

Feasibility and uncertainty of using conceptual rainfall-runoff models in design flood estimation

Qiang Zeng, Hua Chen, Chong-Yu Xu, Meng-Xuan Jie and Yu-Kun Hou

ABSTRACT

Hydrological models are developed for different purposes including flood forecasting, design flood estimation, water resources assessment, and impact study of climate change and land use change, etc. In this study, applicability and uncertainty of two deterministic lumped models, the Xinanjiang (XAJ) model and the Hydrologiska Byråns Vattenbalansavdelning (HBV) model, in design flood estimation are evaluated in a data rich catchment in southern China. Uncertainties of the estimated design flood caused by model equifinality and calibration data period are then assessed using the generalized likelihood uncertainty estimation (GLUE) framework. The results show that: (1) the XAJ model is likely to overestimate the design flood while HBV model underestimates the design flood; (2) the model parameter equifinality has significant impact on the design flood estimation results; (3) with the same length of calibration period, the results of design flood estimation are significantly influenced by which period of the data is used for model calibration; and (4) 15–20 years of calibration data are suggested to be necessary and sufficient for calibrating the two models in the study area.

Key words | continuous simulation, flood frequency, GLUE, HBV model, L-moments, Xinanjiang model

Qiang Zeng
Hua Chen
Chong-Yu Xu (corresponding author)
Meng-Xuan Jie
Yu-Kun Hou
State Key Laboratory of Water Resources and
Hydropower Engineering Science,
Wuhan University,
Wuhan 430072,
China
E-mail: chongyu.xu@geo.uio.no

Chong-Yu Xu
Department of Geosciences,
University of Oslo,
Norway

INTRODUCTION

Precise determination of design flood in hydrologic design and risk assessment of hydraulic engineering is of significant importance. In designing hydraulic schemes, the estimation of the design flood value is usually accomplished through the calculation of flood frequency by utilizing long-term historical hydrological data, whose length significantly influences the estimation performance. However, due to the lack of observed hydrological data of sufficient length and quality in many basins, it is hard to achieve the required level of consistency, homogeneity and stationarity of the estimated results. Under these circumstances, rainfall-runoff models can be used to simulate the required runoff series for frequency analysis. It is well known that rainfall-runoff models are widely used in flood forecasting (e.g. Refsgaard *et al.* 1988; Mwale *et al.* 2014), water resources assessment (e.g. Xu *et al.* 1996; Kizza *et al.* 2013; Hailegeorgis & Alfredsen 2016), regional and global water balance calculation (e.g. Arnell

1999; Li *et al.* 2013a, 2013b), impacts of climate change and land use change assessment (e.g. Bastola *et al.* 2011; Lawrence & Haddeland 2011; Gosling 2014; McIntyre *et al.* 2014; Emam *et al.* 2015; Singh *et al.* 2015; Yan *et al.* 2016) and streamflow simulation in ungauged catchments (e.g. Xu, 1999; McIntyre *et al.* 2005; Murray & Blöschl 2011; Mwale *et al.* 2014). Among these applications most emphasis has been put on flood control and water resources management (Pechlivanidis *et al.* 2011), while only a few applications to estimate design flood have been reported and discussed (Boughton & Droop 2005; Ngongondo *et al.* 2013). As precipitation, potential evapotranspiration and temperature are more widely available with longer records than runoff observations (Blazkova & Beven 2002), they are used as the main inputs to rainfall-runoff models for runoff simulations to estimate the design floods.

The use of rainfall-runoff models for design flood estimation can be classified into two categories (Boughton &

Droop 2003). One is the event-based simulation, in which rainfall-runoff models are fed by the design rainfall event and the assumed antecedent conditions of the catchment. Another one is the continuous simulation, in which rainfall-runoff models are fed by the historical or simulated rainfall to make a continuous runoff simulation, from which the design flood estimation can be drawn. The event-based simulation approach often assumes the T-year design rainfall event will generate the T-year flood event (Bradley & Potter 1992; Smithers *et al.* 1997). However, this pragmatic assumption clearly does not give good representation of the complex relationship between the design rainfall and design flood (Brigode *et al.* 2014). In the event-based simulation approach the input design rainfall can be the one derived from historical records or from stochastic rainfall event simulations. Saghafian *et al.* (2014) keeps the idea that the simulated runoff values by rainfall-runoff models from precisely observed rainfall may suffer less uncertainty than the runoff transformed from the measured water level through water level-discharge relation curves. He applied this event-based method to the Tangrah watershed located in north-eastern Iran to analyse the flood frequency, and the results showed that the design flood derived from simulated flood peaks were less than that from observations (Saghafian *et al.* 2014). With the fast development of computing facilities, continuous simulation has been used for the design flood estimation (Beven 1987; Calver & Lamb 1995; Lamb 1999; Suman & Bhattacharya 2015). Calver & Lamb (1995) performed the continuous simulation to 10 UK catchments by using the historical rainfall records as the input of two models, the probability-distributed model and the time-area topographic extension model, for flood frequency analysis. Lamb (1999) discussed the calibration method of rainfall-runoff models used for design flood estimation by continuous simulation. The continuous simulation approach can also be used for data limited or ungagged sites if good correlations can be made between the model parameters and the characteristics of the river basin (Saliha *et al.* 2011). Discussions and applications of the continuous simulation in ungagged sites are presented in the works by Blazkova & Beven (2002) and Smithers *et al.* (2013).

The merits and drawbacks of these two design flood estimation approaches have been analysed by many researchers

(Lamb 1999; Boughton & Droop 2003). The continuous runoff simulation approach is regarded to be more promising where the problem of antecedent condition is avoided (Calver & Lamb 1995), while in the event-based approach assumptions about the antecedent condition must be made, which will cause uncertainty.

The existence of the four important sources of uncertainties in hydrological modelling, i.e. uncertainties in input data, uncertainties in output data used for calibration, uncertainties in model parameters and uncertainties in model structure (Refsgaard & Storm 1996), means the modelling results are uncertain and the equifinality problem has been universally found in hydrological models. Uncertainty of rainfall-runoff models for flood forecasting has been discussed in many studies (Cameron *et al.* 2000; Beven & Freer 2001; Blazkova & Beven 2004; Li *et al.* 2009, 2010; Wang *et al.* 2015; Tian *et al.* 2016), however uncertainty that is inherent in the continuous simulation for design flood estimation has not drawn much attention in the hydrological literature. Brigode *et al.* (2014) used the bootstrap method to analyse the uncertainties in a semi-continuous design flood estimation method, the SCHADEX extreme flood estimation method, which was developed by Paquet *et al.* (2013). The results showed that the variability of observed rainfall and the difference of the rainfall-runoff model calibration periods had significant impact on the design flood estimation.

In recent years, many researchers have focused on the coupling of stochastic rainfall models and rainfall-runoff models to give a long runoff simulation in order to reduce the uncertainty caused by the extrapolation of the fitted distribution curve of observed peak discharges (Boughton & Hill 1997; Cameron *et al.* 1999; Blazkova & Beven 2002). Cameron *et al.* (1999) used TOPMODEL coupled with a stochastic rainfall generator to estimate 1,000-year flood in the Wye catchment in UK and used the generalized likelihood uncertainty estimation (GLUE) framework to analyse the uncertainty of the continuous simulation approach. The above studies indicate that compared with the large amount of studies conducted on using hydrological models for flood forecasting, water resources assessment, and impact studies of climate and land use changes, more effort needs to be directed to the usefulness and uncertainty of using hydrological models for design flood study.

The objectives of this paper are: (i) to investigate the applicability of lumped conceptual deterministic hydrological models in design flood estimation; and (ii) to analyse the uncertainties that are inherent in the hydrological model-design flood model approach. To achieve these objectives, the following tasks are performed: (1) the historical rainfall data are used to run the rainfall-runoff models (Xinjiang (XAJ) model and Hydrologiska Byråns Vattenbalansavdelning (HBV) model) to simulate runoff data with the same length of observed runoff data; (2) design flood estimations derived from runoff simulations and runoff observations are compared to analyse the feasibility of using runoff produced by hydrological models to estimate design flood; and (3) to evaluate the uncertainty caused by model equifinality by using the GLUE method, and the uncertainty induced by the difference of model calibration periods by using different sub-records of runoff observations to calibrate hydrological models.

DATA AND METHODS

Study area and data

The study was conducted in Xiangjiang basin, a tributary basin of the Yangtze River, which is located between 24–29 °N and 110 °30′–114 °E, central-south China with a total area of 94,660 km² and total river length of 856 km. Mountainous landscapes are found in the east, south, and west of the basin. Its northern part is made up of plains and hills. Xiangjiang River originates from the southern mountainous region and runs to the north plains, and finally inlets into the Dongting Lake. The climate of this basin is controlled by the Mongolia high pressure system in winter and influenced by southeast monsoon in summer, which results in the inhomogeneous spatial and temporal distribution of precipitation. The mean annual precipitation is 1,450 mm, of which 60–70% occurs in the rainy season from April to September. The annual mean temperature is about 17 °C. The mean temperature of the coldest month (January) is about 4 °C. The study area is a sub-catchment of Xiangjiang basin with a drainage area of 52,150 km². The boundary of the study area is shown in Figure 1. It is located in the upper part of Xiangjiang basin with a control runoff station, Hengyang station, at the centre of Xiangjiang basin.

Forty-one years of daily hydro-meteorological data (from 1965 to 2005), including precipitation, pan evaporation and observed runoff, are used in this study. Among them, the precipitation data come from 97 precipitation stations, and the evaporation data come from 12 evaporation stations. The distribution of these stations is shown in Figure 1. The quality of these data is controlled by the Hydrology and Water Resources Bureau of Hunan Province, China. Many other studies (Xu *et al.* 2013, 2015a, 2015b) have used these data for various research purposes.

Xinjiang model

The Xinjiang model was developed in the 1970s (Zhao *et al.* 1980). The main characteristic of this deterministic lumped model is the concept of runoff formation on repletion of storage, which means that runoff is not generated until the soil moisture content reaches filled capacity. This characteristic guarantees good performance of this model in continuous hydrological simulation in humid and semi-humid regions. The XAJ model has been successfully and widely used in China and many other countries for hydrological forecasting and flow simulations (Zhao 1992; Zhao *et al.* 1995; Li *et al.* 2013a, 2013b). The flowchart of this deterministic lumped model is shown in Figure 2. Symbols inside the boxes of Figure 2 are variables including model inputs, model outputs and state variables. The inputs to the model include daily precipitation, P , and measured pan evaporation, EM . The outputs are the simulated runoff, Q_{sim} , of the whole basin and the actual evapotranspiration, E , which contains three components, upper soil layer evaporation (EU), lower soil layer evaporation (EL), and deep soil layer evaporation (ED). Symbols outside the boxes of Figure 2 are parameters of this model and their explanations are listed in Table 1.

In the XAJ model, the catchment is represented by a stack of horizontal soil layers with total water storage capacity of WM. They are the upper soil layer, the lower soil layer, and the deep soil layer with certain water storage capacity, UM, LM, and DM, respectively. The potential evapotranspiration rate (EP) is derived by the product of the measured pan evaporation and the model parameter KE. Water stored in the upper layer is evaporated first with a rate of EU. If it is not able to meet the remaining

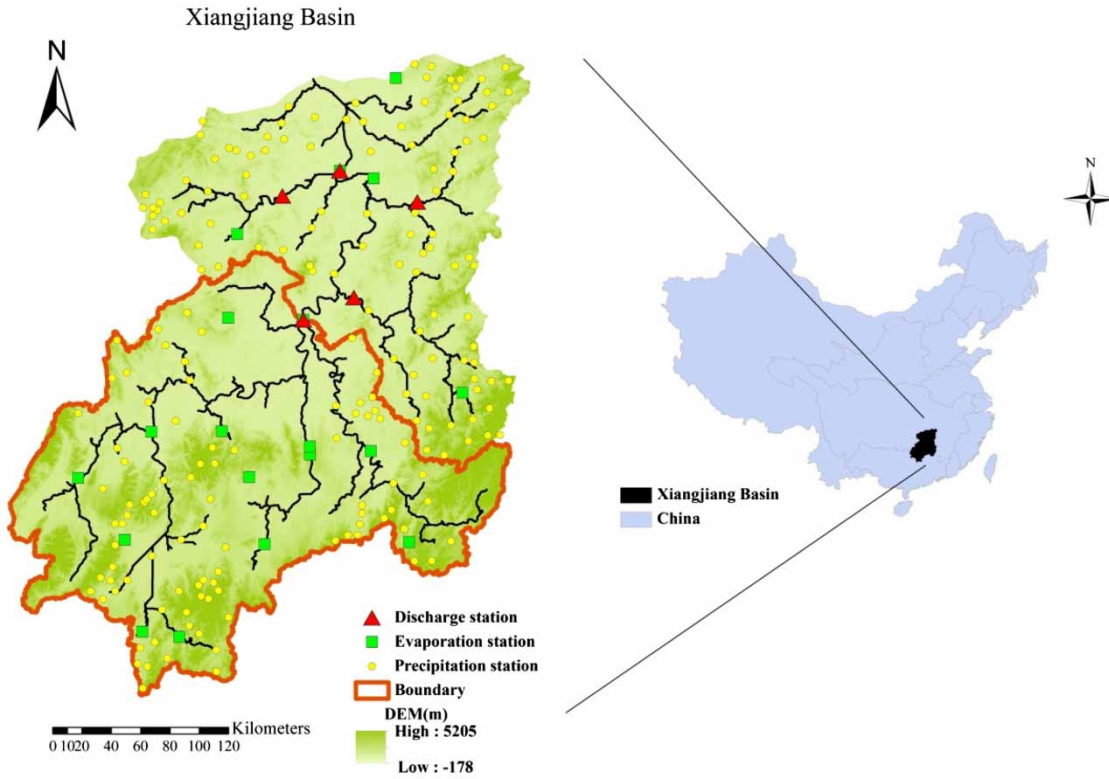


Figure 1 | Distribution diagram of runoff station, evaporation station and precipitation station of Xiangjiang basin.

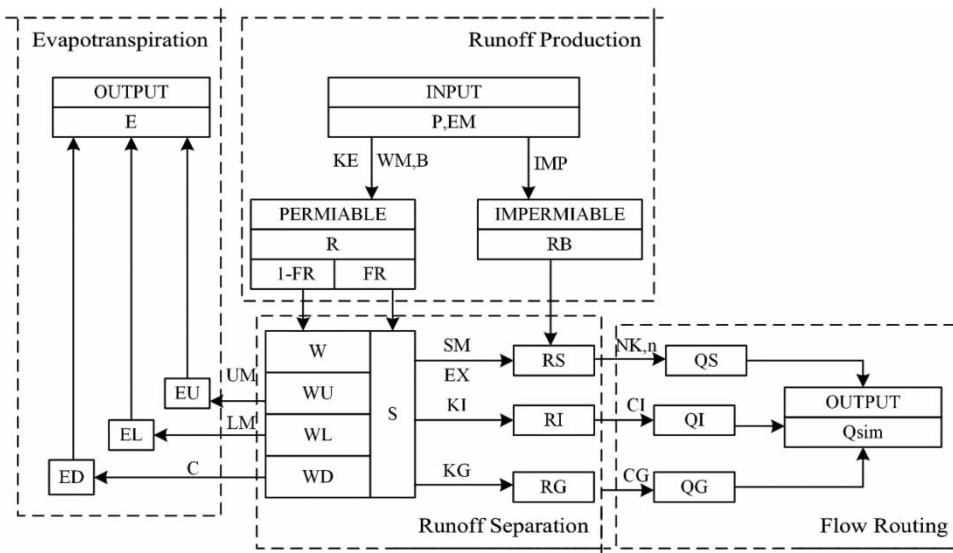


Figure 2 | Flowchart of the XAJ model.

evaporation capacity (EP-P), evaporation from the lower layer occurs at the rate of the remaining evaporation capacity (EP-P-EU) multiplied by the ratio between the

lower layer water storage and its water storage capacity when this ratio is larger than parameter C. Otherwise the whole water stored in the lower layer is evaporated, and

Table 1 | Parameters of XAJ model

Number	Parameter	Explanation	Range	Unit
1	KE	Ratio of potential evapotranspiration to pan evaporation	0.5–1.5	–
2	WM	Areal mean tension water storage capacity	–	mm
3	UM	Upper layer tension water storage capacity	–	mm
4	LM	Lower layer tension water storage capacity	–	mm
5	B	Tension water distribution index	–	–
6	IMP	Impermeable coefficient	–	–
7	SM	Areal mean free water storage capacity	0–100	mm
8	EX	Free water distribution index	–	–
9	KI	Out flow coefficient of free water storage to interflow	–	–
10	KG	Out flow coefficient of free water storage to groundwater flow	0.1–0.7	–
11	C	Deep layer evapotranspiration coefficient	–	–
12	CI	Interflow recession coefficient	0.5–1.0	–
13	CG	Groundwater recession coefficient	0.5–1.0	–
14	n	Parameter of Nash unit hydrograph	–	–
15	NK	Parameter of Nash unit hydrograph	2.0–3.0	–

the evaporation from the deep layer occurs at the rate of the remaining evaporation capacity (EP-P-EU) multiplied by parameter *C* when water stored in the deep layer is sufficient enough or else all the water in deep layer is evaporated.

When the precipitation is larger than the potential evapotranspiration, runoff is generated in those areas whose soil water content reaches field capacity. The amount of runoff is derived according to the rainfall and soil storage deficit. The total runoff is then treated as the input to a free water reservoir whose storage capacity is non-uniformly distributed over the area. And its distribution is described by the parameter EX. Through this free water reservoir, the total runoff is subdivided into three components, the surface runoff, RS, the interflow, RI, and the groundwater runoff, RG. After that the surface runoff is routed to the outlet of the study area through a Nash-cascade model with parameters *n* and NK, while the interflow and the groundwater runoff is routed through single linear reservoirs with recession coefficients CI and CG, respectively. The sum of these three routing results is the simulated runoff of the XAJ model.

HBV model

For comparison purpose, another widely used deterministic lumped conceptual model, the HBV model, which

was developed by the Swedish Meteorological and Hydrological Institute in early 1970s, is used in this study. The version of the model used in this study is HBV light (Seibert 1997). The snow routine of HBV model is not considered in this study because snowfall is rare in the study region. The flowchart of this model is shown in Figure 3. Input data of this model are daily precipitation, *P*, and measured pan evaporation, EM. The output data include daily actual evapotranspiration, *E*, and runoff at the outlet of the study basin, Q_{sim} . Model parameters are in bold in Figure 3, and the meanings of these parameters are listed in Table 2.

In the HBV model, actual evaporation equals to the input evaporation if the ratio between soil moisture storage, SM, and field capacity, FC, is larger than parameter LP. Otherwise, a linear reduction is used to calculate the actual evaporation (Figure 3). The input precipitation *P* is divided into soil moisture storage and groundwater recharge according to the value of SM/FC. Groundwater recharge is then added to the upper groundwater box whose water storage is denoted as SUZ. The upper groundwater box has three outlets, the fast flow outlet (Q_0), the interflow outlet (Q_1) and percolation (PERC) to the lower groundwater box. Q_0 occurs only when SUZ is larger than a threshold UZL. The lower groundwater box has only one linear

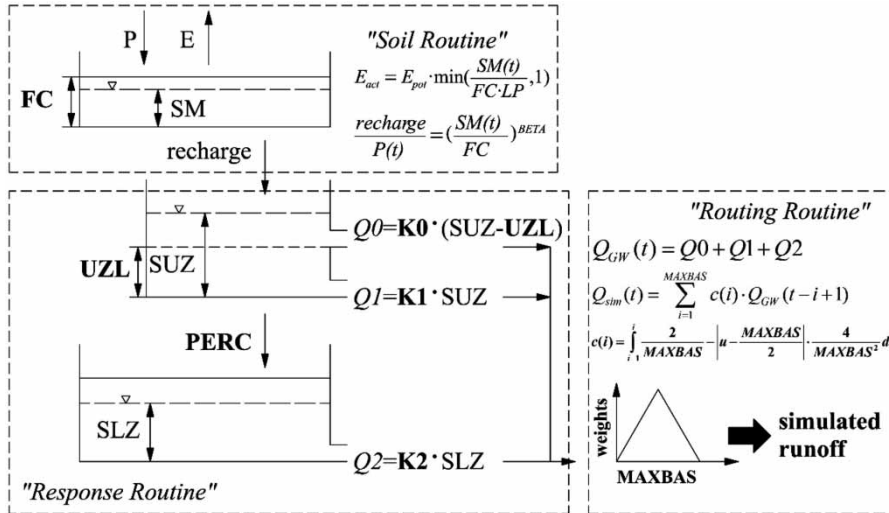


Figure 3 | Flowchart of the HBV model.

Table 2 | Parameters of HBV model

Number	Parameter	Explanation	Range	Unit
1	FC	Maximum soil moisture storage	100–400	mm
2	LP	Soil moisture value above which actual ET reaches potential ET	0.5–1.0	–
3	BETA	Recharge-soil moisture curve coefficient	1.0–5.0	–
4	PERC	Maximum flow from upper to lower response box	0–4.0	mm/day
5	UZL	Threshold value for upper response box to runoff from K0	0–70	mm
6	K0	Recession coefficient	0.1–0.5	day ⁻¹
7	K1	Recession coefficient	0.01–0.5	day ⁻¹
8	K2	Recession coefficient	5E-05-0.1	day ⁻¹
9	MAXBAS	Parameter of triangular weighting function	3.0–6.0	day

outflow, $Q2$. The sum of the three groundwater outflows, $Q0$, $Q1$, and $Q2$ is then transformed by a triangular weighting function defined by the parameter MAXBAS to derive the simulated runoff.

Model calibration method

The genetic algorithm is used to optimize the 15 free parameters of the XAJ model and the nine free parameters of the HBV model. For both models, the commonly used Nash–Sutcliffe coefficient (NS) and the relative error (RE) are used to evaluate the performance of the models. The

functions are expressed as follows:

$$NS = 1 - \frac{\sum (Q_{obs}^t - Q_{sim}^t)^2}{\sum (Q_{obs}^t - \bar{Q}_{obs})^2} \quad (1)$$

$$RE = \frac{\sum (Q_{obs}^t - Q_{sim}^t)}{\sum Q_{obs}^t} \quad (2)$$

where Q_{obs}^t , Q_{sim}^t , and \bar{Q}_{obs} are observed, simulated and the mean of the observed discharge series, respectively.

The split-sample test (Klemeš 1986) is used to assess the performance of both models in simulating the rainfall-runoff relationship of the study area. In this test, these two models

are calibrated and validated by using the 41-year historical runoff series of Hengyang runoff station. The 41-year historical runoff records are split into two sub-records, 1965–1985 and 1986–2005, for calibration and validation, respectively. When these models are used for design flood estimation, the 41-year observed data are all used for model calibration to achieve the best runoff simulation.

To evaluate the influence of using different model calibration period and calibration length on the results of design flood estimation, the 41 years of historical runoff records are separated into several sub-records (overlap exists) with the length of 2 years (40 sub-records), 5 years (37 sub-records), 10 years (32 sub-records), 15 years (27 sub-records) and 20 years (22 sub-records). Model parameters are calibrated based on these sub-records using genetic algorithm. Then the calibrated models are used to simulate the runoff series of the whole period, 1965–2005 for comparison.

Pearson Type III distribution and L-moments approach

Annual maximum runoff series are selected from the 41-year daily runoff observations and model simulations to estimate design flood for a given return period. These annual maximum runoff series are analysed by fitting the Pearson Type III distribution. Pearson Type III distribution was first applied in hydrology by Foster (1924) to describe the probability distribution of annual maximum flood peaks (Chow et al. 1988). This distribution is recommended for official use for frequency analysis of annual maximum floods in China, is well tested and widely used across the country for design flood estimation (Hua 1987). The L-Moments approach as defined by Hosking (1990), is used to derive the three statistical parameters (the arithmetic mean E_x , the coefficient of variance C_v , and the coefficient of skewness C_s) of the annual maximum floods. Comparing with conventional moments, L-moments are more robust to the presence of outliers and can make more accurate inferences when the sample volume is small.

The GLUE method

Resulting from the uncertainties that are inherent in the observed data, hydrological model parameters, model structure and the difference of model calibration periods, the design flood estimation will unavoidably have some

uncertainty. The uncertainty caused by model parameter equifinality is evaluated by the GLUE method. This method was developed by Beven & Binley (1992). The main idea of this method is that there is no optimal model structure and model parameter set that can perfectly represent a river basin. Instead, this method keeps the idea that there are many combinations of parameter values that can reproduce observed runoff of a given river basin with the same efficiency (Freer et al. 1996; Beven & Freer 2001).

General steps of the GLUE method are as follows: Initially, a large number of model runs are conducted for a particular model with many model parameter sets randomly selected from subjectively determined parameter space with priori probability distribution. As no information is available about the parameter space, uniform distribution is chosen to be the priori distribution. Then the likelihood value of each model run is calculated by comparing the model simulations and observations. Higher likelihood values indicate better model simulation. After that, a cut-off threshold value is subjectively chosen. Model runs with likelihood values less than the cut-off threshold value are considered to be ‘non-behavioural’ and will not be considered in the further analysis. Finally, the 95% confidence interval of the runoff simulations can be drawn from the remaining ‘behavioural’ model simulations regarding their likelihood values as relative weights. The annual maximum discharge series are selected from the ‘behavioural’ model simulations. Then Pearson Type III distribution is fitted to each series, and the 95% confidence intervals of the design floods of different return periods are derived considering the likelihood function value as weighting factor.

The choice of likelihood function has been discussed extensively (Beven & Binley 1992; Romanowicz et al. 1994) since the outcome of the GLUE method. In this study, the standardized Nash–Sutcliffe value is chosen to be the likelihood measure.

Two indices are used to give a quantitative evaluation of the GLUE simulation, the percentage of observations that are contained in the 95% confidence interval (P-95CI) (Li et al. 2009), and the average relative interval length (ARIL) proposed by Jin et al. (2010). These two indices are expressed as follows:

$$P - 95CI = \frac{NQ_{in}}{N} \times 100\% \quad (3)$$

Table 3 | Model performance by split-sample test

Model	Calibration			Validation			Simulation		
	Period	NS	RE (%)	Period	NS	RE (%)	Period	NS	RE (%)
XAJ	1965–1984	0.92	−0.33	1985–2005	0.72	20.58	1965–2005	0.83	10.59
	1985–2005	0.90	0.11	1965–1984	0.79	−24.12			
HBV	1965–1984	0.91	1.49	1985–2005	0.76	13.96	1965–2005	0.83	−14.06
	1985–2005	0.88	0.09	1965–1984	0.79	−29.51			

where NQ_{in} is the number of observations contained in the 95% confidence interval; N is the total number of observations.

$$ARIL = \frac{1}{n} \sum \frac{\text{Limit}_{\text{Upper},t} - \text{Limit}_{\text{Lower},t}}{R_{\text{obs},t}} \quad (4)$$

where $\text{Limit}_{\text{Upper},t}$ and $\text{Limit}_{\text{Lower},t}$ are the upper and lower boundary of the 95% confidence interval, respectively, and $R_{\text{obs},t}$ is the runoff observation.

RESULTS AND DISCUSSION

Model calibration results

Table 3 summarizes the split-sample test results. It is seen from Table 3 that both models have excellent performance

in this study area in the calibration period. Somehow reduced performance is seen for both models in the validation period in terms of NS and RE values. Considering the total simulation period, the Nash–Sutcliffe value is larger than 0.83 for both models, and the absolute value of RE of XAJ model is less than 11.56% and that of HBV model is less than 14.06%. This indicates that both models performed equally well in representing the flow characteristics of the study area.

When the 41-year observed data are all used for model calibration, the Nash–Sutcliffe value of XAJ model is 0.88 and that of the HBV model is 0.87; the RE of the XAJ model is 0.01%, and that of the HBV model is 5.23%. This indicates that both models are well calibrated and give good simulation results. As the runoff simulations are used for flood frequency analysis, the daily flow duration curves of the runoff observations and runoff simulations are

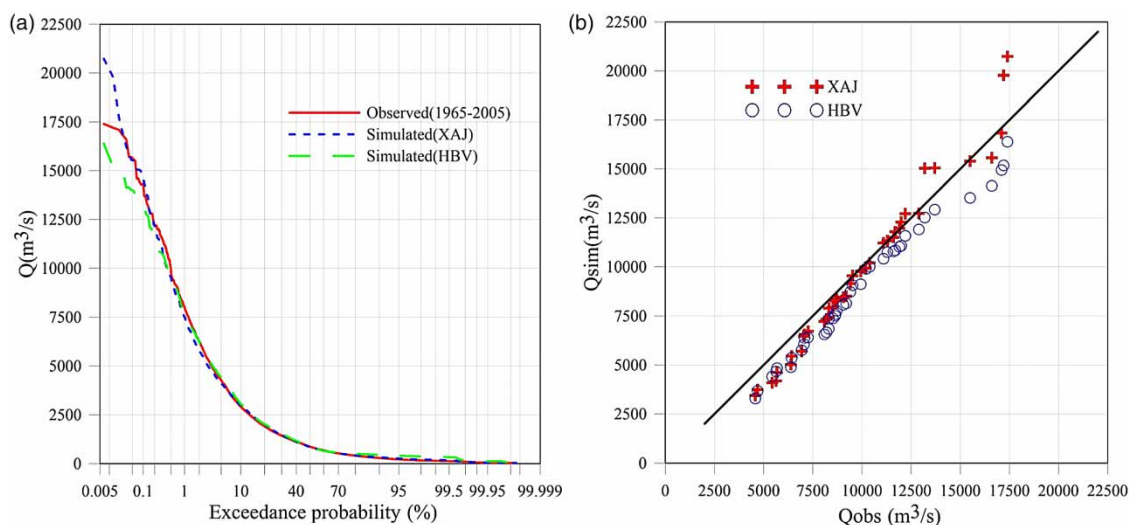


Figure 4 | (a) Daily flow duration curve for observed, XAJ model simulated and HBV model simulated runoff series of the upper Xiangjiang River at Hengyang runoff station. The abscissa is in probability scale to clearly demonstrate floods with lower exceedance probability. (b) The deviation of simulated annual maximum flood peaks to the observed ones. Points over the solid line indicate overestimation and vice versa.

drawn and shown in Figure 4(a). As the annual maximum runoff series are analysed to estimate design flood, special attention was given to the simulation of the annual flood peaks. The 41 simulated annual flood peaks were selected and plotted against the corresponding observed values to evaluate the deviation of the simulated flood peaks. The result is shown in Figure 4(b). It is seen that the difference between the flow duration curve of the runoff observations and that of the runoff simulations is small except in the left hand tail, as expected. The mismatches to the extreme left are mainly because of the imperfection of the models in simulating floods with high return periods as well as the errors in the model input data. Figure 4(b) further shows that the XAJ model is likely to overestimate high flood peaks but underestimate low flood peaks, and that HBV model underestimates both high flood peaks and low flood

peaks. The impact of these imperfect simulations of peak floods on the estimation of design flood is analysed in the next section.

Design flood estimation

Annual maximum runoff series of observations and simulations are selected. Their statistical parameters, the arithmetic mean E_X , the coefficient of variance C_v and the coefficient of skewness C_s , are calculated using the L-moments approach. The statistical parameters of observed runoff are chosen to be the reference values. The results are listed in Table 4, which shows that the mean value of annual flood peaks is underestimated by both models. The coefficient of variance is overestimated by both models due to the errors that are inherent in the simulated peaks. The C_s value is overestimated by XAJ model, and underestimated by HBV model. Just from these three statistical parameters, it seems XAJ model gives better simulation of annual flood peaks than HBV model.

Parameters of Pearson Type III distribution are calculated based on these statistical parameters. Then this distribution is fitted to the annual maximum runoff series. The results are shown in Figure 5(a). The estimates of design flood of different return periods are listed in Table 5 and shown in Figure 5(b) to give a clear comparison.

Table 4 | Statistic parameters of simulated and observed 41-year annual flood peak series

Data	Mean value (m^3/s)	Coefficient of variance value	Coefficient of skewness value
Observations (1965–2005)	9,944	0.359	0.883
Simulations (XAJ)	9,713	0.448	0.935
Simulations (HBV)	8,886	0.389	0.574

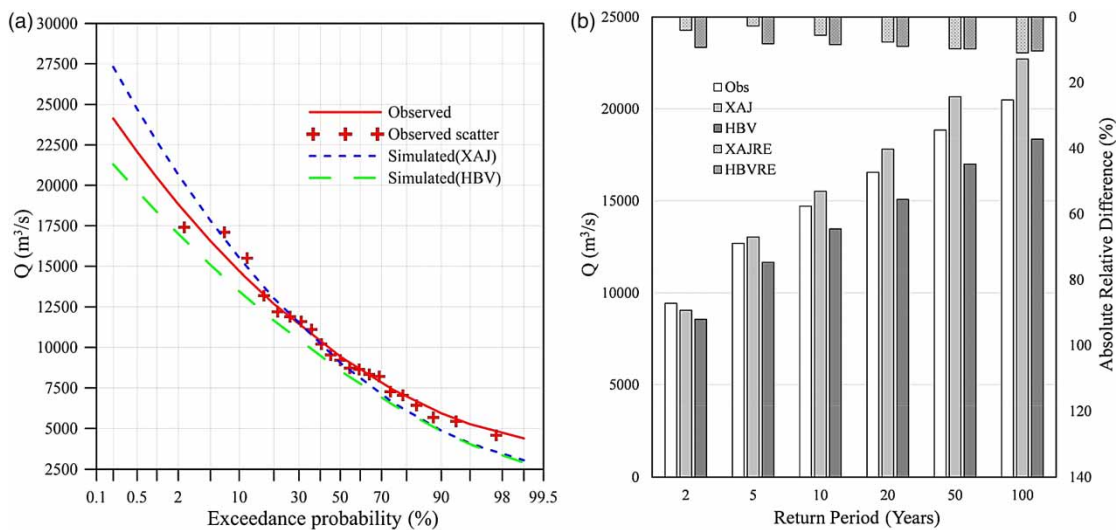


Figure 5 | (a) Frequency curves derived from runoff observations and simulations. (b) Design flow estimates and corresponding RE in percentage (absolute value) with return period $T = 2, 5, 10, 20, 50$ and 100 years.

Table 5 | Design flood estimations of different return periods from runoff observations and model simulations

Return period (Years)	Design flood from observation (m ³ /s)	Design flood from XAJ simulation		Design flood from HBV simulation	
		Value (m ³ /s)	Difference (%)	Value (m ³ /s)	Difference (%)
2	9,431	9,051	-4.0	8,558	-9.3
5	12,685	13,027	2.7	11,652	-8.1
10	14,706	15,523	5.6	13,468	-8.4
20	16,554	17,816	7.6	15,075	-8.9
50	18,837	20,661	9.7	17,004	-9.7
100	20,482	22,719	10.9	18,362	-10.4

Since the length of available data is 41 years, design flood estimates are presented up to 100 years only.

Figure 5(a) reveals that the XAJ model is likely to overestimate the design floods of long return period, while HBV model underestimates them, though both models are well calibrated based on the same hydrometeorological data by genetic algorithms. Figure 5(a) also shows that the slope of the frequency curve derived from XAJ model simulation is larger than the other two curves, and that the frequency curve derived from HBV model simulation lies under the frequency curve of observed annual maximum runoff series though their slopes are similar. The reason is that, for the XAJ model, though it gives good estimation of the mean value of annual flood peaks, it overestimates the C_v value; for the HBV model, it gives a better estimation of C_v , however it underestimates mean value (Table 4). Figure 5(b) and Table 5 show that, for a low return period, the difference between the design flood estimates of the observed series and that of the simulated series is small, while this difference increases with the increase of return period. Figure 5(b) and Table 5 also demonstrate that the absolute relative difference of design floods estimated by the XAJ model is less than that estimated by HBV model, especially for a low return period. Table 5 shows that, for both models, the absolute values of relative difference of the design flood estimate increases with the increase of return period, and that their maximum value is less than 11%. The difference between the design flood estimates derived from runoff observations and simulations is an illustration of the uncertainties that are inherent in the continuous simulation approach. These uncertainties are further analyzed in the next section.

Uncertainty caused by model parameter equifinality

No 'optimal' parameter set could be found in hydrological modelling practice, and many parameter sets may result in similar results, which is the well known equifinality problem in hydrological modelling. In this study, the effect of the equifinality problem on flow simulation as well as on the design flood estimation is studied by using the GLUE method. The likelihood threshold value required in the GLUE method is subjectively determined to be 0.8. To reduce the number of total model runs, only sensitive parameters (KE, SM, KG, CI, CG, and NK) of XAJ model are considered to be free parameters, other parameters retain their optimized values derived in the section, 'Model calibration results'. In this study area, it is unnecessary to consider snow routine of HBV model, so the number of free parameters of HBV model reduces to 9. The ranges of free parameter for both models when applying the GLUE method are listed in Tables 1 and 2, and 100,000 times of model runs are conducted. After selection, 12,849 times of XAJ model runs and 14,258 times of HBV model runs are considered to be 'behavioural' and are used to derive the 95% confidence interval of runoff simulation and that of design flood estimation considering the likelihood value to be the weighting factor. For illustrative purposes, the 95% confidence intervals of the flood year (1978), the normal year (1987), and the dry year (1979) are plotted in Figure 6.

Figure 6 shows that, for both models, 95% confidence interval contains most of the runoff observations and no significant difference between the HBV model simulation and XAJ simulation can be found. Table 6 lists the values of indices used for assessing the 95% confidence interval. It also shows that the P-95CI and ARIL of both models are rather close, with HBV model having slightly higher values of P-95CI.

The 95% confidence intervals of the design floods estimation of both models are shown in Figure 7 and the relative interval lengths of the 95% confidence intervals are listed in Table 7. Figure 7 shows that for XAJ model, the confidence interval is almost symmetrically distributed around the frequency curve of the runoff observations which is quite close to the interval median. However, for the HBV model, most of the confidence interval is beneath the frequency curve of the runoff observations, which indicates that the HBV model is likely to underestimate the

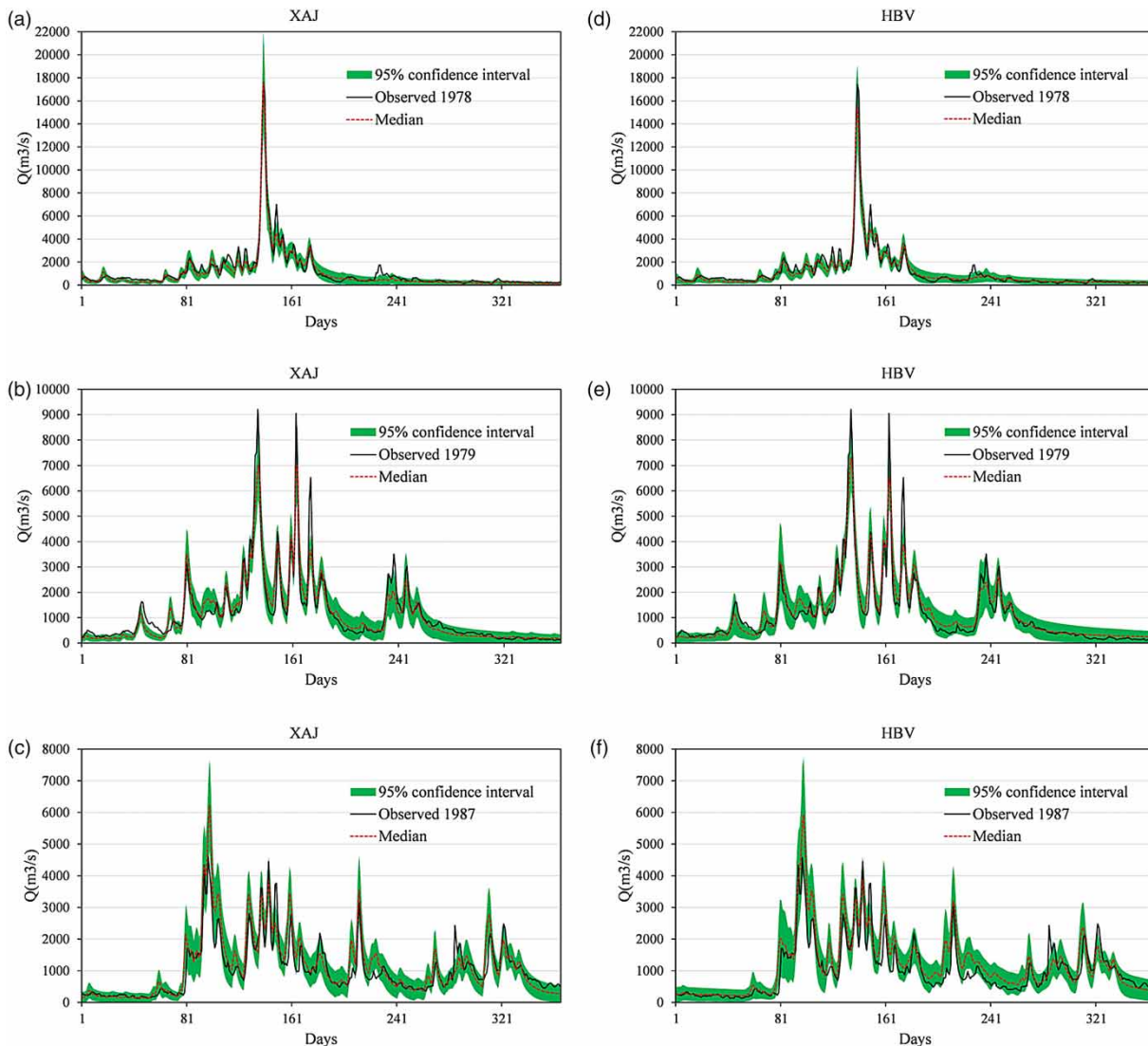


Figure 6 | 95% confidence interval of simulated runoff series in flood year 1978, normal year 1979 and dry year 1987: (a), (b), (c) using XAJ model; (d), (e), (f) using HBV model.

Table 6 | Indices for assessing the 95% runoff confidence interval

Indices	Period	Models	
		XAJ	HBV
P-95CI (%)	1965–2005	79.84	81.22
	Flood (1978)	83.84	87.12
	Dry (1979)	83.29	90.41
	Normal (1987)	84.93	87.12
ARIL	1965–2005	1.07	1.14
	Flood (1978)	1.13	1.29
	Dry (1979)	1.17	1.43
	Normal (1987)	1.24	1.12

hydrological extreme values in this study basin. Table 7 shows that considering the same return period, the XAJ model gives a wider 95% confidence interval though there are less ‘behavioural’ model simulations of the XAJ model. It also shows that for both models the relative interval length is quite large even for the return period of just 2 years, which reveals that the uncertainty of model parameters and model structure have a great impact on the results of design flood estimation by continuous simulation.

Figure 8 shows the 95% confidence intervals of different return periods. It reveals that the interval length increases

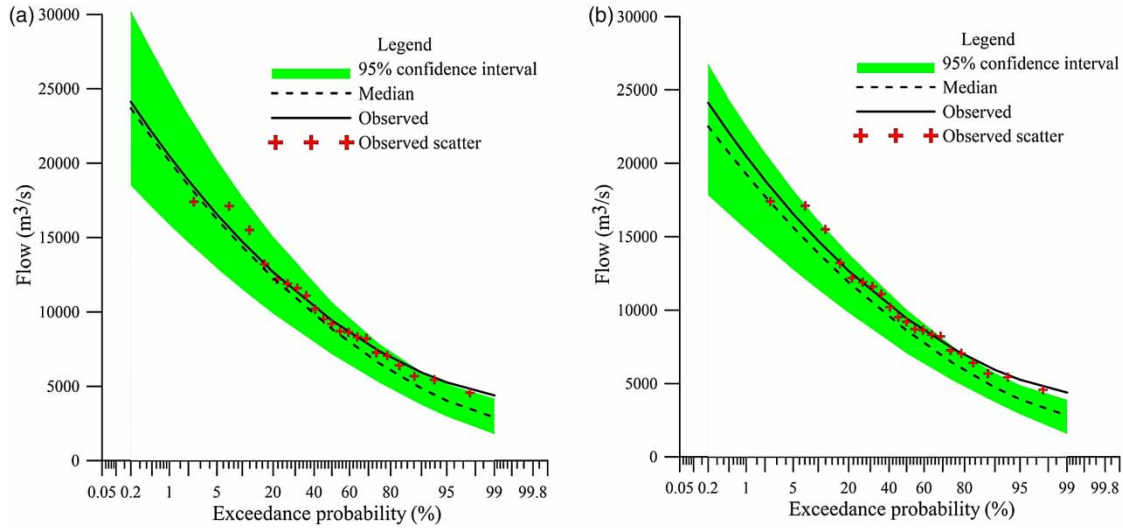


Figure 7 | The 95% confidence interval of frequency curve derived from runoff simulations (1965 to 2005). (a) Simulations of the XAJ model; (b) simulations of the HBV model.

Table 7 | Relative length of 95% confidence interval of frequency curves derived from XAJ model simulations and HBV model simulations

Return period (years)		2	5	10	20	50	100
Relative interval length (%)	XAJ	37.1	40.7	42.3	43.9	45.6	46.7
	HBV	31.5	31.6	32.0	32.7	33.7	34.5

Relative interval length = (upper bound - lower bound) / estimation from observed runoff series.

with the increase of the return period and for each return period the interval length derived from the XAJ model simulation is larger than that derived from HBV simulation. In

summary, no matter how well the model parameters are calibrated, the frequency curve derived from the deterministic model simulations is only a single realization of the process to be studied and not reliable due to the significant influence of the model equifinality.

Uncertainty caused by using different model calibration period and data length

Figure 9 demonstrates the frequency curves derived from the runoff observations and runoff simulations with different

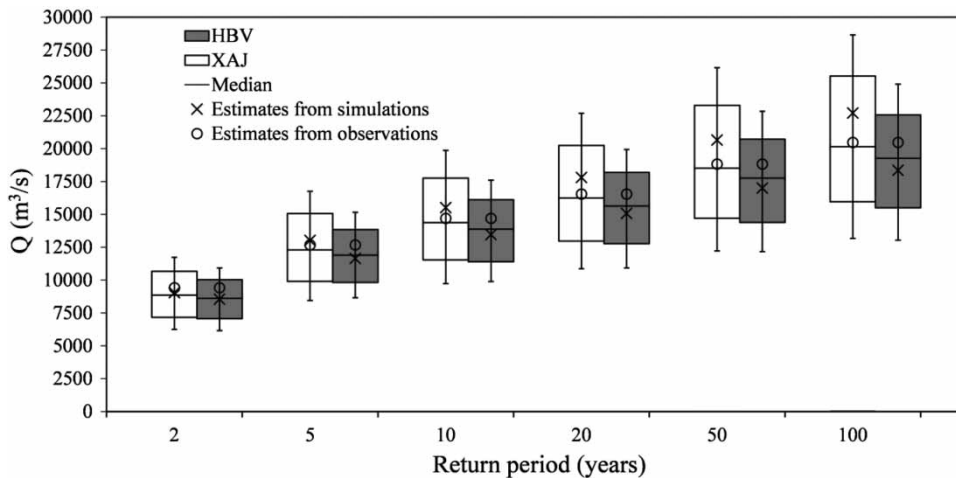


Figure 8 | Boxplots of 95% confidence intervals of different return periods. Cross shows the design flood estimates derived from the model simulations using the whole data (1965–2005) to do model calibration without GLUE approach.

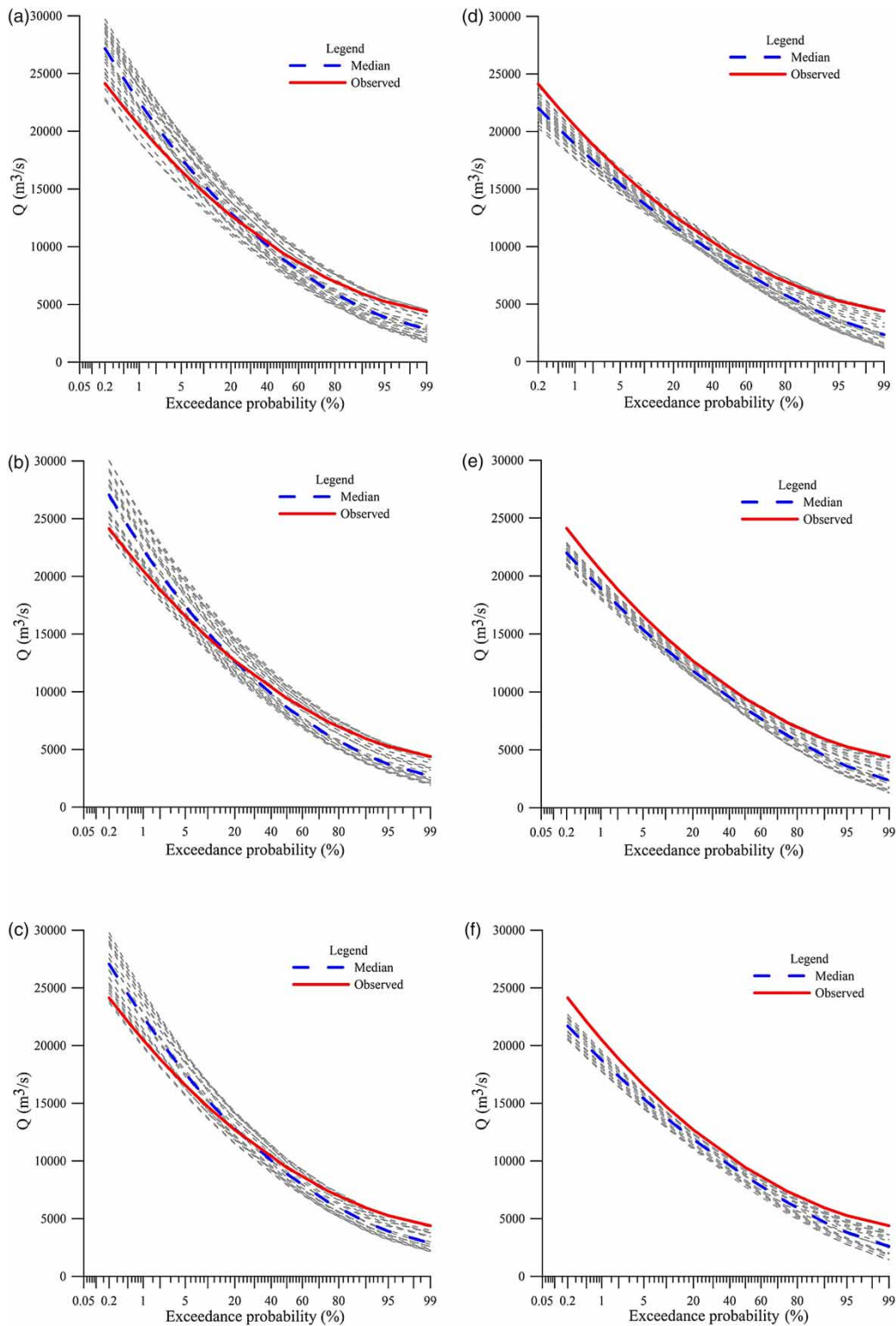


Figure 9 | Frequency curves derived from observations and model simulations with different model calibration periods and calibration length. The gray short dash line is the frequency curves from model simulations; the thick long dash line is their median; the continuous line is the frequency curve from observations: (a), (b) and (c) are the 10-, 15- and 20-year calibration lengths for the XAJ model, respectively; (d), (e) and (f) are the 10-, 15- and 20-year calibration lengths for the HBV model, respectively.

Table 8 | Relative width of spread band of frequency curves derived from XAJ model simulations and HBV model simulations with different model calibration periods

Model	XAJ					HBV					
	CL (years)	2	5	10	15	20	2	5	10	15	20
RP (years)	Relative band width (%)										
2	47.9	42.7	35.7	32.0	24.1	42.9	28.0	21.7	17.4	16.0	
5	46.1	39.6	31.5	28.8	22.0	37.9	23.7	15.6	11.0	10.8	
10	46.1	39.1	29.9	27.8	21.8	36.3	23.4	14.5	9.7	9.6	
20	46.4	39.4	28.8	27.3	22.1	35.5	23.5	14.0	8.9	9.0	
50	46.9	40.1	28.2	26.9	23.0	34.8	23.9	13.7	8.5	8.7	
100	47.4	40.7	28.4	26.8	23.7	34.6	24.2	14.0	8.4	8.8	

Relative band width = (upper bound–lower bound)/estimation from observed runoff series. CL is calibration length, RP is return period.

model calibration periods and calibration data length. It is seen from Figure 9 that a wide spread of design flood estimates is produced by using the different calibration periods and data lengths. With the same calibration data length, the spread of design flood estimates of XAJ model is larger than that of the HBV model, which indicates XAJ model is more sensitive to the difference of the calibration period when used for design flood estimation by continuous simulation. It should also be noted that the XAJ model is likely to overestimate the design floods of longer return period while HBV model will underestimate the design floods. Figure 9 also shows that the value of exceedance probability has no significant influence on the band width compared with Figure 7, though the band is relatively a little thinner in the middle and wider in the tails. This means the uncertainty of model parameters has a greater influence on the value of design flood estimation with a high return period but less influence on the design flood estimation of low return period. However, the influence of the difference of model calibration period on design flood estimation does not significantly vary with the return period of design flood. This phenomenon can be quantitatively demonstrated in Table 8.

It is seen from Table 8 that the relative band widths derived from XAJ model simulations are larger than that from HBV model simulations. Comparing Table 8 and Table 7, it is clear that for a large return period, the uncertainty caused by model parameter equifinality is more significant than the uncertainty caused by using different

model calibration periods and data length, when applying hydrological models to design flood estimation by continuous simulation.

The change of band width of design flood estimation with data length of model calibration is visually demonstrated by Figure 10. It shows that the band width of design flood with a certain return period decreases with the increase of model calibration data length, which indicates the increase of model calibration data length will reduce the uncertainty caused by the difference of model calibration period. However, when the model calibration length is larger than a certain value (between 10 to 15 years), the further increase of calibration data length will not result in a significant reduction of the uncertainty caused by using different model calibration periods, especially in the HBV model with less parameters. For the study area, 15–20 years' calibration data length for the conceptual hydrological models seems necessary and sufficient, and is recommended in design flood estimation by continuous simulation to minimize the uncertainty caused by the difference of model calibration period.

The above analyses demonstrate that the part of the data used for model calibration has significant influence on further design flood estimation, though it is not as significant as the uncertainty caused by model parameter equifinality. And once the hydrological model is well calibrated the further increase of calibration period length beyond a certain threshold has little improvement on design flood estimation.

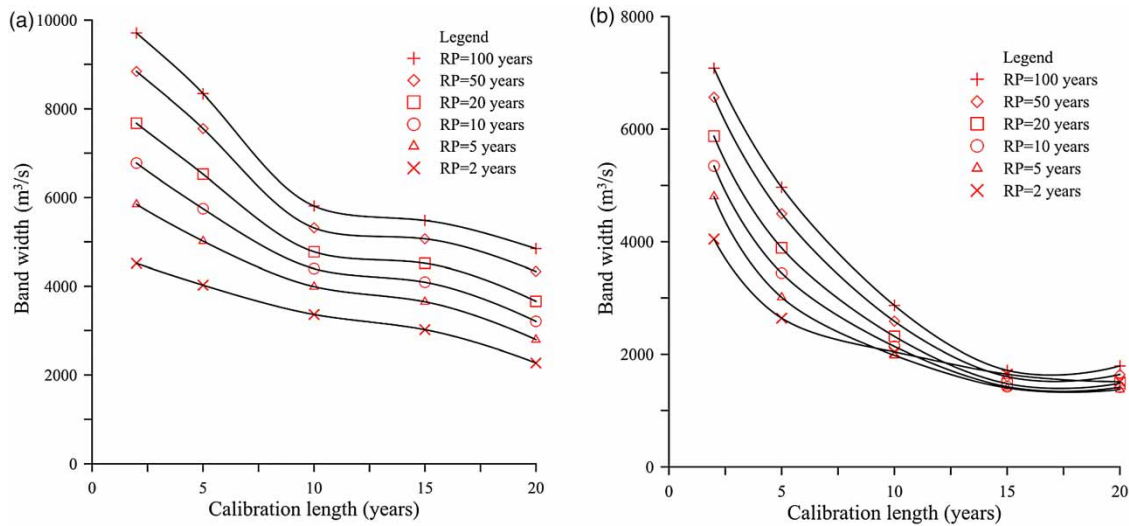


Figure 10 | Variation of design flood band width with model calibration length. (a) XAJ model, (b) HBV model. Band width = maximum estimation–minimum estimation of design flood with a certain return period (RP).

CONCLUSIONS

In this study, two hydrological models are used for the design flood estimation by continuous simulation. Uncertainty that is inherent in this methodology caused by model parameter equifinality is further analysed by using the GLUE method. Uncertainty caused by using different model calibration periods and calibration data length is also assessed. The study was conducted in a humid catchment in southern China. The following conclusions are drawn from the study.

Both models can give design flood estimations similar to those derived from runoff observations with maximum absolute RE of 11% for the design flood with a return period less than 100 years.

The large width of the 95% confidence interval of the flood frequency curve indicates significant uncertainty caused by model parameter equifinality. This means no matter how well the model parameters are calibrated, the frequency curve derived from the deterministic model simulations is only a single realisation of the possible values and is not reliable due to the significant influence of the model equifinality.

Different model calibration periods generate a wide spread of design flood estimation and that the error band width varies little with the exceedance probability, which

indicates that the uncertainty induced by the difference in the model calibration period will have significant influence on the estimation of design flood of all return periods. The study also shows that once the model calibration data length is long enough for model calibration, the further increase of calibration data length has little improvement on design flood estimation.

This study systematically investigates the applicability of hydrological models to design flood estimation by continuous simulation. The results show that great uncertainty exists in this methodology. When applying this method to design flood estimation, the uncertainty should be treated seriously, or significant error may exist in the estimation results. However, this study is only a preliminary study because only two deterministic lumped models are used and only one runoff station in a humid region is considered. More studies need to be done on other river basins using other hydrologic models to draw a generalised conclusions and guidance for design flood estimations using hydrological models.

ACKNOWLEDGEMENTS

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

REFERENCES

- Arnell, N. W. 1999 A simple water balance model for the simulation of streamflow over a large geographic domain. *J. Hydrol.* **217** (3), 314–335.
- Bastola, S., Murphy, C. & Sweeney, J. 2011 The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. *Adv. Water Resour.* **34**, 562–576.
- Beven, K. 1987 Towards the use of catchment geomorphology in flood frequency predictions. *Earth Surf. Proc. Land.* **12**, 69–82.
- Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Proc.* **6**, 279–298.
- Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **249**, 11–29.
- Blazkova, S. & Beven, K. 2002 Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty). *Water Resour. Res.* **38**, 14–11.
- Blazkova, S. & Beven, K. 2004 Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic. *J. Hydrol.* **292**, 153–172.
- Boughton, W. C. & Droop, O. 2003 Continuous simulation for design flood estimation – a review. *Environ. Modell. Softw.* **18**, 309–318.
- Boughton, W. C. & Hill, P. I. 1997 *A Design Flood Estimation Procedure Using Data Generation and Daily Water Balance Model*. CRC for Catchment Hydrology, Clayton, Vic.
- Bradley, A. A. & Potter, K. W. 1992 Flood frequency analysis of simulated flows. *Water Resour. Res.* **28**, 2375–2385.
- Brigode, P., Bernardara, P., Paquet, E., Gailhard, J., Garavaglia, F., Merz, R. & Ribstein, P. 2014 Sensitivity analysis of SCHADEX extreme flood estimations to observed hydrometeorological variability. *Water Resour. Res.* **50**, 353–370.
- Calver, A. & Lamb, R. 1995 Flood frequency estimation using continuous rainfall-runoff modelling. *Phys. Chem. Earth* **20**, 479–483.
- Cameron, D. S., Beven, K. J., Tawn, J., Blazkova, S. & Naden, P. 1999 Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). *J. Hydrol.* **219**, 169–187.
- Cameron, D., Beven, K., Tawn, J. & Naden, P. 2000 Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation). *Hydrol. Earth Syst. Sci.* **4**, 23–34.
- Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology*. McGraw-Hill, New York.
- 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.
- Foster, H. A. 1924 Theoretical frequency curves and their application to engineering problems. *Trans. Am. Soc. Civ. Eng.* **87** (1), 142–173.
- Freer, J., Beven, K. & Ambrose, B. 1996 Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resour. Res.* **32**, 2161–2173.
- Gosling, R. 2014 Assessing the impact of projected climate change on drought vulnerability in Scotland. *Hydrol. Res.* **45**, 806–886.
- 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.
- Hosking, J. R. M. 1990 L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc. B Met.* **1**, 105–124.
- Hua, S. Q. 1987 A general survey of flood-frequency analysis in China. *J. Hydrol.* **96**, 15–25.
- Jin, X., Xu, C. Y., Zhang, Q. & Singh, V. P. 2010 Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *J. Hydrol.* **383**, 147–155.
- 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.
- Klemeš, V. 1986 Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **31**, 13–24.
- Lamb, R. 1999 Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. *Water Resour. Res.* **35**, 3103–3114.
- Lawrence, D. & Haddeland, I. 2011 Uncertainty in hydrological modelling of climate change impacts in four Norwegian catchments. *Hydrol. Res.* **42**, 457–471.
- Li, L., Xia, J., Xu, C. Y., Chu, J. T. & Wang, R. 2009 Analyse the sources of equifinality in hydrological model using glue methodology. Hydroinformatics in hydrology, hydrogeology and water resources. In: *Proceedings of Symposium JS.4 at the Joint Convention of the International Association of Hydrological Sciences (IAHS) and the International Association of Hydrogeologists (IAH)*. Hyderabad, India, 6–12 September 2009.
- Li, L., Xia, J., Xu, C. Y. & Singh, V. P. 2010 Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. *J. Hydrol.* **390**, 210–221.
- Li, L., Ngongondo, C., Xu, C. Y. & Gong, L. 2013a 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, Z. J., Xin, P. L. & Tang, J. H. 2013b Study of the Xinanjiang model parameter calibration. *J. Hydrol. Eng.* **18**, 1513–1521.

- McIntyre, N., Lee, H., Wheeler, H., Young, A. & Wagener, T. 2005 Ensemble predictions of runoff in ungauged catchments. *Water Resour. Res.* **41**, W12434. Doi:10.1029/2005WR004289.
- 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.
- Murray, C. P. & Blöschl, G. 2011 Hydrological modelling in a changing world. *Prog. Phys. Geog.* **35**, 249–261.
- Mwale, F. D., Adeloye, A. J. & Rustum, R. 2014 Application of self-organising maps and multi-layer perceptron-artificial neural networks for streamflow and water level forecasting in data-poor catchments: the case of the Lower Shire floodplain, Malawi. *Hydrol. Res.* **45**, 838–854.
- Ngongondo, C., Li, L., Gong, L., Xu, C. Y. & Alemaw, B. F. 2013 Flood frequency under changing climate in the upper Kafue River basin, southern Africa: a large scale hydrological model application. *Stoch. Env. Res. Risk A.* **27**, 1883–1898.
- Paquet, E., Garavaglia, F., Garçon, R. & Gailhard, J. 2013 The SCHADEX method: a semi-continuous rainfall-runoff simulation for extreme flood estimation. *J. Hydrol.* **495**, 23–37.
- Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R. & Wheeler, H. S. 2011 Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST J.* **13**, 193–214.
- Refsgaard, J. C. & Storm, B. 1996 Construction, calibration and validation of hydrological models. In: *Distributed hydrological modelling* (Michael B. Abbott & Jens Christian Refsgaard, eds). Springer, The Netherlands, pp. 41–54.
- 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.
- Romanowicz, R., Beven, K. J. & Tawn, J. 1994 Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach. *Stat. Environ.* **2**, 297–317.
- Saghafian, B., Golian, S. & Ghasemi, A. 2014 Flood frequency analysis based on simulated peak runoffs. *Nat. Hazards* **71**, 403–417.
- Saliha, A. H., Awulachew, S. B., Cullmann, J. & Horlacher, H. B. 2011 Estimation of flow in ungauged catchments by coupling a hydrological model and neural networks: case study. *Hydrol. Res.* **42**, 386–400.
- Seibert, J. 1997 Estimation of parameter uncertainty in the HBV model. *Nord. Hydrol.* **28**, 247–262.
- Singh, H. V., Kalin, L., Morrison, A., Srivastava, P., Lockaby, G. & Pan, S. 2015 Post-validation of SWAT model in a coastal watershed for predicting land use/cover change impacts. *Hydrol. Res.* **46** (6), 837–853.
- Smithers, J., Schulze, R. & Kienzle, S. 1997 Design flood estimation using a modelling approach: a case study using the ACRU model. *IAHS Publications – Series of Proceedings and Reports – Intl Assoc. Hydrolog. Sci.* **240**, 365–376.
- Smithers, J. C., Chetty, K. T., Frezghi, M. S., Knoesen, D. M. & Tewolde, M. H. 2013 Development and assessment of a daily time-step continuous simulation modelling approach for design flood estimation at ungauged locations: ACRU model and Thukela Catchment case study. *Water SA* **39**, 467–475.
- Suman, A. & Bhattacharya, B. 2015 Flood characterisation of the Haor region of Bangladesh using flood index. *Hydrol. Res.* **46** (5), 824–835.
- Tian, Y., Xu, Y. P., Booij, M. J. & Cao, L. 2016 Impact assessment of multiple uncertainty sources on high flows under climate change. *Hydrol. Res.* **47** (1), 61–74.
- 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. 1999 Estimation of parameters of a conceptual water balance model for ungauged catchments. *Water Resour. Manag.* **13** (5), 353–368.
- 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. & 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.
- Xu, H. L., Xu, C. Y., Sælthun, N. R., Zhou, B. & Xu, Y. 2015a Evaluation of reanalysis and satellite-based precipitation datasets in driving hydrological models in a humid region of Southern China. *Stoch. Env. Res. Risk A.* **29**, 2003–2020.
- Xu, H. L., Xu, C. Y., Sælthun, N. R., Xu, Y., Zhou, B. & Chen, H. 2015b Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling – A case study of humid area in southern China. *J. Hydrol.* **525**, 138–151.
- 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.
- 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 Xinanjiang model. *Proc., Oxford Symposium on Hydrological Forecasting*. International Association of Hydrological Sciences, Wallingford, UK, 351–356.
- Zhao, R. J., Liu, X. R. & Singh, V. P. 1995 The Xinanjiang model. In: *Computer models of watershed hydrology* (Vijay P. Singh, ed.). Water Resources Publications, Highlands Ranch, Colorado, USA, pp. 215–232.