzy model for short-term flood forecasting. Dixon (2005) examined the sensitivity of neuro-fuzzy models used to predict ground-water vulnerability in a spatial context by integrating GIS and neuro-fuzzy techniques. Shu and Ouarda (2008) used ANFIS for flood quantile estimation at ungauged sites. Other researchers reported good results in applying ANFIS in hydrological prediction (Cheng et al., 2005; Firat and Gungor,2008; Keskin et al., 2006; Zounemat-Kermaniand and Teshnehlab, 2008).

Genetic Programming (GP), an extension of the well known field of genetic algorithms (GA) belonging to the family of evolutionary computation, is an automatic programming technique for evolving computer programs to solve problems (Koza, 1992). GP model was used to emulate the rainfall–runoff process (Whigam and Crapper, 2001) and was evaluated in terms of root mean square error and correlation coefficient (Liong et al., 2002; Whigam and Crapper, 2001). It was shown to be a viable alternative to traditional rainfall–runoff models. The GP approach was also employed by Johari et al. (2006) to predict the soil–water characteristic curve of soils. GP is employed for modeling and prediction of algal blooms in Tolo Harbour, Hong Kong (Muttil and Chau, 2006) and the results indicated good predictions of long-term trends in algal biomass. The Darwinian theory-based GP approach was suggested for improving fortnightly flow forecast for a short time series (Sivapragasam et al., 2007). Guven et al. (2008) used GP-based empirical model for daily reference evapotranspiration estimation.

The support vector machine (SVM) is based on structural risk minimization (SRM) principle and is an approximation implementation of the method of SRM with a good generalization capability (Vapnik, 1998). Although SVM has been used in applications for a relatively short time, this learning machine has been proven to be a robust and competent algorithm for both classification and regression in many disciplines. Recently, the use of the SVM in water resources engineering has attracted much attention. Dibike et al. (2001) demonstrated its use in rainfall–runoff modeling. Liong and Sivapragasam (2002) applied SVM to flood stage forecasting in Dhaka, Bangladesh and concluded that the accuracy of SVM exceeded that of ANN in one-lead-day to seven-lead-day forecasting. Yu et al. (2006) successfully explored the usefulness of SVM based modeling technique for predicting of real-time flood stage forecasting on Lan-Yang river in Taiwan 1–6 h ahead. Khan and Coulibaly (2006) demonstrated the application of SVM to time series modeling in water resources engineering for lake water level prediction. Wu et al. (2008) used a distributed support vector regression for river stage prediction. The SVM method has also been employed for stream flow predictions (Asefa et al., 2006; Lin et al., 2006).

A number of publications have addressed hydrological time series (Carlson et al., 1970; Chang et al., 2002; Chen and Rao, 2002, 2003; Cheng et al., 2005; Firat and Gungor, 2008; Hu et al., 2001; Keskin et al., 2006; Lin et al., 2006; Nayak et al., 2004; Salas, 1993; Sivapragasam et al., 2007). Komornik et al. (2006) compared the forecasting performance of several nonlinear time series models with respect to their capabilities of forecasting monthly and seasonal flows in the Tatry region. Lin et al. (2006) used the support vector machine for hydrological prediction. Jain and Kumar (2007) presented the use of combining ANNs and traditional time series approaches for achieving improved accuracies in hydrological time series forecasting. Cheng et al. (2005), Zounemat-Kermani and Teshnehlab (2008) used adaptive neuro-fuzzy inference system for hydrological time series prediction. But none of the previous publications provide comparison for using these techniques.

Therefore, the major objectives of the study presented in this paper are to investigate several AI techniques for modeling monthly discharge time series, which include ANN approaches, ANFIS techniques, GP models and SVM method, and to compare their performance with other traditional time series modeling techniques such as ARMA. Four quantitative standard statistical performance evaluation measures, i.e., coefficient of correlation (R), Nash–Sutcliffe efficiency coefficient (E), root mean squared error (RMSE), mean absolute percentage error (MAPE), are employed to validate all models. Brief introduction and model development of these AI methods are also described before discussing the results and making concluding remarks. The performances of various models developed are demonstrated by forecasting monthly river flow discharges in Manwan Hydropower and Hongjiadu Hydropower.

Description of selected models

Several AI techniques employed in this study include ANNs, ANFIS techniques, GP models and SVM method. A brief overview of these techniques is presented here.

Artificial neural networks (ANNs)

Since early 1990s, ANNs, and in particular, feed-forward back-propagation perceptrons have been used for forecasting in many areas of science and engineering (Chau and Cheng, 2002). An ANN is an information processing system composed of many nonlinear and densely interconnected processing elements or neurons, which is organized as layers connected via weights between layers. An ANN usually consists of three layers: the input layer, where the data are introduced to the network; the hidden layer or layers, where data are processed; and the output layer, where the results of given input are produced. The structure of a feed-forward ANN is shown in Fig. 1.

A multi-layer feed-forward back-propagation network with one hidden layer has been used throughout the study (Haykin, 1999). In a feed-forward back-propagation network, the weighted connections feed activations only in the forward direction from an input layer to the output layer. These interconnections are adjusted using an error convergence technique so that the network's response best matches the desired response. The main advantage of the ANN technique over traditional methods is that it does not require information about the complex nature of the underlying process under consideration to be explicitly described in mathematical form.

Adaptive neural-based fuzzy inference system (ANFIS)

The ANFIS used in the study is a fuzzy inference model of Sugeno type, and is a composition of ANNs and fuzzy logic ap-

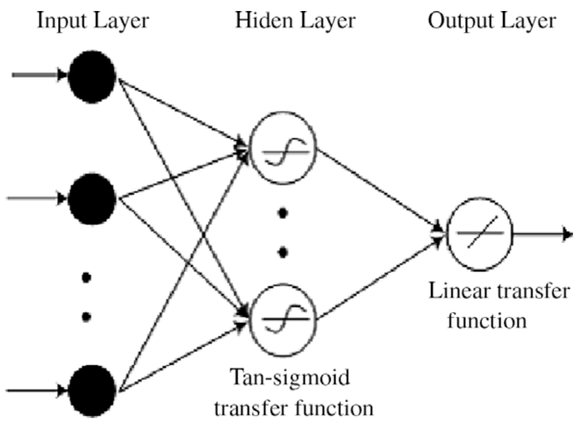

Fig. 1. Architecture of three layers feed-forward back-propagation ANN.

proaches (Jang, 1993; Jang et al., 1997). The model identifies a set of parameters through a hybrid learning rule combining the back-propagation gradient descent and a least-squares method. It can be used as a basis for constructing a set of fuzzy If-Then rules with appropriate membership functions in order to generate the previously stipulated input–output pairs (Keskin et al., 2006).

The Sugeno fuzzy inference system is computationally efficient and works well with linear techniques, optimization and adaptive techniques. As a simple example, we assume a fuzzy inference system with two inputs x and y and one output z. The first-order Sugeno fuzzy model, a typical rule set with two fuzzy If-Then rules can be expressed as:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$
Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$

The resulting Sugeno fuzzy reasoning system is shown in Fig. 2. It illustrates the fuzzy reasoning mechanism for this Sugeno model to derive an output function (f) from a given input vector $[x, y]$. The corresponding equivalent ANFIS architecture is a five-layer feed-forward network that uses neural network learning algorithms coupled with fuzzy reasoning to map an input space to an output space, and nodes are associated with membership functions. It is shown in Fig. 3, and an introduction of the model is as follows:

Layer 1: input nodes. Each node of this layer generates membership grades based on the appropriate fuzzy set they belong to using membership functions. The node output $O_{1,i}$ is defined by:

$$\begin{aligned} O_{1,i} &= \mu_{A_i}(x) \quad \text{for} \quad i = 1,2 \\ O_{1,i} &= \mu_{B_{i-2}}(y) \quad \text{for} \quad i = 3,4 \end{aligned} \tag{1}$$

where x (or y) is the input to the node, and A_i, (or B_{i-2}) is a fuzzy set associated with this node, characterized by the shape of the MFs in this node and can be any appropriate functions that are continuous and piecewise differentiable such as Gaussian, generalized bell shaped, trapezoidal shaped and triangular shaped functions. In this study, the generated bell-shaped membership function given below was used:

$$\mu_{A_i} = \frac{1}{1 + \left|\frac{x-c_i}{a_i}\right|^{2b_i}} \quad \mu_{B_{i-2}} = \frac{1}{1 + \left|\frac{x-c_i}{a_i}\right|^{2b_i}} \tag{2}$$

where $\{a_i, b_i, c_i\}$ is the parameter set of the membership functions in the premise part of fuzz If-Then rules that changes the shapes of the membership function with the maximum equal to 1 and the minimum equal to 0; and $\{a_i, b_i, c_i\}$ are called premise parameters.

Layer 2: rule nodes. Nodes in this layer are labeled II, whose output represents a firing strength of a rule. The node generates the output (firing strength) by cross multiplying all the incoming signals:

$$O_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_{i-2}}(x), \quad i = 1,2 \tag{3}$$

Layer 3: average nodes. Every node in this layer is labeled as N, and computes the normalize firing strength as,

$$O_{3,i} = \bar{w} = \frac{w_i}{w_1 + w_2}, \quad i = 1,2 \tag{4}$$

Layer 4: consequent nodes. Node i in this layer computes the contribution of the ith rule towards the model output, with the following node function:

$$O_{4,i} = \bar{w}_i f = \bar{w}_i(p_i + q_i + r_i), \quad i = 1,2 \tag{5}$$

where $\bar{w}_i$ is the output of layer 3 and $\{p_i, q_i, r_i\}$ is the consequent parameter set.

Layer 5: output nodes. The single node in this layer computes the overall output of the ANFIS as:

$$O_{5,i} = \sum_{i=1}^{4} \bar{w}_i f = \frac{\sum_{i=1}^{4} w_i f}{\sum_{i=1}^{4} w_i} \tag{6}$$

In this paper, the hybrid learning algorithm, which combines the least-squares method and the back-propagation, is used to rapidly train and adapt the FIS. This algorithm converges much faster since it reduces the dimension of the search space of the original back-propagation (Jang, 1993). The more comprehensive presentation of ANFIS for forecasting hydrological time series can be found in the literature (Cheng et al., 2005; Keskin et al., 2006; Nayak et al., 2004; Shu and Ouarda, 2008).

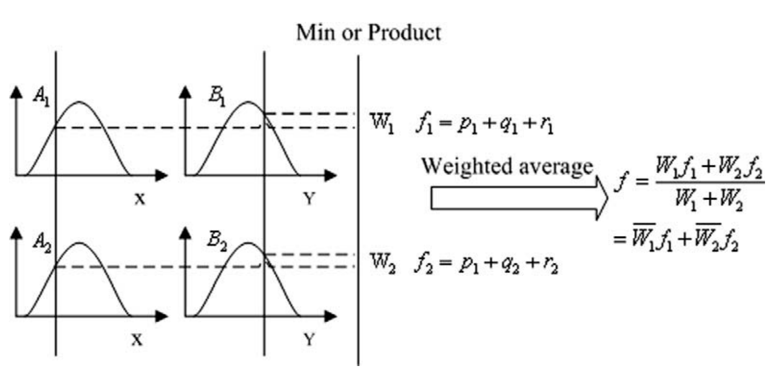

Fig. 2. Two inputs first-order Sugeno fuzzy model with two rules.

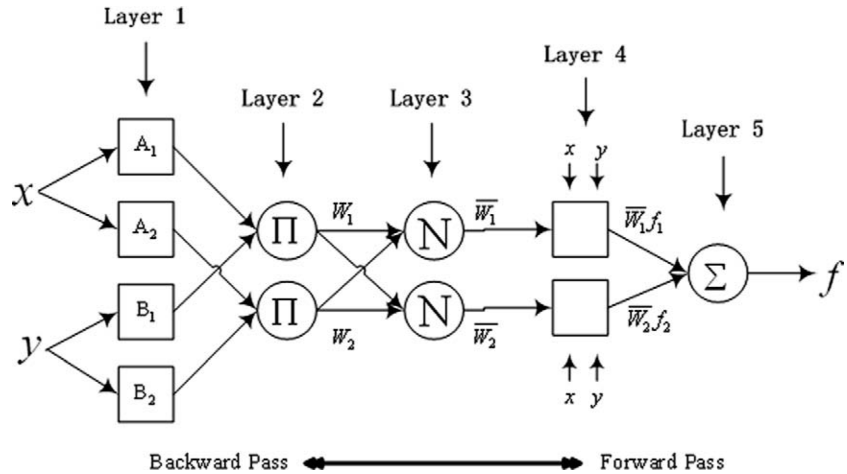

Fig. 3. Architecture of ANFIS.

Genetic programming (GP)

GP is a search methodology belonging to the class of 'intelligent' methods which allows the solution of problems by automatically generating algorithms and expressions. These expressions are codified or represented as a tree structure with its terminals (leaves) and nodes (functions). GP, similar to GA, initializes a population that compounds the random members known as chromosomes (individual). Afterward, fitness of each chromosome is evaluated with respect to a target value. GP works with a number of solution sets, known collectively as a "population", rather than a single solution at any one time; the possibility of getting trapped in a "local optimum" is thus avoided. GP differs from the traditional GA in that it typically operates on parse trees instead of bit strings. A parse tree is built up from a terminal set (the variables in the problem) and a function set (the basic operators used to form the function). GP is provided with evaluation data, a set of primitives and fitness functions. The evaluation data describe the specific problem in terms of the desired inputs and outputs. They are used to generate the best computer program to describe the relationship between the input and output very well (Koza, 1992).

The representation of GP can be viewed as a parse tree-based structure composed of the function set and terminal set. The function set is the operators, functions or statements such as arithmetic operators ({+, -, $*$, /}) or conditional statements (If… Then…) which are available in the GP. The terminal set contains all inputs, constants and other zero-argument in the GP tree. An example of such a parse tree can be found in Fig. 4. Once a population of the GP tree is initialized, the following procedures are similar to GAs including defining the fitness function, genetic operators such as crossover, mutation and reproduction and the termination criterion. In GP, the crossover operator is used to swap the sub-tree from the parents to reproduce the children using mating selection policy rather than exchanging bit strings as in GAs.

The genetic programming introduced here is one of the simplest forms available. A more comprehensive presentation of GP can be found in the literature (Borrelli et al., 2006; Koza, 1992).

Support vector machine (SVM)

SVM is the state-of-the-art neural network technology based on statistical learning (Vapnik, 1995, 1998). The basic idea of SVM is to use linear model to implement nonlinear class boundaries through some nonlinear mapping of the input vector into the high-dimensional feature space. The linear model constructed in the new space can represent a nonlinear decision boundary in

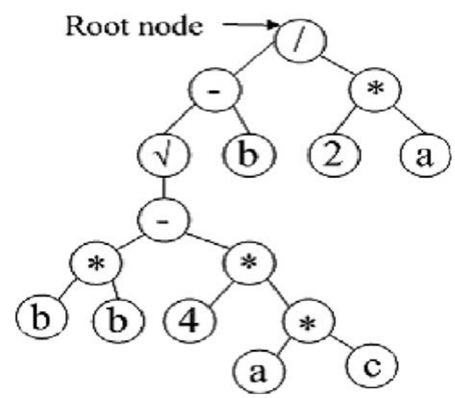

Fig. 4. GP parse tree representing function $\left(\sqrt{(b^2-4ac)}-b\right)/2a$.

the original space. In the new space, SVM constructs an optimal separating hyperplane. If the data is linearly separated, linear machines are trained for an optimal hyperplane that separates the data without error and into the maximum distance between the hyperplane and the closest training points. The training points that are closest to the optimal separating hyperplane are called support vectors. Fig. 5 exhibits the basic concept of SVM. There exist uncountable decision functions, i.e. hyperplanes, which can effectively separate the negative and positive data set (denoted by '*x*' and '*o*', respectively), that has the maximal margin. This indicates that the distance from the closest positive samples to a hyperplane and the distance from the closest negative samples to it will be maximized.

Given a set of training data $\{(x_i, d_i)\}_i^N$ (x_i is the input vector, d_i is the desired value and N is the total number of data patterns), the regression function of SVM is formulated as follows:

$$y = f(x) = w_i\phi_i(x) + b \tag{7}$$

where $\phi_i(x)$ is the feature of inputs, and w_i and b are coefficients. The coefficients (w_i and b) are estimated by minimizing the following regularized risk function (Vapnik, 1995, 1998):

$$r(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2}\|\omega\|^2 \tag{8}$$

where

$$L_\varepsilon(d, y) = \begin{cases} |d-y| - \varepsilon & if |d-y| \geqslant \varepsilon \\ 0 & otherwise \end{cases} \tag{9}$$

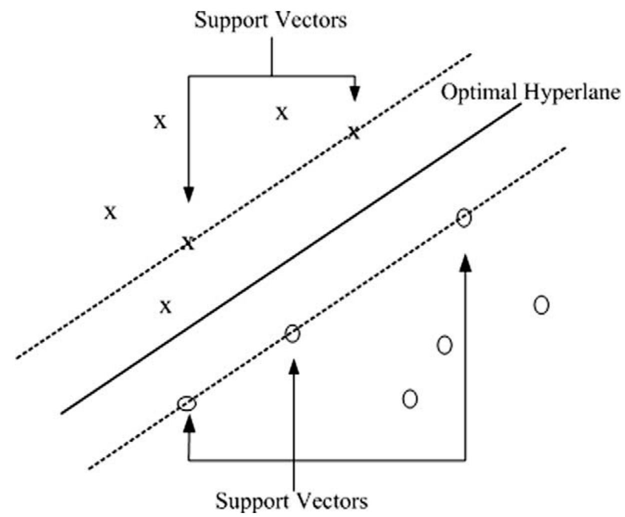

Fig. 5. The basis of the support vector machines.

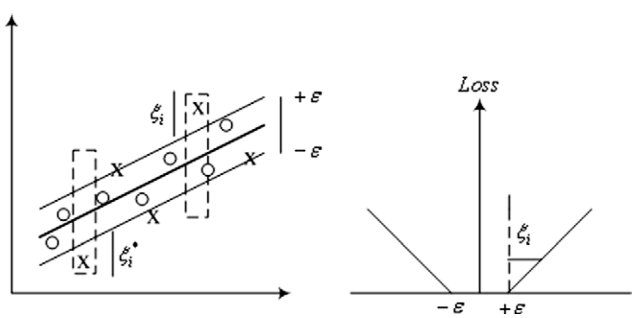

Fig. 6. The soft margin loss setting for a linear SVM and ε-insensitive loss function.

In Eq. (2), the first term is the empirical error (risk). They are measured by Eq. (3). $L_\varepsilon(d, y)$ is called the ε-insensitive loss function, the loss equals zero if the forecast value is within the ε-tube and Fig. 6. The second term is used as a measure of the flatness of the function, Hence, C is referred to as the regularized constant and it determines the trade-off between the empirical risk and the regularization term. Increasing the value of C will result in an increasing relative importance of the empirical risk with respect to the regularization term. ε is called the tube size and it is equivalent to the approximation accuracy placed on the training data points. Both C and ε are user-prescribed parameters, two positive slack variables ξ and ξ^*, which represent the distance from actual values to the corresponding boundary values of ε-tube (Fig. 6), are introduced. Then, Eq. (2) is transformed into the following constrained form.

$$\text{Minimize}: \quad \frac{1}{2}\|\omega\|^2 + C\left(\sum_{i}^{N}(\xi_i + \xi_i^*)\right) \tag{10}$$

$$\text{Subject to} \begin{cases} \omega_i\phi(x_i) + b_i - d_i \leqslant \varepsilon + \xi_i^*, & i = 1, 2, \ldots, N \\ d_i - \omega_i\phi(x_i) - b_i \leqslant \varepsilon + \xi_i, & i = 1, 2, \ldots, N \\ \xi_i, \xi_i^*, \quad i = 1, 2, 3, \ldots, N \end{cases}$$

This constrained optimization problem is solved using the following primal Lagrangian form:

$$L = \frac{1}{2}\|\omega\|^2 + C\left(\sum_{i}^{N}(\xi_i + \xi_i^*)\right) - \sum_{i}^{N}\alpha_i[\omega_i\phi(x_i) + b - d_i + \varepsilon + \xi_i] - \sum_{i=1}^{N}\alpha_i^*[di - \omega_i\phi(x_i) - b + \varepsilon + \xi_i^*] - \sum_{i}^{N}(\beta_i\xi_i + \beta_i^*\xi_i^i) \tag{11}$$

Eq. (5) is minimized with respect to primal variables ω_i, b, ξ and ξ^*, and maximized with respect to the nonnegative Lagrangian multipliers α_i^* and β_i^*, Finally, Karush–Kuhn–Tucker conditions are applied to the regression, and Eq. (5) has a dual Lagrangian form:

$$\upsilon(\alpha_i, \alpha_i^*) = \sum_{i=1}^{N} d_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \tag{12}$$

with the constraints,

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad a_i, \alpha_i^* \in [0, C], \quad i = 1, 2, \ldots, N$$

In Eq. (6), the Lagrange multipliers satisfy the equality $\alpha_i^*\alpha_i^* = 0$. The Lagrange multipliers α_i and α_i^* are calculated, and the optimal desired weight vector of the regression hyperplane is

$$\omega^* = \sum_{i=1}^{N}(\alpha_i - \alpha_i^i)K(x, x_i) \tag{13}$$

Therefore, the regression function can be given as

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x, x_i) + b \tag{14}$$

Here, $K(x, x_i)$ is called the Kernel function. The value of the Kernel is inner product of the two vectors x_i and x_j in the feature space $\varphi(x)$ and $\varphi(x_j)$, so $K(x, x_j) = \phi(x)^*\phi(x_j)$, and function that satisfies Mercer's condition (Vapnik, 1998) can be used as the Kernel function. In general, three kinds of Kernel function are used as follows:

Polynomial:

$$K(x, x_j) = (x \cdot x_j + 1)^n \tag{15}$$

Radial basis function (RBF)

$$K(x, x_j) = \exp(-\|x - x_j\|^2/2\sigma^2) \tag{16}$$

Two-layer neural networks

$$K(x, x_j) = \tanh(kx \cdot x_j - \delta)^n \tag{17}$$

Study area and data

In this study, Manwan Hydropower in Lancangjiang River is selected as a study site. The monthly flow data from January 1953 to December 2004 are studied. The data set from January 1953 to December 1999 is used for calibration whilst that from January 2000 to December 2004 is used for validation (Fig. 7). Lancangjiang River is a large river in Asia, which originates from Qinghai-Tibet Plateau, penetrates Yunnan from northwest to the south and passes through Laos, Burma, Thailand, Cambodia and Vietnam, ingresses into South China Sea finally. The river is about 4500 km long and has a drainage area of 744,000 km^2. Manwan Hydropower merges on the middle reaches of Lancang River and at borders of Yunxian and Jingdong counties. The catchment area at Manwan dam site is 114,500 km^2, the length above Manwan is 1579 km, and the mean elevation is 4000 m. The average yearly runoff is 1230 m^3/s 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. Locations of Lancang River and Manwan Hydropower are shown in Fig. 8A.

The second study site is at Hongjiadu Hydropower on Wujiang River in southwest China. The monthly flow data from January 1951 to December 2004 are studied. The data set from January 1951 to December 1994 is used for calibration whilst that from January 1995 to December 2004 is used for validation (Fig. 9).

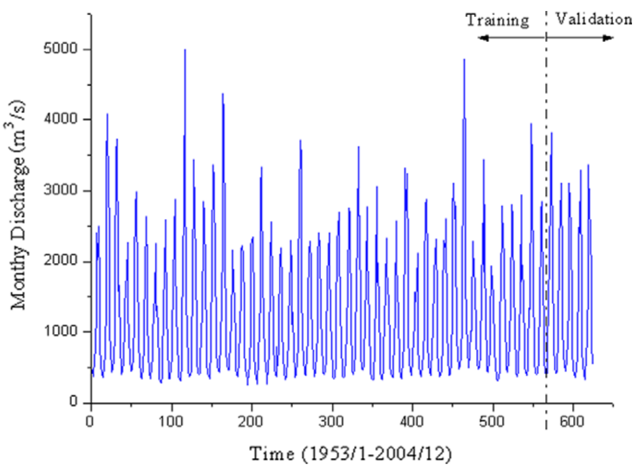

Fig. 7. Monthly discharge at Manwan reservoir.

Wujiang River, originating from Wumeng foothill of Yun-Gui Plateau, is the biggest branch at the southern bank of Yangtze River, which covers 87,920 km^2, total length of 1037 km, centralized fall of 2124 m, and with approved installed capacity 8800 MW. Nowadays, Hongjiadu Hydropower station is the master regulation reservoir for the cascade hydropower stations on Wujiang River. The catchment area at Hongjiadu dam site is 9900 km^2 and the average yearly runoff is 155 m^3/s at the dam site. Rainfall provides most of the runoff. Locations of Wujiang River and Hongjiadu Hydropower are shown in Fig. 8B.

In ANN, ANFIS and SVM modeling processes, large attribute values might cause numerical problems because the neurons in ANN and ANFIS are combined Sigmoid function as excitation function, and the Kernel values in SVM usually depend on the inner products of feature vectors, such as the linear Kernel and the polynomial Kernel. There are two main advantages to normalize features before applying ANN, ANFIS and SVM to prediction. One advantage is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges, and another advantage is to avoid numerical difficulties during the calculation. It is recommended to linearly scale each attribute to the range [−1, +1] or [0, 1]. In the modeling process, the data sets of river flow were scaled to the range between 0 and 1 as follow:

$$q'_i = \frac{q_i - q_{\min}}{q_{\max} - q_{\min}} \tag{18}$$

where q'_i is the scaled value, q_i is the original flow value and $q_{\min}$, $q_{\max}$ are, respectively, the minimum and maximum of flow series.

Prediction modeling and input selection

We are interested in hydrological forecasting model that predict outputs from inputs based on past records. There are no fixed rules for developing these AI techniques (ANN, ANFIS, GP, SVM), even

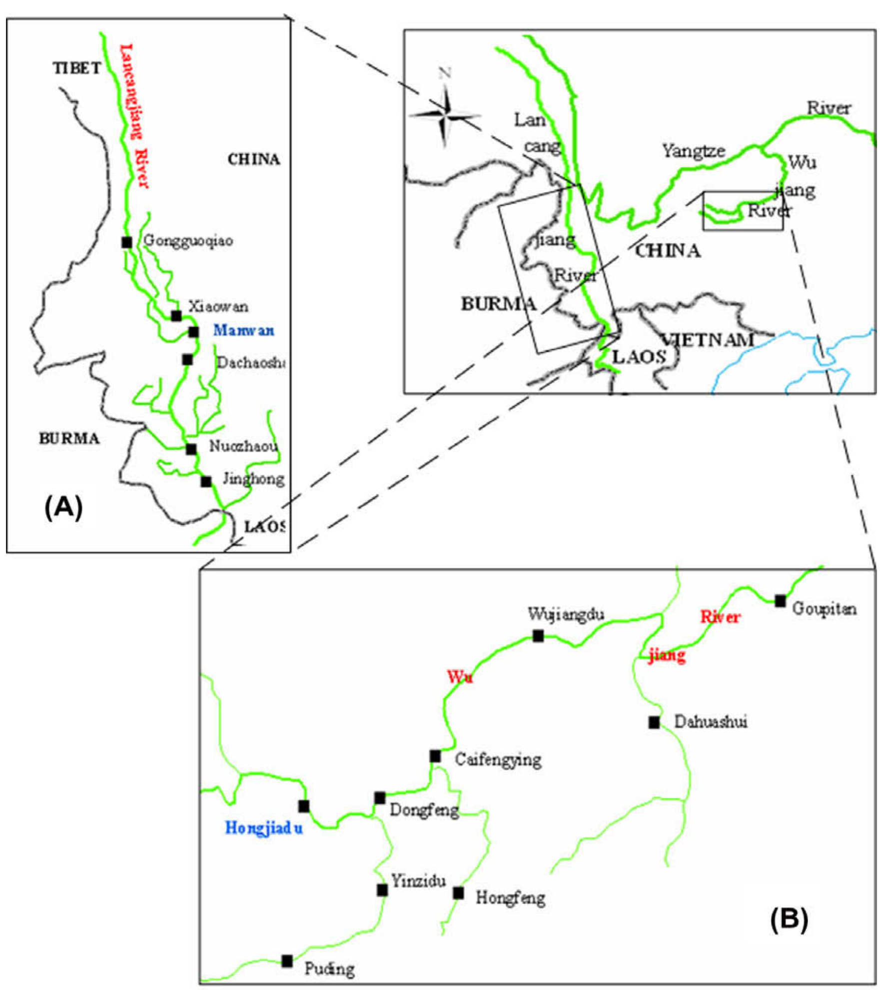

Fig. 8. Location of study sites.

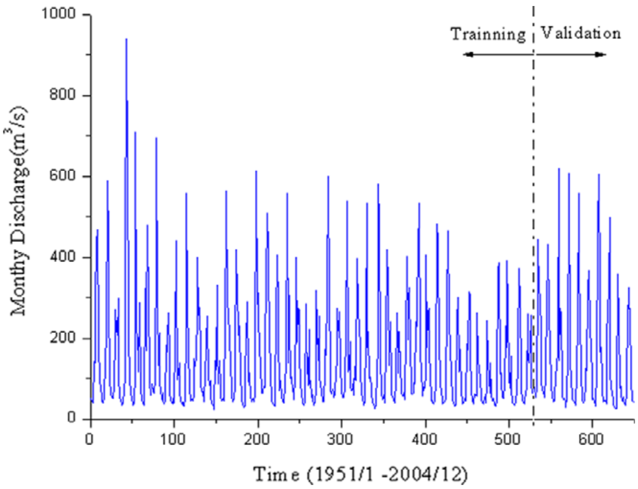

Fig. 9. Monthly discharge at Hongjiadu reservoir.

though a general framework can be followed based on previous successful applications in engineering (Cheng et al., 2005; Lin et al., 2006; Nayak et al., 2004; Sudheer et al., 2002). The univariate time series is also employed for developing inflow forecasting model (Sudheer et al., 2002; Nayak et al., 2004; Cheng et al., 2005; Lin et al., 2006; Keskin et al., 2006). The objective on predicting discharges using antecedent values is to generalize a relationship of the following form:

$$Y = f(X^m) \tag{19}$$

where X^m is a m-dimensional input vector consisting of variables $x_1, \ldots, x_i, \ldots, x_m$, and Y is the output variable. In discharge modeling, values of x_i may be flow values with different time lags and the value Y is generally the flow in the next period. Generally, the number of antecedent values included in the vector X^m is not known a priori.

In these AI techniques, being typical in any data-driven prediction models, the selection of appropriate model input vector would play an important role in their successful implementation since it provides the basic information about the system being modeled. The parameters determined as input variables are the numbers of flow values for finding the lags of runoff that have a significant influence on the predicted flow. These influencing values corresponding to different lags can be very well established through a statistical analysis of the data series. Statistical procedures were suggested for identifying an appropriate input vector for a model (Lin et al., 2006; Sudheer et al., 2002). In this study, two statistical methods (i.e. the autocorrelation function (ACF) and the partial autocorrelation function (PACF)) are employed to determine the number of parameters corresponding to different antecedents values. The influencing antecedent discharge patterns can be suggested by the ACF and PACF in the flow at a given time. The ACF and PACF are generally used in diagnosing the order of the autoregressive process and can also be employed in prediction modeling (Lin et al., 2006). The values of ACF and PACF of monthly flow sequence (1953/1–1999/12) is calculated for lag 0–24 in Manwan, which are presented in Fig. 10. Similarly, the values of ACF and PACF of monthly flow sequence (1951/1–1994/12) is calculated for lag 0–24 in Hongjiadu, which are presented in Fig. 11. From Figs. 10a and 11a, the ACF exhibits the peak at lag 12. In addition, Figs. 10b and 11b showed a significant correlation of PACF at 95% confidence level interval up to 12 months of flow lag. Therefore twelve antecedent flow values have the most information to predict future flow and are considered as input for monthly discharge time series modeling.

Model performance evaluation

Some techniques are recommended for hydrological time series forecasting model performance evaluation according to published literature related to calibration, validation, and application of hydrological models. Four performance evaluation criteria used in this study are computed as in the following section.

The coefficient of correlation (R) or its square, the coefficient of determination (R^2)

It describes the degree of collinearity between simulated and measured data, which ranges from −1 to 1, is an index of the degree of linear relationship between observed and simulated data. If $R = 0$, no linear relationship exists. If $R = 1$ or −1, a perfect positive or negative linear relationship exists. Its equation is

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n}(Q_0(i) - \overline{Q_0})(Q_f(i) - \overline{Q_f})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_0(i) - \overline{Q_0})^2} * \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_f(i) - \overline{Q_f})^2}} \tag{20}$$

R and R^2 have been widely used for model evaluation (Lin et al., 2006; Santhi et al., 2001; Van Liew et al., 2003), though they are oversensitive to high extreme values (outliers) and insensitive to additive and proportional differences between model predictions and measured data (Legates and McCabe, 1999).

Nash–Sutcliffe efficiency coefficient (E)

The Nash–Sutcliffe model efficiency coefficient is used to assess the predictive power of hydrological models (Nash and Sutcliffe, 1970). It is a normalized statistic that determines the relative magnitude of the residual variance (“noise”) compared to the measured

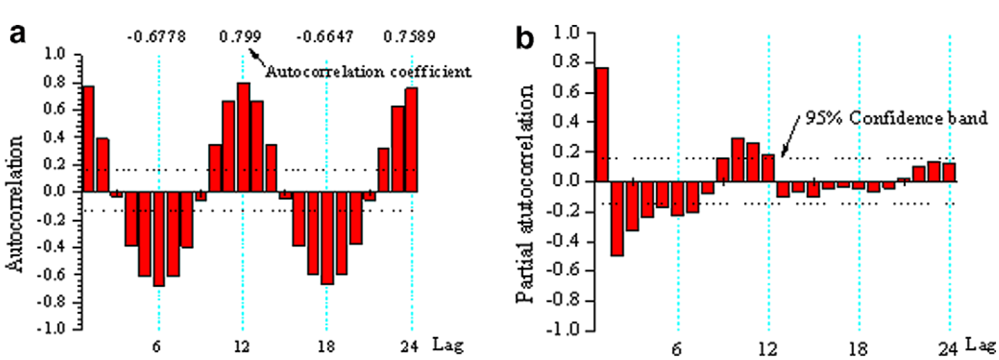

Fig. 10. (a) The autocorrelation function of flow series. (b) The partial autocorrelation function of flow series in Manwan.

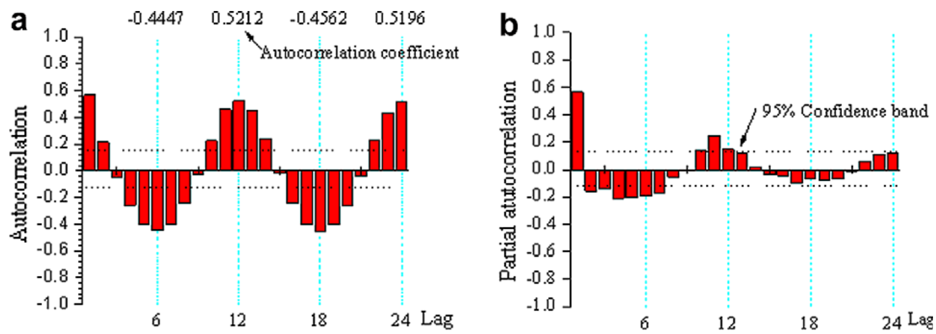

Fig. 11. (a) The autocorrelation function of flow series. (b) The partial autocorrelation function of flow series in Hongjiadu.

Table 1
AIC value and performance indices of alternative ARMA models for Manwan Hydropower.

(p, q)	AIC	Training				Validation			
		R	E	RMSE	MAPE	R	E	RMSE	MAPE
(5, 8)	12.043	0.916	0.839	365.60	17.56	0.927	0.878	359.22	15.72
(6, 7)	12.045	0.915	0.838	366.78	17.42	0.925	0.874	355.18	15.56
(8, 7)	11.786	0.922	0.849	354.27	16.77	0.928	0.869	354.35	15.43
(9, 8)	11.813	0.921	0.847	356.98	16.47	0.923	0.856	380.69	15.89
(11, 8)	11.817	0.921	0.848	355.95	16.13	0.928	0.859	376.04	15.26

data variance and indicates how well the plot of observed versus simulated data fits the 1:1 line (Moriasi et al., 2007). It is defined as:

$$E = 1 - \frac{\sum_{i=1}^{n}(Q_0(i) - Q_f(i))}{\sum_{i=1}^{n}(Q_0(i) - \overline{Q_0})} \tag{21}$$

Nash–Sutcliffe efficiencies ranges between $(-\infty, 1)$: $E = 1$ corresponds to a perfect match of forecasting discharge to the observed data; $E = 0$ shows that the model predictions are the same as the mean value of the observed data; and $-\infty < E < 0$ occurs when the mean observed value is a better predictor than the model simulated value, which indicates unacceptable performance.

Root mean squared error (RMSE)

It is an often used measure of the difference between values predicted by a model and those actually observed from the thing being modeled. RMSE is one of the commonly used error index statistics (Lin et al., 2006; Nayak et al., 2004) and is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_f(i) - Q_0(i))^2} \tag{22}$$

Mean absolute percentage error (MAPE)

The MAPE is computed through a term-by-term comparison of the relative error in the prediction with respect to the actual value of the variable. Thus, the MAPE is an unbiased statistic for measuring the predictive capability of a model. It is a measure of the accuracy in a fitted time series value in statistics and has been used for river flow time series prediction evaluation (Hu et al., 2001). It usually expresses accuracy as a percentage and is defined as:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Q_f(i) - Q_0(i)}{Q_0(i)}\right| \times 100 \tag{23}$$

where $Q_0(i)$ and $Q_f(i)$ are, respectively, the observed and forecasted discharge and $\overline{Q_0}, \overline{Q_f}$ denote their means, and n is the number data points considered.

Development of models

ARMA model uses the direct dependence of the previous measurements and depends on the previous innovation of the process in a moving-average form. The monthly discharge series, which do fit a normal distribution with respect to the skewness coefficient, can be normalized using a log-transformation function in order to remove the periodicity in the original record (Keskin et al., 2006). In order to choose the appropriate ARMA (p, q) model, the Akaike information criteria (AIC) are used to select the value of p and q, which represent, respectively the number of autoregressive orders and the number of moving-average orders of the ARMA model. In this study, the models ARMA (5, 8), (6, 7), (8, 7), (9, 8) and (11, 8), have a relatively minimum AIC value based on flow series in Manwan, and the models ARMA (5, 9), (6, 10), (7, 9), (8, 9) and (10, 11) have a relatively minimum AIC value based on flow series in Hongjiadu. Tables 1 and 2, respectively, show their AIC values and the performance of alternative ARMA models. Hence, according to their performance indices, ARMA (8, 7) is selected as the ARMA model in Manwan, and ARMA (6, 10) is selected as the ARMA model in Hongjiadu.

In this study, a typical three-layer feed-forward ANN model (Fig. 1) is constructed for forecasting monthly discharge time series. The conventional back-propagation training algorithm is a supervised training mechanism and is normally adopted in most of the engineering application. The scaled conjugate gradient algorithm introduced by (Moller (1993)) is a more effective learning algorithm than back-propagation training algorithm. The primary goal is to minimize the error at the output layer by searching for a set of connection strengths that cause the ANN to produce outputs that are equal to or closer to the targets. In this paper, a scaled conjugate gradient algorithm (Moller, 1993) is employed for training, and the training epoch is set to 500. The neurons of hidden layer use the tan–sigmoid transfer function, and the linear transfer function for output layer. The optimal number of neuron in the hidden layer was identified using a trial and error procedure by varying the number of hidden neurons from 2 to 13. The training data are further divided into the training set and the testing set. The cross-validation method is used and the number of hidden neurons was selected based on the RMSE. The effect of changing

Table 2
AIC value and performance indices of alternative ARMA models for Hongjiadu Hydropower.

(p, q)	AIC	Training				Validation			
		R	E	RMSE	MAPE	R	E	RMSE	MAPE
(5, 9)	9.231	0.722	0.523	91.57	44.06	0.760	0.557	97.32	49.76
(6, 10)	9.221	0.725	0.521	91.57	46.42	0.786	0.584	94.34	48.03
(7, 9)	9.242	0.724	0.520	91.89	44.91	0.748	0.538	99.39	48.50
(8, 9)	9.252	0.726	0.516	92.24	45.56	0.754	0.540	99.21	47.60
(10, 11)	9.268	0.722	0.501	93.68	42.30	0.760	0.540	99.22	46.29

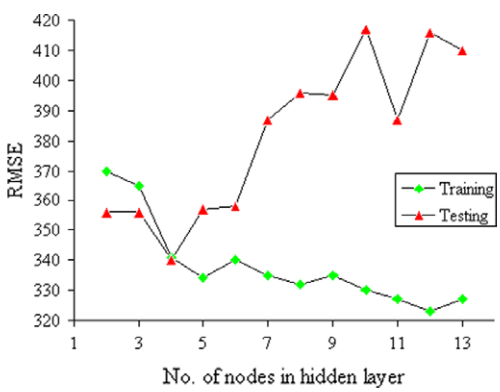

Fig. 12. Sensitivity of the number of nodes in the hidden layer on the RMSE of the neural network for Manwan Hydropower.

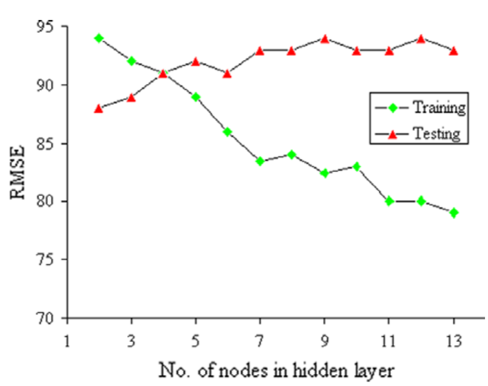

Fig. 13. Sensitivity of the number of nodes in the hidden layer on the RMSE of the neural network for Hongjiadu Hydropower.

the number of hidden neurons on the RMSE of the data set is shown in Figs. 12 and 13. It can be observed that the effect of the number of neurons assigned to the hidden layer has insignificant effect on the performance of the feed-forward model. When the numbers of hidden neurons are four, the training error is the closest to the testing error. Therefore, the numbers of hidden neurons are selected to be four and four for Manwan and Hongjiadu, respectively.

The ANFIS applies a hybrid learning algorithm that combines the back-propagation gradient descent and the least-squares estimate method, which outperforms the original back-propagation algorithm. Jang et al. (1997) provided the detailed description and mathematical background of the hybrid learning algorithm. An essential part of fuzzy logic is fuzzy sets defined by membership functions and rule bases. Shapes of the fuzzy sets are defined by the membership functions. The adjustment of adequate membership function parameters is facilitated by a gradient vector. After determining a gradient vector, the parameters are adjusted and the performance function is minimized via least-squares estimation. For the proposed Sugeno-type model, when the premise parameters are fixed, the overall output is expressed as linear combinations of the resulting parameters. The output f in Fig. 3 can be rewritten as:

$$\begin{aligned} f &= \overline{w_1}f_1 + \overline{w_2}f_2 \\ &= (\overline{w_1}x)p_1 + (\overline{w_1}y)q_1 + (\overline{w_1})r_1 + (\overline{w_2}x)p_2 + (\overline{w_2}y)q_2 + (\overline{w_2})r_2 \end{aligned} \quad (24)$$

The consequent parameters (p_1, q_1, r_1, p_2, q_2, r_2) are computed by the least-squares method, and the antecedent parameters are computed by back-propagation. Specifically, in the forward pass of the hybrid learning algorithm, node outputs go forward until layer 4 and the consequent parameters are identified by the least-squares method. In the backward pass, the error signals propagate backward and the premise parameters are updated by gradient descent. Consequently, the optimal parameters of the ANFIS model can be estimated using the hybrid learning algorithm. The training parameters of this model are given in Table 3.

Table 3
The training parameters of the ANFIS.

Parameters	Study sites	
	Manwan Hydropower	Hongjiadu Hydropower
Numbers of the rules	5	3
Epoch	15	10
Membership function	gbellmf	gbellmf
AND method	Prod	Prod
Imp. method	Minimum	Minimum
Aggr. method	Maximum	Maximum
Defuzzification method	wtaver	wtaver

GP has the ability to generate the best computer program to describe the relationship between the input and output. In this study, in order to find the optimal monthly flow series forecasting model, the selection of the appropriate parameters of GP evolution is necessary. Although the fine-tuning of algorithm was not the main concern of this paper, we investigated various initialization and run approaches and the adopted GP parameters are presented in Table 4. This setup furnished stable and effective runs throughout experiments. The evolutionary procedures are similar to GAs including defining the fitness function, genetic operators such as crossover, mutation and reproduction and the termination

Table 4
Values of primary parameters used in GP runs.

Parameter	Value
Terminal set	Variable x, random (0, 1)
Function set	+, −, ∗, /, sin, cos, ^
Population	2000 individuals
The maximum number of generations	100
Crossover rate	0.9
Mutation rate	0.05
Selection	Tournament with elitist strategy
Initial population	Ramped-half-and-half
The maximum depth of tree representation	9

criterion. In GP, the crossover operator is used to swap the sub-tree from the parents to reproduce the children using mating selection policy rather than exchanging bit strings as in GAs.

A Kernel function has to be selected from the qualified functions in using SVM. Dibike et al. (2001) applied different Kernels in SVR to rainfall–runoff modeling and demonstrated that the radial basis function (RBF) outperforms other Kernel functions. Also, many works on the use of SVR in hydrological modeling and forecasting have demonstrated the favorable performance of the RBF (Khan and Coulibaly, 2006; Lin et al., 2006; Liong and Sivapragasam, 2002; Yu et al., 2006). Therefore, the RBF is used as the Kernel function for prediction of discharge in this study. There are three parameters in using RBF Kernels: C, ε and σ. The accuracy of a SVM model is largely dependent on the selection of the model parameters. However, structured methods for selecting parameters are lacking. Consequently, some kind of model parameter calibration should be made. Recently, there are several methods developed to identify the parameters, such as the simulated annealing algorithms (Pai and Hong, 2005), GA (Pai, 2006) and the shuffled complex evolution algorithm (SCE-UA) (Lin et al., 2006; Yu et al., 2004). The SCE-UA method belongs to the family of evolution algorithm and was presented by Duan et al. (1993). In this study, the SCE-UA is employed as the method of optimizing parameters of SVM and a more comprehensive presentation can be found by Lin et al. (2006). To reach at a suitable choice of these parameters, the RMSE was used to optimize the parameters. Optimal parameters $(C, \varepsilon, \sigma) = (19.9373,\ 8.7775e-004,\ 1.2408)$ and $(C, \varepsilon, \sigma) = (0.5045,\ 5.0814e-004,\ 0.6623)$ were obtained for Manwan and Hongjiadu, respectively.

Results and discussion

The Manwan Hydropower, has been studied by Cheng et al. (2005) using ANFIS with discharges of monthly river flow discharges during 1953–2003, and by Lin et al. (2006) using SVM with discharges of monthly river flow discharges during 1974–2003. In their study, the R and RMSE were employed for evaluation model performance. In this paper, in order to identify more suitable models for forecasting future monthly inflows to hydropower reservoirs, the monthly discharge time series data of two study sites

Table 5
Forecasting performance indices of models for Manwan Hydropower.

Model	Training				Validation					
	R	RMSE	MAPE	E	R	RMSE	MAPE	E	Min	Max
Observed									334.0	3821.0
ARMA	0.922	354.27	16.77	0.849	0.928	354.35	15.63	0.869	373.4	3115.7
ANN	0.925	346.31	16.16	0.856	0.932	345.37	14.01	0.867	369.6	3307.8
ANFIS	**0.9322**	**329.77**	15.02	**0.869**	0.9405	335.02	14.30	0.883	**343.7**	**3509.3**
GP	0.918	360.96	17.79	0.843	0.9408	334.04	14.69	**0.8838**	360.1	3321.0
SVM	0.9315	334.07	**12.49**	0.866	**0.9410**	**332.86**	**12.49**	**0.8836**	369.0	3333.6

Notes: Min means minimum peak flows, and Max means maximum peak flows.

Table 6
Forecasting performance indices of models for Hongjiadu Hydropower.

Model	Training				Validation					
	R	RMSE	MAPE	E	R	RMSE	MAPE	E	Min	Max
Observed									25.5	619.0
ARMA	0.727	91.56	46.42	0.521	0.786	94.34	48.03	0.584	11.1	357.0
ANN	0.725	91.16	46.25	0.526	0.786	91.07	46.15	0.612	39.1	358.7
ANFIS	0.751	**87.38**	47.41	**0.564**	0.801	88.71	46.67	0.632	17.8	416.9
GP	0.734	90.28	50.29	0.535	0.815	**86.07**	50.81	**0.654**	27.6	**430.1**
SVM	**0.753**	89.89	**28.25**	0.539	**0.823**	87.57	**33.77**	0.641	**24.6**	382.8

Notes: Min means minimum peak flows, and Max means maximum peak flows.

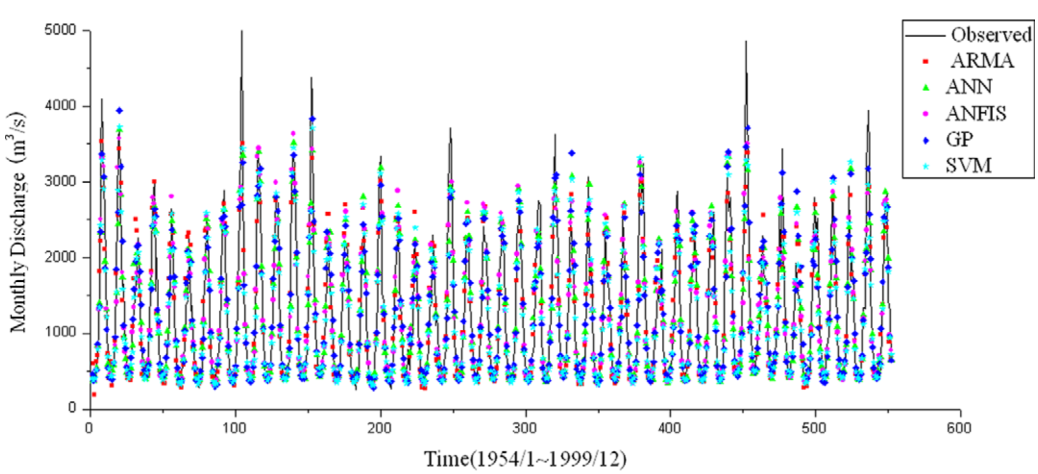

Fig. 14. Forecasted and observed flow during training period by ARMA, ANN, ANFIS, GP and SVM for Manwan Hydropower.

in different rivers are applied. For the same basis of comparison, the same training and verification sets, respectively, are used for all the above models developed, whilst the four quantitative standard statistical performance evaluation measures are employed to evaluate the performances of various models developed. Tables 5 and 6 present the results of Manwan and Hongjiadu study sites, respectively, in terms of various performance statistics.

It can be observed from Tables 4 and 5 that various AI methods have good performance during both training and validation, and they outperform ARMA in terms of all the standard statistical measures. For Manwan Hydropower, in the training phase, the ANFIS model obtained the best *R*, RMSE, and *E* statistics of 0.932, 329.77, and 0.869, respectively; while the SVM model obtained the best MAPE statistics of 12.49. Analyzing the results during validation, it can be observed that the SVM model outperforms all other models. Similarly, for Hongjiadu Hydropower, in the training phase, the ANFIS model obtained the best RMSE and *E* statistics of 887.38 and 0.564, respectively; while the SVM model obtained the

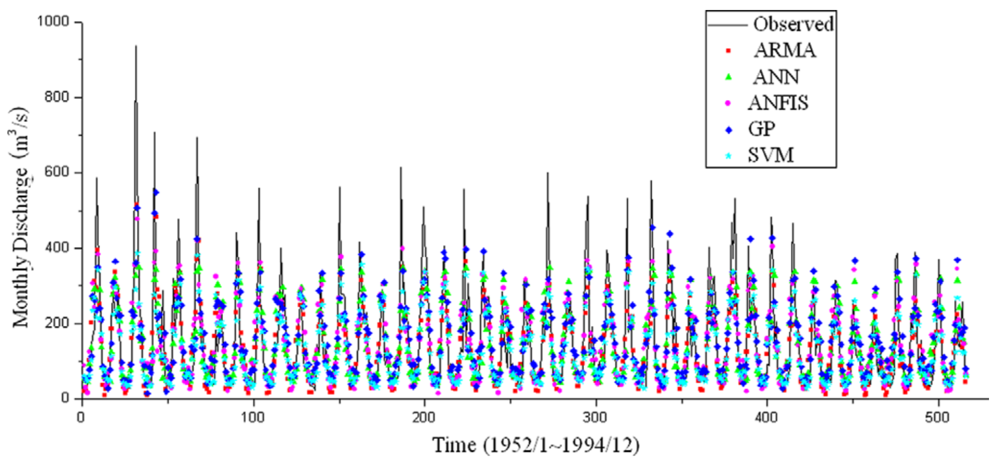

Fig. 15. Forecasted and observed flow during training period by ARMA, ANN, ANFIS, GP and SVM for Hongjiadu Hydropower.

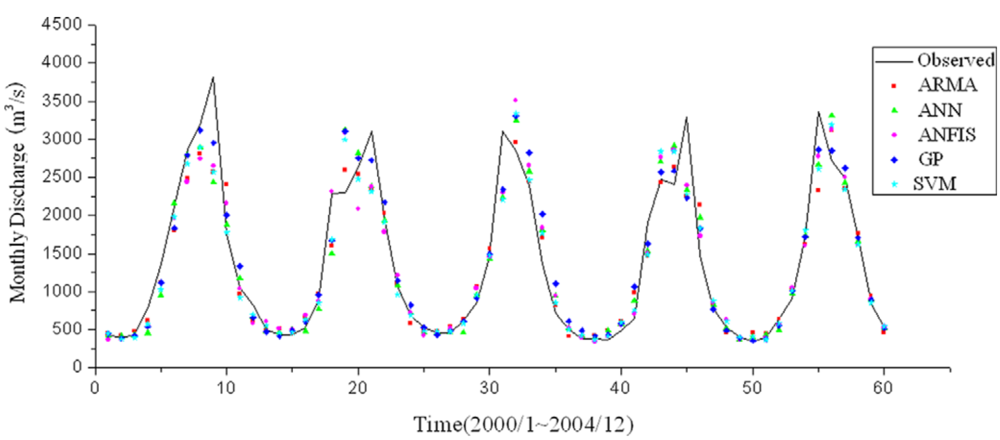

Fig. 16. Forecasted and observed flow during validation period by ARMA, ANN, ANFIS, GP and SVM for Manwan Hydropower.

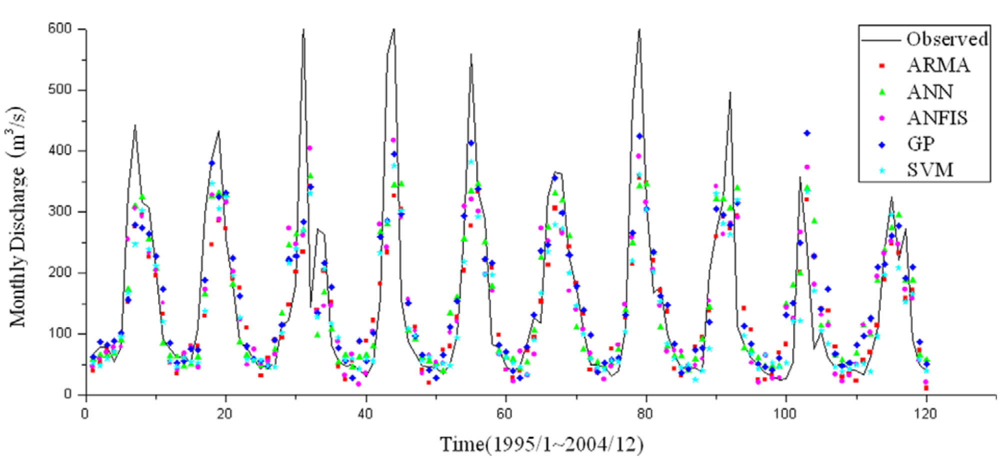

Fig. 17. Forecasted and observed flow during validation period by ARMA, ANN, ANFIS, GP and SVM for Hongjiadu Hydropower.

best R and MAPE statistics of 0.753 and 28.25, respectively. Analyzing the results during validation, the SVM model obtained the best R and MAPE statistics of 0.823 and 33.77, respectively, while the GP model obtained the best RMSE, and E statistics of 86.07 and 0.654, respectively. RMSE evaluates the residual between observed and forecasted flow, and MAPE measures the mean absolute percentage error of the forecast. R evaluates the linear correlation between the observed and computed flow, while E evaluates the capability of the model in predicting flow values away from the mean. According to the figures in Tables 4 and 5, we can conclude that the best performance of all AI methods developed in this paper is different in terms of the different statistical measures.

In addition, in the validation phase as seen in Tables 4 and 5, the values with the ANFIS, GP and SVM model prediction were able to produce a good, near forecast, as compared to those with ARMA and ANN model, whilst it can be concluded that the ANFIS model obtained the best minimum absolute error between the observed and modeled maximum and minimum peak flows in Manwan Hydropower, and the GP and SVM model obtained the best minimum absolute error between the observed and modeled maximum and minimum peak flows, respectively, in Hongjiadu Hydropower. In the validation phase, the SVM model improved the ARMA forecast of about 6.06% and 20.12% reduction in RMSE and MAPE values, respectively; Improvements of the forecast results regarding the R and E were approximately 1.22% and 1.69%, respectively, in Manwan Hydropower. In Hongjiadu Hydropower, the GP model obtained the best value of RMSE during the validation phase decreases by 8.77% and the best value of E increases by 11.99% comparing with ARMA; while, the SVM model obtained the best value of R during the validation phase increases by 4.71% and the best value of MAPE decreases by 29.69% comparing with ARMA. Thus the results of this analysis indicate that the ANFIS or SVM is able to obtain the best result in terms of different evaluation measures during the training phase, and the GP or SVM is able to obtain the best result in terms of different evaluation measures during the validation phase. Furthermore, as can be seen from Tables 4 and 5 that the virtues or defect degree of forecasting accuracy is different in terms of different evaluation measures during the training phase and the validation phase. SVM model is able to obtain the better forecasting accuracy in terms of different evaluation measures during the validation phase not only during the training phase but also during the validation phase. The forecasting results of ANFIS model during the validation phase are inferior to the results during the training phase. GP is in the middle or lower level in training phases, but the GP model is able to obtain the better forecasting result in validation phases, and especially the GP model is able to obtain the maximum peak flows among all models developed in Hongjiadu Hydropower. The performances of all prediction models developed in this paper during the training and validation periods in the two study sites are shown in Figs. 14–17.

Conclusions

An attempt was made in this study to investigate the performance of several AI methods for forecasting monthly discharge time series. The forecasting methods investigated include the ANNs ANFIS techniques, GP models and SVM method. The conventional ARMA is also employed as a benchmarking yardstick for comparison purposes. The monthly discharge data from actual field observed data in the Manwan Hydropower and Hongjiadu Hydropower were employed to develop various models investigated in this study. The methods utilize the statistical properties of the data series with certain amount of lagged input variables. Four standard statistical performance evaluation measures are adopted to evaluate the performances of various models developed.

The results obtained in this study indicate that the AI methods are powerful tools to model the discharge time series and can give good prediction performance than traditional time series approaches. The results indicate that the best performance can be obtained by ANFIS, GP and SVM, in terms of different evaluation criteria during the training and validation phases. SVM model is able to obtain the better forecasting accuracy in terms of different evaluation measures during the validation phase during both the training phase and the validation phase. The forecasting results of ANFIS model during the validation phase are inferior to the results during the training phase. GP is in the middle or lower level in training phases, but the GP model is able to obtain the better forecasting result in validation phases. The ANFIS and GP model obtain the maximum peak flows among all models developed in different studies sites, respectively. Therefore, the results of the study are highly encouraging and suggest that ANFIS, GP and SVM approaches are promising in modeling monthly discharge time series, and this may provide valuable reference for researchers and engineers who apply AI methods for modeling long-term hydrological time series forecasting. It is hoped that future research efforts will focus in these directions, i.e. more efficient approach for training multi-layer perceptrons of ANN model, the increased learning ability of the ANFIS model, the fine-tuning of algorithm for selecting more appropriate parameters of GP evolution, saving computing time or more efficient optimization algorithms in searching optimal parameters of SVM model etc to improve the accuracy of the forecast models in terms of different evaluation measures for better planning, design, operation, and management of various engineering systems.

Acknowledgements

This research was supported by the Central Research Grant of Hong Kong Polytechnic University (G-U265), the National Natural Science Foundation of China (No. 50679011), the public welfare industry research project of Ministry of Water Resources, PR China (No. 200801015) and the Scientific Research Foundation of North China Institute of Water Conservancy and Hydroelectric Power for High-level Talents (No. 200821). We gratefully acknowledge the critical comments and corrections of the anonymous reviewers, which improved the presentation of the paper considerably.

References

Alvisi, S., Mascellani, G., Franchini, M., Bardossy, A., 2006. Water level forecasting through fuzzy logic and artificial neural network approaches. Hydrology and Earth System Sciences 10 (1), 1–17.

Arena, C., Cannarozzo, M., Mazzola, M.R., 2006. Multi-year drought frequency analysis at multiple sites by operational hydrology – a comparison of methods. Physics and Chemistry of the Earth. Parts A/B/C 31 (18), 1146–1163.

ASCE Task Committee, 2000a. Artificial neural networks in hydrology – I: preliminary concepts. Journal of Hydrologic Engineering, ASCE 5 (2), 115–123.

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

Asefa, T., Kemblowski, M., McKee, M., Khalil, A., 2006. Multi-time scale stream flow predictions: the support vector machines approach. Journal of Hydrology 318 (1–4), 7–16.

Bazartseren, B., Hildebrandt, G., Holz, K.P., 2003. Short-term water level prediction using neural networks and neuro-fuzzy approach. Neurocomputing 55 (3–4), 439–450.

Borrelli, A., De Falco, I., Della Cioppa, A., Nicodemi, M., Trautteur, G., 2006. Performance of genetic programming to extract the trend in noisy data series. Physica A – Statistical Mechanics and Its Applications 370 (1), 104–108.

Box, G.E.P., Jenkins, G.M., 1970. Times Series Analysis Forecasting and Control. Holden-Day, San Francisco.

Campolo, M., Soldati, A., Andreussi, P., 2003. Artificial neural network approach to flood forecasting in the River Arno. Hydrological Sciences Journal 48 (3), 381–398.

Carlson, Robert F., MacCormick, A.J.A., Watts, Donald G., 1970. Application of linear random models to four annual streamflow series. Water Resources Research 6 (4), 1070–1078.

Chang, L.C., Chang, F.J., 2001. Intelligent control for modelling of real-time reservoir operation. Hydrological Processes 15 (9), 1621–1634.
Chang, F.J., Chang, L.C., Huang, H.L., 2002. Real-time recurrent learning neural network for stream-flow forecasting. Hydrological Processes 16 (13), 2577–2588.
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., Cheng, C.T., 2002. Real-time prediction of water stage with artificial neural network approach. Lecture Notes in Artificial Intelligence 2557, 715.
Chen, H.L., Rao, A.R., 2002. Testing hydrologic time series for stationarity. Journal of Hydrologic Engineering 7 (2), 129–136.
Chen, H.L., Rao, A.R., 2003. Linearity analysis on stationary segments of hydrologic time series. Journal of Hydrology 277 (1–2), 89–99.
Cheng, C.T., Lin, J.Y., Sun, Y.G., Chau, K.W., 2005. Long-term prediction of discharges in Manwan Hydropower using adaptive-network-based fuzzy inference systems models. Advances in Natural Computation, Pt 3, Proceedings. Lecture Notes in Computer Science. Springer-Verlag, Berlin. pp. 1152–1161.
Dibike, Y.B., Velickov, S., Solomatine, D., Abbott, M.B., 2001. Model induction with support vector machines: introduction and applications. Journal of Computing in Civil Engineering 15 (3), 208–216.
Dixon, B., 2005. Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: a GIS-based sensitivity analysis. Journal of Hydrology 309 (1–4), 17–38.
Duan, Q.Y., Gupta, V.K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient minimization. Journal of Optimization Theory and Applications 76 (3), 501–521.
Firat, M., Gungor, M., 2008. Hydrological time-series modelling using an adaptive neuro-fuzzy inference system. Hydrological Processes 22 (13), 2122–2132.
Freiwan, M., Cigizoglu, H.K., 2005. Prediction of total monthly rainfall in Jordan using feed forward back propagation method. Fresenius Environmental Bulletin 14 (2), 142–151.
Guven, A., Aytek, A., Yuce, M.I., Aksoy, H., 2008. Genetic programming-based empirical model for daily reference evapotranspiration estimation. Clean-Soil Air Water 36 (10–11), 905–912.
Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, 2nd ed. Prentice-Hall, Upper Saddle River, NJ.
Hipel, K.W., Mcleod, A.I., 1994. Time series modelling of water resources and environmental systems. Elsevier, Amsterdam. pp. 463–465.
Hu, T.S., Lam, K.C., Ng, S.T., 2001. River flow time series prediction with a range-dependent neural network. Hydrological Sciences Journal 46 (5), 729–745.
Jain, A., Kumar, A.M., 2007. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing 7 (2), 585–592.
Jang, J.-S.R., 1993. ANFIS: adaptive-network-based fuzzy inference systems. IEEE Transactions on Systems, Man and Cybernetics 23 (3), 665–685.
Jang, J.-S.R., Sun, C.-T., Mizutani, E., 1997. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Prentice-Hall, Upper Saddle River, NJ.
Johari, A., Habibagahi, G., Ghahramani, A., 2006. Prediction of soil–water characteristic curve using genetic programming. Journal of Geotechnical and Geoenvironmental Engineering 132 (5), 661–665.
Keskin, M.E., Taylan, D., Terzi, O., 2006. Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. Hydrological Sciences Journal 51 (4), 588–598.
Khan, M.S., Coulibaly, P., 2006. Application of support vector machine in lake water level prediction. Journal of Hydrologic Engineering 11 (3), 199–205.
Komornik, J., Komornikova, M., Mesiar, R., et al., 2006. Comparison of forecasting performance of nonlinear models of hydrological time series. Physics and Chemistry of the Earth. Parts A/B/C 31 (18), 1127–1145.
Koza, J., 1992. Genetic Programming: On the Programming of Computers by Natural Selection. MIT Press, Cambridge, MA.
Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research 35 (1), 233–241.
Lin, J.Y., Cheng, C.T., Chau, K.W., 2006. Using support vector machines for long-term discharge prediction. Hydrological Sciences Journal 51 (4), 599–612.
Liong, S.Y. et al., 2002. Genetic Programming: a new paradigm in rainfall–runoff modeling. Journal of the American Water Resources Association 38 (3), 705–718.
Liong, S.Y., Sivapragasam, C., 2002. Flood stage forecasting with support vector machines. Journal of the American Water Resources Association 38 (1), 173–186.
Loukas, Y.L., 2001. Adaptive neuro-fuzzy inference system: an instant and architecture-free predictor for improved QSAR studies. Journal of Medicinal Chemistry 44 (17), 2772–2783.
Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6 (4), 525–533.
Moriasi, D.N. et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50 (3), 885–900.
Muttil, N., Chau, K.W., 2006. Neural network and genetic programming for modelling coastal algal blooms. International Journal of Environment and Pollution 28 (3–4), 223–238.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. Journal of Hydrology 10 (3), 282–290.
Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., 2004. A neuro-fuzzy computing technique for modeling hydrological time series. Journal of Hydrology 291 (1–2), 52–66.
Nayak, P.C., Sudheer, K.P., Rangan, D.M., Ramasastri, K.S., 2005. Short-term flood forecasting with a neuro-fuzzy model. Water Resources Research 41 (4) (Art. NO. W04004).
Olsson, J., Uvo, C., Jinno, B., et al., 2004. Neural networks for rainfall forecasting by atmospheric downscaling. Journal of Hydrologic Engineering 9 (1), 1–12.
Pai, P.-F., 2006. System reliability forecasting by support vector machines with genetic algorithms. Mathematical and Computer Modelling 43 (3–4), 262–274.
Pai, P.-F., Hong, W.-C., 2005. Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Conversion and Management 46 (17), 2669–2688.
Salas, J.D., 1993. Analysis and modeling of hydrologic time series. In: Maidment, D.R. (Ed.), Handbook of Hydrology. The McGraw Hill, New York, pp. 19.5–19.9.
Santhi, C. et al., 2001. Validation of tbe swat model on a large river basin with point and nonpoint sources. Journal of the American Water Resources Association 37 (5), 1169–1188.
Shu, C., Ouarda, T.B.M.J., 2008. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. Journal of Hydrology 349 (1–2), 31–43.
Silverman, D., Dracup, J.A., 2000. Artificial neural networks and long-range precipitation prediction in California. Journal of Applied Meteorology 39 (1), 57–66.
Sivapragasam, C., Vincent, P., Vasudevan, G., 2007. Genetic programming model for forecast of short and noisy data. Hydrological Processes 21 (2), 266–272.
Srikanthan, R., McMahon, T.A., 2001. Stochastic generation of annual, monthly and daily climate data: a review. Hydrology and Earth System Sciences 5 (4), 653–670.
Sudheer, K.P., Jain, A., 2004. Explaining the internal behaviour of artificial neural network river flow models. Hydrological Processes 18 (4), 833–844.
Sudheer, K.P., Gosain, A.K., Ramasastri, K.S., 2002. A data-driven algorithm for constructing artificial neural network rainfall–runoff models. Hydrological Processes 16 (6), 1325–1330.
Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology 239 (1–4), 132–147.
Van Liew, M.W., Arnold, J.G., Garbrecht, J.D., 2003. Hydrologic simulation on agricultural watersheds: choosing between two models. Transactions of the ASAE 46 (6), 1539–1551.
Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, New York.
Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.
Wang, T., Yang, K.L., Guo, Y.X., 2008. Application of artificial neural networks to forecasting ice conditions of the Yellow River in the Inner Mongolia Reach. Journal of Hydrologic Engineering 13 (9), 811–816.
Whigam, P.A., Crapper, P.F., 2001. Modelling rainfall–runoff relationships using genetic programming. Mathematical and Computer Modelling 33, 707–721.
Wu, C.L., Chau, K.W., Li, Y.S., 2008. River stage prediction based on a distributed support vector regression. Journal of Hydrology 358 (1–2), 96–111.
Yu, X.Y., Liong, S.Y., Babovic, V., 2004. EC-SVM approach for real-time hydrologic forecasting. Journal of Hydroinformatics 6 (3), 209–223.
Yu, P.S., Chen, S.T., Chang, I.F., 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology 328 (3–4), 704–716.
Zounemat-Kermani, M., Teshnehlab, M., 2008. Using adaptive neuro-fuzzy inference system for hydrological time series prediction. Applied Soft Computing 8 (2), 928–936.