Stoch Environ Res Risk Assess (2014) 28:1755–1767
DOI 10.1007/s00477-013-0838-3

ORIGINAL PAPER

Copula entropy coupled with artificial neural network for rainfall–runoff simulation

Lu Chen · Vijay P. Singh · Shenglian Guo · Jianzhong Zhou · Lei Ye

Published online: 24 December 2013

Abstract The rainfall–runoff relationship is not only nonlinear and complex but also difficult to model. Artificial neural network (ANN), as a data-driven technique, has gained significant attention in recent years and has been shown to be an efficient alternative to traditional methods for hydrological modeling. However, for different input combinations, ANN models can yield different results. Therefore, input variables and ANN types need to be carefully considered, when using an ANN model for stream flow forecasting. This study proposes the copula-entropy (CE) theory to identify the inputs of an ANN model. The CE theory permits to calculate mutual information (MI) and partial MI directly which avoids calculating the marginal and joint probability distributions. Three different ANN models, namely multi-layer feed (MLF) forward networks, radial basis function networks and general regression neural network, were applied to predict stream flow of Jinsha River, China. Results showed that the inputs selected by the CE method were better than those by the traditional linear correlation analysis, and the MLF ANN model with the inputs selected by CE method obtained the best predicted results for the Jinsha River at Pingshan gauging station.

Keywords Rainfall–runoff simulation · Input variables selection · Copula entropy · Artificial neural networks

L. Chen (✉) · J. Zhou · L. Ye
College of Hydropower & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China
e-mail: chl8505@126.com

V. P. Singh
Department of Biological and Agricultural Engineering, Texas A&M University, TAMU, College Station, TX 77843-2117, USA

V. P. Singh
Department of Civil and Environmental Engineering, Texas A&M University, TAMU, College Station, TX 77843-2117, USA

S. Guo
State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

1 Introduction

Flooding is the most common natural hazard and third most damaging globally after storms and earthquakes. Extreme flood events have been and continue to be one of the most important natural hazards responsible for deaths and economic losses. In China, a number of floods have occurred over the last 100 years and caused great economic losses (Zhang and Hall 2004). For example, floods in the Yangtze River (Chang Jiang) basin in central and eastern China have occurred periodically and often have caused considerable destruction of property and loss of life. Among the most major flood events are those of 1870, 1931, 1954, 1998, and 2010. In 1998, the Yangtze River basin suffered from tremendous flooding—the largest flood since 1954, which led to the economic loss of 166 billion Chinese Yuan (or nearly 20 billion US$; Yin and Li 2001). With the use of precipitation data, a rainfall–runoff model is used to forecast flow for periods ranging from a few hours to days ahead, depending on the size of the watershed or river basin. The forecast values are of great importance for the reduction of disasters and the optimal use of water resources.

The rainfall–runoff relation, needed for flood forecasting, is not only complex and nonlinear but is also difficult to model. There are many models used for rainfall–runoff simulation. A data-driven technique that has gained significant attention for its effectiveness in function approximation characteristics is artificial neural network (ANN) modeling (de Vos and Rientjes 2005; Kasiviswanathan and Sudheer 2013). Many studies focusing on stream flow predictions have shown that ANN is superior to traditional regression techniques and time-series models, including autoregressive (AR) and AR moving average (ARMA) (Raman and Sunilkumar 1995; Jain et al. 1999; Thirumalaiah and Deo 2000; Abrahart and See 2002; Castellano-Méndeza et al. 2004). Hsu et al. (1995) showed that the ANN model provided a better representation of the rainfall–runoff relationship than the autoregressive moving average with exogenous inputs (ARMAX) time series model or the conceptual Sacramento soil moisture accounting models. Shamseldin (1997) examined the effectiveness of rainfall–runoff modeling with ANNs by comparing their results with the Simple Linear Model, the seasonally based Linear Perturbation Model (LPM) and the Nearest Neighbor LPM, and concluded that ANNs provided more accurate discharge forecasts than some of the traditional models. Birikundavyi et al. (2002) investigated the ANN models for daily stream flow prediction and also showed that ANNs outperformed the classic AR model coupled with a Kalman filter (ARMAX-KF) and a conceptual model (PREVIS). Therefore, ANNs have proved to be an excellent tool for rainfall–runoff simulation.

However, in most ANN applications to rainfall–runoff simulation, little attention has given to the task of selecting appropriate model inputs (Maier and Dandy 2000). Usually, not all of the potential input variables are equally informative, since some may be correlated, noisy or have no significant relationship with the output variable being modeled (Bowden et al. 2005a). Presenting a large number of inputs to ANN models and relying on the network to determine the critical model inputs usually increases the network size (Maier and Dandy 2000). This also has a number of disadvantages, such as decreasing processing speed and increasing the amount of data required to estimate the connection weights efficiently (Lachtermacher and Fuller 1994).

Bowden et al. (2005a) reviewed the methods for input determination in water resources ANN applications. Three most commonly used approaches are methods that rely on the use of a priori knowledge of the system being modeled, methods based on linear correlation, and methods that utilize a heuristic approach. The priori knowledge method which depends on an expert's knowledge is very subjective and case dependent. The drawbacks of the linear correlation method are summarized as: (a) it only applies to linear correlation, and (b) it tends to focus on the degree of dependence, and ignore the structure of dependence (Zhao and Linb 2011). For a heuristic approach, various ANN models are trained using different subsets of inputs. The main disadvantage of these approaches is that they are based on trial-and-error, and as such there is no guarantee that they will find the globally best subsets. Another disadvantage of step wise approaches is that they are computationally intensive (Bowden et al. 2005a).

Maier and Dandy (2000) indicated that there were distinct advantages in using analytical techniques to help determine the inputs for multivariate ANN models. Mutual information (MI) is an analytical and non-linear method to measure the dependences, which has been successfully employed by many researchers (e.g. Mishra and Singh 2009; Angulo et al. 2011; Jeong et al. 2012; Tongal et al. 2013; Mishra et al. 2013). However, there is a disadvantage when using MI to select inputs of ANN. Although a candidate model input might have a strong relationship with the model output, this information might be redundant if the same information is already provided by another input (Fernando et al. 2009). Sharma (2000) proposed an input determination method based on the partial MI (PMI) criterion. In a review of approaches used to select the inputs to ANN models, Bowden et al. (2005a) concluded that the PMI algorithm of Sharma was superior to methods commonly used to determine the inputs to ANN models, as it is model-free, uses a non-linear measure of dependence (MI), is able to cater to input redundancy and has a well-defined stopping criterion (Fernando et al. 2009). May et al. (2008) used this model for forecasting water quality in water distribution systems. Furthermore, Fernando et al. (2009) modified PMI input selection algorithm in order to increase its computational efficiency, while maintaining its accuracy. They introduced the average shifted histograms as an alternative to kernel based methods for the estimation of MI.

However, there are several disadvantages of the methods of PMI algorithm mentioned above. First, hydrological events, such as rainfall and runoff, are continuous. Actually, some methods mentioned above used the discrete version to calculate PMI. Therefore, a method for continuous variable should be used. Second, these methods need to estimate both the marginal and joint probability density distributions, which involve a product of two terms and lead to a complex calculation work. In order to overcome this problem, this paper will apply the copula and entropy based method, named copula entropy (CE) method, to calculate the MI and PMI values.

Ma and Sun (2011) took into account the copula function and entropy theory together and introduced the concept of CE. CE is defined as the entropy of copula function, which is related to the joint entropy, marginal entropy and

MI. Research on CE has received significant attention recently. Calsaverini and Vicente (2009) discussed a couple of consequences yielded by connections between the copula and entropy methods, which involve CE. Zhao and Linb (2011) applied CE models with two and three variables to measure the dependence in stock markets. The advantage of the CE is summarized as follows: (a) it makes no assumptions about the marginal distributions and can be used for higher dimensions, and (b) the MI can be obtained from the calculation of the CE instead of the marginal or joint entropy, which estimates the MI more directly and avoids the accumulation of systematic bias. Until now the CE method has not been widely used in the hydrological field.

Furthermore, when using an ANN model, the appropriate ANN model needs to be selected. Lekkas et al. (2004) indicated that it is preferable for every new application to test different types of ANNs rather than using a pre-selected one. A large number of ANN architectures and algorithms have been developed so far, such as multilayer feed (MLF) forward networks (Rumelhart et al. 1986), self-organising feature maps (Kohonen 1982), Hopfield networks (1987), counterpropagation networks (Hecht-Nielsen 1987), radial basis function (RBF) networks (Powell 1987), general regression neural network, GRNN (Specht 1991), and recurrent ANNs (Elman 1988). Of these networks, the most commonly used are feedforward networks and RBF networks (Karunanithi et al. 1994). MLF forward networks have been found to perform best when used in hydrological applications (Hsu et al. 1995) and as such they are by far the most commonly used (Maier and Dandy 2000; Lekkas et al. 2004). Jayawardena et al. 1997 compared multilayer perceptron (MLP) and RBF ANN approaches in flood forecasting applications. Results showed that the RBF network based models gave predictions comparable in accuracy to those from the MLP based models. Park and Sandberg (1991) proved theoretically that the RBF type ANNs were capable of universal approximations and learning without local minima, thereby guaranteeing convergence to globally optimum parameters. In addition, Bowden et al. (2005a, b) recommended the GRNN, a class of ANN that was first introduced by Specht (1991), for hydrological prediction, because of its advantages of nonlinear modeling between inputs and output, fixed network architecture and quicker training than other ANNs. Therefore, the MLF forward networks, RBF networks and GRNN were considered in this study.

This study aims to build and improve the accuracy of the hydrological rainfall–runoff model established by an ANN method. The input technique, CE method, was first proposed and used to select optimal model inputs. Three representative models, namely MLF forward networks, RBF networks and GRNN, were then applied to conduct rainfall–runoff prediction. The upper Jinsha River was selected as a case study. The performance of these models was analyzed and compared. Finally, the model with best predicted results for Jinsha River was determined.

2 Determination of model inputs of ANN based on CE

Following the discussion in the introduction section, an analytical technique, namely CE method, was used for input selection. The whole process is described in three steps. First, the CE is introduced. Second, theory of PMI is given. Third, according to the CE and PMI theories, the input selection procedure using CE method is introduced.

2.1 Copula entropy

Recent studies have shown that copula modeling can provide a simple, yet powerful framework for modeling interdependence among hydrological data (Nazemi and Elshorbagy 2012). The copula function has been widely used in hydrological field, such as flood frequency analysis (Favre et al. 2004; Chen et al. 2010, 2012) and drought analysis (Song and Singh 2010a, b; Chen et al. 2013). However, those researches mainly focused on making use of the copula function to establish the joint distribution. In this study, the entropy of the copula function, named CE, was proposed to measure the dependence between variables. The definition of CE was as follows.

Let X_1, X_2 be random variables with marginal functions $F(x_1)$, $F(x_2)$, and $U_1 = F(x_1)$, $U_1 = F(x_2)$. Then U_1 and U_2 are uniformly distributed random variables; and u_1 and u_2 will denote a specific value of U_1 and U_2, respectively. We define the entropy of copula function as CE in this study, which can be expressed as:

$$H_C(U_1, U_2) = -\int_0^1 \int_0^1 c(u_1, u_2)\log(c(u_1, u_2))du_1du_2, \quad (1)$$

where $c(u_1, u_2)$ is the probability density function of copulas and is expressed as $\frac{\partial C(u_1, u_2)}{\partial u_1 \partial u_2}$.

The relationship between CE and MI was investigated as follows.

The joint probability density function of variable x_1, x_2 can be defined as (Grimaldi and Serinaldi 2006):

$$f(x_1, x_2) = c(u_1, u_2)f(x_1) \cdot f(x_2). \quad (2)$$

Based on Eqs. (1) and (2), the joint entropy can be expressed as:

$$
\begin{aligned}
H(X_1,\ X_2) &= -\int_0^\infty\int_0^\infty f(x_1,\ x_2)\log[f(x_1,\ x_2)]dx_1dx_2 \\
&= -\int_0^\infty\int_0^\infty c(u_1,\ u_2)f(x_1)f(x_2)\log[c(u_1,\ u_2)f(x_1)f(x_2)]dx_1dx_2 \\
&= -\int_0^\infty\int_0^\infty c(u_1,\ u_2)f(x_1)f(x_2)\{\log[c(u_1,\ u_2)] + \log[f(x_1)] + \log[f(x_2)]\}dx_1dx_2 \\
&= -\int_0^\infty\int_0^\infty c(u_1,\ u_2)f(x_1)f(x_2)\cdot\log[c(u_1,\ u_2)]dx_1dx_2 \\
&\quad -\int_0^\infty\int_0^\infty c(u_1,\ u_2)f(x_1)f(x_2)\cdot\{\log[f(x_1)] + \log[f(x_2)]\}dx_1dx_2 \\
&= A + B,
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
A &= -\int_0^\infty\int_0^\infty c(u_1,\ u_2)f(x_1)f(x_2)\{\log[f(x_1)] + \log[f(x_2)]\}dx_1dx_2 \\
&= -\int_0^\infty\int_0^\infty f(x_1,\ x_2)\cdot\{\log[f(x_1)] + \log[f(x_2)]\}dx_1dx_2 \\
&= -\int_0^\infty\int_0^\infty f(x_1,\ x_2)\cdot\log[f(x_1)]dx_1dx_2 - \int_0^\infty\int_0^\infty f(x_1,\ x_2)\cdot\log[f(x_2)]dx_1dx_2 \\
&= -\int_0^\infty \log[f(x_1)]\left[\int_0^\infty\int_0^\infty f(x_1,\ x_2)\cdot dx_2\right]dx_1 - \int_0^\infty \log[f(x_2)]\left[\int_0^\infty\int_0^\infty f(x_1,\ x_2)\cdot dx_1\right]dx_2 \\
&= -\int_0^\infty f(x_1)\log[f(x_1)]dx_1 - \int_0^\infty f(x_2)\log[f(x_2)]dx_2 \\
&= H(X_1) + H(X_2).
\end{aligned}
\tag{4}
$$

Using equation $du_1 = dx_1{\cdot}f(x_1)$ and $du_2 = dx_2{\cdot}f(x_2)$,

$$
\begin{aligned}
B &= -\int_0^\infty\int_0^\infty c(u_1,\ u_2)f(x_1)f(x_2)\cdot\log[c(u_1,\ u_2)]dx_1dx_2 \\
&= -\int_0^\infty\int_0^\infty c(u_1,\ u_2)\cdot\log[c(u_1,\ u_2)]du_1du_2 \\
&= H_C(U_1,\ U_2).
\end{aligned}
\tag{5}
$$

Therefore, the joint CE can be expressed as the sum of the d univariate marginal entropies and the CE as follows:

$$H(X_1,\ X_2) = H(X_1) + H(X_2) + H_C(U_1,\ U_2). \tag{6}$$

Equation (6) indicates that the joint entropy $H(X_1, X_2)$ is expressed as the sum of the marginal entropies $H(X_1)$, $H(X_2)$ and the CE $H_C(U_1, U_2)$.

The MI can be expressed as:

$$T(X_1,\ X_2) = H(X_1) + H(X_2) - H(X_1,\ X_2). \tag{7}$$

Therefore, from Eqs. (6) and (7), one can write

$$
\begin{aligned}
T(X_1,\ X_2) &= H(X_1) + H(X_2) - H(X_1,\ X_2) \\
&= -H_C(X_1,\ X_2).
\end{aligned}
\tag{8}
$$

Equation (8) indicates that the MI is equal to the negative CE.

2.2 Partial MI

MI can be used to identify the non-linear dependence between candidate input and output variables (Fernando et al. 2009). However, this method is not directly able to deal with the issue of redundant inputs (Bowden et al. 2005a). To overcome this problem, Sharma (2000) introduced the concept of PMI, which is a more general approach because it relates also to nonlinear dependencies, and it needs no explicit modeling. It represents the information between two observations that is not contained in a third one and provides a measure of the partial or additional dependence the new input can add to the existing prediction model (Bowden et al. 2005a). The PMI between the output (dependent variable) y and the input (independent variable) x, for a set of pre-existing inputs $\mathbf{z}$, can be given by Bowden et al. (2005a):

$$\mathrm{PMI} = \int\int f_{X',Y'}(x', y') \ln\left[\frac{f_{X',Y'}(x', y')}{f_{X'}(x')f_{Y'}(y')}\right] dx' dy', \tag{9}$$

where $x' = x - E[x|\mathbf{z}]$; $y' = y - E[y|\mathbf{z}]$, where E denotes the expectation operation. Variables x' and y' only contain the residual information in variables x and y after considering the effect of already selected input $\mathbf{z}$ (Fernando et al. 2009).

The calculation of PMI can be divided into two parts, the first of which is the calculation of the variables x' and y', and the second is the calculation of PMI. For the first step, one kind of ANN model, namely, the GRNN, can be used to estimate the variables x' and y', since GRNN is an estimate for $E[y|\mathbf{X}]$, which is the conditional expectation of y given x. For the second step, the CE method can be used to calculate the PMI value.

According to the relationship between CE and MI (or PMI), the PMI can be obtained based on the CE method. Therefore, the calculation of PMI can be converted to the calculation of the CE. The relation between PMI and CE can be expressed as:

$$\begin{aligned}\mathrm{PMI} &= \int\int f_{X',Y'}(x', y') \ln\left[\frac{f_{X',Y'}(x', y')}{f_{X'}(x')f_{Y'}(y')}\right] dx' dy' \\ &= -H_C(x', y').\end{aligned} \tag{10}$$

Therefore, the calculation of PMI is converted to calculate the negative CE. The input variables can be determined based on the CE method.

2.3 Input selection based on CE method

First, the copula function between the potential inputs and output are built. Parameters of the copula function need to be estimated, and then the copula probability density function can be determined.

Second, the values of CE are calculated. According to Eq. (1), the CE can be derived using the multiple integration method. The integrand function has been known based on step one. The multiple integration method, proposed by Berntson et al. (1991), was applied to do multiple integrations. In order to test this multiple integration method, we used the copula probability density function as an integrand. The result of integration should be 1.

Third, the CE algorithm requires a reliable and efficient criterion to decide when to stop the addition of new inputs to the list of selected inputs (Fernando et al. 2009). Hampel identifier is an outlier detection method for determining whether a given value x is significantly different from others within a set of values X. Assume that a set of candidates will initially contain some proportion of redundant variables, and significant variable will be detected. The Hampel distance begins by calculating the absolute deviation from the median PMI for all candidates and defined as (May et al. 2008; Fernando et al. 2009):

$$d_j = \left|\mathrm{CE}_j - \mathrm{CE}^{(50)}\right|, \tag{11}$$

where d_j is the absolute deviation of the jth candidate, CE_j represents the CE values of the jth candidate, and $\mathrm{CE}^{(50)}$ denotes the median CE value for candidate set.

Then the Hampel distance is calculated by Fernando et al. (2009) and May et al. (2008):

$$HD_j = \frac{d_j}{1.4826 d_j^{(50)}}, \tag{12}$$

where d_j denotes the Hampel distance, and $d_j^{(50)}$ denotes the median absolute deviation d_j. If the Hampel distance is greater than 3, namely $HD_j > 3$, then add the candidates to the selected input set.

3 Determination of ANN models

3.1 ANN models

The majority of studies on ANNs employed rainfall and previous flow (or water level) as potential inputs (Wu and Chau 2011). Poor forecasting results were obtained, when using only the rainfall data (Campolo et al. 1999). The necessity of using the previous flow data is widely recognized. Furthermore, the frequently used inputs to ANNs also include observed runoff at nearby sites or neighboring sub-basins. Therefore, in this study, a representative ANN model for rainfall–runoff model can be defined as:

$$\hat{Q}_t = f(Q_{t-l_1}, R_{t-l_2}, X_{t-l_3}), \tag{13}$$

where $\hat{Q}_t$ stands for the predicted flow at time instance t, Q_{t-l_1} is the antecedent flow (up to $t - l_1$ time steps), R_{t-l_2} is the antecedent rainfall, and X_{t-l_3} represents the observed runoff at a neighboring sub-basin in this study.

In this study, three ANN models were used. They are MLF forward networks, RBF networks, and GRNN. Details of these models are introduced in the following.

3.1.1 MLF-forward networks

The feedforward neural network was the first and is arguably the simplest type of ANN devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any), to the output nodes. There are no cycles or loops in the network. MLF neural networks, trained with a back-propagation learning algorithm, are the most popular neural networks (Zupan and Gasteiger 1993).

An MLF neural network consists of neurons, which are ordered into layers. The first layer is called the input layer, the last layer is called the output layer, and the layers between are hidden layers (Svozil et al. 1997). Each neuron in a particular layer is connected with all neurons in the next layer. The connection between the ith and jth neuron is characterised by the weight coefficient w_{ij} and the ith neuron by the threshold coefficient θ_i. The weight coefficient reflects the degree of importance of the given connection in the neural network. The output value (activity) of the ith neuron x_i is determined by Eqs. (14) and (15) in the following:

$$x_i = f(\xi_i), \tag{14}$$

$$\xi_i = \vartheta_i + \Sigma\omega_{ij}x_j, \tag{15}$$

where ξ_i is the potential of the ith neuron, and $f(\cdot)$ is so-called transfer function. In MLF forward neural networks, the most popular non-linear transfer function used in neural network studies is the logistic function, defined by $f(\xi) = 1/(1 + e^{-\xi})$.

The supervised adaptation process varies the threshold coefficients ξ and weight coefficients w_{ij} to minimize the sum of the squared differences between the computed and required output values (Svozil et al. 1997).

3.1.2 Radial basis function (RBF)

An RBF network is a three-layer feed-forward type network. The three layers include the input layer, the hidden layer and the output layer. The input of RBF is transformed by the basic functions at the hidden layer. At the output layer, linear combinations of the hidden layer node responses are added to form the output (Jayawardena et al. 1997).

The name RBF comes from the fact that the basic functions in the hidden layer nodes are radially symmetric. The most common choice, however, is the Gaussian function which can be defined by a mean U and a standard deviation σ. For an input X, the jth hidden node produces a response given as (Jayawardena et al. 1997):

$$h_j = \exp\left\{\frac{-||X^i - U_j||}{2\sigma_j^2}\right\}, \tag{16}$$

where $X^i - U_j$ is the distance between the point representing the input X and the center of the hidden node as measured by some norm. In RBF networks, the connections between input and hidden layers are not weighted. The inputs therefore reach the hidden layer nodes unchanged.

The output y_i of the network at the output node is given as:

$$y_i = \sum_{i=1}^{m} h_i w_{ij}. \tag{17}$$

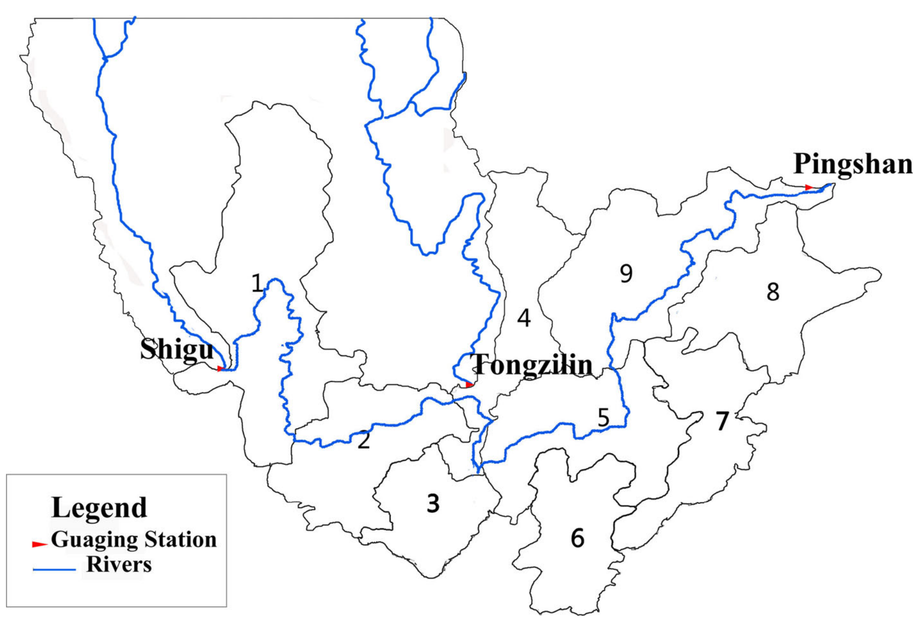

Fig. 1 Jinsha River basin

Parameters of an RBF type neural network are the mean U and standard deviation σ of the basic functions at the hidden layer nodes, and the synaptic weights w_{ij} of the output layer nodes.

3.1.3 GRNN model

The GRNN, developed by Specht (1991), is a simple yet very effective local approximation based neural network in the sense of estimating a probability distribution function (Islam et al. 2001). The GRNN network is a three-layer network with one hidden layer. The GRNN paradigm is briefly outlined below and details can be found in Specht (1991). Assume that $f(\mathbf{x}, y)$ represents a known joint continuous probability density function of a vector random variable, $\mathbf{x}$, and a scalar random variable, y. Let $\mathbf{X}$ be a particular measured value of the random variable $\mathbf{x}$. The conditional mean of y given $\mathbf{X}$, also called the regression of y on $\mathbf{X}$, is given by Specht (1991):

$$E[y|\mathbf{X}] = \frac{\int_{-\infty}^{\infty} yf(\mathbf{X}, y)dy}{\int_{-\infty}^{\infty} f(\mathbf{X}, y)dy}, \tag{18}$$

where $f(\mathbf{X}, y)$ is not known, then a sample of observations of $\mathbf{X}$ and y is used to obtain an estimate $\hat{f}(\mathbf{X}, y)$. The GRNN is an estimate of $E[y|\mathbf{X}]$, which is the conditional expectation of y given $\mathbf{X}$.

3.2 Performance indexes

The performance indexes were used to evaluate the established ANN model, and the one with the best performance was finally selected for the rainfall–runoff simulation.

The performance of the hydrological forecasting models was assessed in accordance with the criteria specified by the Ministry of Water Resources of China (2006). These are the coefficient of efficiency (i.e., Nash–Sutcliffe efficiency, R^2), which is a measure of the goodness-of-fit between recorded and predicted discharge time series data, and the 'qualified rate' (α) of predicted individual flood event peak discharges and volumes (Li et al. 2010). A forecast peak discharge or flood volume is termed 'qualified', when the difference between the predicted and the recorded values is within ±20 % of the recorded value. The root-mean-square error (RMSE) between observed and predicted flood values was also used as a performance criterion in this study. The formulas of these criteria are given as follows.

Nash–Sutcliffe efficiency R^2 is defined as:

$$R^2 = \left[1 - \frac{\sum_{i=1}^{N}(Q_t - \hat{Q}_t)^2}{\sum_{i=1}^{N}(Q_t - \bar{Q})^2}\right] \times 100\,\%, \tag{19}$$

where N represents total number of observations, Q_t and $\hat{Q}_t$ represent the observed and predicted discharges at time t, respectively, and $\bar{Q}$ means the mean value of Q_t.

RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Q_t - \hat{Q}_t)^2}{N}}. \tag{20}$$

Table 1 Rainfall stations of the study area

Sub-basins	Stations	Longitude	Latitude	Length of records
1	Shigu	99°56′00″ E	26°54′00″ N	1990–2010
	Daqiaotou	102°54′00″	26°37′00″	
	Luoji	102°54′00″	28°48′00″	2004–2010
	Jinmian	101°35′59″	27°11′0″	2004–2010
	Zongguantian	100°34′33″	26°48′23″	2002–2010
	Huangping	100°23′12″	26°05′25″	2004–2010
2	Jinjiangjie	100°32′32″	26°13′22″	2001–2010
	Dahuizhuang	100°32′11″	25°57′41″	2004–2010
	Renli	101°00′00″	26°29′00″	2000–2010
3	Fengtun	101°22′42″	25°17′51″	2000–2010
	Baihe	101°12′00″	26°40′0″	2004–2010
	Duoke	101°51′44″	25°49′43″	2004–2010
	Doubashi	102°45′40″	25°23′11″	2004–2010
4	Panzhihua	101°40′29″	26°34′50″	1998–2010
	Tongzilin	101°50′15″	26°41′28″	1998–2010
	Xiaohuangua	101°52′00″	25°50′00″	2000–2010
5	Sanduizi	101°50′30″	26°29′0″	2004–2010
	Gaoqiao	102°10′00″	26°38′0″	2004–2010
	Fengguo	102°22′00″	26°9′0″	2004–2010
6	Yunlong	102°23′30″	25°5′16″	2004–2010
	Chemuhe	102°22′00″	25°37′0″	2004–2010
	Caijiacun	102°28′54″	25°18′53″	2004–2010
7	Dacun	103°02′32″	27°04′38″	2004–2010
	Dashuijing	103°34′00″	27°04′02″	2004–2010
	Jinle	103°28′11″	26°29′03″	2004–2010
8	Xiaohe	103°12′00″	27°13′00″	2004–2010
	Huanggeshu	104°34′00″	28°0′0″	2004–2010
	Moshiyi	103°44′06″	27°49′38″	2004–2010
	Malucun	104°02′00″	27°38′36″	1998–2010
	Doushaguan	104°07′42″	28°02′00″	1995–2010
9	Ningnan	102°43′07″	27°02′33″	1992–2010
	Zhaojue	102°51′09″	28°00′41″	1991–2010
	Huapingzi	103°9′7″	27°26′5″	1980–2010
	Xiluodu	103°41′39″	28°13′31″	2004–2010
	Xinhua	103°56′54″	28°34′18″	2000–2010

4 Application

4.1 Study area and data

The Yangtze River rises in the Tanggula Mountains and the Qinghai–Tibet Plateau in southwestern China. The reach from Yushu in Qinghai Province to Yibin in Sichuan Province is called the Jinsha River, lying on the eastern edge of the Plateau and influenced by a variety of monsoons, e.g., tropical monsoon, subtropical monsoon, and Qinghai–Tibetan Plateau Monsoon. Jinsha River is the westernmost of the major headwater streams of the Yangtze River. It flows through the Qinghai, Sichuan and Yunnan Provinces in western China. The Jinsha River basin is divided into nine sub-basins controlled by the Pingshan gauging station, shown in Fig. 1. The rainfall stations in those sub-basins used in this study are listed in Table 1. The areal average rainfall of each sub-basin was calculated.

The first and most important steps for building an ANN model is the determination of potential input variables (Bowden et al. 2005a; Fernando et al. 2009). Figure 2 demonstrates the areal average rainfall and daily runoff time series in the upper Jinsha basin. It can be seen from

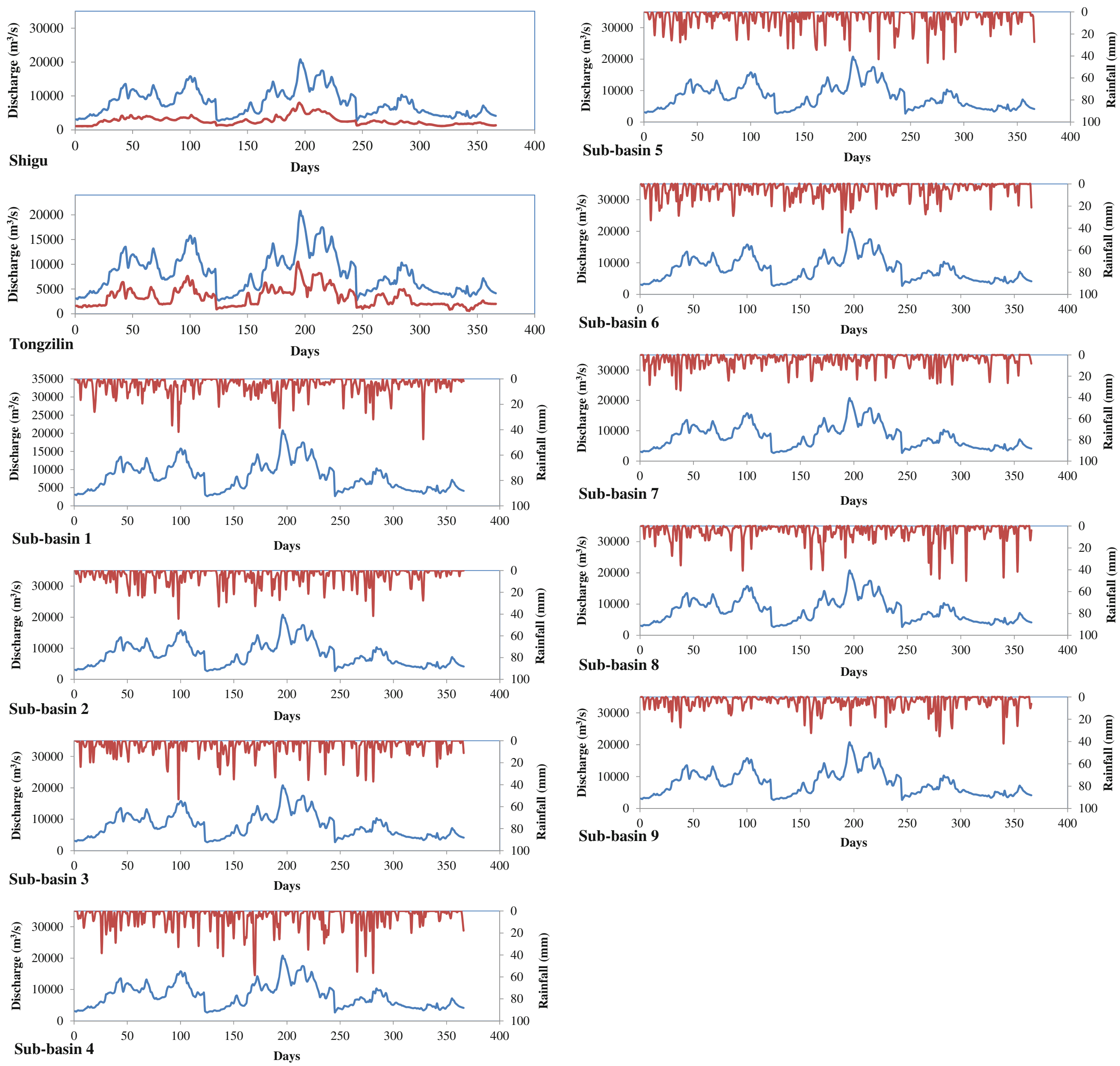

Fig. 2 Daily rainfall–runoff (runoff–runoff) time series of the Jinsha River basin

Fig. 2 that the discharge at Pingshan gauging station has a strong relationship with that at Tongzilin and Shigu gauging stations, and rainfall in some sub-basins has a significant impact on the daily discharge at Pingshan gauging station, such as rainfall in sub-basins 1–4. Therefore, the areal average rainfall of each sub-basin and the previous discharge of Tongziling, Shigu and Pingshan gauging stations with different lags were taken as potential inputs of ANN models. The discharge of Pingshan gauging station was predicted, based on the established rainfall–runoff model.

4.2 Selection of model inputs

Bowden et al. (2005b) proposed a two stage procedure for input selection using PMI. The same method was used in this study. The first step is called bivariate stage, which aims to determine the significant lag of each variable. The second step is called multivariate stage, in which the significant lags selected in the previous step are combined to form a subset of candidates. Then the final set of significance input can be obtained using the same method in step 1.

The details are described as follows. If the number of candidate variables is d (i.e. x_1, x_2, x_i,…,x_d) and the output variable is y_t, their own past values ($x_{i,t-1}$, $x_{i,t-2}$,…,$x_{i,t-k}$) and (y_{t-1}, y_{t-2},…,y_{t-k}) are the potential inputs, where k refers to the maximum lag that has been included as a potential input. k was equal to 15 in this study.

First, the CE values between each of ($x_{i,t-1}$, $x_{i,t-2}$,…, $x_{i,t-k}$) and y_t and (y_{t-1}, y_{t-2},…,y_{t-k}) and y_t were calculated. Then the variable with the maximum negative CE value and its Z value greater than 3 were listed, as shown in Table 2.

During this stage, the original 180 inputs were reduced to 17 inputs. The past runoff (y_{t-1}, y_{t-2},…,y_{t-k}) at Pingshan Station has great impacts on y_t. Therefore, lags $t-1$, $t-2$, $t-3$, $t-4$ and $t-5$ were selected for Pingshan Station. Only one or two inputs were selected for sub-basins 1–9. And the selected lag time of these stations basically matched the flood travel time.

Second, significant lags selected in step one were combined to form a subset of candidates and then the CE was calculated. During this stage, 17 inputs were reduced to 11. The final selected variables for ANN model are given in Table 3.

Bowden et al. (2005a) pointed out that the linear correlation analysis (LCA) method is the most popular analytical technique for selecting appropriate inputs. The Pearson linear correlation coefficients (LCCs) were calculated, as shown in Fig. 3. It can be seen that the dependencies between rainfall of sub-basins 1 and 2, Tongzilin and Shigu and the flow of Pingshan Station are high. The lags of these stations were selected by the CE method as inputs of the model. There are still some differences between these two methods. For both Tongzilin and Shigu gauging stations, the highest correlation coefficient values occurred at lag $t-2$. However, the input selected by the CE method was lag $t-3$ for Shigu gauging station. According to Fig. 1, Shigu gauging station is farther than Tongzilin. From this point of view, the inputs selected by the CE method are more appropriate.

Table 2 Selected input variables in the first stage

Stations	Lags
Pingshan	$t-1$, $t-2$, $t-3$, $t-4$, $t-5$
Tongziling	$t-2$
Shigu	$t-3$
Sub-basin 1	$t-5$
Sub-basin 2	$t-4$, $t-6$
Sub-basin 3	$t-4$
Sub-basin 4	$t-4$
Sub-basin 5	$t-5$
Sub-basin 6	$t-4$
Sub-basin 7	$t-5$
Sub-basin 8	$t-6$
Sub-basin 9	$t-6$

Table 3 Final selected input variables for the ANN model based on the copula entropy method

Stations	Lags
Pingshan	$t-1$, $t-2$, $t-3$, $t-4$, $t-5$
Tongzilin	$t-2$
Shigu	$t-3$
Sub-basin 1	$t-5$
Sub-basin 2	$t-4$
Sub-basin 3	$t-4$
Sub-basin 4	$t-4$
Sub-basin 5	$t-5$

4.3 Identification of models

The inputs obtained by the CE method were used for rainfall–runoff modeling. The identification of a prediction model is to determine the structure by using training data to optimize relevant model parameters, once model inputs are already obtained (Wu and Chau 2011). Three ANN models, namely MLF, RBF and GRNN, were used in this study. The identification of ANN models is to find the model which performs best when the model inputs have been determined. Three performance criteria were used to assess these models. The initial data set consisted of 7 years, from which data for 2004–2008 for model calibration and those for 2009 and 2010 were used for model validation.

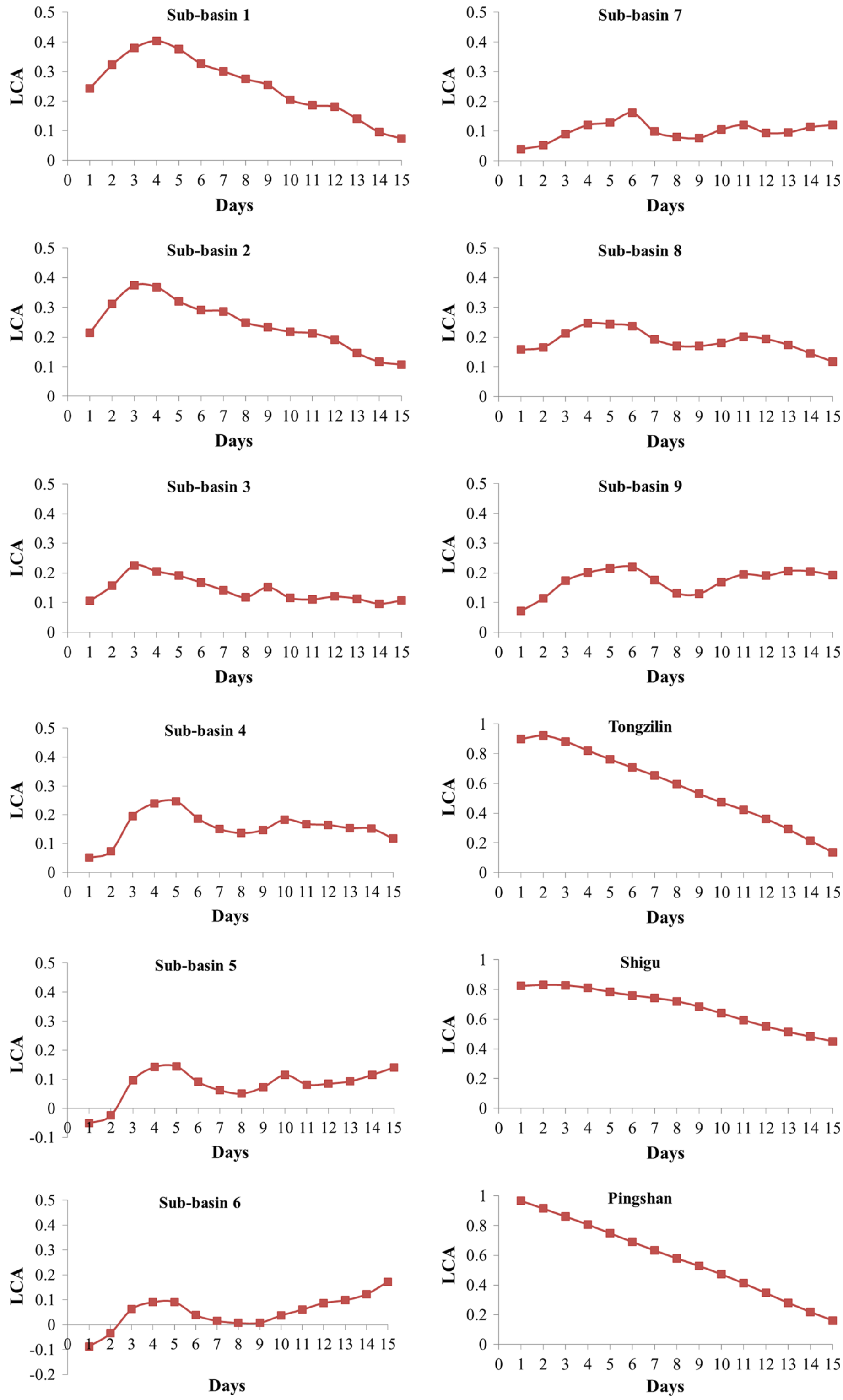

Fig. 3 Linear correlation coefficients between potential inputs and output of ANN model

Table 4 comprises the results obtained using different ANN models. The fitting curves between observed and predicted daily flows at Pingshan Station are given in Fig. 4. It can be seen from Table 4 and Fig. 4 that the MLF ANN model obtained performed better (in terms of small Nash–Sutcliffe efficiency R^2, RMSE and qualified rate) than any other ANN model for predicting the flow at Pingshan Station on Jinsha River. Therefore, the MLF ANN model with

Table 4 Comparison of results obtained with different ANN models

Methods	ANN	Nash–Sutcliffe efficiency R^2		RMSE		Qualified rate	
		Training	Validation	Training	Validation	Training	Validation
Copula entropy	MLF	0.9781	0.9524	548	751	0.9880	0.9786
	GRNN	0.9797	0.9427	528	825	0.9675	0.8846
	RBF	0.9395	0.9008	911	1,086	0.9077	0.8426

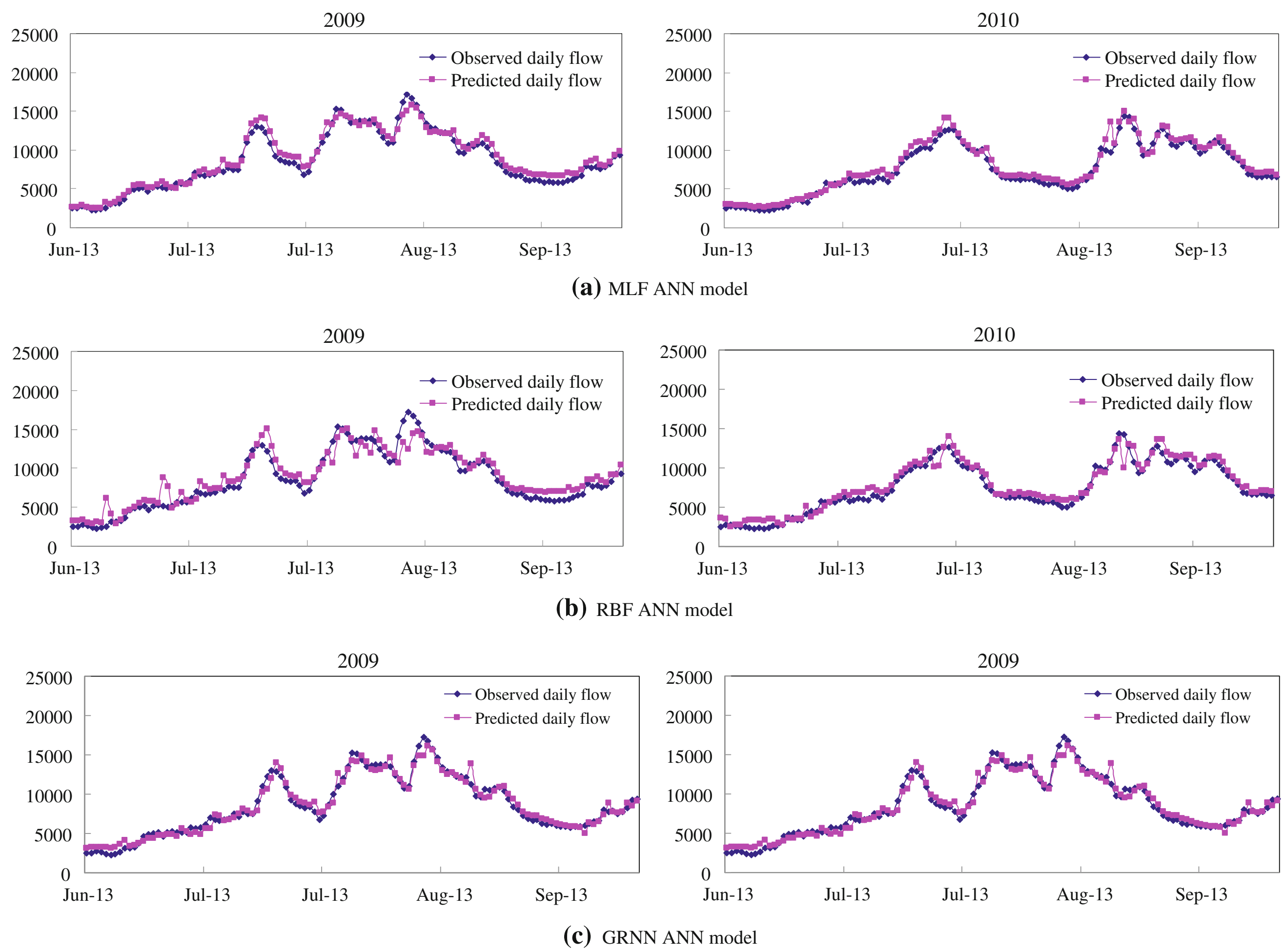

Fig. 4 Comparison between observed and predicted runoff values

Table 5 Comparison of results obtained with different input variables

Methods	ANN	Nash–Sutcliffe efficiency R^2		RMSE		Qualified rate	
		Training	Validation	Training	Validation	Training	Validation
CE	MLF	0.9781	0.9524	548	751	0.9880	0.9786
LCC	BP	0.9674	0.9170	649	906	0.9738	0.9393

the inputs derived by the CE method gave the best results for predicting the flow at Pingshan gauging station. The values of R^2, RMSE and qualified rate for validation period calculated by the MLF ANN model were 0.9524, 751 and 0.9786, which indicate that the proposed model can predict daily flow at Pingshan Station extremely well.

4.4 Comparison of predicted results with different input sets

Two input sets were used for establishing the ANN rainfall–runoff model, one of which was obtained by the LCC method and the other by the CE method. According to the analysis of Sect. 4.3, we know that the MLF ANN model gave the best results for predicting the flow at Pingshan gauging station. Therefore, MLF was used here to compare the results with the two input sets mentioned above. The forecasting performances of those two input sets are shown in Table 5. It can be seen that the RMSE values based on the inputs selected by the CE method are smaller than those based on the inputs selected by LCC method, and the Nash–Sutcliffe efficiency R^2 and qualified rate based on the inputs selected by the CE method are higher than those based on the inputs selected by the LCC method.

5 Conclusions

The present study aims to utilize an ANN to model the rainfall–runoff relationship in a catchment area located in the upper Yangtze River basin, China. The CE method was used to select the appropriate inputs and three different types of ANN models were applied for forecasting stream flow. Finally, the rainfall–runoff model for the Jinsha River with the best predicted results is determined. The main conclusions are summarized as follows.

(a) The CE method, which calculates the PMI and MI directly and avoids a fractional calculation work, can be used to select the inputs of an ANN model. The inputs obtained by the CE method are more appropriate than by the traditional LCA method.
(b) The case study presented above shows that ANNs, as also suggested by other authors, can be used for flow forecasting. Compared with RBF and GRNN models, the MLF ANN model with the inputs selected by the CE method yields the best results.

Acknowledgments The project was financially supported by the National Natural Science Foundation of China (NSFC Grant 51309104, 51239004, and 51190094) and Fundamental Research Funds for the Central Universities (2013QN113), and Open Research Fund Program of State Key Laboratory of Water Resources and Hydropower Engineering Science.

References

Abrahart RJ, See L (2002) Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments. Hydrol Earth Syst Sci 6(4):655–670
Angulo JM, Madrid AE, Ruiz-Medina MD (2011) Entropy-based correlated shrinkage of spatial random processes. Stoch Environ Res Risk Assess 25(3):389–402
Berntson J, Espelid TO, Genz A (1991) An adaptive algorithm for the approximate calculation of multiple integrals. ACM Trans Math Softw 17:437–451
Birikundavyi S, Labib R, Trung HT, Rousselle J (2002) Performance of neural networks in daily streamflow forecasting. J Hydrol Eng 7(5):392–398
Bowden GJ, Dandy GC, Maier HR (2005a) Input determination for neural network models in water resources applications. Part 1—background and methodology. J Hydrol 301(1–4):93–107
Bowden GJ, Maier HR, Dandy GC (2005b) Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. J Hydrol 301(1–4):93–107
Calsaverini RS, Vicente R (2009) An information-theoretic approach to statistical dependence: copula information. Eur Phys Lett 88(6):3–12
Campolo M, Andreussi P, Soldati A (1999) River flood forecasting with a neural network model. Water Resour Res 35(4):1191–1197
Castellano-Méndeza M, González-Manteigaa W, Febrero-Bande M, Prada-Sáncheza MJ, Lozano-Calderón R (2004) Modeling of the monthly and daily behavior of the runoff of the Xallas River using Box–Jenkins and neural networks methods. J Hydrol 296:38–58
Chen L, Guo S, Yan B, Liu P, Fang B (2010) A new seasonal design flood method based on bivariate joint distribution of flood magnitude and date of occurrence. Hydrol Sci J 55(8):1264–1280
Chen L, Singh VP, Guo S, Hao Z, Li T (2012) Flood coincidence risk analysis using multivariate copula functions. J Hydrol Eng 17(6):742–755
Chen L, Singh VP, Guo S, Mishra AK, Guo J (2013) Drought analysis based on copulas. J Hydrol Eng 18(7):797–808
de Vos NJ, Rientjes THM (2005) Constraints of artificial neural networks for rainfall–runoff modelling: trade-offs in hydrological state representation and model evaluation. Hydrol Earth Syst Sci Discuss 2:365–415
Elman JL (1988) Finding structure in time. In: CRL Technical Report 8801. Centre for Research in Language, University of California, San Diego
Favre AC, Adlouni S, Perreault L, Thiémonge N, Bobée B (2004) Multivariate hydrological frequency analysis using copulas. Water Resour Res 40:W01101, 12
Fernando TMKG, Maier HR, Dandy GC (2009) Selection of input variables for data driven models: an average shifted histogram partial mutual information estimator approach. J Hydrol 367:165–176
Grimaldi S, Serinaldi F (2006) Design hyetographs analysis with 3-copula function. Hydrol Sci J 51(2):223–238
Hecht-Nielsen R (1987) Counterpropagation networks. Appl Opt 26:4979–4984
Hopfield JJ (1987) Learning algorithms and probability distributions in feed-forward and feed-back networks. Proc Natl Acad Sci USA 84:8429–8433
Hsu KL, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall–runoff process. Water Resour Res 31(10):2517–2530
Islam MN, Liong SY, Phoon KK, Liaw C-Y (2001) Forecasting of river flow data with a general regression neural network. In: Integrated water resources management. Proceedings of a symposium held at Davis, California. IAHS Publication number 272
Jain SK, Das A, Srivastava DK (1999) Application of ANN for reservoir inflow prediction and operation. J Water Resour Plan Manag 125(5):263–271
Jayawardena DA, Fernando AK, Zhou MC (1997) Comparison of multilayer perceptron and radial basis function networks as tools for flood forecasting. In: Destructive water: water-caused natural

disasters, their abatement and control. Proceedings of the conference held at Anaheim, California, June 1996. IAHS Publication number 239

Jeong DI, St-Hilaire A, Ouarda TBMJ, Gachon P (2012) Comparison of transfer functions in statistical downscaling models for daily temperature and precipitation over Canada. Stoch Environ Res Risk Assess 26(5):633–653

Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. J Comput Civil Eng 8:201–219

Kasiviswanathan KS, Sudheer KP (2013) Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. Stoch Environ Res Risk Assess 27(1):137–146

Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69

Lachtermacher G, Fuller JD (1994) Backpropagation in hydrological time series forecasting. In: Hipel KW, McLeod AI, Panu US, Singh VP (eds) Stochastic and statistical methods in hydrology and environmental engineering. Kluwer Academic, Dordrecht

Lekkas DF, Onof C, Lee MJ, Baltas EA (2004) Application of Artificial Neural networks for flood forecasting. Glob Nest Int J 6(3):205–211

Li X, Guo SL, Liu P, Chen GY (2010) Dynamic control of flood limited water level for reservoir operation by considering inflow uncertainty. J Hydrol 391:124–132

Ma J, Sun Z (2011) Mutual information is copula entropy. Tsinghua Sci Technol 16(1):51–54

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environ Model Softw 15:101–124

May RJ, Maier HR, Dandy GC, Fernando TMK (2008) Non-linear variable selection for artificial neural networks using partial mutual information. Environ Model Softw 23:1312–1326

Mishra AK, Singh VP (2009) Analysis of drought severity–area–frequency curves using a general circulation model and scenario uncertainty. J Geophys Res 114:D06120. doi:10.1029/2008JD010986

Mishra AK, Ines AVM, Singh VP, Hansen JW (2013) Extraction of information content from stochastic disaggregation and bias corrected downscaled precipitation variables for crop simulation. Stoch Environ Res Risk Assess 27(2):449–457

MWR (Ministry of Water Resources) (2006) Standard for hydrological information and hydrological forecasting (SL250-2000) (in Chinese)

Nazemi A, Elshorbagy A (2012) Application of copula modelling to the performance assessment of reconstructed watersheds. Stoch Environ Res Risk Assess 26(2):189–205

Park J, Sandberg IW (1991) Universal approximations using Radial-Basis-Function networks. Neural Comput 3(2):246–257

Powell MJD (1987) Radial basis functions for multivariable interpolation: a review. In: Mason JC, Cox MG (eds) Algorithms for approximation. Clarendon Press, Oxford, pp 143–167

Raman H, Sunilkumar N (1995) Multivariate modeling of water resources time series using artificial neural networks. Hydrol Sci J 40(2):145–163

Rumelhart DE, Hinton E, Williams J (1986) Learning internal representation by error propagation. Parallel Distrib Process 1:318–362

Shamseldin AY (1997) Application of a neural network technique to rainfall–runoff modeling. J Hydrol 199:272–294

Sharma A (2000) Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 a strategy for system predictor identification. J Hydrol 239:232–239

Song S, Singh VP (2010a) Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. Stoch Environ Res Risk Assess 24(3):425–444

Song S, Singh VP (2010b) Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm. Stoch Environ Res Risk Assess 24(5):783–805

Specht DF (1991) A general regression neural network. IEEE Trans Neural Netw 2(6):568–576

Svozil D, Kvasnieka V, Pospichal J (1997) Introduction to multi-layer feed-forward neural networks. Chemom Intell Lab 39:43–62

Thirumalaiah K, Deo MC (2000) Hydrological forecasting using neural networks. J Hydrol Eng 5(2):180–189

Tongal H, Demirel MC, Booij MJ (2013) Seasonality of low flows and dominant processes in the Rhine River. Stoch Environ Res Risk Assess 27(2):489–503

Wu CL, Chau KW (2011) Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. J Hydrol 18:394–409

Yin HF, Li CA (2001) Human impact on floods and flood disasters on the Yangtze River. Geomorphology 41:105–109

Zhang JY, Hall MJ (2004) Regional flood frequency analysis for the Gan-Ming River basin in China. J Hydrol 296(4):98–117

Zhao N, Linb WT (2011) A copula entropy approach to correlation measurement at the country level. Appl Math Comput 218(2):628–642

Zupan J, Gasteiger J (1993) In: Zupan J, Gasteiger J (eds) Neural networks for chemists: an introduction. VCH, Weinheim