Real-time correction of antecedent precipitation for the Xinanjiang model using the genetic algorithm

Sheng-Li Liao, Gang Li, Qian-Ying Sun and Zhi-Fu Li

ABSTRACT

The Xinanjiang model has been successfully and widely applied in humid and semi-humid regions of China for rainfall–runoff simulation and flood forecasting. However, its forecasting precision is seriously affected by antecedent precipitation (Pa). Commonly applied methods relying on the experience of individual modelers are not standardized and difficult to transfer. In particular, the Xinanjiang daily model may result in obvious errors in the determination of Pa. Thus, a practical method for estimating Pa is proposed in this paper, which is based on a genetic algorithm (GA) and is estimated during a rising flood period. In the optimization process of a GA, Pa values form a chromosome, the root-mean-squared error between the observed and simulated streamflow is chosen as the fitness function. Simultaneously, the best individual reserved strategy is adopted between correction periods to avoid complete independence between each optimization process as well as to ensure the stability of the algorithm. Twenty-seven historical floods observed at the gauge station of the Shuangpai reservoir in Hunan Province of China are used to test the presented algorithm for estimation of Pa, and the results demonstrate that the proposed method significantly improves the quality of flood forecasting in the Xinanjiang model.

Key words | antecedent precipitation, flood forecasting, genetic algorithm, real-time correction, Xinanjiang model

Sheng-Li Liao (corresponding author)
Gang Li
Qian-Ying Sun
Zhi-Fu Li
Institute of Hydropower System and Hydroinformatics,
Dalian University of Technology,
Dalian 116024,
China
E-mail: *shengliliao@dlut.edu.cn*

NOTATION

CRRM conceptual rainfall–runoff model
DM daily model
GA genetic algorithm
Pa antecedent precipitation
RMSE root-mean-squared error

INTRODUCTION

The Xinanjiang model (Zhao *et al.* 1980) is a conceptual rainfall–runoff model (CRRM) which has clear definitions of hydrological processes, and interdependent parameters with explicit physical meanings. Owing to these features, it has been extensively applied in humid and semi-humid regions of China since its initial development in the 1970s (Ju *et al.* 2009; Yao *et al.* 2014). The Xinanjiang model introduces the conception of a free-water sluice reservoir by taking into account the aeration zone's vertical storage to replace Horton's conceptualization of infiltration and divides runoff into surface runoff, soil flow, and groundwater flow. Model parameters and antecedent precipitation (Pa) are the main factors influencing the forecasting accuracy of the Xinanjiang model. The model parameters are closely associated with the catchment characteristics and determine whether the model can reflect the runoff generation mechanism correctly. As the calculation accuracy largely relies on precise definition and cannot be directly obtained from measurable quantities of

doi: 10.2166/hydro.2016.168

catchments' characteristics, model parameters usually need to be calibrated. The earlier use of manual methods and the exploitation of computer-based automatic calibration processes later transformed parameter calibrating into a global optimization problem, involving the process of selecting a set of parameter values until a satisfactory agreement between simulated and observed catchment behavior is obtained, which has promoted parameter calibration (Sorooshian & Dracup 1980; Henrik 2003; Feyen *et al.* 2007). A great many metaheuristic techniques and research algorithms, such as the shuffled complex evolution algorithm (Duan *et al.* 1992, 1994; Cooper *et al.* 2007), simulated annealing (Rozos *et al.* 2004), and particle swarm optimization (Gill *et al.* 2006; Goswami & O'Connor 2007) have been employed for parameter calibration in seeking better optimal solutions. Among the metaheuristic techniques, the genetic algorithm (GA) has been frequently used (Cheng *et al.* 2002, 2006; Sharifi *et al.* 2009; Awad & Von 2010; Wang *et al.* 2012) since its first application in the calibration of hydrological conceptual models (Wang 1991). Pa has a great impact on forecasting results, directly influences the simulated streamflow process and the runoff volume of a flood event, which is one key index for evaluating forecasting accuracy. For a specific CRRM, once the model structure and its model parameters are determined, the accuracy of Pa becomes the main decisive factor of flood forecasting accuracy. In the Xinanjiang model, the introduction of a free-water sluice reservoir of the model makes the components of Pa more comprehensive, which also includes soil moisture of the free-water reservoir, S and flow-producing area, FR. Except for the initial upper, lower, and deeper layer soil moisture, all of these variables cannot be directly obtained from measurable quantities of catchment characteristics.

At present, the Pa of a storm flood can be determined by manual experience or be computed recursively using the Xinanjiang model within a daily routing period. However, traditional setting methods based on experiences have not been widely used in the actual calculation, due to the strong subjective consciousness involved and high demand of hydrological professional abilities required. Furthermore, even though the Xinanjiang daily model (DM) is frequently used in practice, the Pa obtained by this method is subject to a certain degree of uncertainty because of a long computation period and the difficulty in obtaining daily data. Thus, the Xinanjiang DM cannot guarantee high precision forecasting results. In the process of flood forecasting in practice, the Pa values need to be constantly adjusted manually at the beginning of the flood to ensure the simulated process is as close to the actual process as possible, which is a tedious real-time correction process. Consequently, it is crucial to seek appropriate ways to obtain Pa or an automatic correction algorithm to modify Pa to improve the forecasting precision.

To improve the precision of Pa, this paper aims to exploit an intelligent algorithm in the real-time correction process of Pa computed by the Xinanjiang DM. In recent research, there has been a large amount of literature related to the application of contemporary soft computing techniques in water resources engineering (Cheng *et al.* 2005; Chen & Chau 2006; Wu *et al.* 2009; Chau & Wu 2010; Taormina & Chau 2015; Wang *et al.* 2015), promoting great development in hydrology forecasting. However, the GA has been declared to be more robust and efficient in achieving global optima and has advantages over classical optimization methods, becoming one of the most widely used techniques for solving hydrology and water resource problems (Kim *et al.* 2008; Bakhtyar & Barry 2009; Bi *et al.* 2015). The advantages of GA in solving large-scale and complex problems make it suitable to modify the Pa which has multi-variable characteristics in a flood season. Herein, the GA is adopted in the real-time correction process to improve the Pa during a rising flood period. In the algorithm, the fitness function is based on the root-mean-squared error (RMSE) between the observed and simulated streamflow from the starting time of the precipitation to the current time, aiming to improve the forecasting accuracy of flood volume as well as the consistency of the predicted process with the actual process.

XINANJIANG MODEL AND PA

Overview of the Xinanjiang model

The Xinanjiang model is a decentralized hydrological model composed of four sub-models: evapotranspiration, runoff generation, water source runoff division, and runoff routing.

To fully consider the effect of the spatial inhomogeneous of underlying surface characteristics and meteorological factors, the basin is divided into several sub-basins, and the total outflow is considered to be a linear superposition of the outlet flow of each unit basin. The structure of the Xinanjiang model is shown in Figure 1 (Zhao *et al.* 1980). The labels inside the blocks refer to inputs, outputs, state variables, and internal variables; those outside the blocks represent parameters of the model, including runoff generating component parameters and runoff routing parameters. More details of the Xinanjiang model are explained in Zhao (1992).

The evapotranspiration sub-model divides the soil into three layers according to the vertical uneven distribution of the soil. The runoff generation mechanism is based on repletion of storage and is applicable to humid regions with sufficient soil moisture. The drainage basin storage capacity curve is the core of the Xinanjiang model, and fully considers the water storage capacity of the aeration zone and the uneven distribution of the soil moisture. As to the water source runoff division, the Xinanjiang model introduces the structure of the free-water sluice reservoir, which fully considers vertical storage of the aeration zone and separates the total runoff into three components: surface runoff, interflow, and groundwater runoff to replace the infiltration of CRRM with two runoff components. The total runoff obtained from the saturation excess model will be divided into different parts after the capacity adjustment of the free-water sluice reservoir. The surface runoff is separated from the total runoff by the overflow of the flow-producing area, interflow, and groundwater runoff by the interflow and groundwater outflow coefficients, respectively. The unit hydrograph or Muskingum method can be used for drainage basin runoff concentration of the outflow, interflow and groundwater flow are routed through linear reservoir methods.

Pa and potential problems in its determination

Pa, as a reflecting variable of the antecedent soil moisture and runoff capacity, plays an important role in flood forecasting. In the Xinanjiang model, Pa involves five variables (Figure 1): the initial upper layer soil moisture, WU; lower layer soil moisture, WL; deeper layer soil moisture, WD; soil moisture of the free-water reservoir, S; and flow-producing area, FR. The three-layer soil moisture determines the total runoff volume, and the soil moisture of the free-water reservoir determines the volume of each runoff component, affecting the drainage basin runoff concentration of the outflow. For a flood, the five variables codetermine the streamflow process and the total runoff volume. Although the Pa cannot be directly obtained from measurable quantities of catchment characteristics, they can be ascertained by the empirical method or derived from the Xinanjiang DM in practice. A well-known knowledge of basin characteristics and model structure is required in the empirical

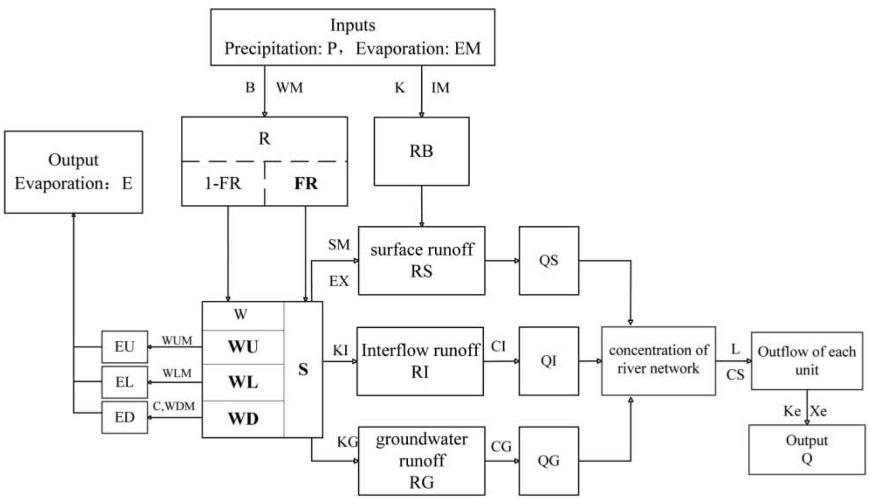

Figure 1 | Structure of the Xinanjiang model. Modified after Zhao (1992).

method, which has proved to be an efficient and accurate tool. The Xinanjiang DM is the most widely accepted method for calibrating Pa because of its simplicity and applicability.

Despite the unique advantages of the empirical method and the wide application of the DM, their potential problems should not be ignored. The empirical method has rigid requirements for researchers and experts. A long-term combination of theory and practice is required to become deeply acquainted with the characteristics of the studied area and a great deal of rainfall runoff data is essential for obtaining an accurate Pa value based on the analysis of the model structure, physical significance, and sensitivity of the parameters. Therefore, it cannot be employed universally. The DM is an easy-to-implement method with a clear structure for determining the Pa value. However, the calculated results are affected by many factors resulting in great uncertainty. The main two factors are summarized as follows:

- **Long time span increases prediction error.** For the first storm flood of each year, Pa is usually calculated from the last storm flood of last year with a long calculation period. Furthermore, a complete set of input data of the study basin is difficult to obtain, resulting in great deviation for the preceding affected precipitation of the first flood in the year. Without amendment, all the other Pa values in each period are unreliable, which deeply affects the accuracy of flood forecasting.
- **Long routing period homogenizes data distribution.** The actual forecast flood process typically employs 1 hour or 3 hours as a routing period. However, the DM method expands the routing period to 1 day, decreasing the accuracy of the Xinanjiang model because of the homogenization of rainfall and evaporation data in the days, which in turn brings great uncertainty of Pa.

Owing to the above limitations, the computation of Pa remains a critical question for the Xinanjiang model. In the process of flood forecasting in practice, Pa needs to be revised gradually according to the observed and simulated streamflow to ensure the precision of the Xinanjiang model.

CORRECTION OF PA USING GA

GA

The GA is a global searching technique that attempts to find the optimal solution in a given decision space based on the mechanics of biological evolution such as natural selection and genetic variation (Goldberg 1989). It is free from auxiliary gradient information and is widely applied because of its robustness and self-adaptability (Ndiritu & Daniell 2001; Hakimi-Asiabar et al. 2010; Hıncal et al. 2011; Stanislawska et al. 2012; Ariadji et al. 2014). The GA uses the idea of fitness to analyze the variety of solutions and generate a new and better solution. The fitness function is based on the optimality of the objective function, and each individual has its own fitness value. Beginning with a randomly generated initial set of solutions and progressing to enhance the fitness of each solution through interaction by conduction operators, including selection, crossover, and mutation, the GA continues to seek better solutions until optimal results are obtained.

The real-time correction process based on GA

Correcting Pa is a multi-variable, nonlinear, and single-objective optimization problem. There are two key points in the correction procedure: one is the introduction of the best individual reserved strategy when initializing the population, the other is the construction of the fitness function. In addition, the selection of the coding strategy, the determination of genetic operation, and the design of control parameters are important factors. The key technologies and skills for modifying Pa by GA are listed as follows.

Generating initial population and the best individual reserved strategy

GA starts with a population in a series of possible individuals generated randomly. The initial individuals are generated in a feasible region, which depends on the constraints of the problem.

Compared with conventional GA, the real-time correction algorithm needs to reinitialize the population at each optimal period. To avoid the loss of excellent individuals and the complete independence between adjacent periods, the optimal individual reserved strategy is employed, reserving the optimal individual in the previous optimization period as an initial individual for the next optimization period (Pulido & Coello 2003). The strategy fully considers the optimum solution of the former correction process, reserving the possible excellent individual and resulting in a more stable and faster discovery for an optimal solution. Figure 2 shows the reserved strategy and the migration process of the best individual.

t_b is the beginning time of the real-time correction process, and t_e is the ending time of the real-time correction process. The optimum result of the correction time t is reserved and migrated into the next correction process, after the last optimal process of time t_e, the optimum result turns out to be the final result, and the optimal process comes to an end.

Fitness function

The fitness function of GA, which is the sole criterion to evaluate individuals, determines the evolutionary directions of the optimal solution. The design of the fitness function is based on practical matters associated with the objective function of the problem. The correction of Pa of the Xinanjiang CRRM is an iterative procedure, in which Pa variable values are adjusted based on the real-time information. The main objective is to look for the best Pa to provide the best agreement between simulated and observed values. To measure the closeness of the simulated and observed results, the minimum RMSE criteria has been most commonly employed and selected as the optimal objective (Boyle et al. 2000; Cooper et al. 2007). In this paper, the Pa values are adjusted based on the minimum RMSE criterion to estimate the performance of the method.

As for the current time t_c, the objective function is described as:

$$\min RMSE_{t_c} = \min\left(\sqrt{\frac{1}{t_c - t_0}\sum_{t=t_1}^{t_c}(Q_t^s - Q_t^o)^2}\right) \quad t_c \in (t_b, t_e) \tag{1}$$

where t_0 is the beginning time of flood forecasting, and it is also the beginning rainfall time; t_c is the current time; Q_t^o is the observed discharge at time t; Q_t^s is the simulated discharge at time t and $RMSE_{t_c}$ is the RMSE at time t_c.

When $t = t_0$, $Q_t^s = Q_t^0$, so in Equation (1), t ranges from t_1 to t_c.

The fitness function of the GA is designed as the following equation:

$$f_{t_c} = \begin{cases} C - RMSE_{t_c} & RMSE_{t_c} < C \\ 0 & RMSE_{t_c} \geq C \end{cases} \tag{2}$$

where C is a positive constant that ensures the reasonableness of the fitness values. Here $C = \max\limits_{t_0 \leq t \leq t_c}(Q_t^o)$.

It is apparent that the fitness function changes over time, and its amplitude contrasts with the error between the simulated and observed discharge process; the smaller the error of the simulated and observed discharge process, the higher the fitness of the individual.

Real number coding strategy

It is necessary to make the coding space consistent with the solution space appropriately. The selection of the encoding strategy affects the design of the GA operation, especially the crossover and mutation operator, which consequently determines the searching speed and convergent effect.

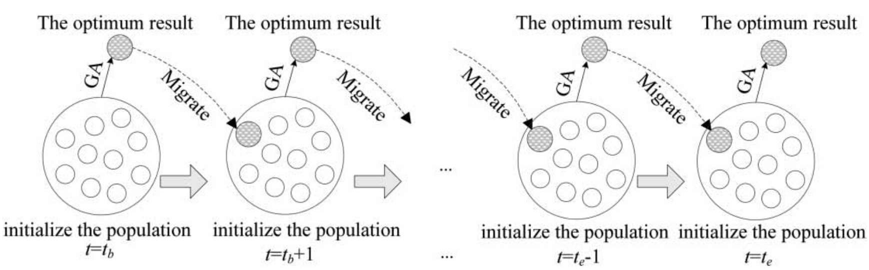

Figure 2 | Diagram of the best individual reserved migration.

Binary coding and real coding are the most common strategies. The former strategy is easy to encode and decode but does not easily solve the problem of high dimensions or the continuum function. Compared with the binary coding, the real coding method is free from encoding and decoding procedures and capable of easily handling complex, high dimension, and multi-constraint problems. Thus, the real coding strategy is adopted in this paper. Five variable values of Pa are used to form a chromosome representing a dependent individual. The individual x_i is described as:

$$x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5})(i = 1, 2, \cdots, n) \tag{3}$$

where n is the population size. $x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}$ are respectively the initial upper layer soil moisture, lower layer soil moisture, deeper layer soil moisture, the soil moisture of the free-water reservoir, and the flow-producing area of individual x_i.

Genetic operation

The GA is an evolutionary process involving selection, crossover, and mutation operations on the genes of individuals based on their fitness to generate better offspring, eventually attaining the optimal or near-optimal solution.

Selection: The offspring receives the superior genes from its parents by the selection operation according to the fitness value. Different selection operators may induce different progressive processes. The roulette wheel selection approach, as a genetic operator used in GAs for selecting potentially useful solutions for recombination, is adopted in the paper. Ranking schemes of the approach operate by sorting the population on the basis of fitness values and then assigning a probability of selection based upon the rank. The selection strategy ensures that the individual with the best fitness is most likely to be duplicated and the worst individual is also of a certain possibility to be selected. The probability of individual x_i being selected is described as:

$$P(x_i) = \frac{f(x_i)}{\sum_{i=1}^{n} f(x_i)} \tag{4}$$

where $f(x_i)$ is the fitness of individual x_i. $P(x_i)$ is the selection probability of individual x_i.

Crossover: The crossover operation is the key to convergence in the GA and is performed on each mated pair of recombination genes with crossover probability. The crossover operation efficiently improves the global search ability of the algorithm. The common crossover operation includes single point, two points, uniform and arithmetic crossover. Compared with other crossover strategies, arithmetic crossover is more simple and effective for real code. In this study, an arithmetic crossover is selected and is described as:

$$x'_1 = \alpha * \mathrm{x}_1 + (1 - \alpha) * \mathrm{x}_2 \tag{5}$$

$$x'_2 = (1 - \alpha) * \mathrm{x}_1 + \alpha * \mathrm{x}_2 \tag{6}$$

where x_1 and x_2 are the parent individuals chosen to make pairs, x'_1 and x'_2 are the offspring, α is a random number between 0 and 1.

Mutation: Mutation is an important tool to maintain diversity and avoid entrapment in local optima. Mutation probability affects the efficiency and accuracy of the algorithm. A non-uniform mutation is employed in this paper. The non-uniform mutation keeps the population from stagnating in the early stages of the evolution, seeking the global solution randomly in the initial space and explores solution in later stages of evolution, seeking the local optimal solution. If gene $x_{ij}(j = 1, 2, 3, 4, 5)$ of individual x_i is selected to perform the mutation operation, then the new individual is $x'_i = (x_{i1}, \cdots, x'_{ij}, \cdots, x_{i5})$ and

$$x'_{ij} = \left\{ \begin{array}{l} x_{ij} + \Delta(k, x_{j_{\max}} - x_{ij}), \, if \; random(0, 1) = 0 \\ x_{ij} - \Delta(k, x_{ij} - x_{j_{\min}}), \, if \; random(0, 1) = 1 \end{array} \right\} \tag{7}$$

where $random(0, 1)$ is the random result in the range [0, 1], the function $\Delta(k, y)$ returns a value in the range [0, y] and it approaches zero of an increasing probability with the increase of k.

$$\Delta(k, y) = y(1 - r)^{(1-k/K)^{\lambda}} \tag{8}$$

where r is a random number in the range [0, 1], k is the index of current evolutional generation, K is the maximum evolutional generation, and λ is a parameter that determines the non-consistent level, which usually ranges from 2 to 5.

Design of GA parameters

The initial GA parameters are first preset and can be performed interactively based on the visual interfaces, the parameters include the population size, n; crossover probability, Pc; mutation probability, Pm; maximum iterations, K; and λ. The population size 'n' is designed in an appropriate scope to provide enough decision space as well as limit the computational burden, which determines the speed of the algorithm and the accuracy of the solution. Pc is the crossover probability, which is generally set between 0.5 and 0.8, Pm is the mutation probability, a large Pm is good for acquiring the optimum solution in an extensive search, but it may lose the best solution, so Pm is often set as a small number, between 0.001 and 0.1, λ is a parameter that determines the degree of dependency with the number of iterations. $n = 300$, $Pc = 0.8$ and $Pm = 0.04$, $K = 500$ and $\lambda = 2$ are employed for each correction during the real-time correction process.

Consequently, the optimal process of each current period can be summarized as follows:

- Step 1. Set the GA control parameters: the population size, n; crossover probability, Pc; mutation probability, Pm; maximum iterations, K; and λ.
- Step 2. Initialize the population randomly. For current time t_c, the optimal solution of the last period is reserved as an initial chromosome for the new optimal process.
- Step 3. Evaluate the fitness value of each individual.
- Step 4. If k is equal to K, go to step 6; else, go to step 5.
- Step 5. Execute the selection, crossover, and mutation operations of the individuals. $k = k + 1$, and then return to step 3.
- Step 6. The individual with the best fitness value is selected as the optimal solution.

The flowchart of the modified procedure for Pa is shown in Figure 3.

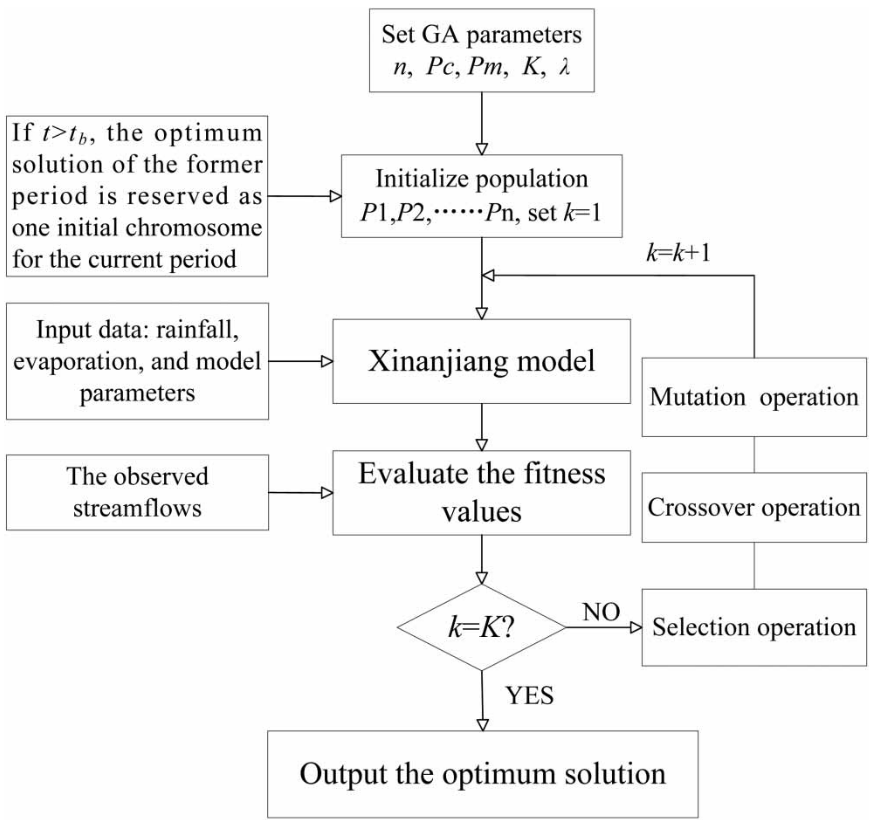

Figure 3 | Flowchart of the real-time correction of Pa at time t_c.

CASE STUDY

Study area and the design of controlling parameters of the GA

The proposed algorithm is applied to Shuangpai Reservoir (Figure 4) in Hunan province of southern China, at the downstream of Xiaoshui Stream, which is one of the tributaries of the Xiangjiang River. The reservoir is used for flood control, power generation, and irrigation, as well as for aquiculture purposes, with a drainage of 10,594 km^2 and a water holding capacity of up to 373.8 million m^3. The involved basin is located in a subtropical monsoon zone with rich rainfalls and dense vegetation cover, and it is suitable for flood forecasting and flood simulation using the Xinanjiang model. The annual rainfall is 1,500 mm; the averaged depth of runoff is 893 mm and the averaged discharge is 300 $m^3\ s^{-1}$. The flood events in this area are mainly caused by thunderstorms, so the temporal distribution of the rainfall during a given year is significantly uneven in this area. 45.9% of the total rainfall occurs in April and June, and 34% of the total rainfall occurs in September and October, and these months are regarded as the high flow periods. For details of Shuangpai Reservoir refer to the literature (Xu *et al.* 2013).

In this paper, the region is divided into 12 sub-areas according to each part's unique characteristics, and the sub-areas share the same set of Pa values. Each sub-area is shown in Figure 4 'Rain gauge'. Twenty-seven historical floods for six years between 2005 and 2010 are employed to test the effectiveness and correctness of the proposed algorithm, and one typical flood of the year 2010 is especially described to present the performance of the proposed algorithm.

According to the national criteria for flood forecasting in China (National Center of Hydrological Information (NCHI) 2000), the result of forecasting is qualificatory relative to peak value, total runoff volume, and peak time for this flood if the absolute percentage error of peak discharge and total runoff volume between the simulated and observed floods is less than 20%, and if the difference in peak time is within a routing period, respectively. The comparisons of three characteristics of the selected floods have been adopted to test the proposed algorithm. A typical flood in

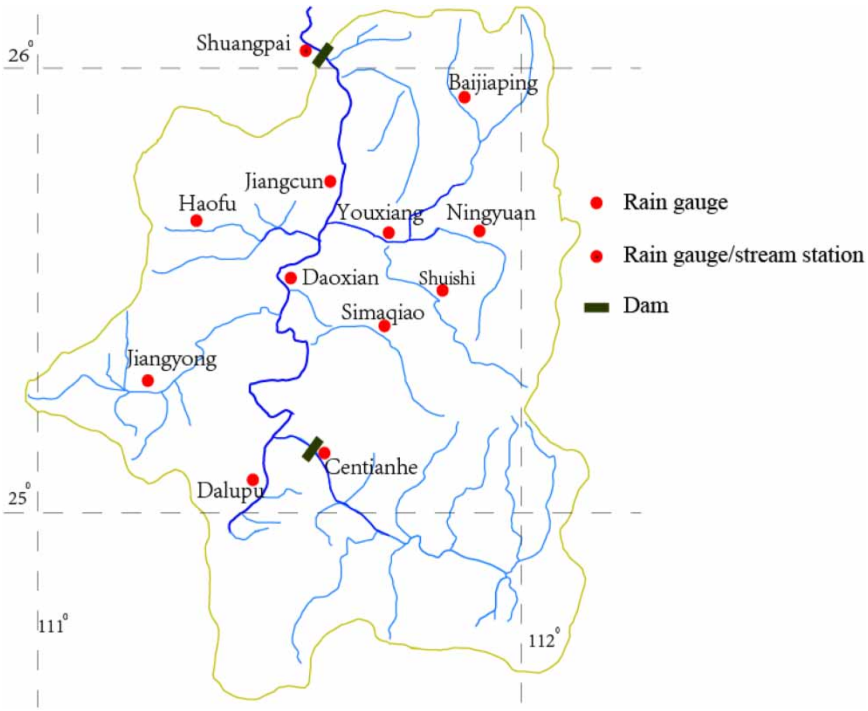

Figure 4 | Map of the Shuangpai area with locations of rain gauge stations.

Table 1 | The ranges of Pa variables

Variable	WU (mm)	WL (mm)	WD (mm)	S (mm)	FR
Upper limit	18.0	63.0	46.0	12.9	1.0
Lower limit	0	0	0	0	0

2010 and a total of 27 historical floods with 3-hour routing periods between 2005 and 2010 are extracted for the study. Table 1 lists the initial ranges of Pa variable values and Table 2 shows the parameter values of the Xinanjiang model which are calibrated from historical flood data (Cheng *et al.* 2002, 2006; Wang *et al.* 2012).

Analysis of a typical flood

The flood 20100621 (date: 21/06/2010), a typical thunderstorm flood of the year 2010, is selected to evaluate the performance of the proposed algorithm. The rainfall began at 11:00 on June 20 and ended at 8:00 on June 21, with a duration time of 21 hours involving seven periods, and after 17:00 of June 21, intermittent light rain continued. The maximum 3-hour precipitation is 30 mm and the 24-hour precipitation is 84.5 mm of the rainfall. The values of the three important characteristics among the observed data are as follow: the peak discharge is 5 m 829.2 $m^3\ s^{-1}$, the total runoff volume is 770 million m^3, and the peak time occurred at 8:00 on June 21.

The forecasting process starts with the rainfall beginning; the observed and simulated rainfall–runoff processes are shown in Figure 5, including the observed hydrographs and one simulated process for which the Pa values are calculated by Xinanjiang DM without correction: simulated (DM) shown in Figure 5, and three simulated processes with real-time corrected Pa by GA: simulated (GA-17:00); simulated (GA-20:00); simulated (GA-23:00) shown in Figure 5. It can be seen from Figure 5 that all of the simulated hydrographs share the same peak time with the observed, occurring at

Table 2 | Parameter values of the Xinanjiang model

Parameter	Um	Lm	Dm	B	Im	K	C	Sm
Value	18	63	46	0.55	0.02	0.74	0.14	12.9
Parameter	**Ex**	**Kg**	**Ki**	**Cg**	**Ci**	**Cs**	**Ke**	**Xe**
Value	1.4	0.27	0.38	0.87	0.21	0.15	1.93	0.21

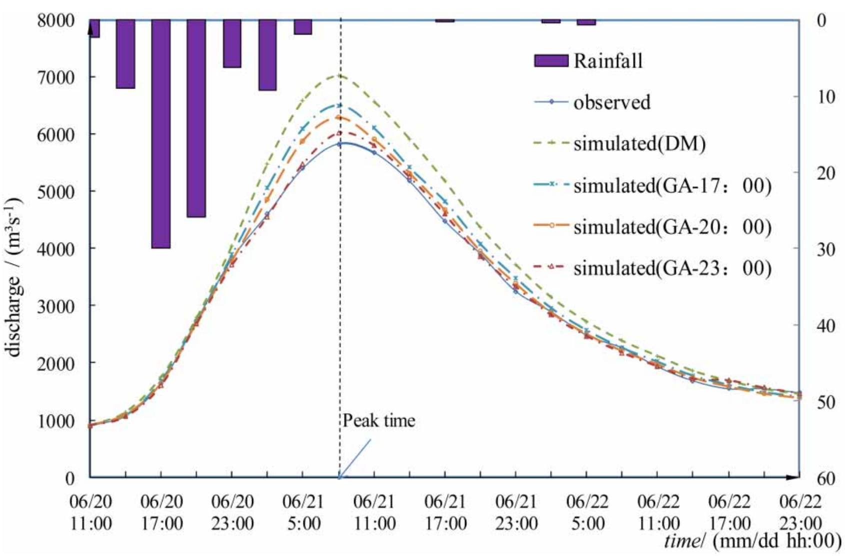

Figure 5 | The rainfall, observed hydrographs, and simulated hydrographs (DM and GA) for flood 20100621.

Table 3 | Pa values based on DM and the real-time correction Pa values based on GA

Method			GA		
Correction time (mm/dd/hh:00)		DM –	06/20/17:00	06/20/20:00	06/20/23:00
Pa	WU (mm)	**17.3**	17.0	16.9	**16.8**
	WL (mm)	**63.0**	58.1	54.6	**51.4**
	WD (mm)	**46.0**	44.0	44.0	**44.0**
	S (mm)	**2.3**	2.4	3.2	**3.6**
	FR	**0.95**	0.92	0.9	**0.85**

8:00 on June 21. However, the simulated rainfall–runoff process exhibits a big gap with the actual process if Pa directly obtained by Xinanjiang DM is not modified, meaning low forecast accuracy. Otherwise, in the forecasting process, if the real-time correction of Pa by GA is considered and performed, a high accuracy of forecasting rainfall–runoff process would be achieved. As well, with the increase of the observed data and the simulated data participating in the real-time correction of Pa by GA, a higher accurate Pa is obtained resulting in a higher forecast accuracy, as shown by the simulated processes GA-17:00, GA-20:00, and GA-23:00 in Figure 5.

Table 3 shows the initial Pa calculated by DM and the real-time correction results modified by GA, respectively. Pa values are corrected by GA at each period according to the RMSE using the observed and simulated discharge. As can be seen in Table 3, compared with the initial values computed by DM, the majority of the Pa values optimized at 23:00 on June 20 are obviously changed with a reduction to be consistent with the forecasting rainfall–runoff process. The value of WU decreases from 17.3 to 16.8 mm, WL with a reduction of 11.6 mm, and WD derives a lower value of 44.0 mm and the total value of WU, WL, and WD reduces from 126.3 to 112.2 mm after the real-time correction. Such reduction demonstrates that Pa modified by GA turns out to conform better to the actual process. Consequently, the simulated rainfall–runoff process with a lower deviation matches better with the actual process, as shown in Figure 5 and Table 4.

Table 4 lists the statistical information of the result for the simulated flood streamflows based on the Xinanjiang DM and the final modified results by GA for comparison. From Table 4, we can see that although the simulated result using Pa by DM is eligible to the total runoff volume and peak time, it is incompetent to the peak discharge for its percentage error of 20.34%, exceeding the limitation of 20%. On the other hand, after the gradual correction on Pa by GA, all the simulated results have smaller errors than the results by DM, and are qualificatory to the criteria. Seen from the optimum result after the correction at 23:00 on June 20, it has the absolute peak discharge percentage error of only 3.20%, the total runoff volume decreases from 13.38 to 1.18%, and the same peak time of that as Pa calculated by DM. Thus, the simulated result based on the modified Pa values optimized by GA are qualificatory related to the peak discharge, total runoff volume, and peak time.

Table 4 | Comparison between DM and GA

Method			GA		
Correction time (mm/dd/hh:00)		DM –	06/20/17:00	06/20/20:00	06/20/23:00
Peak discharge	Observed ($m^3 s^{-1}$)	5,829.2	5,829.2	5,829.2	5,829.2
	Simulated ($m^3 s^{-1}$)	7,014.6	6,502.9	6,288.6	6,015.1
	Error ($m^3 s^{-1}$)	1,185.4	673.7	459.4	185.9
	Error (%)	***20.34***	11.56	7.88	3.19
Total runoff volume	Observed ($10^8 m^3$)	7.70	7.70	7.70	7.70
	Simulated ($10^8 m^3$)	8.73	8.24	8.03	7.79
	Error ($10^8 m^3$)	1.03	0.54	0.33	0.09
	Error (%)	13.38	7.02	4.22	1.18
Peak-time	Observed (mm/dd/hh:00)	06/21/08:00	06/21/08:00	06/21/08:00	06/21/08:00
	Simulated (mm/dd/hh:00)	06/21/08:00	06/21/08:00	06/21/08:00	06/21/08:00
	Error number	0	0	0	0

Analysis of 27 historical floods

Twenty-seven historical floods from 2005 to 2010 are employed to verify the performance of the proposed algorithm. As the algorithm has little influence on the peak time, here we list only the results of peak discharge and total runoff volume for comparison. Table 5 presents the simulated results of peak discharge and total runoff volume based on Pa calculated by the Xinanjiang DM and modified by GA, and the observed results of peak discharge and total runoff volume for all floods. Table 6 shows the statistical results of the DM and GA.

The statistical results from Tables 5 and 6 indicate that only 16 of the 27 floods simulated with Pa calibrated by Xinanjiang DM are qualificatory considering the error of peak discharge, and 21 floods are qualificatory considering the error of the total runoff volume; while after the real-time correction by GA, all of the 27 floods have smaller errors than those without correction of Pa. Moreover, for 26 of the 27 floods, the simulation results have an absolute

Table 5 | Comparisons of peak discharge and total runoff volume based on Pa from DM and GA

	Peak discharge					Total runoff volume				
		DM		GA			DM		GA	
Floods	Observed ($m^3\,s^{-1}$)	Simulated ($m^3\,s^{-1}$)	Error (%)	Simulated ($m^3\,s^{-1}$)	Error (%)	Observed ($10^8\,m^3$)	Simulated ($10^8\,m^3$)	Error (%)	Simulated ($10^8\,m^3$	Error (%)
20050213	2,089.9	1,961.2	−6.16	2,156.9	3.21	3.38	3.63	7.21	3.44	1.70
20050619	3,956.9	4,852.5	***22.63***	4,243.7	7.25	7.75	8.89	14.71	7.74	0.13
20060608	3,353.0	4,256.0	***26.93***	3,353.2	0.01	5.66	6.69	18.20	5.90	4.24
20060613	2,257.6	2,264.5	0.31	2,251.6	0.27	3.77	4.33	14.85	4.35	15.38
20060804	2,499.7	2,184.6	−12.61	2,361.8	5.52	4.32	4.11	4.86	4.39	1.62
20070429	1,947.7	1,261.8.3	− ***35.22***	1,903.2	−2.28	3.11	2.09	***−32.72***	2.99	−4.00
20070603	3,398.4	3,126.9	−7.99	3,221.5	5.21	4.20	4.12	1.90	4.22	0.48
20070607	7,326.6	9,111.6	***24.36***	7,419.8	1.27	8.86	11.16	***25.92***	9.35	5.56
20070613	3,398.4	4,367.0	***28.50***	3,388.5	−2.9	5.02	6.26	***24.72***	5.31	5.88
20080509	1,511.2	1,424.0	−5.77	1,446.7	4.27	2.19	2.08	5.02	2.12	3.20
20080613	7,642.2	10,053.7	***31.56***	7,663.0	2.72	14.58	16.39	12.45	13.12	10.02
20090424	875.3	864.3	−1.26	868.1	0.82	0.90	0.88	2.78	0.89	1.77
20090519	3,138.0	2,928.6	−6.67	2,874.1	8.41	4.64	4.12	11.16	4.28	7.89
20090602	896.5	793.7	−11.47	890.0	0.73	1.32	1.27	−3.30	1.37	3.79
20090611	1,569.9	1,880.2	19.77	1,666.4	6.15	1.97	2.77	***40.33***	2.26	14.72
20090728	1,399.7	1,679.8	***20.01***	1,413.2	0.96	1.72	2.33	***35.46***	1.59	−7.56
20100416	3,139.1	3,300.9	5.15	3,192.2	1.69	6.78	6.86	1.18	6.58	2.95
20100421	3,454.2	3,507.4	1.54	3,233.6	6.39	4.60	5.43	18.04	5.12	11.30
20100426	1,351.4	1,444.1	6.86	1,269.1	6.09	1.14	1.38	***21.05***	1.26	10.53
20100513	2,706.5	3,339.9	***23.40***	2,758.4	1.92	3.96	4.22	6.57	3.47	12.37
20100519	1,602.9	1,552.5	−3.14	1,602.9	0.00	1.56	1.53	1.92	1.58	1.28
20100522	2,691.4	2,366.6	−12.07	2,795.4	3.86	2.66	2.30	13.53	2.66	0.00
20100531	3,270.0	4,693.6	***43.54***	3,109.2	4.92	5.79	6.65	14.85	5.10	11.92
20100609	1,036.8	1,307.6	***26.12***	1,038.0	0.12	1.30	1.48	13.85	1.26	3.08
20100613	3,270.1	3,778.6	15.55	2,919.4	−10.72	4.92	5.54	12.60	4.61	6.30
20100617	2,683.1	3,209.5	19.62	2,680.2	−0.11	3.51	4.12	17.38	3.65	3.99
20100621	5,829.2	7,014.6	***20.34***	6,015.1	3.19	7.70	8.73	13.38	7.79	1.18

Table 6 | Statistical result comparison of DM and GA

Method	Qualificatory peak discharge		Qualificatory total runoff volume	
	Number	Ratio (%)	Number	Ratio (%)
DM	16	69.57	21	91.30
GA	27	100	27	100

percentage error ratio of peak discharge within 10% illustrating excellent performance of the proposed algorithm. All simulation results are qualificatory relative to the error of total runoff volume and only one is greater than 15%, but still under 20%. Furthermore, 20 of the 27 floods have an absolute percentage error ratio less than 10%. The results from Tables 5 and 6 show that the proposed real-time correction algorithm for Pa based on GA performs efficiently, the proposed algorithm are able to improve the forecast accuracy of the Xinanjiang model and ensure the reliability and applicability of the forecast result.

CONCLUSIONS

Antecedent precipitation has a significant impact on the process of runoff production during a rainstorm, especially for peak discharge and total runoff volume. Thus, to improve the performance and accuracy of the Xinanjiang model for rainfall–runoff simulation and flood forecasting, one key step is to appropriately determine the value of Pa. This paper proposes a solution for determining the Pa values in the Xinanjiang model.

The proposed algorithm considers Pa values as a chromosome, employs the RMSE between the observed and simulated streamflow as the fitness function, and adopts the best individual reserved strategy in the optimization process. This method can obtain more reasonable Pa values than conventional methods such as the empirical method and DM. Finally, 27 historical floods with 3-hour routing periods in Shuangpai Reservoir are employed to test the effectiveness and correctness of the algorithm; the results prove that the algorithm can significantly improve the flood forecasting quality of the Xinanjiang model. However, rainfall data acting as the main input of the CRRM, directly determine the model output precision. In order to meet the demand for timely and accurate forecasts in hydrology, further research focused on atmospheric coupling is an important guidance for flood forecasting.

ACKNOWLEDGEMENTS

This study is supported by the National Natural Science Foundation of China (No. 51209029), the Fundamental Research Funds for the Central Universities (20120041120002) and the Fundamental Research Funds for the Central Universities (No. DUT14QY15). The authors would like to thank the three anonymous reviewers for their extremely valuable comments and suggestions that greatly helped us improve our paper and the future research.

REFERENCES

Ariadji, T., Haryadi, F., Rau, I. T., Aziz, P. A. & Dasilfa, R. 2014 A novel tool for designing well placements by combination of modified genetic algorithm and artificial neural network. *J. Pet. Sci. Eng.* **122**, 69–82.

Awad, A. R. & Von Poser, I. 2010 Calibrating conceptual rainfall-runoff models using a real genetic algorithm combined with a local search method. In: *Latest Trends on Computers*, Corfu, Greece, pp. 174–181.

Bakhtyar, R. & Barry, D. A. 2009 Optimization of cascade stilling basins using GA and PSO approaches. *J. Hydroinform.* **11**, 119–132.

Bi, W., Dandy, G. C. & Maier, H. R. 2015 Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge. *Environ. Model. Software* **69**, 370–381.

Boyle, D. P., Gupta, H. V. & Sorooshian, S. 2000 Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resour. Res.* **36**, 3663–3674.

Chau, K. W. & Wu, C. L. 2010 A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. *J. Hydroinform.* **12**, 458–473.

Chen, W. & Chau, K. W. 2006 Intelligent manipulation and calibration of parameters for hydrological models. *Int. J. Environ. Pollut.* **28**, 432–447.

Cheng, C. T., Ou, C. P. & Chau, K. W. 2002 Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. *J. Hydrol.* **268**, 72–86.

Cheng, C. T., Chau, K. W., Sun, Y. G. & Lin, J. Y. 2005 Long-term prediction of discharges in Manwan Reservoir using artificial neural network models. *LNCS* **3498**, 1040–1045.

Cheng, C. T., Zhao, M. Y., Chau, K. W. & Wu, X. Y. 2006 Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *J. Hydrol.* **316**, 129–140.

Cooper, V. A., Nguyen, V.-T.-V. & Nicell, J. A. 2007 Calibration of conceptual rainfall–runoff models using global optimization methods with hydrologic process-based parameter constraints. *J. Hydrol.* **334**, 455–466.

Duan, Q. Y., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28**, 1015–1031.

Duan, Q. Y., Sorooshian, S. & Gupta, V. K. 1994 Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.* **158**, 265–284.

Feyen, L., Vrugt, J. A., Nuallain, B. O., van der Knijff, J. & De Roo, A. 2007 Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *J. Hydrol.* **332**, 276–289.

Gill, M. K., Kaheil, Y. H., Khalil, A., McKee, M. & Bastidas, L. 2006 Multiobjective particle swarm optimization for parameter estimation in hydrology. *Water Resour. Res.* **42**, W07417.

Goldberg, D. E. 1989 *Genetic Algorithm in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, USA.

Goswami, M. & O'Connor, K. M. 2007 Comparative assessment of six automatic optimization techniques for calibration of a conceptual rainfall-runoff model. *Hydrolog. Sci. J.* **52**, 432–449.

Hakimi-Asiabar, M., Ghodsypour, S. H. & Kerachian, R. 2010 Deriving operating policies for multi-objective reservoir systems: application of self-learning genetic algorithm. *Appl. Software Comput.* **10**, 1151–1163.

Henrik, M. 2003 Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. Water Resour.* **26**, 205–216.

Hıncal, O., Altan-Sakarya, A. B. & Ger, A. M. 2011 Optimization of multireservoir systems by genetic algorithm. *Water Resour. Manage.* **25**, 1465–1487.

Ju, Q., Yu, Z. B., Hao, Z. C., Ou, G. X. & Zhao, J. 2009 Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. *Neurocomputing* **72**, 2873–2883.

Kim, T., Heo, J. H., Bae, D. H. & Kim, J. H. 2008 Single-reservoir operating rules for a year using multiobjective genetic algorithm. *J. Hydroinform.* **10**, 163–179.

National Center of Hydrological Information (NCHI) 2000 *The National Criteria for Hydrological Forecasting*. Hydroelectric Press, Beijing, China.

Ndiritu, J. G. & Daniell, T. M. 2001 An improved genetic algorithm for rainfall-runoff model calibration and function optimization. *Math. Comput. Modell.* **33**, 695–706.

Pulido, G. T. & Coello, C. A. C. 2003 The microgenetic algorithm 2: towards on-line adaptation in evolutionary multiobjective optimization. In: *Evolutionary Multicriterion Optimization Second International Conference EMO 2003*, 2632, pp. 252–266.

Rozos, E., Efstratiadis, A., Nalbantis, I. & Koutsoyiannis, D. 2004 Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows. *J. Hydrolog. Sci.* **49**, 819–842.

Sharifi, S., Sterling, M. & Knight, D. W. 2009 A novel application of a multi-objective evolutionary algorithm in open channel flow modelling. *J. Hydroinform.* **11**, 31–50.

Sorooshian, S. & Dracup, J. A. 1980 Stochastic parameter estimation procedures for hydrologic rainfall–runoff models: correlated and heteroscedastic error cases. *Water Resour. Res.* **16**, 430–442.

Stanislawska, K., Krawiec, K. & Kundzewicz, Z. W. 2012 Modeling global temperature changes with genetic programming. *Comput. Math. Appl.* **64**, 3717–3728.

Taormina, R. & Chau, K. W. 2015 Neural network river forecasting with multi-objective fully informed particle swarm optimization. *J. Hydroinform.* **17**, 99–113.

Wang, Q. J. 1991 The genetic algorithm and its application to calibrating conceptual rainfall–runoff models. *Water Resour. Res.* **27**, 2467–2471.

Wang, W. C., Cheng, C. T., Chau, K. W. & Xu, D. M. 2012 Calibration of Xinanjiang model parameters using hybrid genetic algorithm based fuzzy optimal model. *J. Hydroinform.* **14**, 784–799.

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**, 2655–2675.

Wu, C. L., Chau, K. W. & Li, Y. S. 2009 Methods to improve neural network performance in daily flows prediction. *J. Hydrol.* **372**, 80–93.

Xu, D. M., Wang, W. C., Chau, K. W., Cheng, C. T. & Chen, S. Y. 2013 Comparison of three global optimization algorithms for calibration of the Xinanjiang model parameters. *J. Hydroinform.* **15**, 174–193.

Yao, C., Zhang, K., Yu, Z. B., Li, Z. J. & Li, Q. L. 2014 Improving the flood prediction capability of the Xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph. *J. Hydrol.* **517**, 1035–1048.

Zhao, R. J. 1992 The Xinanjiang model applied in China. *J. Hydrol.* **135**, 371–381.

Zhao, R. J., Zhang, Y. L. & Fang, L. R. 1980 The Xinanjiang model. In: *Paper Presented at Hydrological Forecasting Proceeding Oxford Symposium*, IASH 129, pp. 351–356.

First received 11 August 2015; accepted in revised form 24 February 2016. Available online 5 April 2016