

# A comparative study of different objective functions to improve the flood forecasting accuracy

Meng-Xuan Jie, Hua Chen, Chong-Yu Xu, Qiang Zeng and Xin-e Tao

## ABSTRACT

In the calibration of flood forecasting models, different objective functions and their combinations could lead to different simulation results and affect the flood forecast accuracy. In this paper, the Xinanjiang model was chosen as the flood forecasting model and shuffled complex evolution (SCE-UA) algorithm was used to calibrate the model. The performance of different objective functions and their combinations by using the aggregated distance measure in calibrating flood forecasting models was assessed and compared. And the impact of different thresholds of the peak flow in the objective functions was discussed and assessed. Finally, a projection pursuit method was proposed to composite the four evaluation indexes to assess the performance of the flood forecasting model. The results showed that no single objective function could represent all the characteristics of the shape of the hydrograph simultaneously and significant trade-offs existed among different objective functions. The results of different thresholds of peak flow indicated that larger thresholds of peak flow result in good performance of peak flow at the expense of bad simulation in other aspects of hydrograph. The evaluation results of the projection pursuit method verified that it can be a potential choice to synthesize the performance of the multiple evaluation indexes.

**Key words** | flood forecasting, multi-objective functions, projection pursuit, single objective function, threshold of peak flow

Meng-Xuan Jie

Hua Chen (corresponding author)

Chong-Yu Xu

Qiang Zeng

Xin-e Tao

State Key Laboratory of Water Resources and  
Hydropower Engineering Science,

Wuhan University,

Wuhan 430072,

China

E-mail: [chua@whu.edu.cn](mailto:chua@whu.edu.cn)

Chong-Yu Xu

Department of Geosciences,

University of Oslo,

Oslo,

Norway

## INTRODUCTION

Hydrological models have been widely used in flood forecasting, water resource management, impact studies of climate change and land-use change, and regional and global water balance calculations, etc. (Refsgaard *et al.* 1988; Xu *et al.* 1996; Kizza *et al.* 2011, 2013; Li *et al.* 2013; Gosling 2014; Li *et al.* 2014; McIntyre *et al.* 2014; Emam *et al.* 2015). Among these applications, flood forecasting is perhaps the most original and widely performed. Generally, parameters of these models are not directly obtained from the measurable catchment characteristics, but calibrated by using various algorithms and procedures so that the simulated hydrologic process matches as close to the hydrological behaviour of the catchment as possible (Duan *et al.* 1992; Cheng *et al.* 2002; Goswami & O'Connor 2007; Jeon *et al.* 2014). The common goal of each optimization

algorithm is to seek the optimal value of model parameters according to numerical measures of the goodness-of-fit such as minimizing or maximizing an objective function. In general, a single overall objective function is often used to measure the performance of the model calibration (Xu 2001; Vrugt *et al.* 2003; Jeon *et al.* 2014), such as the Nash-Sutcliffe efficiency coefficient (NSE), coefficient of determination ( $R^2$ ) and the relative volume error (RE) of the hydrograph. If RE is chosen as an objective function, the simulations may have a good performance on the total volume, but a bad performance on the shape of the hydrograph and peak flow (Moussa & Chahinian 2009). When using the NSE,  $R^2$  (Legates & McCabe 1999) or similar indexes as an objective function, the simulations are prior to matching the high flows and these measures are

oversensitive to extreme values. It is difficult to make the simulation match most of the characteristics of the observed hydrograph simultaneously when calibrated by a single objective function (Vrugt *et al.* 2003). Different objective functions are adopted for different purposes. However, when all the characteristics of a hydrograph demand to be reproduced in real applications, no single objective function is efficient enough to calibrate the selected model.

Many studies have applied multi-objective functions to calibrate the parameters of hydrological models, which have been verified to perform better than single ones (Madsen 2000, 2003; Moussa & Chahinian 2009; Prakash *et al.* 2015; Hailegeorgis & Alfredsen 2016). It would be advisable to take various objective functions into consideration by calibrating the model in a multi-objective framework (Yapo *et al.* 1998; Liu & Sun 2010; Li *et al.* 2010; Moussa *et al.* 2011). Dickinson (1973) uses a minimum variance criterion to yield a composite forecast by estimating the weights of different forecasts. Yapo *et al.* (1998) present an efficient and effective algorithm for solving the multi-objective global optimization problem. Dong *et al.* (2013) combine the results of diverse models by using the Bayesian Model Averaging method and an integrated result is obtained. The most widely used method is to transform a multi-objective problem into a single objective problem by using aggregated distance measure (Yapo *et al.* 1998; Madsen 2000; Moussa & Chahinian 2009). Madsen (2003) aggregates two objective functions in a distributed hydrological modelling system to propose a proper balance between groundwater flow and surface runoff. Moussa & Chahinian (2009) composite two or more objective functions into a balanced objective function.

In most continuous models, the objective functions are used to present different parts of the hydrograph which generally contains volume error, root mean square error (RMSE) of the overall hydrograph, high flow events, middle flow events and low flow events. Further, peak flow events, middle flow events and low flow events are generally separated from the whole hydrograph by using a pre-defined threshold value (Madsen 2000; Van *et al.* 2009; Liu & Sun 2010). Madsen (2000) and Liu & Sun (2010) define the peak flow and low flow events as periods with flow above or below a threshold, and a compromise solution is obtained between peak flow and low flow. Van *et al.* (2009) combine

the middle flow objective function which is calculated using flows between the 30th and 70th percentile of the flow duration curve to other objective functions. While in real-time flood forecasting research, the event-based flood models are universally adopted to investigate multi-objective calibration issues. Peak flow reproduction is essential for flood forecasting exercises (Chahinian *et al.* 2005; Moussa & Chahinian 2009). It is obvious that the selection of threshold value of peak flow events or low flow events in these cases is subjective, whose reasonability needs to be discussed and compared with other issues in flood forecasting; however, the effect of this subjectivity is not well addressed in the literature. Thus, in this study, the reasonability of the threshold of the peak flow events is analysed and discussed by using a series of equally spaced thresholds. It is important to note that trade-offs exist between different objective functions so that a Pareto front or Pareto surface corresponding to a number of parameter sets could be obtained when solving multi-objective optimization problems (Gupta *et al.* 1998; Messac & Mattson 2002; Madsen 2003).

The performance of a simulated hydrograph is evaluated by a series of evaluation indexes such as RE, NSE, the ratio of annual simulated runoff volume to annual observed runoff volume ( $V_{sim}/V_{obs}$ ) (Yu & Yang 2000), and RMSE, etc. Some of them may be highly correlated, mutually opposing or even contradictory to each other. Therefore, integrating multiple evaluation indexes into a comprehensive value is necessary to identify the simulation performance effectively. Many multivariate statistical techniques have been applied for the analysis of high dimensional data (Wang *et al.* 2006; Wang & Ni 2008). Some of them are intractable in data processing, while the projection pursuit algorithm has superior ability on highly dimensional data processing so that it has been widely used in many fields such as forecasting, water resources assessment, environmental protection and so on (Rajeevan *et al.* 2007; Zhang & Dong, 2009). In this study, the projection pursuit method is used to solve the multi-index evaluation problem.

The main goal of this paper is to assess the effects of different objective functions and their combinations in calibrating flood forecasting models, and a new evaluation method is proposed to assess the performance of a flood forecasting model when evaluation indexes are more than two. The main goal is

achieved through the following steps: (1) the Xinanjiang (XAJ) model is chosen as the flood forecasting model and the shuffled complex evolution (SCE-UA) algorithm (Duan *et al.* 1993) is used to calibrate the model; (2) the performances of different single objective functions and multiple objective functions are assessed in model calibration; (3) the impact of different threshold values of the peak flow on flood forecasting is evaluated; and (4) the practicability of the projection pursuit method is verified through the evaluation of the performance of the flood forecasting model.

## STUDY AREA AND DATA

### Study area

The Chongyang River basin with an area of 4,848 km<sup>2</sup> and a river length of 126 km is selected to perform the study, which

is a main upstream tributary of Minjiang River located in southeastern China. The physiography of the basin consists of highly dissected topography with steep slopes and high stream densities (Figure 1). The surface of this area is covered by well drained yellow-red soil derived from a variety of parent materials such as sandstone, shale and granite. The forest coverage of this region is above 80% of the total area, making it one of the regions with the highest percentage of forest cover in China (Hu *et al.* 2014). This also means that human activities have little impact on the runoff of the basin. Its climate is dominated by the southeast Pacific Ocean and the southwest Indian Ocean subtropical monsoons and the regional landforms (Tang *et al.* 2013), which bring abundant long-term and extensive rainfall. The mean annual precipitation is 1,700 mm, of which 70–80% occurs in the rainy season from April to September. The average runoff depth is 1,100 mm and the runoff coefficient is 0.64. The catchment can be classified as humid, with abundant

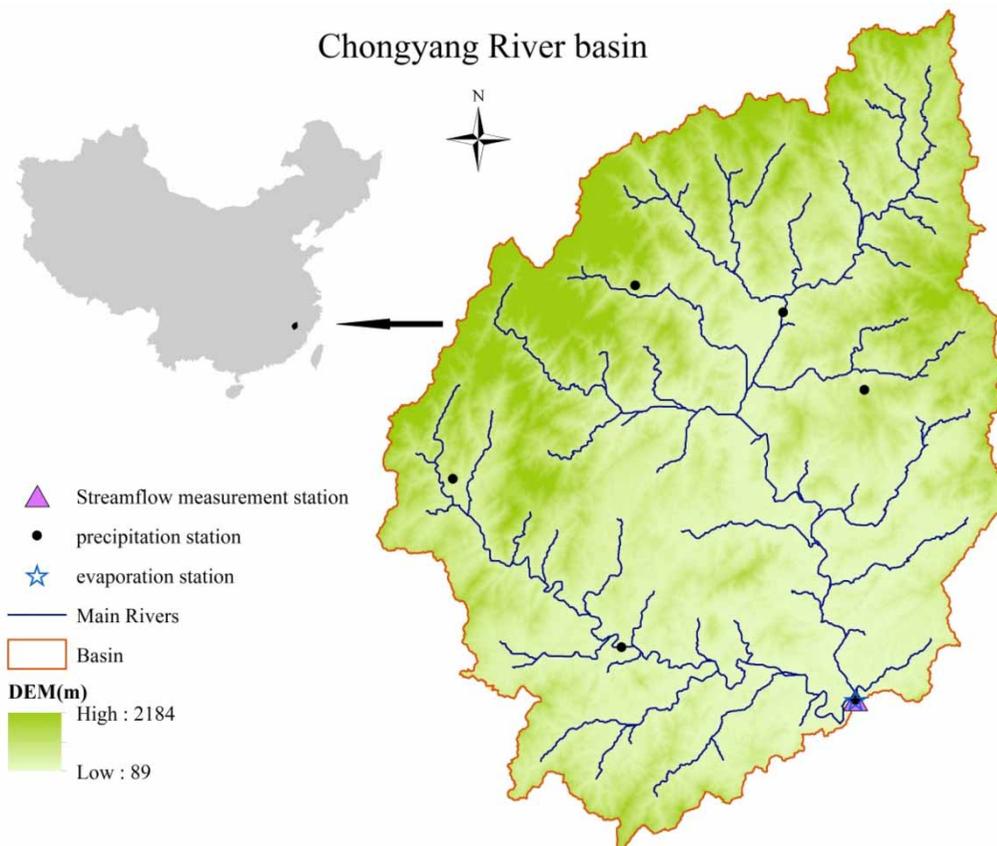


Figure 1 | Location of the study area and gauges.

soil moisture and high vegetation cover, and the flow generation mechanism is saturation excess runoff.

In this basin, flood is generally caused by two kinds of rain type, the typhoon rain and convection rain of short duration and high precipitation intensity. The former is caused by typhoon weather system which occurs mostly from July to September. The latter results from the interaction between the mid-latitude weather system and the low-latitude weather system, which often brings the frontal rainstorm and generally occurs from April to June. The geography and climate condition of this area makes it a central area of high rain and high flood risk, which often greatly threatens the safety of life and property in the basin. In June 1998, there was a very serious flood event with a return period up to 200 years in the basin, which led to huge financial losses to the local residents (Zhang & Hall 2004). As a result, it is very important and necessary to establish a flood forecasting model for public safety and water management in this study area.

### Data and their characteristics

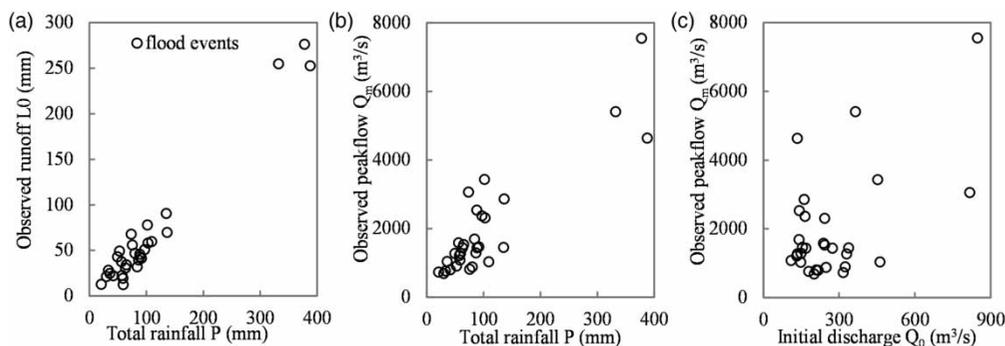
There are six rain gauges, one evaporation station and one stream gauging station to be used in the basin, whose locations can be seen in Figure 1. The hourly data series from 2001 to 2013 were applied to evaluate the model, which were provided by the Hydrology Bureau of Fujian Province, China. All the hydrological data have been recognized as high quality by the local Hydrology Bureau to be published in the local Hydrological Year Book. We have carefully checked all data and did not notice any unexpected

behaviour (result not shown). The continuous hydrological data series were applied as input to simulate the hourly continuous runoff and 30 flood events with different magnitudes were selected and abstracted from the continuous hydrological data series to evaluate the hydrological model. These events represent various hydrological behaviours and display a wide range of durations and rainfall intensity. The relations between the total runoff depth and the total rainfall, the peak flow and the total rainfall, and the peak flow and the initial discharge of the selected historical flood events are shown in Figure 2. It can be seen that there is a close relation between the total rainfall and the total runoff, and a relatively good relation between the total rainfall and the observed peak flow in this basin. It also shows that the relation between initial discharge and peak flow is not good, meaning that the observed peak flow is a result of the combined effect of rainfall and other factors such as the antecedent soil moisture condition, etc. So it is necessary to set a long enough period as warm-up period (in this study it is 2 days) to eliminate the effect of the initial discharge value in calibrating the flood forecasting model.

## MODEL AND OBJECTIVE FUNCTIONS

### Xinjiang model

The Xinjiang model is a conceptual rainfall-runoff model proposed by Zhao in the 1970s (Zhao et al. 1980). It has been extensively and successfully used for flood simulation and operational forecasting in the humid and semi-humid



**Figure 2** | Relationships between: (a) the total runoff depth and the total rainfall; (b) the peak flow and the total rainfall; and (c) the peak flow and the initial discharge of the selected historical flood events.

region in China with good performance (Hu et al. 2005; Li et al. 2009, Yao et al. 2009; Yan et al. 2016). The main feature of the model is the proposal of the concept of runoff formation on repletion of storage and storage capacity curve. It implies that runoff is not produced until the soil water content of the aeration zone reaches its field capacity (Zhao 1992). Zhao et al. (1995) used the storage capacity curve to solve the problem of the unevenly distributed soil moisture deficit.

The Xinanjiang model contained 15 parameters in total, which are listed in Table 1. Each represents the properties of the catchment. Although some insensitive parameters can be obtained from the observed information, the sensitive parameters must be calibrated (Thiemann et al. 2001). The range of each parameter is listed in Table 1, which is used for the calibration of the model's parameters.

### SCE-UA algorithm

The Shuffled Complex Evolution algorithm is a global optimization algorithm called SCE-UA for short (Duan

et al. 1993). The SCE-UA method is considered as the most effective and efficient search algorithm for applying parameter calibration for various conceptual rainfall-runoff models (Gan & Biftu 1996; Madsen 2000; Wang et al. 2010). The SCE-UA method is based on the notion of sharing information and on the concepts extracted from principles of natural biological evolution (Duan et al. 1994). It takes advantage of the Controlled Random Search (CRS) algorithms (i.e., global sampling, complex evolution) and combines them with powerful concepts of competitive evolution and complex shuffling (Nelder & Mead 1965; Duan et al. 1992). Both concepts mentioned above help to ensure that the information of the sample is fully used and does not become degenerate, so that the SCE-UA method has a high probability of succeeding in finding the global optimum (Duan et al. 1993; Jeon et al. 2014).

The SCE-UA method contains various algorithmic parameters in which the number of complexes  $p$  is the most important one. Some studies show that the proper choice of  $p$  depends on the dimensionality of the calibration problem (Duan et al. 1994). The larger the value of  $p$ , the higher the probability of converging into the global optimum but at the expense of a larger number of model simulations, and vice versa (Madsen 2000). Take all the factors into consideration,  $p$  is set equal to the number of calibration parameters, and for the rest of the parameters, the default values are used (Duan et al. 1994).

### Objective functions

In general terms, the objective of model calibration is to select model parameter values such that the model simulates the hydrological behaviour of the catchment as close to the observations as possible. In the course of flood forecasting, the runoff volume, flood hydrograph and flood peak – called the three elements of flood – are considered to be the most important factors in evaluating the performance of flood forecasting. The objectives that measure the three elements of the hydrological response are listed as follows: (i) a good agreement between the simulated and observed water volume such as a good water balance; (ii) a good overall agreement of the flood hydrograph; and (iii) a good agreement of the flood peak (Moussa & Chahinian 2009).

**Table 1** | Parameter ranges used in the simulations for the XAJ model

Parameter	Description	Range
WM	Areal mean tension water storage capacity: $WM = WUM + WLM + WDM$	100–150
X	$WUM = X * WM$ , WUM is the upper layer tension water storage capacity	0.1–0.6
Y	$WLM = Y * (WM - WUM)$ , WLM is the lower layer tension water storage capacity	0.1–0.6
KE	Ratio of potential evapotranspiration to pan evaporation	0.6–1.2
B	Tension water distribution index	0.1–1.2
SM	Areal mean free water storage capacity	20–70
EX	Free water distribution index	0.1–1.2
KI	Out flow coefficient of free water storage to interflow	0.8–0.99
KG	Out flow coefficient of free water storage to groundwater flow	0.95–1
IMP	Impermeable coefficient	0.01–0.1
C	Deep layer evapotranspiration coefficient	0.15–0.2
CI	Interflow recession coefficient	0.01–0.1
CG	Groundwater recession coefficient	0.03–0.15
N	Parameter of Nash unit hydrograph	1.0–10
NK	Parameter of Nash unit hydrograph	5–15

In the procedure of calibration, the quality of the final model parameter values could be affected by many factors, such as the quality of the input data, the simplifications and errors inherent in the model structure, the power of the optimization algorithm, the estimation criteria, the objective functions and so on (Madsen 2003; Feyen et al. 2007). It is difficult to take all the factors into account. In this paper our aim is to analyse the influence of the objective functions on the model parameters and simulations.

The following numerical performance statistics measure the different calibration objectives stated above (Madsen 2000; Moussa & Chahinian 2009):

volume error of the flood events

$$F_1(\theta) = \frac{1}{M_p} \sum_{j=1}^{M_p} \left| \frac{1}{n_j} \sum_{i=1}^{n_j} [Q_{obs,i} - Q_{sim,i}(\theta)] \right| \quad (1)$$

RMSE of the flood events

$$F_2(\theta) = \frac{1}{M_p} \sum_{j=1}^{M_p} \left[ \frac{1}{n_j} \sum_{i=1}^{n_j} [Q_{obs,i} - Q_{sim,i}(\theta)]^2 \right]^{1/2} \quad (2)$$

peak flow error of the flood events

$$F_p(\theta) = \frac{1}{M_p} \sum_{j=1}^{M_p} |Q_{obs \max j} - Q_{sim \max j}| \quad (3)$$

where  $Q_{obs,i}$  is the observed discharge at time  $i$  in each flood event,  $Q_{sim,i}$  the simulated discharge,  $n_j$  the number of time steps in the flood event  $j$ ,  $M_p$  the total number of flood events,  $\theta$  the set of model parameters to be calibrated,  $Q_{obs \max j}$  the observed peak flow of discharge in the flood event  $j$ , and  $Q_{sim \max j}$  the simulated peak flow of discharge in the flood event  $j$ .

Considering that the objective function  $F_p(\theta)$  is based on a small sampling of data as compared with the other two objective functions, the reliability and stability of the calibration result may not be guaranteed. As a result, the average RMSE of peak flow events which include more sample data has been adopted as an objective function to simulate the hydrograph, which is defined as the peak

flow events above a given threshold level (Madsen 2000; Van et al. 2009; Liu & Sun 2010). However, there is no universal method to decide a reasonable threshold for selecting the peak flow events. In this study, selection of a reasonable threshold has been tried by rolling thresholds from 0 to 100% of the peak flow value with equal step of 2%. A new equation of  $F_f(\theta)$ , named root mean square error of peak flow events, is defined as follows:

$$F_f(\theta) = \frac{1}{M_p} \sum_{j=1}^{M_p} \left[ \frac{1}{n p_{fj}} \sum_{i=1}^{n p_{fj}} [Q_{obs,i,p} - Q_{sim,i,p}(\theta)]^2 \right]^{1/2} \quad (4)$$

where  $Q_{obs,i,p}$  and  $Q_{sim,i,p}$  are the observed and simulated discharges at time  $i$  in each peak flow event, respectively;  $f$  is the threshold;  $n p_{fj}$  is the number of peak flow events (hydrograph in which discharge is greater than the threshold in flood event  $j$ ); and other notations are as previously defined.

The objective functions listed above are positive functions, and the parameters corresponding to the minimum value of each function (i.e., as close as to 0) could be regarded as the optimum value. Each of the single-objective calibration procedures is undertaken separately and contains a coupled manual and automatic calibration procedure. The main goal of the manual calibration procedure is to obtain the range of each parameter. The automatic calibration procedure aims at finding the optimum in the range. It is worth noting that trade-offs and equilibrium constraints exist between the different objective functions. So it is necessary to consider the calibration in a multi-objective framework.

In the multi-objective framework, the calibration problem can be stated as follows (Madsen 2003):

$$\text{Min}\{F_1(\theta), F_2(\theta), \dots, F_n(\theta)\}, \quad \theta \in \Theta \quad (5)$$

where  $F_i(\theta)$  ( $i = 1, 2, \dots, n$ ) are the different objective functions. The optimization problem is constrained because  $\theta$  is restricted to the feasible parameter space  $\Theta$ . The parameter space is usually defined as a hypercube by specifying lower and upper limits on each parameter which are chosen according to physical and mathematical constraints in the model or from modelling experiences (Kuczera 1997; Madsen 2000).

In general, the solution of Equation (5) will not be a single unique set of parameters but will consist of the so-called Pareto set of solutions according to various trade-offs between the different objectives (Gupta et al. 1998). A member of the Pareto set will be better than any other sets with respect to some of the objectives instead of all of other objectives because of the trade-off between the different objectives.

It is very challenging to solve the multi-objective calibration problem at present, so some simple ways are proposed, such as transforming the problem into a single-objective optimization problem by defining an equation that aggregates various objective functions. The equation of such an aggregate measure, called the Euclidean distance, is defined as follows (Madsen 2000):

$$F(\theta) = [(F_1(\theta) + A_1)^2 + (F_2(\theta) + A_2)^2 + \dots + (F_n(\theta) + A_n)^2]^{1/2} \quad (6)$$

where  $A_i$  ( $i = 1, 2, \dots, n$ ) are transformation constants corresponding to different objectives so that different relative priorities can be adopted to certain objectives. Different values of  $A_i$  using in the aggregated distance measure can investigate the entire Pareto front. However, it is computationally too expensive to calculate the entire Pareto front. So we can calculate some of the Pareto optimal solutions that people are of interest. In this case, an aggregated objective function was proposed to put equal weights on the different objectives. The value of  $A_i$  can be calculated by Equation (7) (Madsen 2000):

$$A_i = \text{Max}\{F_{j,\min}, j = 1, 2, \dots, n\} - F_{i,\min}, i = 1, 2, \dots, n \quad (7)$$

Equation (7) makes sure that each of the objective functions is transformed to having the same distance from the origin so that equal weights are put on the different objectives.

### Evaluation criteria

Four evaluation indexes are adopted for evaluating the goodness-of-fit of the 30 simulated flood hydrographs, which have been widely applied in flood model calibration and flood forecasting (Madsen 2000; Chahinian et al. 2005;

Moussa & Chahinian 2009; Wang et al. 2015). In order to measure the overall performance of flood events, the absolute values of the evaluation indexes of the 30 flood events are averaged as follows:

relative error of water,

$$RE = \frac{1}{M_p} \sum_{j=1}^{M_p} \left| \frac{\sum_{i=1}^{n_j} Q_{sim,i} - Q_{obs,i}}{\sum_{i=1}^{n_j} Q_{obs,i}} \right| \quad (8)$$

Nash-Sutcliffe coefficient,

$$NSE = \frac{1}{M_p} \sum_{j=1}^{M_p} \left( 1 - \frac{\sum_{i=1}^{n_j} [Q_{sim,i} - Q_{obs,i}]^2}{\sum_{i=1}^{n_j} [Q_{obs,i} - \overline{Q_{obs}}]^2} \right) \quad (9)$$

relative error of peak flow,

$$Q_{re} = \frac{1}{M_p} \sum_{j=1}^{M_p} \left| \frac{Q_{sim \max,j} - Q_{obs \max,j}}{Q_{obs \max,j}} \times 100\% \right| \quad (10)$$

and time lag of peak flow,

$$\Delta T = \frac{1}{M_p} \sum_{j=1}^{M_p} |T_{sim,j} - T_{obs,j}| \quad (11)$$

where  $T_{sim,j}$  is the time of occurrence of the simulated peak flow in the flood event  $j$ ,  $T_{obs,j}$  the time of occurrence of the observed peak flow in the flood event  $j$ , and other notations are as previously defined.

Owing to the contradictions existing among the evaluation results of the four indexes showed above, it is difficult to visually evaluate the performance among different objective-function combinations according to the evaluation results of the four indexes. In this study, a type of statistical technique called the projection pursuit method (Li & Chen 1985) is used to composite the four evaluation indexes so that an aggregated indicator is obtained to assess the performance of simulations calibrated by each objective function. Projection pursuit

aims at locating high dimensional space projections into low dimensional space ones which reveal the major characteristics about the structure of the data sets, and finally the optimal projection direction vector and projection eigenvalues are obtained (Swinson 2005). The bigger the projection eigenvalue, the better the comprehensive performance so that the sample data can be classified based on the comprehensive performance (Huang & Lu 2014). To assess the impact of different threshold of the peak flow on the performance of flood forecasting, the projection pursuit is applied to composite the four evaluation indexes.

## RESULTS AND DISCUSSION

The flood forecasting model was established with single and multi-objective functions to calibrate the model parameters and simulate the flood process. Seven flood forecasting schemes were put forward with three single objective functions and their combination in different ways. The specific description of each scheme is listed in Table 2. In order to compare the performance of single and multi-objective functions, the averages of the absolute evaluation values of the overall 30 flood events are listed in Table 3, and their variations are presented in Figure 3. The results are discussed in the following subsections.

**Table 2** | Specific description of each objective function scheme

Scheme	Objective function	Description
1	$F_1(\theta)$	Volume error of the flood events
2	$F_2(\theta)$	Root mean square error of the flood events
3	$F_p(\theta)$	Peak flow error of the flood events
4	$F_1(\theta)F_2(\theta)$	Combination of volume error and root mean square error of the flood events
5	$F_1(\theta)F_p(\theta)$	Combination of volume error and peak flow error of the flood events
6	$F_2(\theta)F_p(\theta)$	Combination of root mean square error and peak flow error of the flood events
7	$F_1(\theta)F_2(\theta)F_p(\theta)$	Combination of volume error, root mean square error and peak flow error of the flood events

**Table 3** | The performance measures of the 30 flood events with different objective functions

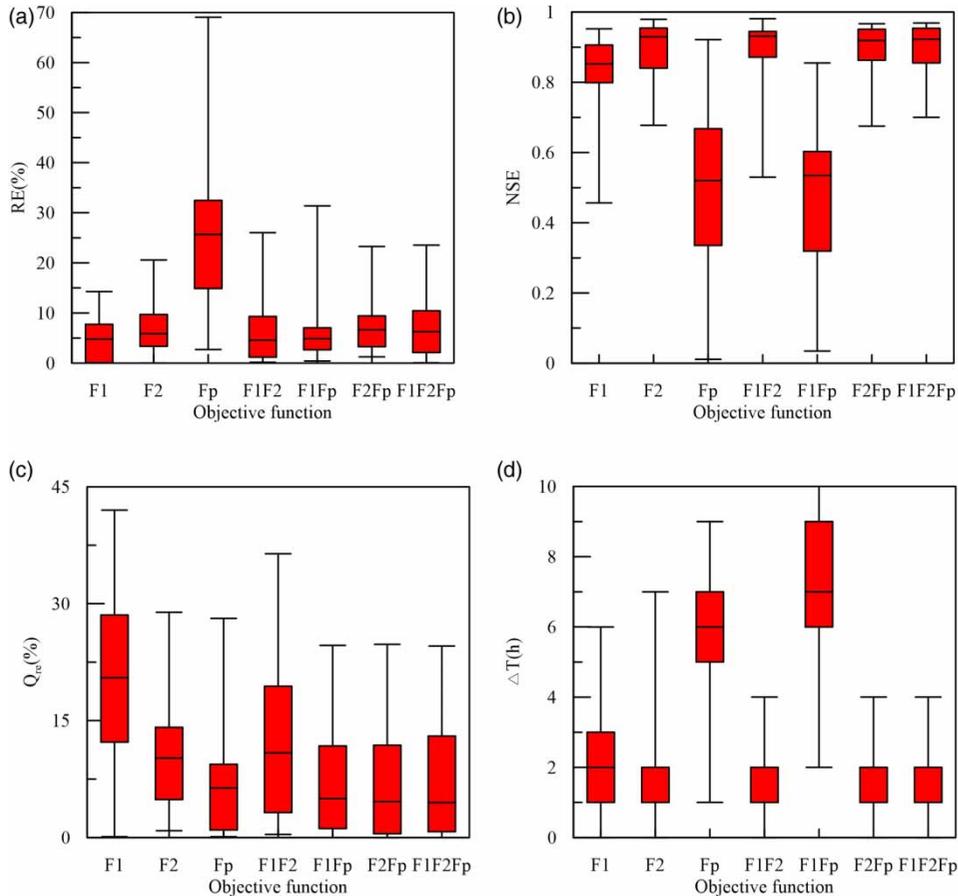
Scheme	RE (%)	NSE	Q <sub>re</sub> (%)	$\Delta T$ (h)
1	4.67	0.8386	19.32	2.03
2	6.98	0.8892	10.87	1.63
3	24.79	0.4217	6.54	5.8
4	5.8	0.8793	12.76	1.4
5	7.06	0.4368	7.11	7.07
6	7.93	0.8894	6.8	1.4
7	7.29	0.8923	6.75	1.4

### Single objective calibration results

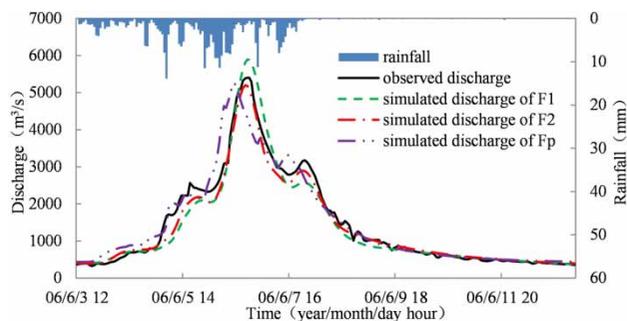
It can be seen from Tables 2 and 3 that there are great differences among the calibrated evaluation indexes for the three objective functions  $F_1(\theta)$ ,  $F_2(\theta)$  and  $F_p(\theta)$ . It is obvious that the objective function  $F_1(\theta)$  has a best value of relative error of water volume (RE) with 4.67%, while worst value of relative error of peak flow (Q<sub>re</sub>) with 19.32%. The objective function  $F_p(\theta)$  has an inverse performance compared with the  $F_1(\theta)$ . The objective function  $F_2(\theta)$  has the highest accuracy of the Nash-Sutcliffe coefficient (NSE) and time lag of peak flow ( $\Delta T$ ) among the three single objective functions.

The same phenomenon can be observed in the variation of the evaluation values of the 30 flood events presented in Figure 3. The objective function  $F_1(\theta)$  has the best average value and the least variation of RE, and the objective function  $F_p(\theta)$  performs the water volume worst. It is obvious that the objective function  $F_1(\theta)$  is to measure the agreement between the simulated and observed water volume, while the objective function  $F_p(\theta)$  puts weight on the simulation of the flood peak and naturally the overall water volume and flood hydrograph do not perform well. Accordingly, the objective function  $F_2(\theta)$  simulates the NSE best and  $F_p(\theta)$  performs the Q<sub>re</sub> best while the values of  $\Delta T$  are a little high.

In order to distinguish the differences of the hydrograph shapes among the three single objective functions, for illustrative purposes, the simulated hydrographs for the event of 6 June 2006 calibrated by the three objective functions are drawn in Figure 4. The numerical results can be found in Table 3 for the three single objectives (Schemes 1–3). In general, the simulations match the observed flood events well, which is closely related to the



**Figure 3** | Box plots of different evaluation indexes of 30 simulated flood event processes with different objective functions: (a) relative error of water volume; (b) Nash-Sutcliffe coefficient; (c) relative error of peak flow; (d) time lag of peak flow.



**Figure 4** | Simulated hydrographs for the event of 6 June 2006 calibrated by the single-objective functions  $F_1$ ,  $F_2$  and  $F_p$ .

humid region of the study area. The results are also in accordance with the findings of *Hu et al. (2005)*, *Li et al. (2009)* and *Xu et al. (2013)*, who applied XAJ model in the humid and semi-humid region in China with good performance. Specifically, the global shape of the hydrograph

performs better by using the objective function  $F_2(\theta)$  than other two single objective functions. The amplitude of the peak flow is simulated better by using the objective function  $F_p(\theta)$  than other two objective functions, however, the time of occurrence of the peak flow does not perform well by using the objective function  $F_p(\theta)$ .

It can be concluded from *Table 3*, *Figures 3* and *4* that the characteristics of the observed hydrograph are difficult to be matched simultaneously when it is calibrated by single objective functions. Naturally, it is necessary to consider multi-objective calibration so as to have a better simulation on the hydrological behaviour of the catchment.

### Multi-objective calibration results

The average values of the evaluation indexes of the 30 flood events with different multi-objective functions are listed in

Table 3 (Schemes 4–7) and their distribution are plotted in Figure 3. For illustrative purposes, the hydrographs for the event of 6 June 2006 are presented in Figure 5, which include the observation and the simulations by the four multi-objective functions listed in Table 2. It is clear that the simulations by multi-objective functions (Figure 5) reveal more comprehensive performance of the hydrological behaviour of the catchment than those by single objective functions (Figure 4). The mean value and variation of RE calibrated by multi-objective  $F_1(\theta)F_2(\theta)$  is smaller than that by  $F_2(\theta)$  and larger than that by  $F_1(\theta)$ , while the value of NSE calibrated by  $F_1(\theta)F_2(\theta)$  performs better than that by  $F_1(\theta)$  and worse than that by  $F_2(\theta)$ . The simulations by the multi-objective function  $F_1(\theta)F_2(\theta)$  incorporate the characteristics of the single functions  $F_1(\theta)$  and  $F_2(\theta)$  and get balanced results. The similar result could be obtained as for other multi-objective functions, except the combination of  $F_1(\theta)$  and  $F_p(\theta)$  (Scheme 5). Analysing the performance of each multi-objective functions, it can be found that compared to Scheme 5 ( $F_1(\theta)F_p(\theta)$ ) and Scheme 6 ( $F_2(\theta)F_p(\theta)$ ), Scheme 4 ( $F_1(\theta)F_2(\theta)$ ) has a better performance on the RE and NSE, while it performs worse on the  $Q_{re}$ . The overall best result is obtained by Scheme 7 ( $F_1(\theta)F_2(\theta)F_p(\theta)$ ).

The results shown above were calibrated by multi-objective functions using aggregated distance measure in which different objective functions were set equal weights. In order to estimate the Pareto front and analyse the trade-offs between different objective functions, a number of tests were carried out.

The outcome of the optimization algorithm and the estimated Pareto front calibrated by two objective functions are shown in Figure 6. The tracks of optimization presented in

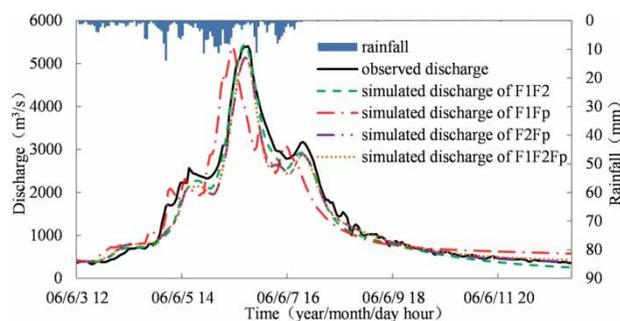


Figure 5 | Simulated hydrographs for the event of 6 June 2006 calibrated by multi-objective functions  $F_1F_2$ ,  $F_1F_p$ ,  $F_2F_p$  and  $F_1F_2F_p$ .

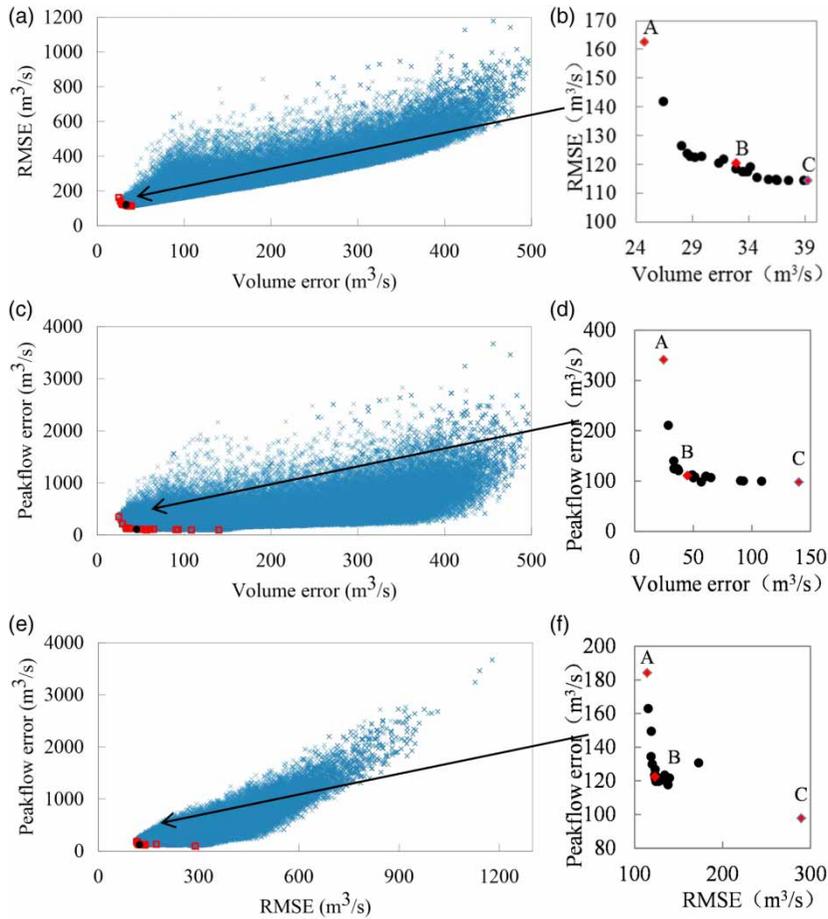
Figure 6 show that a sole optimum solution could not be obtained in multi-objective optimization, while the optimization can be represented by the estimated Pareto front '□'.

With respect to the optimization calibrated by  $F_1(\theta)F_2(\theta)$  (Figure 6(a) and 6(b)), the trade-off between the two objectives ( $F_1(\theta)$  and  $F_2(\theta)$ ) is less significant. This is due to the fact that the range of RMSE is larger than the corresponding relative range of volume error. By moving from C to A only a small reduction of the volume error is obtained at the expense of a large increase in the RMSE (Figure 6(b)). The maximum volume error of 39.21 (corresponding to  $RE = 6.98\%$ ) decreases to 24.74 (corresponding to  $RE = 4.67\%$ ), and at the same time the RMSE is increased from 114.27 (corresponding to  $NSE = 0.8892$ ) to 162.54 (corresponding to  $NSE = 0.8386$ ). This result is in line with the findings of Moussa & Chahinian (2009), who conducted a comparative study of different multi-objective calibration criteria using a conceptual rainfall-runoff model on flood events.

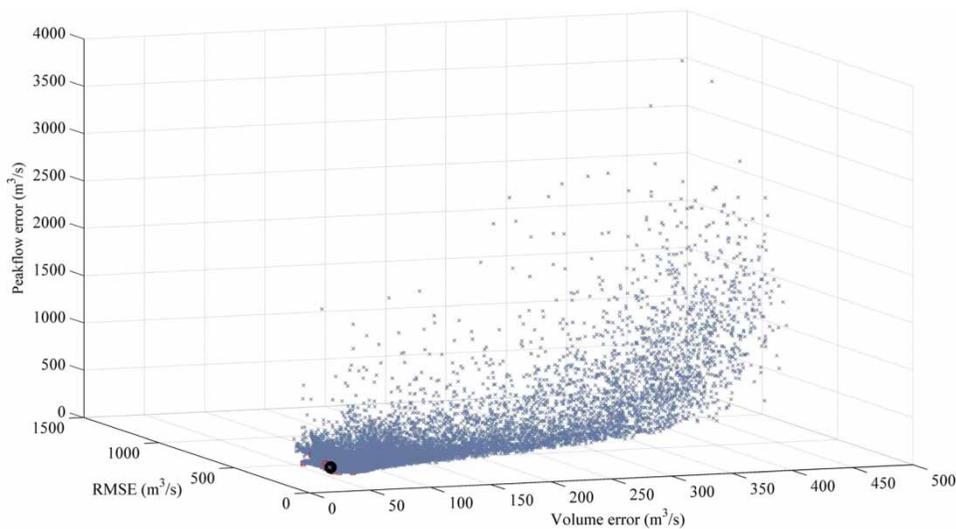
The estimated Pareto front for the calibration results of  $F_1(\theta)F_p(\theta)$  presents significant trade-offs (Figure 6(c) and 6(d)). A good calibration of  $F_1(\theta)$  (corresponding to  $F_1(\theta) = 24.74$ ) provides a bad calibration of  $F_p(\theta)$  (corresponding to  $F_p(\theta) = 341.33$ ), and vice versa ( $F_1(\theta) = 139.96$  for  $F_p(\theta) = 97.83$ ). The aggregated distance measure is seen to provide better balance between the two objectives in the optimization (point B) (corresponding to  $F_1(\theta) = 44.96$  and  $F_2(\theta) = 110.77$ ). Also, a significant trade-off can be observed for the calibration of  $F_2(\theta)F_p(\theta)$  in Figure 6(e) and 6(f):  $F_2(\theta) = 114.27$  when  $F_p(\theta) = 184.20$  and  $F_2(\theta) = 289.30$  when  $F_p(\theta) = 97.83$ . It seems that the single objective optimization provides the tails of the Pareto front and the compromise solution acts as a break point on the Pareto front. When moving along the front in either direction, one of the objective functions would have a small decrease while the other increases dramatically.

Multi-objective function values of Scheme 7 ( $F_1(\theta)F_2(\theta)F_p(\theta)$ ) during the calibration process are shown in Figure 7 in a tridimensional space. It can be found that most of the values center on the place close to the surface composed by smaller volume error and RMSE. The trade-offs between the three objective functions are significant from the result of Pareto front.

As each objective function represents some observed hydrograph characteristics, which have close relations with the model parameters, it is interesting to discuss the effect



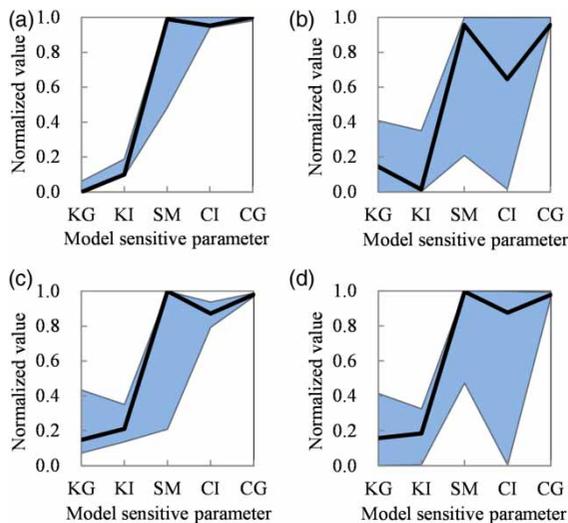
**Figure 6** | Calibration result using  $F_1(\theta)F_2(\theta)$  (a), (b),  $F_1(\theta)F_p(\theta)$  (c), (d), and  $F_2(\theta)F_p(\theta)$  (e), (f). In (a), (c) and (e), 'x', '□' and '●' state for objective function values, Pareto front and balanced aggregated objective function, respectively; (b), (d) and (f) are zoom ins of (a), (c) and (e), respectively. Marked optimum point B corresponds to the balanced aggregated objective function; marked optimum points A and C correspond to the objective functions on the X and Y axes, respectively.



**Figure 7** | Multi-objective function values of Scheme 7 ( $F_1(\theta)F_2(\theta)F_p(\theta)$ ) during the calibration process, where '□' indicates the Pareto front and '●' indicates the balanced aggregated objective function.

of different objective functions on the value of model parameters. The variation of the optimum model parameter sets along the Pareto front is shown in Figure 8. Five sensitive parameters were chosen to evaluate how the objective function affects the parameters of the model. The sensitive parameters are KG (coefficient of free water storage to ground water flow), KI (coefficient of free water storage to interflow), SM (areal mean free water storage capacity), CI (interflow recession coefficient) and CG (groundwater recession coefficient) (Zhang et al. 2012). The parameter values were normalized with respect to the upper and lower limits given in Table 1 so that the feasible range of all parameters is between 0 and 1. The compromised solution using the aggregated distance measure is shown in full bold line on Figure 8. The calibrated parameters of the compromised solution are within the interval delimited by the calibrated parameters of the Pareto front. For the calibration of  $F_1(\theta)$   $F_2(\theta)$  (Figure 8(a)), a narrow variability is observed in the parameter values when moving along the Pareto front which is in accordance with the fact that the trade-off between the two functions is less significant than other function combinations (Figure 8(b)–8(d)).

With respect to the range of the other combinations of objective functions (Figure 8(b)–8(d)), the intervals are



**Figure 8** | Normalized range of sensitive parameter values along the Pareto front using different multi-objective functions; the full bold line indicates the normalized parameter value corresponding to the balanced aggregated objective function: (a) flood events volume error and flood events RMSE; (b) flood events volume error and peak flow error; (c) flood events RMSE and peak flow error; (d) flood events volume error, flood events RMSE and peak flow error.

larger because significant trade-offs existed between each objective function. For instance, the normalized range of SM in Figure 8(a) is from 0.48 to 1, the low boundary values become smaller in the other cases (Figure 8(b)–8(d)). This result is not surprising as a smaller SM contributes to more surface runoff so as to have a good performance on the peak flow. That is to say, the parameter SM is sensitive to the objective function  $F_P(\theta)$ . With regard to the performance of groundwater recession coefficient CG, a narrow span is observed when moving along the Pareto front in the four multi-objective flood-forecasting schemes. Note that CG is closely related to the recession of groundwater and has less effect on high flow. When the trade-offs change between the three objective functions which pay attention to the water balance and high flow, the low flow exhibits good stability so that the variation of CG value is small.

### Comparison of the calibration performance on the flood events and on the continuous flow series

In order to analyse the differences between the simulation performance of flood events and continuous flow series, three sets of calibration results, i.e., the average values of performance measures of 30 individual flood events, performance measures calculated by putting 30 flood events together, and performance measures of continuous flow series are presented in Table 4.

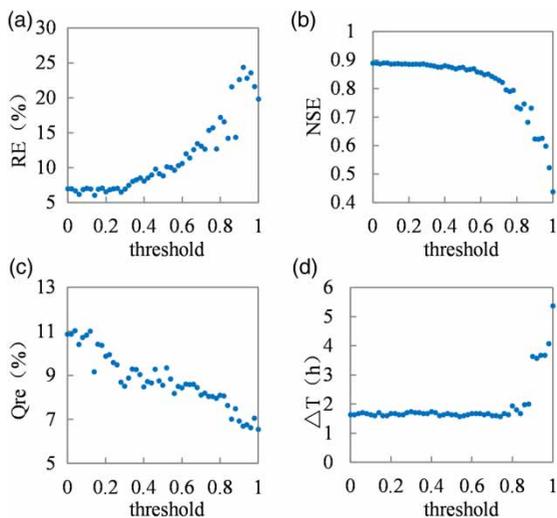
From Table 4 it can be found that the best results are obtained based on the calibration of 30 flood events together, followed by the continuous flow series and the worst results are obtained by averaging the performance measures of 30 individual flood events.

### The impact of different thresholds of the peak flow events on the calibration

Figure 9 shows the performance of the model calibration with single peak flow objective function  $F_f(\theta)$ , whose threshold changes from 0 to 100% of the peak flow in each flood events. It can be seen that  $Q_{re}$  (Figure 9(c)) decreases quickly along with increasing threshold of peak flow, indicating that high threshold results in better performance in the magnitude of the peak flow than small ones. When the threshold is less than 90% of the peak flow, no significant change of  $\Delta T$  is

**Table 4** | The calibration performance of flood events series and continuous flow series with different objective functions

Scheme	Average of 30 individual flood events		Total flood events		Continuous flow series	
	RE (%)	NSE	RE (%)	NSE	RE (%)	NSE
1	4.67	0.8386	-1.05	0.9258	5.98	0.8676
2	6.98	0.8892	-0.12	0.9675	-18.73	0.8809
3	24.79	0.4217	16.55	0.8363	0.92	0.8134
4	5.8	0.8793	-1.98	0.9635	0.49	0.8927
5	7.06	0.4368	0.98	0.7902	-14.59	0.7746
6	7.93	0.8894	0.31	0.9590	-12.66	0.8861
7	7.29	0.8923	0.59	0.9578	-14.73	0.8803

**Figure 9** | Different evaluation indexes of simulated flood process calibrated by objective function  $F_f(\theta)$  under different thresholds of the peak flow, the evaluation indexes are: (a) relative error of water volume; (b) Nash-Sutcliffe coefficient; (c) relative error of peak flow; and (d) time lag of peak flow.

found, however, a dramatic increase of  $\Delta T$  is observed with further increase of threshold. From Figure 9(a) and 9(b), it can be observed that the RE and NSE perform badly as the threshold value increases.

To summarise, a high value of threshold of the peak flow makes good performance on peak flow, however, it performs worse on the volume and the global shape of the flood hydrograph.

Considering the conflicting performance of model calibration by various single objective functions, the objective functions  $F_1(\theta)$ ,  $F_2(\theta)$  and  $F_f(\theta)$  with different thresholds of

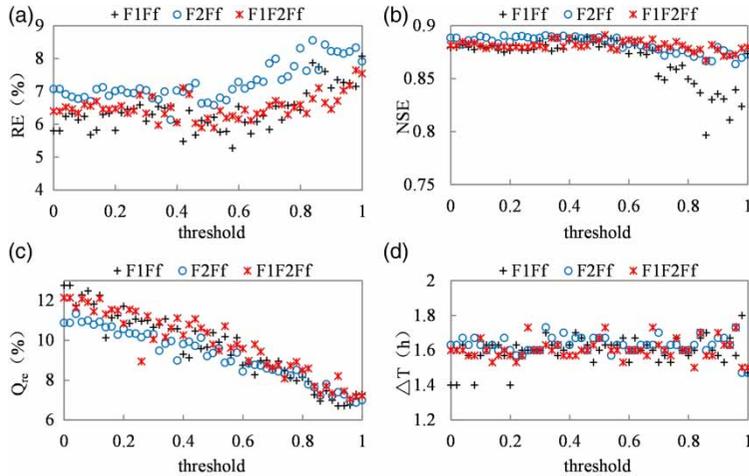
the peak flow are combined in different ways to calibrate the model parameters. Figure 10 shows the change of four evaluation indexes on different objective functions under different thresholds of peak flow respectively.

As for the index RE, it can be seen from Figure 10 that the performances of the calibration by the three multi-objective functions become worse and worse with the increasing threshold. A potential reason for this may be that the objective functions combined with  $F_f(\theta)$  of different thresholds did not take the volume error into consideration. In terms of NSE (Figure 10(b)), when the threshold is smaller than 60% of the peak flow, the NSE values of the multi-objective function  $F_2(\theta)F_f(\theta)$  remain stable and the values are larger than those of other two multi-objective functions. When the threshold is greater than 60% of the peak flow, the values of NSE of the multi-objective functions  $F_2(\theta)F_f(\theta)$  and  $F_1(\theta)F_2(\theta)F_f(\theta)$  decrease slowly, while the value of the multi-objective function  $F_1(\theta)F_f(\theta)$  decreases dramatically. With respect to  $Q_{re}$  (Figure 10(c)), it is obvious that the performance of the calibration becomes better and better for the three multi-objective functions with the increase of threshold. Comparing the values between the three multi-objective functions under different thresholds of peak flow, the value of multi-objective function  $F_2(\theta)F_f(\theta)$  is smaller than that of the other two multi-objective functions when the threshold is less than 80% of the peak flow, while there is no significant difference between them when the threshold increases over 80% of the peak flow. However, there is no obvious change to  $\Delta T$  when using the multi-objective functions with varying threshold of peak flow.

From the results shown above, it can be inferred that too large a threshold will contribute to worse performance of simulations when calibrated by a combination of two objective functions; for a combination of three objective functions, its performance is less influenced by threshold than the other two cases. In general, the threshold values of peak flow between 40 and 70% are good for the calibration of multi-objective function and a better goodness-of-fit is shown in the simulated hydrographs.

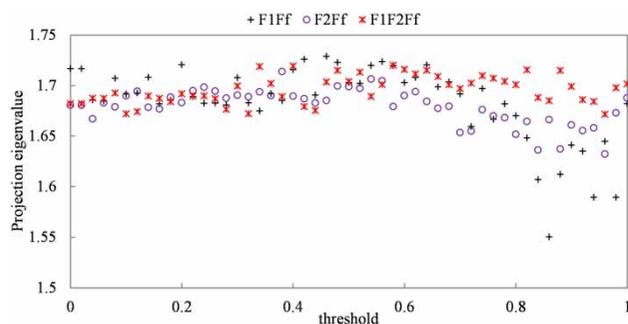
### The results of projection pursuit method

The variations of the projection eigenvalue for each objective function under different thresholds of peak flow are



**Figure 10** | Evaluation index values for different multi-objective functions under different threshold of the peak flow, the evaluation index are: (a) relative error of water volume; (b) Nash-Sutcliffe coefficient; (c) relative error of peak flow; and (d) time lag of peak flow.

shown in Figure 11. It can be found that the projection eigenvalue for the multi-objective function  $F_1(\theta)F_f(\theta)$  is high when the threshold approaches zero in which region the multi-objective function  $F_1(\theta)F_f(\theta)$  is equivalent to the multi-objective function  $F_1(\theta)F_2(\theta)$ , while no obvious differences are found between the projection values for the multi-objective functions  $F_1(\theta)F_2(\theta)$  and  $F_1(\theta)F_2(\theta)F_f(\theta)$  when the threshold value is low. However, the eigenvalue for all of three multi-objective functions increases when the threshold reaches between 40 and 70% of the peak flow. Unfortunately, the value for all objective functions declines to different degrees when the threshold goes over 70% of the peak flow during which the eigenvalues of the multi-objective functions  $F_1(\theta)F_f(\theta)$  and  $F_1(\theta)F_2(\theta)F_f(\theta)$  vary towards minimal and maximum values, respectively.

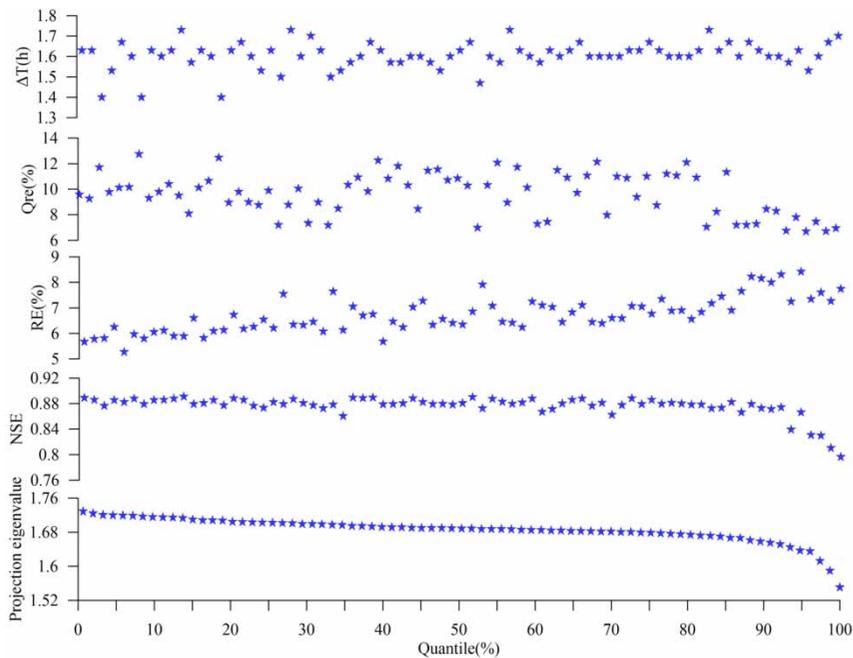


**Figure 11** | Projection eigenvalues for different multiple objective functions under different thresholds of the peak flow.

To verify the efficiency and reasonability of the new method, the projection eigenvalues of each objective function under different thresholds of peak flow were arranged from high to low and the corresponding values of each evaluation index are plotted in Figure 12. The figure suggests that the high projection eigenvalues are those which have low relative volume error (RE) and high value of Nash-Sutcliffe coefficient (NSE), whereas the poor projection eigenvalues are equivalent to big relative volume error (RE) and small Nash-Sutcliffe coefficient (NSE) to some extent. No obvious regularity has been found in the value of  $Q_{re}$  and  $\Delta T$ . That is, the simulation corresponding to high projection eigenvalue fits reasonably well on the observed hydrograph. The results are in accordance with the hypothesis that the bigger the projection eigenvalue the better the comprehensive performance, which in some degree demonstrates the reasonability of the projective pursuit method. Therefore, the projection pursuit method is a reasonable selection in solving multiple evaluation indexes problems which could be applied in performance evaluation of flood forecasting models.

## CONCLUSIONS

The XAJ model and SCE-UA algorithm were applied on the Chongyang River catchment in southeastern China with the purpose of assessing the effects of different objective functions



**Figure 12** | Evaluation index values of corresponding sorted projection eigenvalues for different objective functions under different threshold of the peak flow.

and their combinations in calibrating flood forecasting models. The model was calibrated on 30 flood events with both single objective and multi-objective schemes. Different objective functions and their combinations were chosen to calibrate the model and their performances were assessed and compared. The following conclusions are drawn from this study.

The performance of simulation is dependent on the objective function used in the flood forecasting model. No unique single objective function could predict all the characteristics of the shape of the hydrograph simultaneously. Significant trade-offs exist between different objective functions so that a set of Pareto optimal solutions is adopted to minimize calibration errors. A compromise solution is obtained when equal weights are assigned to different objective functions by using the aggregated distance measure. The results illustrate that the trade-off between flood events volume error and flood events RMSE is less significant. While significant trade-offs between different objective functions are represented in other combinations of objective functions and a wider range of parameters are reflected in these cases. Comparing to the result of total flood events, the performance of continuous flow series and the average of 30 individual flood events are worse.

High value of threshold makes better performance on peak flow, however, it performs worse on the volume and the global shape of the hydrograph. The impact of different thresholds of the peak flow in the multi-objective functions varies from different evaluation index. A larger threshold of peak flow contributes to good performance of peak flow at the expense of bad simulation in other aspects. The extent of the effect to the simulation varies with different objective functions in which the multi-objective function consisting of three objective functions has the minimal impact. The results also indicate that the threshold values between 40 and 70% of the peak flow have better performances for all multi-objective functions.

This study compares the effect of different objective functions in calibrating rainfall-runoff models which is an important early step in flood forecasting. Long-term flow simulation and flood forecasting are two important and different applications of hydrological models, which focus on different aspects of hydrograph, and therefore need different objective functions and evaluation criteria. This study contributes to an improved knowledge and method for accurate forecasting of river flood. It should be noted that although the study is performed using one model and

on one catchment, since the choice of objective functions is governed by the nature of the problem (i.e., the specific aspect of a hydrograph), rather than by the model and the catchment, the findings provide useful reference for other studies. Nevertheless, further studies involving more models and study regions are needed to generalize the findings of our study to other conditions.

## ACKNOWLEDGEMENTS

The study was supported by the National Natural Science Fund of China (51190094; 51339004; 51279138).

## REFERENCES

- Chahinian, N., Moussa, R., Andrieux, P. & Voltz, M. 2005 Comparison of infiltration models to simulate flood events at the field scale. *J. Hydrol.* **306**, 191–214.
- 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.
- Dickinson, J. P. 1973 Some statistical results in the combination of forecasts. *J. Oper. Res. Soc.* **24**, 253–260.
- Dong, L. H., Xiong, L. H. & Yu, K. X. 2013 Uncertainty analysis of multiple hydrologic models using the Bayesian model averaging method. *J. App. Math.* **30**, 701–710.
- Duan, Q. Y., Sorooshian, S. & Gupta, V. K. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28**, 1015–1031.
- Duan, Q. Y., Gupta, V. K. & Sorooshian, S. 1993 Shuffled complex evolution approach for effective and efficient global minimization. *J. Optimiz. Theory App.* **76**, 501–521.
- 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.
- Emam, A. R., Kappas, M. & Hosseini, S. Z. 2015 Assessing the impact of climate change on water resources, crop production and land degradation in a semi-arid river basin. *Hydrol. Res.* **46** (6), 854–870.
- Feyen, L., Vrugt, J. A., Nualláin, B. O., Knijff, J. & Roo, A. D. 2007 Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the IISFLOOD model. *J. Hydrol.* **332**, 276–289.
- Gan, T. Y. & Biftu, G. F. 1996 Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure. *Water Resour. Res.* **32**, 3513–3524.
- Gosling, R. 2014 Assessing the impact of projected climate change on drought vulnerability in Scotland. *Hydrol. Res.* **45**, 806–886.
- 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.
- Gupta, H. V., Sorooshian, S. & Yapo, P. O. 1998 Toward improved calibration of hydrological models: multiple and noncommensurable measures of information. *Water Resour. Res.* **34**, 751–763.
- Hailegeorgis, T. T. & Alfredsen, K. 2016 Multi-basin and regional calibration based identification of distributed precipitation-runoff models for hourly runoff simulation: calibration and transfer of full and partial parameters. *Hydrol. Res.* **47** (2), 239–259.
- Hu, C. H., Guo, S. L., Xiong, L. H. & Peng, D. Z. 2005 A modified Xinanjiang model and its application in northern China. *Nord. Hydrol.* **36**, 175–192.
- Hu, X. S., Wu, C. Z., Hong, W., Qiu, R. Z., Li, J. & Hong, T. 2014 Forest cover change and its drivers in the upstream area of the Minjiang River, China. *Ecol. Indic.* **46**, 121–128.
- Huang, H. & Lu, J. 2014 Identification of river water pollution characteristics based on projection pursuit and factor analysis. *Environ. Earth Sci.* **72**, 3409–3417.
- Jeon, J. H., Park, C. G. & Engel, B. A. 2014 Comparison of performance between genetic algorithm and SCE-UA for calibration of SCS-CN surface runoff simulation. *Water* **6**, 3433–3456.
- Kizza, M., Rodhe, A., Xu, C.-Y. & Ntale, H. K. 2011 Modelling catchment inflows into Lake Victoria: uncertainties in rain-runoff modelling for Nzoia River. *Hydrolog. Sci. J.* **56**, 1210–1226.
- Kizza, M., Guerrero, J. L., Rodhe, A., Xu, C. Y. & Ntale, H. K. 2013 Modelling catchment inflows into Lake Victoria: regionalisation of the parameters of a conceptual water balance model. *Hydrol. Res.* **44**, 789–808.
- Kuczera, G. 1997 Efficient subspace probabilistic parameter optimization for catchment models. *Water Resour. Res.* **33**, 177–185.
- Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**, 233–241.
- Li, G. & Chen, Z. 1985 Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo. *J. Am. Stat. Assoc.* **80**, 759–766.
- Li, H. X., Zhang, Y. Q., Chiew, F. H. S. & Xu, S. G. 2009 Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. *J. Hydrol.* **370**, 155–162.
- Li, X., Weller, D. E. & Jordan, T. E. 2010 Watershed model calibration using multi-objective optimization and multi-site averaging. *J. Hydrol.* **380**, 277–288.
- Li, L., Ngongondo, C. S., Xu, C. Y. & Gong, L. 2013 Comparison of the global TRMM and WFD precipitation datasets in driving a large-scale hydrological model in Southern Africa. *Hydrol. Res.* **44**, 770–788.

- Li, H., Beldring, S. & Xu, C. Y. 2014 Implementation and testing of routing algorithms in the distributed HBV model for mountainous catchments. *Hydrol. Res.* **45**, 322–333.
- Liu, Y. & Sun, F. 2010 Sensitivity analysis and automatic calibration of a rainfall-runoff model using multi-objectives. *Ecol. Inform.* **5**, 304–310.
- Madsen, H. 2000 Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *J. Hydrol.* **235**, 276–288.
- Madsen, H. 2003 Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. Water Resour.* **26**, 205–216.
- McIntyre, N., Ballard, B., Bruen, M., Bulygina, N., Buytaert, W., Cluckie, I., Dunn, S., Ehret, U., Ewen, J., Gelfan, A., Hess, T., Hughes, D., Jackson, B., Kjeldsen, T., Merz, R., Park, J.-S., O'Connell, E., O'Donnell, G., Oudin, L., Todini, E., Wagener, T. & Wheeler, H. 2014 Modelling the hydrological impacts of rural land use change. *Hydrol. Res.* **45**, 737–754. doi:10.2166/nh.2013.145.
- Messac, A. & Mattson, C. A. 2002 Generating well-distributed sets of Pareto points for engineering design using physical programming. *Optim. Eng.* **3**, 431–450.
- Moussa, R. & Chahinian, N. 2009 Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events. *Hydrol. Earth Syst. Sci.* **13**, 519–535.
- Moussa, F., Oudin, L., Plagnes, V., Mangin, A. & Bendjoudi, H. 2011 A multi-objective calibration framework for rainfall-discharge models applied to karst systems. *J. Hydrol.* **400**, 364–376.
- Nelder, J. A. & Mead, R. 1965 A simplex method for function minimization. *Comput. J.* **7**, 308–313.
- Prakash, O., Srinivasan, K. & Sudheer, K. P. 2015 Adaptive multi-objective simulation-optimization framework for dynamic flood control operation in a river-reservoir system. *Hydrol. Res.* **46** (6), 893–911.
- Rajeevan, M., Pai, D. S., Kumar, R. A. & Lal, B. 2007 New statistical models for long-range forecasting of southwest monsoon rainfall over India. *Clim. Dynam.* **28**, 813–828.
- Refsgaard, J. C., Havnø, K., Ammentorp, H. C. & Verwey, A. 1988 Application of hydrological models for flood forecasting and flood control in India and Bangladesh. *Adv. Water Resour.* **11**, 101–105.
- Swinson, M. D. 2005 Statistical modeling of high-dimensional nonlinear systems: a projection pursuit solution. Georgia Institute of Technology Press, Atlanta, pp. 13–88 (dissertation).
- Tang, Q., He, X. B., Bao, Y. H., Zhang, X. B., Guo, F. & Zhu, H. W. 2013 Determining the relative contributions of climate change and multiple human activities to variations of sediment regime in the Minjiang River, China. *Hydrol. Process.* **27**, 3547–3559.
- Thiemann, M., Trosset, M., Gupta, H. & Sorooshian, S. 2001 Bayesian recursive parameter estimation for hydrologic models. *Water Resour. Res.* **37**, 2521–2535.
- Van Werkhoven, K., Wagener, T., Reed, P. & Tang, Y. 2009 Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Adv. Water Resour.* **32**, 1154–1169.
- Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. & Sorooshian, S. 2003 Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour. Res.* **39**, 1214, doi:10.1029/2002WR001746.
- Wang, S. J. & Ni, C. J. 2008 Application of projection pursuit dynamic cluster model in regional partition of water resources in China. *Water Resour. Manag.* **22**, 1421–1429.
- Wang, X., Smith, K. & Hyndman, R. 2006 Characteristic-based clustering for time series data. *Data Min. Knowl. Disc.* **13**, 335–364.
- Wang, Y. C., Yu, P. S. & Yang, T. C. 2010 Comparison of genetic algorithms and shuffled complex evolution approach for calibrating distributed rainfall-runoff model. *Hydrol. Process.* **24**, 1015–1026.
- Wang, C. X., Li, Y. P., Zhang, J. L. & Huang, G. H. 2015 Assessing parameter uncertainty in semi-distributed hydrological model based on type-2 fuzzy analysis – a case study of Kaidu River Basin. *Hydrol. Res.* **46** (6), 969–983.
- Xu, C.-Y. 2001 Statistical analysis of a conceptual water balance model, methodology and case study. *Water Resour. Manag.* **15**, 75–92.
- Xu, C. Y., Seibert, J. & Halldin, S. 1996 Regional water balance modelling in the NOPEX area: development and application of monthly water balance models. *J. Hydrol.* **180**, 211–236.
- Xu, H. L., Xu, C.-Y., Chen, H., Zhang, Z. X. & Li, L. 2013 Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *J. Hydrol.* **505**, 1–12.
- Yan, R. H., Huang, J. C., Wang, Y., Gao, J. F. & Qi, L. Y. 2016 Modeling the combined impact of future climate and land use changes on streamflow of Xinjiang Basin, China. *Hydrol. Res.* **47** (2), 356–372.
- Yao, C., Li, Z. J., Bao, H. J. & Yu, Z. B. 2009 Application of a developed Grid-Xinjiang model to Chinese watersheds for flood forecasting purpose. *J. Hydrol. Eng.* **14**, 923–934.
- Yapo, P. O., Gupta, H. V. & Sorooshian, S. 1998 Multi-objective global optimization for hydrologic models. *J. Hydrol.* **204**, 83–97.
- Yu, P. & Yang, T. 2000 Fuzzy multi-objective function for rainfall-runoff model calibration. *J. Hydrol.* **238**, 1–14.
- Zhang, J. & Hall, M. J. 2004 Regional flood frequency analysis for the Gan-Ming river basin in China. *J. Hydrol.* **296**, 98–117.
- Zhang, C. & Dong, S. H. 2009 A new water quality assessment model based on projection pursuit technique. *J. Environ. Sci.* **21**, S154–S157.
- Zhang, D., Zhang, L., Guan, Y., Chen, X. & Chen, X. 2012 Sensitivity analysis of Xinanjiang rainfall-runoff model parameters: a case study in Lianghui, Zhejiang province, China. *Hydrol. Res.* **43**, 123–134.
- Zhao, R. J. 1992 The Xinanjiang model applied in China. *J. Hydrol.* **135**, 371–381.

Zhao, R. J., Zhang, Y. L., Fang, L. R., Liu, X. R. & Zhang, Q. S. 1980 The Xinangjiang model. In: *Proc. Oxford Symposium on Hydrological Forecasting*. International Association of Hydrological Sciences, Wallingford, UK, pp. 351–356.

Zhao, R. J., Liu, X. R. & Singh, V. P. 1995 The Xinanjiang model. In: *Computer Models of Watershed Hydrology* (V. P. Singh, ed.). Water Resources Publications, Highlands Ranch, CO, pp. 215–232.

First received 3 April 2015; accepted in revised form 26 August 2015. Available online 14 October 2015