

# Model fusion approach for monthly reservoir inflow forecasting

Yun Bai, Jingjing Xie, Xiaoxue Wang and Chuan Li

## ABSTRACT

Considering the complexity of reservoir systems, a model fusion approach is proposed in this paper. According to different inflow information represented, the historical monthly data can be constructed as two time series, namely, yearly-scale series and monthly-scale series. Even grey model (EGM) and adaptive neuro-fuzzy inference system (ANFIS) are adopted for the forecasts at the two scales, respectively. Grey relational analysis (GRA) is subsequently used as a scale-normalized model fusion tool to integrate the two scales' results. The proposed method is evaluated using the data of the Three Gorges reservoir ranging from January 2000 to December 2012. The forecast performances of the individual-scale models are improved substantially by the suggested method. For comparison, two peer models, back-propagation neural network and autoregressive integrated moving average model, are also involved. The results show that, having combined together the small-sample forecast ability of the EGM in the yearly-scale, the nonlinearity of the ANFIS in the monthly-scale, and the grey fusion capability of the GRA, the present approach is more accurate for holistic evaluation than those models in terms of mean absolute percentage error, normalized root-mean-square error, and correlation coefficient criteria, and also for peak inflow forecasting in accordance with peak percent threshold statistics.

**Key words** | adaptive neuro-fuzzy inference system, even grey forecasting model, grey relational analysis inflow forecasts, model fusion, Three Gorges reservoir

**Yun Bai** (corresponding author)  
College of Architecture,  
Anhui Science and Technology University,  
Anhui 233100,  
China  
E-mail: [yunbai@foxmail.com](mailto:yunbai@foxmail.com)

**Jingjing Xie**  
College of Resource and Environment,  
Anhui Science and Technology University,  
Anhui 233100,  
China

**Xiaoxue Wang**  
Nanan District Environmental Monitoring Station of  
Chongqing,  
Chongqing 400060,  
China

**Chuan Li**  
Research Center of System Health Maintenance,  
Chongqing Technology and Business University,  
Chongqing 400067,  
China

## INTRODUCTION

Reservoirs are significant components of water resource systems, providing effective irrigation, water supply, hydropower, flood and drought control, and other functions. Hence the effective operation and scheduling of reservoirs are important tasks in water resource management. An accurate and reliable inflow forecast is a vital reference for making decisions in reservoir management. During the past few decades, a great deal of research has been devoted to the formulations and model developments in reservoir inflow forecasting.

Autoregressive (AR) model, autoregressive moving average and autoregressive integrated moving average (ARIMA) (Box & Jenkins 1976; Pankratz 1983; Brockwell & Davis 1991; Commandeur & Koopman 2007) are all general time

series models in hydrological forecasting (Valipour 2012, 2015; Valipour *et al.* 2012b, 2013; Wang *et al.* 2015). All these approaches are very useful for forecasting changes in the assumption process that the data have an internal structure over time (such as autocorrelation, trend or seasonal variation). However, the real time series of the reservoir inflow often features dynamic, nonlinear and non-stationary characteristics, which limit the applicable scope of these general models.

With the developments in computer technologies and new theories, researchers have proposed various advanced methods, e.g., artificial neural network (Asaad 2010; Othman & Naseri 2011; Kale *et al.* 2012; Sattari *et al.* 2012; Valipour *et al.* 2012a; Prakash *et al.* 2014), support vector

machine (SVM) (Lin *et al.* 2009a; Wang *et al.* 2010), Bayesian regression (Ticlavilca & Mckee 2011), decision support system (Regonda *et al.* 2014), fuzzy inference systems (FIS) (Luna *et al.* 2011; Lohani *et al.* 2011, 2014; Jaiswal *et al.* 2015), model tree (Jothiprakash & Kote 2011; Maestre *et al.* 2013) and grey model (GM) (Lin *et al.* 2012). In addition to the direct time series forecast using the inflow data themselves, external factors such as weather conditions were also taken into account to improve the inflow forecasting accuracy (Zhang *et al.* 2009; Lin *et al.* 2009b, 2010; Wang *et al.* 2012; Tsai *et al.* 2014).

The application of a sole forecast model is often ineffective in forecasting different reservoir inflow series. Hence, some hybrid methods consisting of two or more models or criteria (Li *et al.* 2015) have been reported recently. Chen (2011) proposed a nonlinear model which incorporated improved real-coded grammatical evolution with a genetic algorithm to predict the 10-day inflow of the De-Chi Reservoir in central Taiwan. Budu (2013) studied the capability of two pre-processing techniques, i.e., wavelet transform and moving average, in combination with feed forward neural networks, radial basis and multiple linear regression models for the prediction of the daily inflow values of Malaprabha reservoir, Belgaum, India. Okkan & Ali Serbes (2013) applied wavelet transform technique to remove nonlinear dynamic noise of raw data, and established three hybrid wavelet-based box models, i.e., multiple linear regression, feed forward neural networks and least square SVM, for monthly inflow forecasts. Toro *et al.* (2013) presented a hybrid hydrologic estimation model incorporating both statistical and artificial intelligence techniques for calculating the mean daily inflow of the Salvajina reservoir, Colombia. Li *et al.* (2014) used a couple method, SVM and ensemble Kalman filter, to forecast flooding in the Luo River Basin, China. Awan & Bae (2014) investigated the ability of adaptive neuro-fuzzy inference system (ANFIS) method to improve the accuracy of reservoir inflow forecasting. Bai *et al.* (2014) proposed an additive model for forecasting monthly reservoir inflow, combining AR for trend term, SVM for periodic term and ANFIS for stochastic term, respectively. Wang *et al.* (2015) studied the ARIMA model coupled with the ensemble empirical mode decomposition for forecasting annual runoff time series, improving the accuracy based on a deep insight into the data

characteristics. These hybrid models have shown better forecasting ability than that of the mono-models. These hybrid models are capable of sufficiently reflecting the characteristics of the real time series. Some techniques, such as nonlinear and linear model integration, multiple model combination, data pre-processing, were used for taking advantage of various methods.

Combining data from multiple sources to improve the potential values and interpretation performances of the source data, data fusion is as a general and popular multi-discipline approach to produce a high-quality representation of the data (Zhang 2010). Considering the complexity of hydrologic systems, data fusion approaches have been used in hydrological engineering (Azmi *et al.* 2010; Ababaei *et al.* 2013). Inspired by the idea of data fusion (Liggins *et al.* 2008), we propose in this paper a model fusion approach for monthly reservoir inflow forecasting. Due to the different features in the different forecasting time scales, the time scales of the real time series should be determined first. In each scale, an appropriate model selected by the scale's time characteristics is used to train and forecast the scale's information. The results forecasted at all scales are employed to obtain the final forecasting results through scale-normalized fusion. In this paper, according to different inflow information represented, the historical data of the reservoir are divided into two time series, namely, yearly-scale series and monthly-scale series. To accommodate different characteristics of the two scales, the even grey model (EGM) and the ANFIS are adopted for the forecasts at the two scales, respectively. The grey relational analysis (GRA) is subsequently used as a scale-normalized model fusion tool to integrate the yearly-scale model and the monthly-scale model to obtain final forecasting results. In this way, a model fusion approach is constructed to improve the forecasting accuracy of the monthly reservoir inflow. The proposed approach is applied to the Three Gorges reservoir, China. The forecasting results are also compared to those of some published peer approaches.

---

## METHODOLOGIES

In this study, a data-driven method has been chosen to determine the general structure of the regression model for the monthly reservoir inflow forecasting. The  $t$ th forecasted

value of the reservoir inflow can be associated with the historical data, i.e.,

$$y(t) = F(x(t - \tau), x(t - 2\tau), \dots, x(t - d\tau)), \quad (1)$$

where  $\tau$  denotes the delay time,  $d$  is the embedding dimension,  $F(\cdot)$  stands for the correlation pattern between the future and the historical values, and  $x(t)$  and  $y(t)$  represent the real observation and the forecasted values at time  $t$ , respectively. Generally, the reservoir inflow has special characteristics, i.e., periodicity, seasonality and tendency, in the entire hydrologic times. Moreover, for real hydrologic systems, random disturbances and the weather conditions vary over time. Hence, the reservoir inflow has special patterns in different time scales. This means that a mono model cannot exhibit enough tracking ability for all time scale patterns of the reservoir inflow. To better accommodate the complex natures of the reservoir inflow forecast, we draw lessons from the data fusion approach. From Equation (1), we can find that there exist two relational patterns between the records and the forecasted values: (1) monthly-scale forecasting (the recent monthly data are used to forecast the current monthly value with  $\tau_1 \neq 12$ ); and (2) yearly-scale forecasting (the current monthly data in the historical years are applied to forecast the current monthly value with  $\tau_2 = 12$ ). Different forecasting patterns are capable of representing different time scales' characteristics. The EGM and the ANFIS are used to forecast the monthly inflow series, respectively. The GRA is then used as a fusion tool to calculate the correlation degree between the forecasted series and the target series. Through the distribution of the grey relational degree, we can get the final forecasting result of the monthly reservoir inflow. Using the model fusion, therefore, Equation (1) can be recast as:

$$\begin{cases} y(t) = F_0(y_1(t), y_2(t)) \\ y_1(t) = F_1(x(t - \tau_1), x(t - 2\tau_1), \dots, x(t - d_1\tau_1)), \\ y_2(t) = F_2(x(t - \tau_2), x(t - 2\tau_2), \dots, x(t - d_2\tau_2)) \end{cases} \quad (2)$$

where subscripts '1' and '2' denote the monthly-scale model and the yearly-scale model, respectively,  $F_0(\cdot)$  represents the scale-normalized fusion function for the two constituting models. In this section, the involved methodologies are introduced, respectively.

## ANFIS for the monthly-scale modeling

As mentioned above, the monthly-scale pattern describes the nonlinear changes information of the adjacent months. This calls for a nonlinear model to express the nonlinear relationship  $F_1(\cdot)$  between the input and the output variables. The ANFIS, integrating both the neural networks and the fuzzy logic principles, has capabilities in handling nonlinear, complexity and discontinuity problems (Lohani et al. 2012). It also has some other advantages, such as fast convergence speed, good repeatability and high forecasting precision. Moreover, the results in our previous research (Bai et al. 2014) have shown that the ANFIS has good performance in single application on the monthly inflow forecasting. Hence, the ANFIS is chosen as the monthly-scale regression model.

The ANFIS was proposed by Jang (1993). It can achieve a highly nonlinear mapping and is superior to common linear methods when dealing with nonlinear time series (Jang et al. 1997). The ANFIS is a multilayer feed-forward network that uses the neural network learning algorithms and fuzzy logic to map an input space to an output space (Chang & Chang 2006). For simplicity, assume the fuzzy inference system under consideration has two inputs ( $x_1 = x(t - \tau_1)$ ,  $x_2 = x(t - 2\tau_1)$ ) and one output ( $y_1 = y_1(t)$ ). Its inference system corresponds to a set of fuzzy IF-THEN rules (Takagi & Sugeno 1985; Lohani et al. 2006) that have learning capability to approximate nonlinear functions. The two rules are:

$$\begin{cases} \text{If } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } B_1, \text{ then } f_1 = p_1x_1 + q_1x_2 + r_1 \\ \text{If } x_1 \text{ is } A_2 \text{ and } x_2 \text{ is } B_2, \text{ then } f_2 = p_2x_1 + q_2x_2 + r_2 \end{cases} \quad (3)$$

where  $A_m, B_m$  ( $m = 1, 2$ ) are the linguistic labels,  $p_m, q_m$  and  $r_m$  are the linear parameters in the consequent part of the first order Sugeno fuzzy model, and  $f(\cdot)$  is a first order Sugeno fuzzy model. The architecture of the ANFIS shown in Figure 1 consists of five layers, which can be briefly introduced as follows (Lohani et al. 2007).

*Layer 1:*  $O_{1,m}$  represents the membership function (MF) of  $A_m$  and  $B_m$ , and that is:

$$\begin{aligned} O_{1,m} &= \mu_{A_m}(x_1), \text{ form } = 1, 2; \text{ or } O_{1,m} = \mu_{B_{m-2}}(x_2), \text{ for } m \\ &= 3, 4 \end{aligned} \quad (4)$$

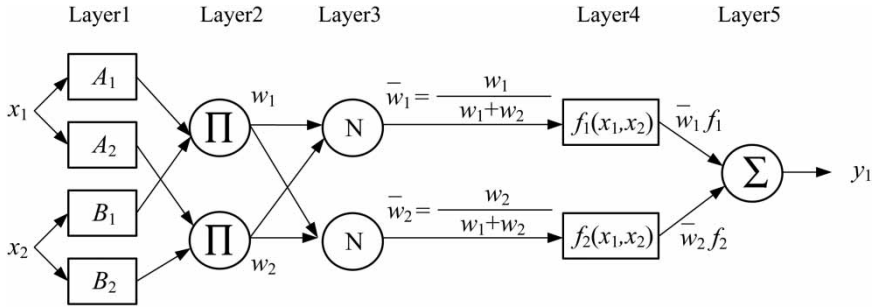


Figure 1 | Architecture of the ANFIS.

The Gaussian membership function is used in this study, so  $\mu_{A_m} = \exp(-(x_1 - \delta_m)^2 / 2\sigma_m^2)$  and  $\mu_{B_m} = \exp(-(x_2 - \delta_m)^2 / 2\sigma_m^2)$ , where  $\{\delta_m, \sigma_m\}$  is the parameter set of the MF in the premise part of fuzzy IF-THEN rules that change the shapes of the MF parameters in this layer and are referred to as the premise parameters.

**Layer 2:** Firing strength  $O_{2,m}$  is the output of this layer and results from multiplying the two MFs obtained in the previous layer using an AND operator:

$$O_{2,m} = w_m = \mu_{A_m}(x_1)\mu_{B_m}(x_2). \quad (5)$$

**Layer 3:** The main objective of the normalized firing strength layer is to calculate the ratio of each  $m$ th rule's firing strength:

$$O_{3,m} = \bar{w}_m = \frac{w_m}{\sum w_m}. \quad (6)$$

**Layer 4:** The square nodes in this layer correspond to the consequent nodes that compute the contribution of each  $m$ th rule aimed at the model output:

$$O_{4,m} = \bar{w}_m f_m = \bar{w}_m(p_m x_1 + q_m x_2 + r_m). \quad (7)$$

**Layer 5:** The single node computes the overall output by summing all the incoming signals. Accordingly, the defuzzification process transforms each rule's fuzzy results into a crisp output in this layer:

$$O_{5,m} = y_1 = \sum_m \bar{w}_m f_m = \frac{\sum_m w_m f_m}{\sum_m w_m}. \quad (8)$$

From the Introduction above, one can find that the ANFIS has two kinds of parameter sets, i.e., the premise parameters  $\{\delta_m, \sigma_m\}$  identifying the shape of the MF, and the consequent parameters  $\{p_m, q_m, r_m\}$  determining the overall output of the system. In this paper, we use the hybrid-learning algorithm (Jang 1993) to identify these parameters. More descriptions of this algorithm can be found by referencing Jang et al. (1997).

As shown in Equation (2), for the monthly-scale forecasting pattern (ANFIS), there are two parameters ( $\tau_1$  and  $d_1$ ) to be specified first. As pointed out in the literature (Sivakumar & Berndtsson 2010), hydrological sequence has generally chaotic characteristics. Thus, chaos phase space reconstruction technique is chosen to determine the ANFIS inputs. In this paper, mutual information (MI) method and false nearest neighbor (FNN) method are applied to determine  $\tau_1$  and  $d_1$ , respectively.

According to the theory of Shannon entropy, the MI can be used to calculate the nonlinear correlation between observation time series  $\mathbf{X}_1 = \{x(t)\}$  and its delay time series  $\mathbf{X}_{1,\tau_1} = \{x(t - \tau_1)\}$ . The MI variable  $I(\tau_1)$  with different  $\tau_1$  is described as:

$$I(\tau_1) = H(\mathbf{X}_1) + H(\mathbf{X}_{\tau_1}) - H(\mathbf{X}_1, \mathbf{X}_{1,\tau_1}), \quad (9)$$

where  $H(\mathbf{X}_1)$  and  $H(\mathbf{X}_{1,\tau_1})$  are the marginal entropies, and  $H(\mathbf{X}_1, \mathbf{X}_{1,\tau_1})$  the joint entropy of  $\mathbf{X}_1$  and  $\mathbf{X}_{1,\tau_1}$ . The three variables are, respectively, given by

$$\begin{aligned} H(\mathbf{X}_1) &= - \sum P(x(t)) \ln P(x(t)), \quad H(\mathbf{X}_{1,\tau_1}) \\ &= - \sum P(x(t - \tau_1)) \ln P(x(t - \tau_1)) \quad \text{and} \quad H(\mathbf{X}_1, \mathbf{X}_{1,\tau_1}) \\ &= - \sum P(x(t), x(t - \tau_1)) \ln P(x(t), x(t - \tau_1)) \end{aligned} \quad (10)$$

where  $P(x(t))$  and  $P(x(t-\tau_1))$  represent the marginal probability distribution functions of  $\mathbf{X}_1$  and  $\mathbf{X}_{1,\tau_1}$ , respectively.

After determining  $\tau_1$ , the FNN (Kennel et al. 1992) is applied to choose  $d_1$ . In a  $d_1$ -dimensional space, each phase vector  $\mathbf{X}_1(t)$  has a nearest neighbor  $\mathbf{X}_{1,n}(t)$ , whose distance  $R_{d_1}$  is expressed as:

$$R_{d_1}(t) = \|\mathbf{X}_1(t) - \mathbf{X}_{1,n}(t)\|. \quad (11)$$

For each  $\mathbf{X}_1(t)$  in the time series, the distance between its nearest neighbor  $\mathbf{X}_{1,n}(t)$  from  $d_1$  to  $(d_1 + 1)$ -dimensional space can be computed as:

$$R = \frac{|R_{d_1}(t) - R_{d_1+1}(t)|}{R_{d_1}(t)}. \quad (12)$$

If  $R > R_{tol} = 15$  (Abarbanel & Gollub 1996), and the loneliness tolerance threshold  $A_{tol}$  satisfies (Kennel et al. 1992):

$$\begin{aligned} \frac{R_{d_1+1}(t)}{R_A} > A_{tol} = 2, \quad R_A = \frac{1}{N} \sum_{t=1}^N (x(t) - \bar{X}), \quad \text{and} \\ \bar{X} = \frac{1}{N} \sum_{t=1}^N x(t), \end{aligned} \quad (13)$$

this point is marked as having a FNN.

### EGM for the yearly-scale modeling

The yearly-scale pattern reflects the similarity of the same month's inflows in the adjacent years. Generally, the weather and the hydrologic conditions are roughly the same in a certain month for different years. Hence, the observations of the history series in the yearly-scale have a quasi-linear and quasi-stationary nature. Moreover, the modeling samples of the history series in the year's scale are usually less (only one sample for one year). Therefore, we adopt the EGM to model the yearly-scale pattern ( $F_2(\cdot)$ ) shown in Equation (2).

The EGM, first proposed by Deng (2005), has been widely used in forecasting fields. The EGM can be utilized to forecast the time series with only four or more observations. It overcomes the disadvantage that the forecasting

accuracy of the statistical model depends on the data amount. In this paper, we employ the EGM (1, 1) algorithm that is described as follows (Liu et al. 2011).

Assume that the yearly-scale series  $\mathbf{X}_2 = (x(t-\tau_2), x(t-2\tau_2), \dots, x(t-d_2\tau_2))$  as the modeling series  $\mathbf{X}^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$  ( $n = 1, 2, \dots, d_2$ ), and  $\mathbf{X}^{(1)}$  is the 1-AGO sequence of  $\mathbf{X}^{(0)}$ ,  $\mathbf{X}^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ , where

$$x^{(1)}(n) = \sum_{i=1}^n x^{(0)}(i).$$

Define  $\mathbf{Z}^{(1)}$  as a sequence obtained by applying the mean operation to  $\mathbf{X}^{(1)}$ ,  $\mathbf{Z}^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$ , where  $z^{(1)}(k) = (x^{(1)}(k) + x^{(1)}(k-1))/2$ ,  $k = 2, 3, \dots, d_2$ , then:

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad (14)$$

is called an even GM (1,1) model, where  $a, b$  mean development coefficient and grey input, respectively. The parameters vector  $\hat{\mathbf{a}} = [a, b]^T$  can be solved by the least square method:

$$\hat{\mathbf{a}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y}, \quad (15)$$

where  $\mathbf{Y} = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix}$ ,  $\mathbf{B} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}$

After  $\mathbf{B}, \mathbf{Y}, \hat{\mathbf{a}}$  obtained, and taking the inverse AGO on  $\hat{\mathbf{X}}^{(1)} = (\hat{x}^{(1)}(n))$ , we can get the time response of EGM (1,1) which gives the forecasting results:

$$y_2(t) = (x^{(0)}(1) - \frac{b}{a})e^{-a(t-1)} + \frac{b}{a}. \quad (16)$$

### GRA for scale-normalized model fusion

Data fusion integrates multiple data and knowledge representing the same real-world object into a consistent, accurate and useful representation. Its operating principle is using various corresponding methods to process data which reflect different characteristics of the time series. In this paper, we employ the GRA (Liu et al. 2011) as a fusion tool to normalize monthly-scale and yearly-scale models for the reservoir inflow forecast. The GRA method makes up the defects of the system analysis caused by the mathematical

statistics, and inconsistencies will not exist between the qualitative analysis and the quantitative results. In addition, it is applicable to handling problems with less samples with the characteristics of generality and less calculation burden. The basic principle of the GRA is based on the geometrical relation comparison between multiple data sequences in the system. The much more similar geometry shapes mean the greater correlation degree. In other words, the smaller the forecasting error is, the higher the grey correlation degree will be. Therefore, model fusion based on the GRA can avoid amplification of different forecasting sequences' errors, thus improving the final forecasting precision.

According to the theory of the data fusion, the reservoir inflow  $\mathbf{X} = (x(1), x(2), \dots, x(N))$  ( $N$  represents the length of time) can be divided into two relevant series, named monthly-scale series  $\mathbf{X}_1 = (x(L)|L \in [1, N])$ , and yearly-scale series  $\mathbf{X}_2 = (x(T + 12d_2)|T \in [1, 12])$ .

Assume that  $\mathbf{X}_0 = (x(1), x(2), \dots, x(M))$  is the target series ( $M \leq N$ ),  $\mathbf{Y}_1 = (y_1(1), y_1(2), \dots, y_1(M))$ ,  $\mathbf{Y}_2 = (y_2(1), y_2(2), \dots, y_2(M))$  are the forecasting results of the relevant series (monthly-scale series  $\mathbf{X}_1$  and yearly-scale series  $\mathbf{X}_2$ ) using corresponding models, respectively, then the absolute relational degree between  $\mathbf{Y}_1$ ,  $\mathbf{Y}_2$  and  $\mathbf{X}_0$  can be expressed as follows:

$$\varepsilon_{0i} = \frac{1 + |s_0| + |s_i|}{1 + |s_0| + |s_i| + |s_i - s_0|}, \quad (17)$$

$$\text{where } |s_0| = \left| \sum_{l=2}^{M-1} x(l) + \frac{1}{2}x(M) \right|, |s_i| = \left| \sum_{l=2}^{M-1} Y_i(l) + \frac{1}{2}Y_i(M) \right|,$$

$$|s_i - s_0| = \left| \sum_{l=2}^{M-1} (Y_i(l) - x(l)) + \frac{1}{2}(Y_i(M) - x(M)) \right|,$$

$Y(l) = x(l) - x(1)$ ,  $Y_i(l) = y_i(l) - y_i(1)$ , and  $y_i(l)$  represents  $l$ th forecasted value of  $i$ th forecasted series,  $i = 1, 2$ .

The integrating process of the data fusion is described as follows:

$$\mathbf{Y} = \sum \frac{\varepsilon_{0i}}{\sum \varepsilon_{0i}} \mathbf{Y}_i, \quad (18)$$

where  $\mathbf{Y}$  means the final forecasting series.

## Overview of the proposed model fusion approach

Having addressed the constituents separately, the present model fusion approach for forecasting reservoir inflow can be concluded as follows and is illustrated by Figure 2.

*Step 1:* Analysis of original data series  $\mathbf{X}$ , and construct new series containing different characteristics (in this paper, we construct two relevant series, monthly-scale series  $\mathbf{X}_1$  and yearly-scale series  $\mathbf{X}_2$ ).

*Step 2:* Forecast the different series' reservoir inflow using corresponding methods (in this paper,  $\mathbf{Y}_1$  was obtained by ANFIS, and  $\mathbf{Y}_2$  was achieved using EGM (1,1)).

*Step 3:* Scale-normalized model fusion through GRA between  $\mathbf{Y}_1$ ,  $\mathbf{Y}_2$  and  $\mathbf{X}_0$ , respectively.

*Step 4:* Output the forecasting results according to Equation (18). End.

## DATA AND PERFORMANCE CRITERIA

The study area, the Three Gorges reservoir shown in Figure 3, is located in the upstream of the Yangtze River at the boundary of Chongqing municipality and Hubei province ( $106^\circ 50' - 110^\circ 50' \text{E} / 29^\circ 16' - 31^\circ 25' \text{N}$ ), China. The largest hydropower station in the world is fed by the Three Gorges reservoir. Besides generating electricity, the dam of

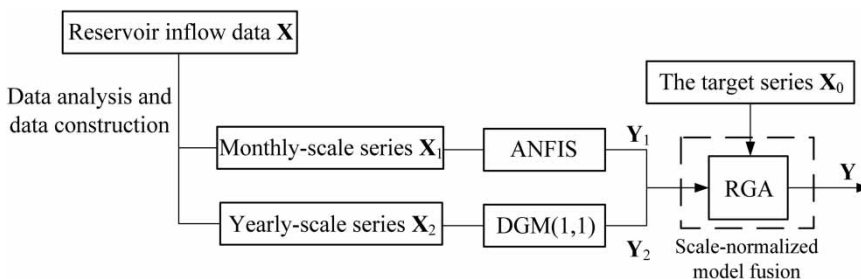


Figure 2 | Architecture of scale-normalized model fusion approach for forecasting reservoir inflow.

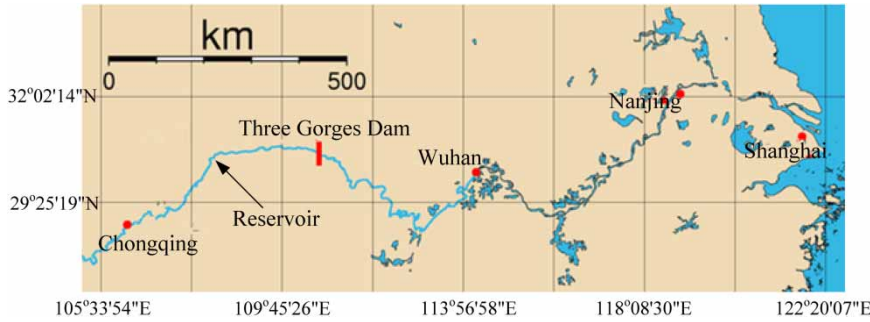


Figure 3 | Location of the Three Georges reservoir and main cities along the Yangtze River.

the reservoir was designed to increase the Yangtze River’s shipping capacity and to reduce the potential for floods downstream by providing flood storage space. The reservoir has a total area of 59,900 km<sup>2</sup>, which is the largest water conservancy for irrigation works in China. The valley of the Three Gorges reservoir below 500 m has an annual temperature of 17–19 °C, with an annual frost-free period of 300–340 days. The annual runoff flow at the site of the dam of the Three Gorges Project is 451 billion cubic metres with an annual sediment discharge of 530 million tons.

In this study, 156 monthly inflow records of the Three Gorges reservoir ranging from year 2000 to year 2012 were collected and are plotted in Figure 4. It is noted that the monthly value (m<sup>3</sup>/month) is generated by summing up every day’s data (m<sup>3</sup>/d) which is averaged by the instantaneous record (m<sup>3</sup>/s) at 8 a.m.

From Figure 4, it can be seen that the reservoir inflow has obvious periodicity in monthly-scale with the wet

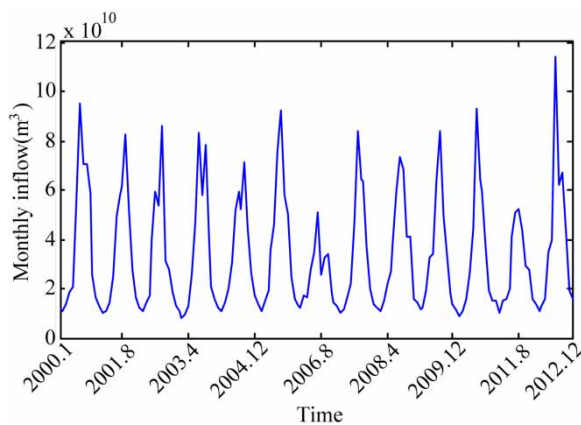


Figure 4 | The Three Gorges reservoir inflow from 2000 to 2012.

season from June to September, the dry season from December to March, and normal season in the rest of the months. In addition, the reservoir inflow also features the seasonality in yearly-scale, that the same monthly inflow in each year is maintained at a certain amount because of the same hydrologic condition and weather changes. One can also observe singular inflow patterns in year 2006 and year 2011 (when drought events occurred in the Yangtze River watershed).

To assess the performance of the forecasting approaches, three traditional criteria, mean absolute percentage error (MAPE), normalized root-mean-square error (NRMSE) and correlation coefficient (R), are used in this paper. These criteria can be formulated as follows:

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{x(t) - y(t)}{x(t)} \right|, \tag{19}$$

$$NRMSE = \frac{\sqrt{\sum_{t=1}^N (x(t) - y(t))^2 / N}}{\sum_{t=1}^N x(t) / N}, \tag{20}$$

$$R = \frac{\sum_{t=1}^N (x(t) - \bar{x})(y(t) - \bar{y})}{\sqrt{\sum_{t=1}^N (x(t) - \bar{x})^2 \sum_{t=1}^N (y(t) - \bar{y})^2}}, \tag{21}$$

where  $x(t)$  and  $y(t)$  represent the recorded and forecasted value of the  $t$ th monthly reservoir inflow, respectively,  $\bar{x}$  and  $\bar{y}$  are the mean value, respectively.

In addition, a new performance criterion, evaluating the forecasting ability of the model precisely from higher to low flow region, peak percent threshold statistics (PPTS)

(Lohani et al. 2014), is adopted in this paper. The equation is:

$$PPTS(lo, up) = \frac{1}{k_{lo} - k_{up} + 1} \sum_{t=k_{lo}}^{k_{up}} |\xi_t|, \quad (22)$$

where  $k_{lo} = \frac{lo \times N}{100}$ ,  $k_{up} = \frac{up \times N}{100}$ ,  $l$  and  $u$  represent the lower and higher limits in percentage, respectively, and  $\xi_t$  is the relative error of the  $t$ th data. Before computing Equation (22), the records are arranged in descending order. Note that the value of  $up = 100\%$ , the PPTS ( $lo$ ,  $up$ ) can be regarded as PPTS ( $lo$ ), which shows the PPTS of top  $lo\%$  data. In this paper, the PPTS (5) and PPTS (10) are all taken into consideration.

## RESULTS AND DISCUSSION

In this section, the proposed model fusion approach is evaluated using the reservoir inflow data of the Three Gorges, as shown in Figure 4. The forecasting performances are also compared with the sole application of the ANFIS, the EGM (1,1) and other peer methods to demonstrate the superiority.

### Inflow forecasting using the proposed approach

In this subsection, the proposed model fusion will be employed to forecast the monthly inflow of the Three Gorges reservoir. First, the two scales' constituting models (the ANFIS and the EGM (1,1)) are established for the monthly- and yearly-scales, respectively. Then, the scale-normalized model fusion based on the GRA is subsequently used to integrate the yearly-scale and the monthly-scale results.

The first constituting model for the proposed approach is the monthly-scale component. As introduced above, the inputs can be determined by the MI and the FNN, respectively. Through recursive algorithm (Fraser & Swinney 1986), the changes of  $I(\tau_1)$  over  $\tau_1$  are shown in Figure 5. As the optimal delay time  $\tau_{1opt}$  is the time of the minimum in  $I(\tau_1)$  found first, one can see that  $\tau_{1opt} = 4$  according to Figure 5.

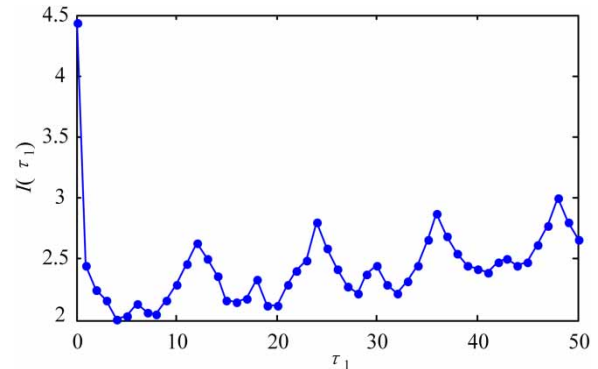


Figure 5 | The change of  $I(\tau_1)$  over  $\tau_1$ .

The minimum embedding dimension, after eliminating all false-nearest neighbors, is just the optimal embedding dimension. Figure 6 shows the FNN results over different embedding dimensions for the data set as shown in Figure 4. According to the two thresholds above, the optimal embedding dimension  $d_{1opt} = 3$  is obtained.

The ANFIS can be trained with the optimal input variables (three inputs and one output) of the model. After training, an ANFIS model with forecasting function will be obtained for output forecasting. In this paper, the monthly-scale series are divided into two sets, the training set and the testing set. The first 132 records, as shown in Figure 4, are employed for training, and the rest for testing. Subtractive clustering is used for training the number of MFs and rules with different value of influence range, squash factor, accept ratio, and reject ratio. A total of 25 models, designed by the orthogonal design, are trained as shown in Table 1.

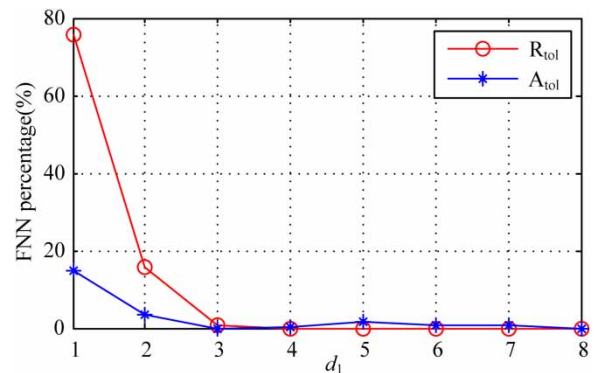


Figure 6 | The percentages of FNN with different embedding dimensions ( $\tau_{1opt} = 4$ ).



**Table 1** | The orthogonal design for the four parameters

Model	Parameters (range of influence, squash factor, accept ratio, reject ratio)	Model	Parameters (range of influence, squash factor, accept ratio, reject ratio)
M1	[0.2, 0.5, 0.2, 0.1]	M14	[0.5, 1.25, 0.2, 0.3]
M2	[0.2, 0.75, 0.4, 0.15]	M15	[0.5, 1.5, 0.4, 0.4]
M3	[0.2, 1, 0.6, 0.3]	M16	[0.8, 0.5, 0.8, 0.15]
M4	[0.2, 1.25, 0.8, 0.4]	M17	[0.8, 0.75, 1, 0.3]
M5	[0.2, 1.5, 1, 0.5]	M18	[0.8, 1, 0.2, 0.4]
M6	[0.3, 0.5, 0.4, 0.3]	M19	[0.8, 1.25, 0.4, 0.5]
M7	[0.3, 0.75, 0.6, 0.4]	M20	[0.8, 1.5, 0.6, 0.1]
M8	[0.3, 1, 0.8, 0.5]	M21	[1, 0.5, 1, 0.4]
M9	[0.3, 1.25, 1, 0.1]	M22	[1, 0.75, 0.2, 0.5]
M10	[0.3, 1.5, 0.2, 0.15]	M23	[1, 1, 0.4, 0.1]
M11	[0.5, 0.5, 0.6, 0.5]	M24	[1, 1.25, 0.6, 0.15]
M12	[0.5, 0.75, 0.8, 0.1]	M25	[1, 1.5, 0.8, 0.3]
M13	[0.5, 1, 1, 0.15]		

Note that the initial values of the four parameters are 0.5, 1.25, 0.5, 0.15, respectively. In addition, ANFIS training needs at least two rules. Therefore, we chose the range of four parameters [0.2, 0.3, 0.5, 0.8, 1], [0.5, 0.75, 1, 1.25, 1.5], [0.2, 0.4, 0.6, 0.8, 1], [0.1, 0.15, 0.3, 0.4, 0.5], respectively.

The different FIS are trained by hybrid method and all performance indices are computed. To determine the effective networks, the testing data are used for validation. Therefore, only nine models shown in Table 2 are suitable

according to the criteria. According to the performances of both training and testing data as shown in Table 2, one can determine the optimal ANFIS networks (M22) with three inputs and one output in this paper. Each input has seven Gaussian membership functions with seven rules.

The second constituting model is the EGM (1,1). For a given sequence, the feasibility test for the EGM modeling is a key process that supplies information about whether an accurate EGM model can be built. Usually, if ratio  $\sigma(k) \in (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ , then the data of the reservoir inflow can be used to establish the EGM model, where 
$$\sigma(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}.$$

In this study, the yearly-scale series are divided into two sets: the records from year 2000 to year 2010, as shown in Figure 4, are employed for training and the rest for testing (that is to say  $d_2 = 11$ ). In practice, we transform the data of the training set into the logarithm form before modeling. The ratio  $\sigma(k) \in [0.964, 1.076] \in (e^{-\frac{2}{11+1}}, e^{\frac{2}{11+1}}) = (0.8458, 1.1824)$ , so the training set conforms to the requirements of the EGM (1,1) modeling. Through modeling 12 groups' series, respectively, we can obtain the yearly-scale's results in the periods from January 2011 to December 2012 (model parameters shown in Table 3).

The two scales' models are then fused by the GRA. With modeling the reservoir inflow pattern using the ANFIS and the EGM (1,1) at the monthly and the yearly-scales,

**Table 2** | Performances of the ANFIS model selected

Model	No. of MFs	Performance criteria					
		Training data			Testing data		
		MAPE	NRMSE	R	MAPE	NRMSE	R
M5	8	15.8013	0.2254	0.9438	20.3637	0.4150	0.8306
M11	16	9.6811	0.1408	0.9785	30.0328	0.5963	0.6541
M15	3	18.3139	0.2498	0.9304	19.2551	0.4771	0.7601
M18	6	14.7392	0.2131	0.9499	17.0150	0.4093	0.8273
M19	3	13.4378	0.1884	0.9614	18.6444	0.5002	0.7414
M20	3	13.4378	0.1884	0.9614	18.6444	0.5002	0.7414
<b>M22</b>	<b>7</b>	<b>9.2670</b>	<b>0.1409</b>	<b>0.9787</b>	<b>12.1322</b>	<b>0.2990</b>	<b>0.9115</b>
M23	4	11.1684	0.1596	0.9727	17.6465	0.4271	0.8095
M25	2	12.3738	0.1791	0.9653	14.4375	0.4519	0.7877

**Table 3** | Parameters trained using the EGM (1,1) in different months

Month	Parameters
Jan.	$a = -0.0011, b = 5.1299$
Feb.	$a = -0.0003, b = 5.0759$
Mar.	$a = -0.0005, b = 5.1626$
Apr.	$a = -0.0005, b = 5.2755$
May	$a = 0.0008, b = 5.5474$
Jun.	$a = 0.0022, b = 5.7944$
Jul.	$a = -0.002, b = 5.8141$
Aug.	$a = 0.0003, b = 5.8658$
Sept.	$a = 0.0005, b = 5.8341$
Oct.	$a = 0.0009, b = 5.686$
Nov.	$a = 0.0003, b = 5.4327$
Dec.	$a = 0.0001, b = 5.2469$

respectively, the GRA of the training data can be calculated according to Equation (12) as  $\epsilon_{01} = 0.9658$  and  $\epsilon_{02} = 0.9309$ . Equation (13) is then used to compute the weigh distribution for integrating the constituting results of both training and testing data. Figure 7(a) displays the comparison between the forecasted and the recorded data using the proposed model fusion method, and the residual error (RE) distribution is shown in Figure 7(b) and 7(c). To study the correlation between the recorded and forecasted data, we generate the scatter plots as shown in Figure 7(d) and 7(e), demonstrating that the higher the agreement between the two data sets, the more the points tend to concentrate in the vicinity of the identity line, marked as 'ideal fit' in this context.

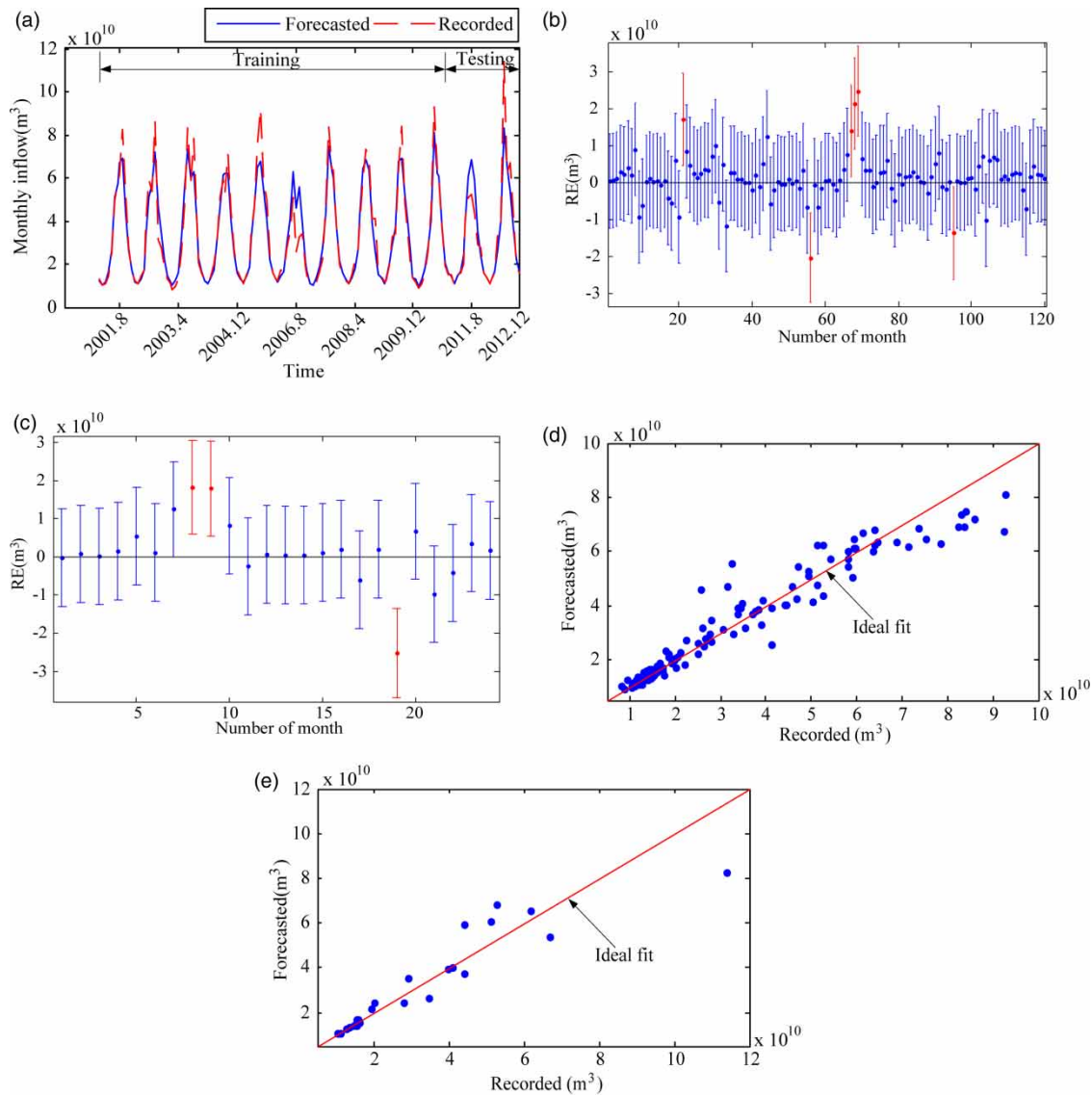
From Figure 7(a), it can be seen that the modeling results can follow the trend of the training data successfully except for a time in the period 2006, and the forecasting results of the testing data are roughly good except for that of the inflow peak. As shown in Figure 7(b) and 7(c), the REs have a little fluctuation both in the training and testing processes, except for several values. There only exist six and three forecasted outliers in the training and the testing processes, respectively. Since the interval around the RE does not contain zero, this indicates that the nine REs are caused by the unfortunate forecasting, beyond the 95% confidence interval. In other words, the number of effective forecasted values

accounts for 95% of the training data, and 87.5% of the testing data, respectively. Hence the thorough processes of the training and the testing are successful, and the results are acceptable. Furthermore, the scatter plot as shown in Figure 7(d) and 7(e), demonstrate that the correlations between the recorded and the forecasted data of the peak inflow are relatively poor, and other data are concentrated in the ideal fit. Besides the intuitive observation, Table 4 shows the quantitative evaluations using the MAPE, NRMSE, R, and PPTS of the training and testing data, respectively.

### Comparisons with the ANFIS and the EGM (1,1)

To show the superiority of the proposed approach, the ANFIS (M22) and the EGM (1,1) are solely applied to forecast the same data for comparison. Figure 8 displays the forecasting results using the ANFIS solely. Compared to Figure 7(a), the modeling results of the training data express higher accuracy, as shown in Figure 8(a). The RE results shown in Figure 8(b) and 8(c) also display better performances (four forecasted outliers) in the training process, and poorer performances (four forecasted outliers) in the testing process. The REs' fluctuation of the ANFIS is the same as the proposed model. There exist four forecasted outliers both in the training and the testing processes, respectively. Compared with the REs of the proposed model, the differences are finite. Figure 8(d) also reveals the correlations between the recorded and the forecasted data of the training data are much more concentrated. However, the correlations of the testing data do not display superiority, shown in Figure 8(e). It is shown in Figure 8(e) that the scatter plots of the inflow peak is much farther from the ideal fit.

Figure 9 plots the forecasting results using the EGM (1,1) solely. It is shown that, compared to Figure 7, the modeling and forecasting results of the training and testing data shown in Figure 9(a) have significant errors in the peak inflow, six forecasted outliers in Figure 9(b) and two forecasted outliers (one less than that in Figure 7(c)) in Figure 9(c) beyond the 95% confidence interval, and the scatter plots of the inflow peak are much farther from the ideal fit, as shown in Figure 9(d) and 9(e).

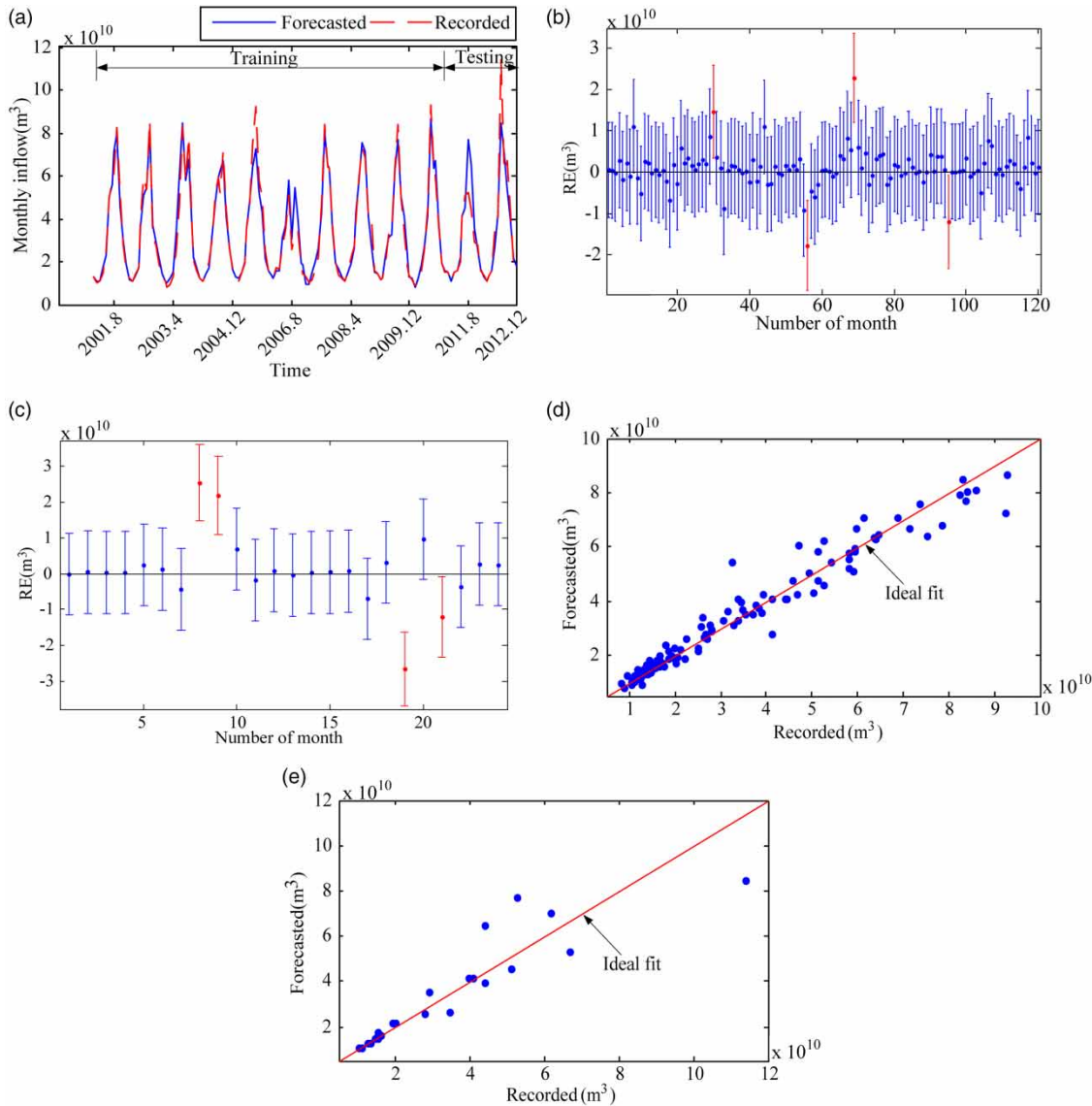


**Figure 7** | Inflow forecasting based on the proposed model fusion approach: (a) plots the forecasting results for the training and the testing data; (b) and (c) display the residual error distribution ( $RE \sim t$ ) of the training and testing data, respectively; (d) and (e) are the scatters for the training and testing data, respectively.

**Table 4** | Performances of the proposed model fusion approach

Grey relation degree	Performance criteria	Training data	Testing data
ANFIS: 0.9658; EGM (1,1): 0.9309	MAPE	10.4973	11.4588
	NRMSE	0.1653	0.2528
	R	0.9641	0.9273
	PPTS(5)	11.0813	12.1885
	PPTS(10)	10.9624	11.4839

For quantitative evaluation, Table 5 shows the forecasting results using the ANFIS and the EGM (1,1), respectively. Compared with Table 4, it more obviously illustrates the fact that the proposed model fusion method has the best forecasting performance in the MAPE, NRMSE, and R. For the PPTS, the lower the value is, the better the capability for forecasting peak inflow will be (Lohani et al. 2014). Compared with Table 4, the PPTS values (PPTS (5) of the ANFIS,



**Figure 8** | Forecasting results and scatter plots based on ANFIS: (a) plots the forecasting results for the training and the testing data; (b) and (c) display the residual error distribution ( $RE \sim t$ ) of the training and testing data, respectively; (d) and (e) are the scatters for the training and testing data, respectively.

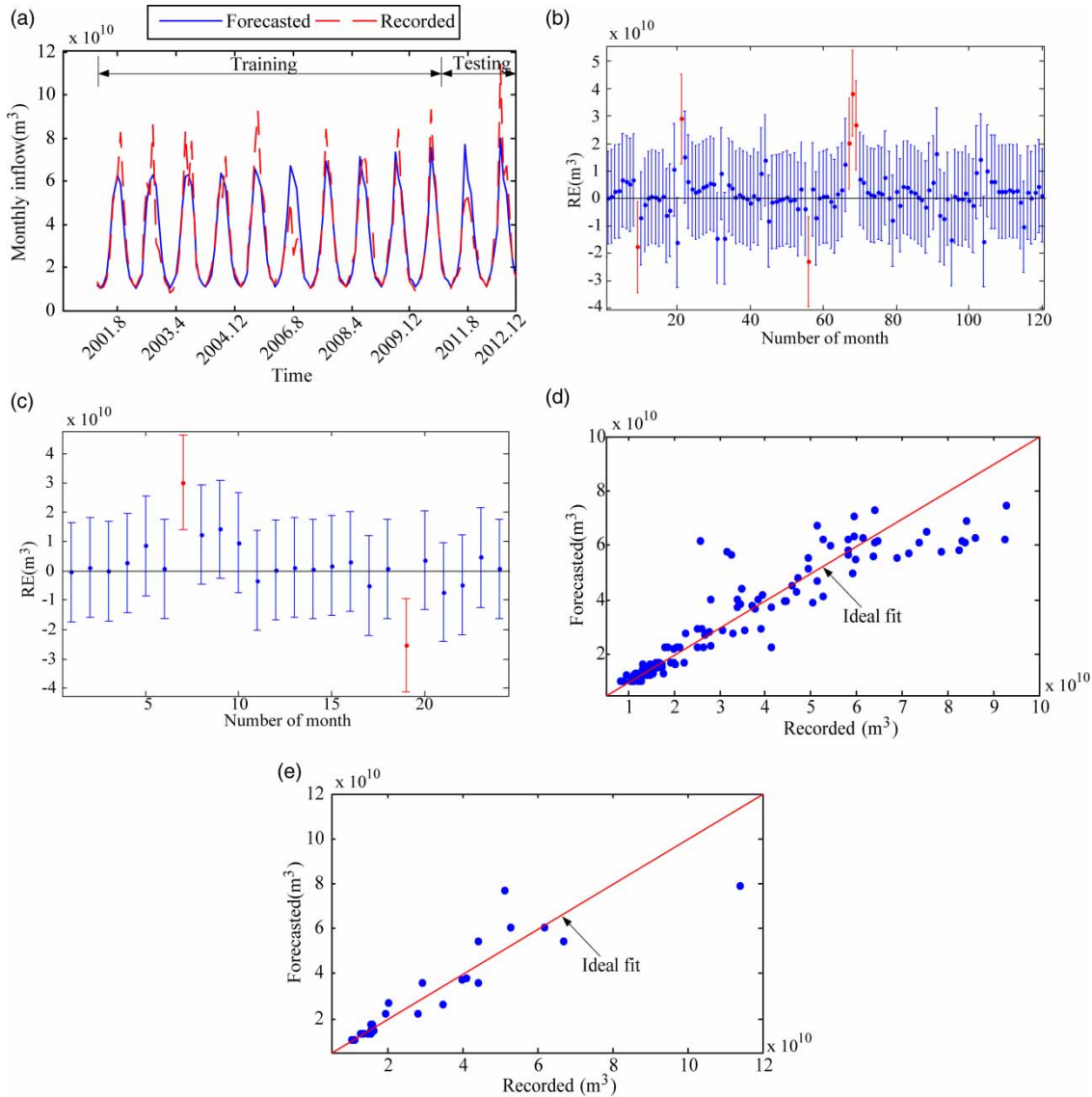
and PPTS (5) and PPTS (10) of the EGM (1,1) confirm that the proposed model is capable of forecasting the peak inflow more accurately.

### Comparisons with other peer methods

In addition to the comparisons with the ANFIS and the EGM (1,1), the proposed approach is also compared with two methods researched and applied more widely, i.e., back propagation neural network (BPNN) and ARIMA,

and the authors' peer method (Bai *et al.* 2014), using the same data set as shown in Figure 4.

Figure 10 plots the forecasting results of the BPNN model for the same input-output structure as the ANFIS. That is, input layer 4, output layer 1, and hidden layer 6. The optimal parameters of the BPNN are obtained by 10-fold cross-validation. As shown in Figure 10(a), the modeling results are generally consistent with the recorded data. However, greater deviation appears in the testing period for the peak inflow. From Figure 10(b) and 10(c), the



**Figure 9** | Forecasting results and scatter plots based on EGM (1,1): (a) plots the forecasting results for the training and the testing data; (b) and (c) display the residual error distribution ( $RE \sim t$ ) of the training and testing data, respectively; (d) and (e) are the scatters for the training and testing data, respectively.

distributions of the REs in the training process display the better tracking capacities, and in this network, the REs' of the testing data have a large fluctuation with five REs of the forecasted values out of the 95% confidence interval. In addition, the scatter plots shown in Figure 10(d) and 10(e) reveal that the modeling results are comparable to those shown in Figure 7(d), and the forecasting results are worse than those in Figure 7(e).

Figure 11 plots the forecasting results of the ARIMA model. Through training, we can construct an optimal

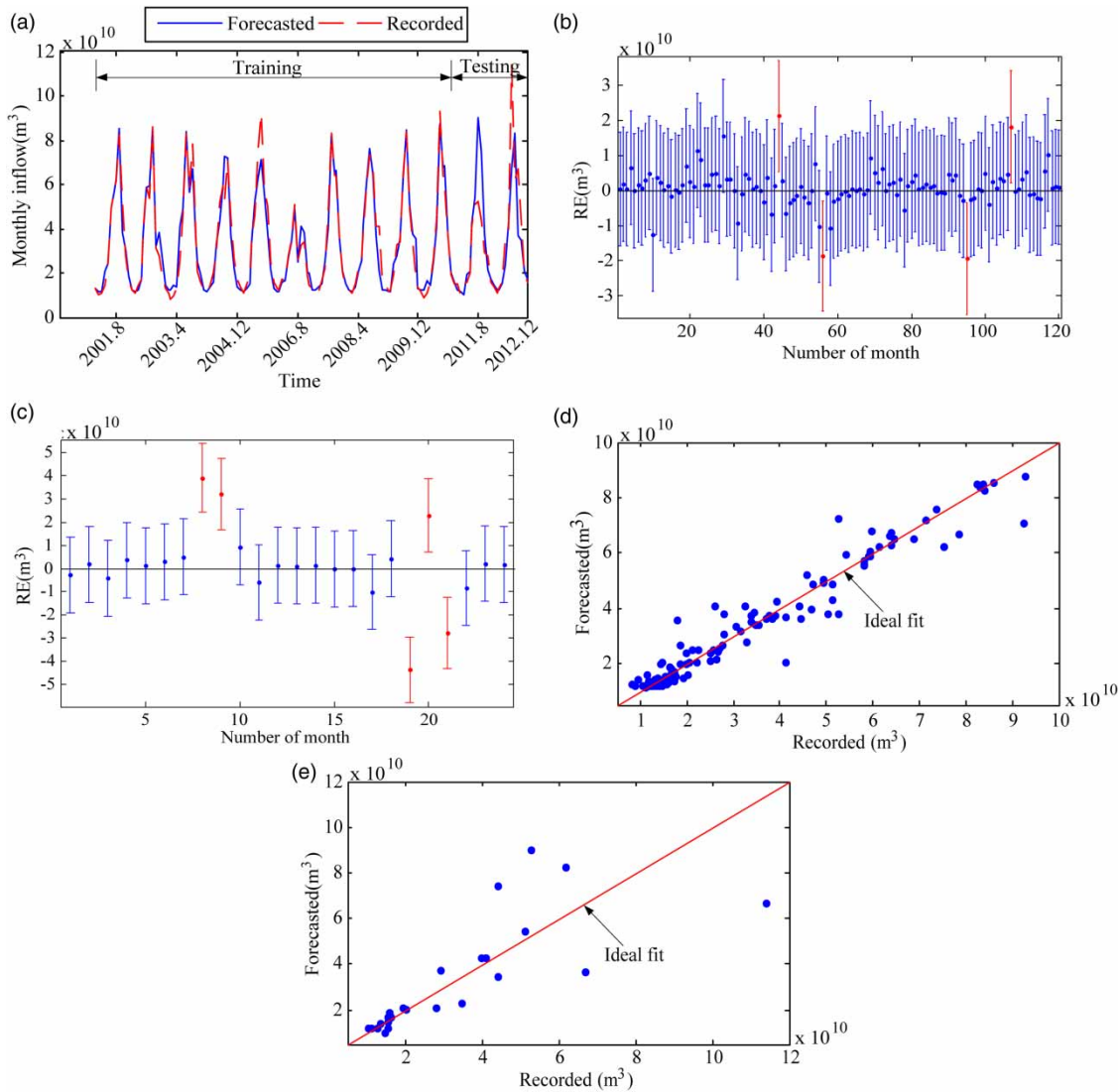
forecasting structure: AR  $P=8$ , moving average  $q=11$ , and difference time  $dt=1$ . As shown in Figure 11(a), the forecasted trends are generally consistent with those of the recorded data. However, there exist significant errors between forecasted and recorded values in all the studied periods. The REs, shown in Figure 11(b), have a bigger fluctuation than in Figure 7(c), but one less forecasted outlier than the presented method. The scatter plots, as shown in Figure 11(c), prove dispersion of the data set.

**Table 5** | Comparison of the forecasting performances using ANFIS and EGM (1,1)

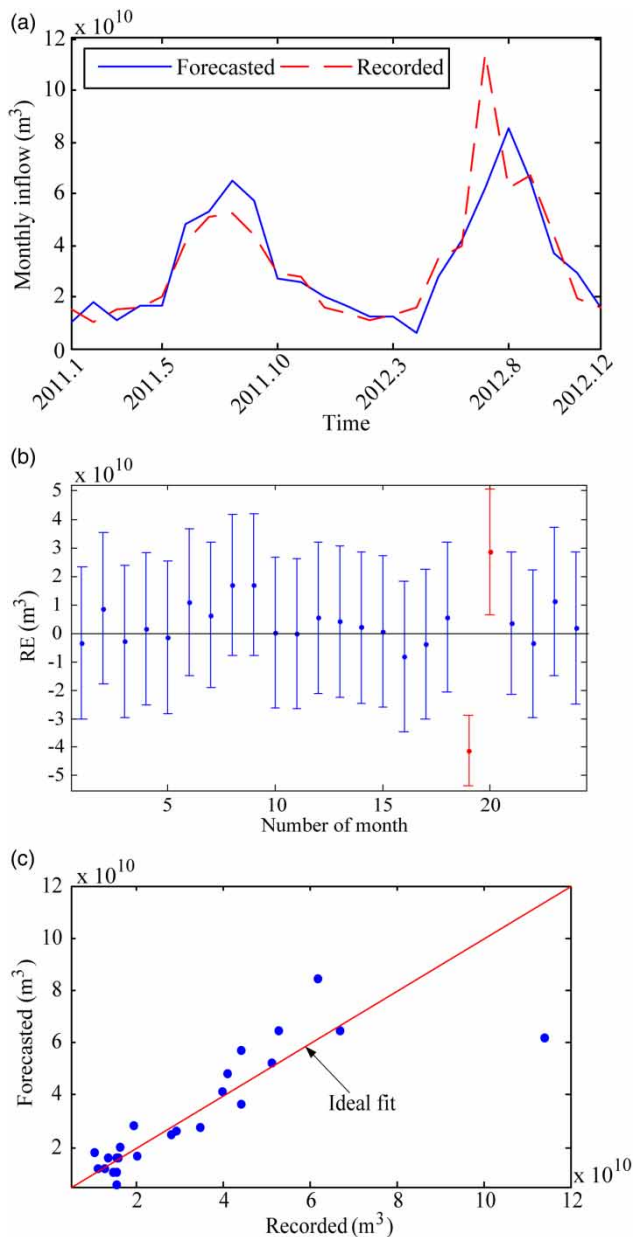
Model	Performance criteria	Training data	Testing data
ANFIS	MAPE	9.2670	12.1322
	NRMSE	0.1409	0.2990
	R	0.9787	0.9115
	PPTS(5)	9.6801	12.8642
	PPTS(10)	9.6136	10.6497
EGM(1,1)	MAPE	15.0627	14.7769
	NRMSE	0.2666	0.3114
	R	0.9221	0.9037
	PPTS(5)	15.0951	14.8831
	PPTS(10)	15.2256	13.2658

Table 6 shows the performance criteria of the BPNN and the ARIMA methods. Compared with Table 4, the BPNN and the ARIMA can provide forecasting accuracy up to 20% in MAPE (11.2858%, 11.3975% reductions for the testing data, respectively), more than 0.4 in NRMSE (0.2424, 0.1411 reductions, respectively), and lower index in R (0.1603, 0.0851 reductions, respectively). Moreover, the PPTS values also demonstrate that the present model is capable of forecasting the peak inflow more accurately.

As above, these criteria shown in Tables 4–6 clearly illustrate that the proposed method in this paper has the best forecasting performance among all the peer methods.



**Figure 10** | Forecasting results and scatter plots using the BPNN (4-6-1): (a) plots the forecasting results for the training and the testing data; (b) and (c) display the residual error distribution ( $RE \sim t$ ) of the training and testing data, respectively; (d) and (e) are the scatters for the training and testing data, respectively.



**Figure 11** | Forecasting results and scatter plots based on ARIMA (8,1,11): (a) plots the forecasting results for the testing data; (b) displays the residual error distribution ( $RE \sim t$ ) of the testing data; and (c) is the scatter for the testing data.

In addition, compared with the additive model proposed by the authors (Bai et al. 2014), the differences are evident: (a) the combination structures (sum-up strategy in the peer paper; GRA in this paper); (b) the features extraction methods (ensemble empirical mode decomposition in the peer paper; scale feature analysis and classification in this paper); (c) the temporal scales (three sub-terms all in

**Table 6** | Forecasting performances using BPNN and ARIMA

Model	Performance criteria	Training data	Testing data
BPNN	MAPE	12.4655	21.7831
	NRMSE	0.1648	0.4952
	R	0.9703	0.7670
	PPTS(5)	16.3414	19.1710
	PPTS(10)	16.0903	17.2658
ARIMA	MAPE		22.8563
	NRMSE		0.3939
	R		0.8422
	PPTS(5)		18.5966
	PPTS(10)		18.3779

monthly-scale in the peer paper; yearly-scale and monthly-scale in this paper); (d) the computational complexities (the peer paper involves four forecasting models and two time-frequency analysis methods, numerous model parameters; this paper contains only two models and one fusion method, especially, the structures of the GRA and EGM (1,1) are simple).

In the case of little gaps in evaluation criteria (compared with the fusion model in this paper, the MAPE, NRMSE, and R are better than 0.0985%, 0.1023%, and 0.0526% in the peer paper, respectively), the computational burden has obviously dropped.

## CONCLUSIONS

In this paper, a model fusion method has been constructed for forecasting monthly reservoir inflow. Considering the characteristics of the reservoir inflow in different time scales (nonlinearity in monthly-scale and quasi-stability in yearly-scale), the ANFIS and the EGM (1,1) are applied to follow the monthly-scale and the yearly-scale inflow patterns, respectively. Using the GRA, the patterns at the two scales are normalized and integrated to generate the final forecasting results. The present method is applied to forecast the Three Georges reservoir inflow. The proposed method is also compared with four peer methods, the ANFIS, the EGM (1,1), the ARIMA, and the BPNN. The results show that, having integrated both monthly-scale and yearly-scale patterns into a grey fashion, the proposed approach exhibits the best forecasting performance in terms of the criteria.

Due to the hydrological similarities of the reservoirs, the proposed model fusion method can also be applied for other reservoirs' inflow forecasts.

## ACKNOWLEDGEMENTS

This work is supported in part by the Natural Science Foundation of China (51375517, 71271226), and the Project of Key Discipline Construction of Anhui Science and Technology University (AKZDXK2015B01). The authors would also like to thank the editors/reviewers for their valuable suggestions and comments that have helped improve this paper.

## REFERENCES

- Ababaei, B., Mirzaei, F., Sohrabi, T. & Araghinejad, S. 2013 Reservoir daily inflow simulation using data fusion method. *Irrig. Drain.* **62** (4), 468–476.
- Abarbanel, H. D. & Gollub, J. P. 1996 Analysis of observed chaotic data. *Physics Today* **49**, 86.
- Asaad, Y. S. 2010 Artificial neural network model for river flow forecasting in a developing country. *J. Hydroinform.* **12** (1), 22–35.
- Awan, J. A. & Bae, D. H. 2014 Improving ANFIS based model for long-term dam inflow prediction by incorporating monthly rainfall forecasts. *Water Resour. Manage.* **28** (5), 1185–1199.
- Azmi, M., Araghinejad, S. & Kholghi, M. 2010 Multi model data fusion for hydrological forecasting using k-nearest neighbour method. *Iran. J. Sci. Technol. Trans. B Eng.* **34**, 81–92.
- Bai, Y., Wang, P., Xie, J. J., Li, J. T. & Li, C. 2014 An additive model for monthly reservoir inflow forecast. *J. Hydrol. Eng.* **20** (7), 04014079.
- Box, G. E. & Jenkins, G. M. 1976 *Time Series Analysis, Control, and Forecasting*. Holden Day, San Francisco, CA, USA.
- Brockwell, P. J. & Davis, R. A. 1991 *Time Series: Theory and Methods*, 2nd edn. Springer, New York, USA.
- Budu, K. 2013 Comparison of wavelet based ANN and regression models for reservoir inflow forecasting. *J. Hydrol. Eng.* **19** (7), 1385–1400.
- Chang, F. J. & Chang, Y. T. 2006 Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Adv. Water Resour.* **29** (1), 1–10.
- Chen, L. 2011 Macro-grammatical evolution for nonlinear time series modeling – a case study of reservoir inflow forecasting. *Eng. Comput.* **27** (4), 393–404.
- Commandeur, J. J. F. & Koopman, S. J. 2007 *An Introduction to State Space Time Series Analysis*. Oxford University Press, Oxford, UK.
- Deng, J. L. 2005 *The Primary Methods of Grey System Theory*. Huazhong University of Science and Technology Press, Wuhan, China.
- Fraser, A. M. & Swinney, H. L. 1986 Independent coordinates for strange attractors from mutual information. *Physical Rev. A* **33** (2), 1134–1140.
- Jaiswal, R. K., Ghosh, N. C., Lohani, A. K. & Thomas, T. 2015 Fuzzy AHP based multi criteria decision support for watershed prioritization. *Water Resour. Manage.* **29** (12), 4205–4227.
- Jang, J. S. R. 1993 ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cyber.* **23** (2), 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. *IEEE Trans. Autom. Control* **42** (10), 1482–1484.
- Jothiprakash, V. & Kote, A. S. 2011 Effect of pruning and smoothing while using M5 model tree technique for reservoir inflow prediction. *J. Hydrol. Eng.* **16** (7), 563–574.
- Kale, M., Nagdeve, M. & Wadatkar, S. 2012 Reservoir inflow forecasting using artificial neural network. *Hydrol. J.* **35**, 52–61.
- Kennel, M. B., Brown, R. & Abarbanel, H. D. 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45** (6), 3403–3411.
- Li, X., Lu, H., Robert, H., An, T. & Yu, Z. 2014 Real-time flood forecast using the coupling support vector machine and data assimilation method. *J. Hydroinform.* **16** (5), 973–985.
- Li, C., Liang, M. & Wang, T. 2015 Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals. *Mech. Syst. Signal Process.* **64–65**, 132–148.
- Liggins, M. E., Hall, D. D. L. & Llinas, J. 2008 *Handbook of Multisensor Data Fusion: Theory and Practice*. CRC Press, Boca Raton, FL, USA.
- Lin, G. F., Chen, G. R., Huang, P. Y. & Chou, Y. C. 2009a Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *J. Hydrol.* **372** (1–4), 17–29.
- Lin, G. F., Wu, M. C., Chen, G. R. & Tsai, F. Y. 2009b An RBF-based model with an information processor for forecasting hourly reservoir inflow during typhoons. *Hydrol. Process.* **23** (25), 3598–3609.
- Lin, G. F., Chen, G. R. & Huang, P. Y. 2010 Effective typhoon characteristics and their effects on hourly reservoir inflow forecasting. *Adv. Water Resour.* **33** (8), 887–898.
- Lin, Y. H., Chiu, C. C., Lee, P. C. & Lin, Y. J. 2012 Applying fuzzy grey modification model on inflow forecasting. *Eng. Appl. Artif. Intell.* **25** (4), 734–745.
- Liu, S. F., Forrest, J. Y. L. & Lin, Y. 2011 *Grey Systems: Theory and Applications*. Springer, Berlin, Heidelberg.
- Lohani, A. K., Goel, N. K. & Bhatia, K. K. S. 2006 Takagi-Sugeno fuzzy inference system for modeling stage-discharge relationship. *J. Hydrol.* **331**, 146–160.
- Lohani, A. K., Goel, N. K. & Bhatia, K. K. S. 2007 Deriving stage-discharge-sediment concentration relationship using fuzzy logic. *Hydrol. Sci. J.* **52** (4), 793–807.



- Lohani, A. K., Goel, N. K. & Bhatia, K. K. S. 2011 Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modelling under different input domains. *Hydrol. Process.* **25** (2), 175–193.
- Lohani, A. K., Kumar, R. & Singh, R. 2012 Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *J. Hydrol.* **442–443**, 23–35.
- Lohani, A. K., Goel, N. K. & Bhatia, K. K. S. 2014 Improving real time flood forecasting using fuzzy inference system. *J. Hydrol.* **509**, 25–41.
- Luna Huamani, I., Ballini, R., Hidalgo, I. G., Franco Barbosa, P. & Francato, A. L. 2011 Daily reservoir inflow forecasting using fuzzy inference systems. In: *IEEE International Conference on Fuzzy Systems*, 27–30 June, Taipei, pp. 2745–2751.
- Maestre, J. M., Raso, L., van Overloop, P. J. & De Schutter, B. 2013 Distributed tree-based model predictive control on a drainage water system. *J. Hydroinform.* **15** (2), 335–347.
- Okkan, U. & Ali Serbes, Z. 2013 The combined use of wavelet transform and black box models in reservoir inflow modeling. *J. Hydrol. Hydromech.* **61** (2), 112–119.
- Othman, F. & Naseri, M. 2011 Reservoir inflow forecasting using artificial neural network. *Int. J. Phys. Sci.* **6**, 434–440.
- Pankratz, A. 1983 *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*. John Wiley & Sons, New York, USA.
- Prakash, O., Sudheer, K. P. & Srinivasan, K. 2014 Improved higher lead time river flow forecasts using sequential neural network with error updating. *J. Hydrol. Hydromech.* **62** (1), 60–74.
- Regonda, S., Zagona, E. & Rajagopalan, B. 2014 Prototype decision support system for operations on the Gunnison basin with improved forecasts. *J. Water Resour. Plann. Manage.* **137** (5), 428–438.
- Sattari, M. T., Yurekli, K. & Pal, M. 2012 Performance evaluation of artificial neural network approaches in forecasting reservoir inflow. *Appl. Math. Modell.* **36** (6), 2649–2657.
- Sivakumar, B. & Berndtsson, R. 2010 *Advances in Data-based Approaches for Hydrologic Modeling and Forecasting*. World Scientific Publishing, Singapore.
- Takagi, T. & Sugeno, M. 1985 Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cyber.* **15** (1), 116–132.
- Ticlavilca, A. M. & McKee, M. 2011 Multivariate Bayesian regression approach to forecast releases from a system of multiple reservoirs. *Water Resour. Manage.* **25** (2), 523–543.
- Toro, C. H. F., Meire, S. G., Gálvez, J. F. & Fdez-Riverola, F. 2013 A hybrid artificial intelligence model for river flow forecasting. *Appl. Soft Comput.* **13** (8), 3449–3458.
- Tsai, M. J., Abrahart, R. J., Mount, N. J. & Chang, F. J. 2014 Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. *Hydrol. Process.* **28** (3), 1055–1070.
- Valipour, M. 2012 Critical areas of Iran for agriculture water management according to the annual rainfall. *Eur. J. Sci. Res.* **84** (4), 600–608.
- Valipour, M. 2015 Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteor. Appl.* **22** (3), 592–598.
- Valipour, M., Banihabib, M. E. & Behbahani, S. M. R. 2012a Monthly inflow forecasting using autoregressive artificial neural network. *J. Appl. Sci.* **12** (20), 2139–2147.
- Valipour, M., Banihabib, M. E. & Behbahani, S. M. R. 2012b Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting. *J. Math. Stat.* **8** (3), 330–338.
- Valipour, M., Banihabib, M. E. & Behbahani, S. M. R. 2013 Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.* **476**, 433–441.
- Wang, W., Nie, X. & Qiu, L. 2010 Support vector machine with particle swarm optimization for reservoir annual inflow forecasting. In: *International Conference on Artificial Intelligence and Computational Intelligence*, 23–24 October, Sanya, China, pp. 184–188.
- Wang, F., Wang, L., Zhou, H., Valeriano, S., Oliver, C., Koike, T. & Li, W. 2012 Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions. *Water Resour. Res.* **48** (7), W07520.
- Wang, W. C., Chau, K. W., Xu, D. M. & Chen, X. Y. 2015 Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour. Manage.* **29** (8), 2655–2675.
- Zhang, J. 2010 Multi-source remote sensing data fusion: status and trends. *Int. J. Image Data Fusion* **1**, 5–24.
- Zhang, J., Cheng, C., Liao, S., Wu, X. & Shen, J. 2009 Daily reservoir inflow forecasting combining QPF into ANNs model. *Hydrol. Earth Syst. Sci. Discuss.* **6** (1), 121–150.

First received 3 July 2015; accepted in revised form 14 November 2015. Available online 4 January 2016