Water Resour Manage (2013) 27:1393–1409
DOI 10.1007/s11269-012-0244-5

Uncertainty Intercomparison of Different Hydrological Models in Simulating Extreme Flows

Xi Chen · Tao Yang · Xiaoyan Wang · Chong-Yu Xu · Zhongbo Yu

Received: 1 May 2012 / Accepted: 16 December 2012 /
Published online: 28 December 2012

Abstract A growing number of investigations on uncertainty quantification for hydrological models have been widely reported in past years. However, limited studies are found on uncertainty assessment in simulating streamflow extremes so far. This article presents an intercomparison of uncertainty assessment of three different well-known hydrological models in simulating extreme streamflows using the approach of generalized likelihood uncertainty estimation (GLUE). Results indicate that: (1) The three modified hydrological models can reproduce daily streamflow series with acceptable accuracy. When the threshold value used to select behavioral parameter sets is 0.7, XAJ model generates the best GLUE estimates in simulating daily flows. However, the percentage of observations contained in the calculated 95 % confidence intervals (P-95CI) is low (<50 %) when simulating the high-flow index (Q10). (2) Decreasing average relative length (ARIL), P-95CI and increasing average asymmetry degree (AAD) are found, when the threshold value increases for both daily-flows and high-flows. However, there is a significant inconsistence between sensitivity of daily-flows and high-flows to various threshold values of the likelihood function. Uncertainty sources from parameter sets, model structure and inputs collectively accounts for above sensitivity. (3) The best hydrological model in simulating daily-flows is not identical under different threshold values. High P-95CIs of GLUE estimate for high-flows (Q10 and Q25) indicate that TOPMODEL generally performs best under different threshold values, while XAJ model produces the smallest ARIL under different threshold values. The

X. Chen · T. Yang (✉)
State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, 818, Beijing Nanlu, Urumqi, Xinjiang 830011, People's Republic of China
e-mail: yang.tao@ms.xjb.ac.cn

T. Yang · X. Wang · Z. Yu
State Key Laboratory of Hydrology-Water Resources and Hydraulics Engineering, Hohai University, Nanjing 210098, China

C.-Y. Xu
Department of Geosciences, University of Oslo, P.O. Box 1047, Blindern 0316 Oslo, Norway

Z. Yu
Department of Geoscience, University of Nevada Las Vegas, Las Vegas, NV 89154-4010, USA

results are expected to contribute toward knowledge improvement on uncertainty behaviors in simulating streamflow extremes by a variety of hydrological models.

Keywords Hydrological model · Uncertainty · Intercomparison · GLUE method · Streamflow extremes · Threshold

1 Introduction

Hydrological models have been widely used to investigate many practical and pressing issues during planning, design, operation and management of water resources systems (Benke et al. 2007; Lin et al. 2007). Lumped precipitation-runoff models, wherein a catchment is treated as a single homogeneous unit, remain in widespread use, because such models tend to be parametrically parsimonious while yielding good model performance after calibration using historical watershed input–output data. On the other hand, distributed models are considered to provide a more realistic representation of the spatial heterogeneity of hydrological processes. However, the shift from lumped to distributed hydrological modeling significantly increases the complexity of the parameter estimation problem. It also raises many important questions related to parameter identifiability and equifinality (Beven 2001), thereby counterbalancing the theoretical advantages of distributed models. The practical consequence is that lumped models, which have much lower data requirements, are often preferred over distributed models.

The assessment of uncertainty of hydrological models is of major importance in hydrologic modeling. Generally, there are three principal sources contributing to modeling uncertainty: errors associated with input data and data for calibration, imperfection in model structure, and uncertainty in model parameters (e.g., Refsgaard and Storm 1996). Xu et al. (2006) demonstrated that the quality of precipitation data influences both simulation errors and calibrated model parameters. Diaz-Ramirez et al. (2012) and Wu et al. (2008) estimated the uncertainty in simulating hydrologic processes induced by the input data (e.g. rainfall and elevation). Engeland et al. (2005) showed that the effect of the model structural uncertainty on the total simulation uncertainty of a conceptual water balance model was larger than parameter uncertainty. Marshall et al. (2007) stated that the uncertainty in model structure requires developing alternatives, where outputs from multiple models are pooled together in order to generate an ensemble of hydrographs that are able to represent uncertainty. Bahremand and Smedt (2010) performed sensitivity and predictive analysis of the model parameters in the Torysa River basin using WetSpa hydrological model, proving that the parameter uncertainty of the model does not result in a significant level of predictive uncertainty. Kavetski et al. (2002) and Chowdhury and Sharma (2007) investigated input data uncertainty by artificially adding noise to input data and then formulating an empirical relationship between this noise and parameter error. Many other examples of the methods dealing with model and data uncertainty are available in the hydrological literatures (e.g., Georgakakos et al. 2004; Carpenter and Georgakakos 2004; Kavetski et al. 2006a, b;Talebizadeh et al. 2010).

A variety of methods have been developed to deal with parameter uncertainty and modeling uncertainty, i.e., to provide posterior distributions for parameters and runoff modelling. Among these methods, the generalized likelihood uncertainty estimation (GLUE) method, developed by Beven and Binley (1992), and the formal Bayesian method using Metropolis-Hastings (MH) algorithm, a Markov Chain Monte Carlo (MCMC) methodology, are extensively used (Freer et al. 1996; Beven and Freer 2001a; Kuczera and Parent 1998; Bates and Campbell 2001). GLUE is easy to implement, requiring no modifications to existing source codes of simulation models. Therefore, many users are attracted by GLUE. Stedinger et al. (2008) pointed out that GLUE methodology can be useful for model calibration and uncertainty analyses only if a proper likelihood function for normal independent distribution error can be found. Other studies show that threshold values and ranges of parameters are important factors influencing the results from the GLUE method (Yang et al. 2008; Vazquez et al. 2009; Jin et al. 2010; Li et al. 2010, 2011).

Recently, public awareness of hydrological extreme events has increased sharply, with a number of mortalities and damages triggered by disasters such as floods and storms (Yang et al. 2010). Therefore, simulation and prediction of streamflow extremes has become a hot spot, both for hydrological science research, water and watershed management and development strategies. However, the uncertainty sensitivity using the GLUE approach with different measures has not been evaluated in the past. Meanwhile, studies addressing uncertainty of different hydrological models in simulating extreme streamflows are still very limited so far. Therefore, this study strives to offer an intercomparison of uncertainty assessment of three hydrological models (i.e. HBV model, XAJ model and TOPMODEL) in simulating streamflow extremes using the well-tested GLUE approach. The main objectives of the paper are hereby to: (1) quantify different uncertainty behaviors of three typical lumped and semi-distributed hydrological models in simulating runoff processes with particular highlights of their potential capabilities in reproducing high-flows; and (2) evaluate and compare uncertainty sensitivity using the GLUE approach with different choices of threshold values and various uncertainty meansures using sensitivity analysis. The results are expected to contribute toward knowledge improvement on simulation and projection of hydrological extremes in response to climate change. Thereafter, the work is potentially beneficial for policymakers and stakeholders in local water hazard mitigation management.

2 Study Region and Data

2.1 Study Area

The headwater region of the Yellow River refers to the catchment above the Longyang Gorge (Fig. 1, IWMI and YRCC 2003). Administratively, this region belongs to the Qing-Tibetan Plateau of China. The headwater region of the Yellow River has a drainage area of 121,000 km^2. It covers approximately 15 % of the entire Yellow River's drainage basin, while it supplies 38 % of the River's total runoff. This region is a very important source of streamflow for the entire Yellow River basin, and is referred to as "the cistern of Yellow River". The mean altitude of this area is about 4,000 m.. The prevailing climate in the region is cold and dry without obvious seasonal variations. The annual average precipitation is about 450 mm. More than 70 % of the total annual precipitation falls in the flood season from July to October. Annual average air temperature varies between −4 °C and 2 °C from southeast to northwest (Xu et al. 2006). Plentiful perennial and seasonal frozen grounds exit in the region.

The increase in temperature in the high altitude area is more sensitive than that in the low altitude due to global warming. In recent years, problems have emerged with increasing pressure due to human activity and economic development, such as degradation of environment, degradation of meadows, and acceleration of soil erosion, and therefore have attracted increasing attention. Subsequently, there have been many studies on the evaluation of climate change, such as temperature increase, annual precipitation change, and induced runoff change (Arora 2002). However, knowledge on change of streamflow extremes over the region is still insufficient so far.

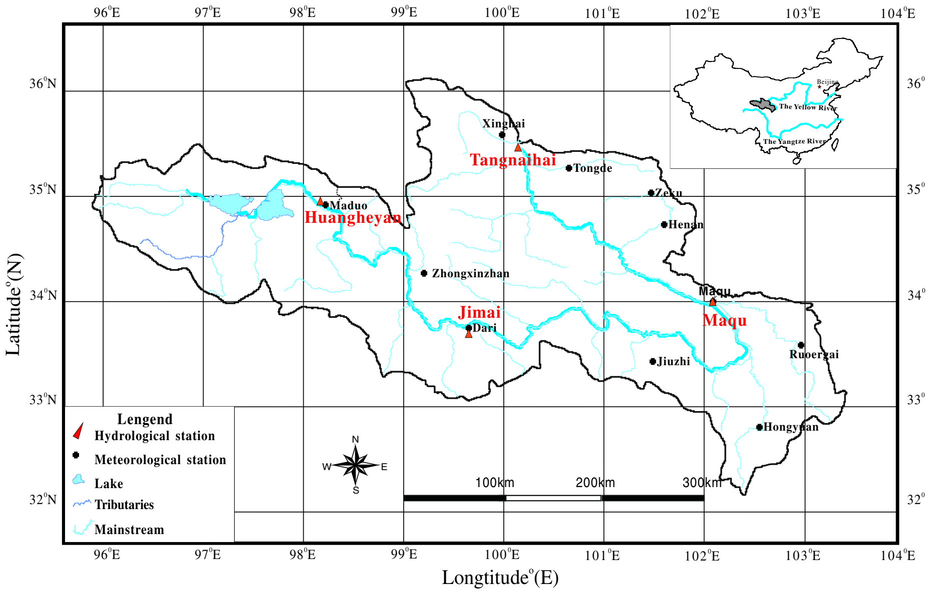

Fig. 1 Map of the Upper Yellow River in China

2.2 Data

The hydrological and meteorological observations (1960–2011) including: precipitation, pan evaporation, cloud covering, temperature, wind speed, humidity, radiation, water vapor pressure at a total of 11 stations (Table 1), and discharge at Tangnaihe gauge are collectively used in the study. These datum were collected from the Hydrology Bureau, Yellow River Conservancy Commission (YRCC) of China and the National Climate Center. The quality of observational data in China meets the International Association of Hydrological Science and

Table 1 List of 11 meteorological gauges (1960–2011) in the head region of the Yellow River (Source of data: The National Center of Climate, China)

Site name	Site number	Longitude	Latitude	Average annual temperature (°C)	Average annual precipitation (mm)
Xinghai	52943	99.98°E	35.58°N	1.36	353.6
Tongde	52957	100.65°E	35.27°N	0.38	429.7
Zeke	52968	101.47°E	35.03°N	−2.12	474.3
Maduo	56033	98.22°E	34.92°N	−3.72	308.2
Zhongxin zhan	56041	99.20°E	34.27°N	−3.79	459.3
Dari	56046	99.65°E	33.75°N	−0.88	543.3
Henan	56065	101.60°E	34.73°N	0.31	580.4
Jiuzhi	56067	101.48°E	33.43°N	0.61	762.2
Maqu	56074	102.08°E	34.00°N	1.53	602.3
Ruoergai	56079	102.97°E	33.58°N	1.11	647.6
Hongyuan	56173	102.55°E	32.80°N	1.43	753.0

World Meteorological Organization's standards. In this study, 1960–1986 is calibration period and 1987–2011 is validation period.

3 Methodology

3.1 Three Hydrological Models

The three models included in the study (e.g. HBV model, XAJ model and TOPMODEL) are well-known models in the hydrology community and details of the model structure and equations were widely presented in the hydrological literatures in past. Hereby, the below sections only provide brief introduction. In the following section, modifications of the TOPMODEL and XAJ model are addressed, which presents how a snow-routine was added to the models.

3.1.1 HBV Model

The HBV model (Bergström 1976, 1995) is a precipitation-runoff model, which includes conceptual numerical descriptions of hydrological processes at the catchment scale. In different model versions, HBV has been applied in more than 40 countries all over the world (Bergström 1995; Seibert 2003). The model is normally run on daily values of precipitation and air temperature, and daily or monthly estimates of potential evaporation. The model includes three main modules: snow accumulation and melt, soil moisture routing, and river routing and response modules (Abebe et al. 2010). Detailed description of the model can be found in above mentioned references, it is therefore not presented here.

3.1.2 TOPMODEL

TOPMODEL, proposed by Beven and Kirby (1979), is a physically-based watershed model that simulates the variable-source-area concept of streamflow generation (Quinn et al. 1995). Since the model has been widely reported across the world (Beven and Kirby 1979; Beven 1997; Huang and Jiang 2002), details will not be addressed here.

3.1.3 Xin'anjiang (XAJ) Model

XAJ model (Zhao et al. 1980) is a well-known lumped watershed model. The model has been widely used in humid and semi-humid regions of China for continuous hydrological simulation on a daily time scale and for rainstorm flood forecasting on an hourly time scale in past years. In such regions, runoff is generated only when the storage exceeds the field capacities. Runoff processes in XAJ model includes surface-, subsurface-runoff and groundwater base-flow. The evapotranspiration processes in the soil column are simulated in three layers. The routing procedures in stream network or channel system are modelled by means of Muskingum routing scheme. Recently, more practices to project future scenarios of hydrological processes and water resources in response to climate change have been widely reported in a variety of regions in China (e.g. Jiang et al. 2007; Ju et al. 2009).

3.2 Snow Accumulation/melt Routines

Generally, TOPMODEL and XAJ model do not have snow accumulation and melt simulation components. To enable simulating these processes in the study region, the two models are modified by adding the snow accumulation and melting routines as follows. Snow accumulation is estimated from precipitation by using atmospheric temperature records to separate the precipitation into snow and rainfall (Davies 1997). In the existing approaches to model snowmelt, the simple degree-day approach is the most commonly used method in operational hydrology and has been successfully verified word-wide over a range of catchments, physical characteristics and climate (e.g. Davies 1997; Rango 1992), which is also used in this study. The basic equation of the degree day method is:

$$M = C_m(T - T_{melt}) \tag{1}$$

Where M is the daily snowmelt (mm/day), C_m is the so called degree-day factor (mm/°C per day), T is the mean daily temperature (°C), and T_{melt} is the critical temperature for melting to occur (°C). The degree–day factor is an empirical constant that accounts for all the physical factors not included in the model, which varies with the land cover.

3.3 The Generalized Likelihood Uncertainty Estimation (GLUE) Method

The Generalized Likelihood Uncertainty Estimation (GLUE) method was proposed by Beven and Binley (1992) and has been widely used to model conditioning and uncertainty estimation in a variety of models of complex environmental systems (e.g. Beven and Binley 1992; Beven 1993; Aronica et al. 1998; Lamb et al. 1998; Cameron et al. 1999; Beven and Freer 2001a, b; Blazkova and Beven 2004; Iorgulescu et al. 2005). The GLUE methodology aims to identify behavioural models from a large sample of model runs with different parameter sets, chosen randomly from the specified ranges for each parameter by Monte Carlo simulation (Beven and Binley 1992; Freer et al. 1996). Here, the performance of each model is evaluated by multiple performance or likelihood measures. In most applications of GLUE, parameter sampling is carried out using non-informative uniform sampling without prior knowledge of individual parameter distribution other than a feasible range of values (Beven and Freer 2001b). Moreover, the likelihood function and the threshold are subjectively determined and this was discussed by Freer et al. (1996).

In this study, the Nash-Sutcliffe efficiency (NS, Nash and Sutcliffe 1970,) was chosen as the likelihood function as in many other studies:

$$NS = 1 - \frac{\sum(\mathrm{Q_{obs}} - \mathrm{Q_{sim}})}{\sum\left(Q_{obs} - \overline{Q}_{obs}\right)} = 1 - \frac{\sigma_i^2}{\sigma_{obs}^2} \tag{2}$$

Where $\overline{Q}_{obs}$ represents mean values of observed streamflows. σ_i^2 is the error variance for the i_{th} model (i.e. the combination of the model and the i_{th} parameter set) and σ_{obs}^2 is the variance of observations.

Tables 2, 3, and 4 show the parameters considered in the Monte Carlo simulation of this study, together with their respective ranges. Each parameter value is drawn uniformly and independently from the ranges and a total of 20000 parameter sets were chosen to drive TOPMODEL, XAJ and HBV models.

3.4 Uncertainty Measures in GLUE-based Hydrological Simulation

In order to provide a quantitative evaluation of the difference among the results of the three different models, the percentages of observations that are contained in the calculated 95 % confidence intervals (Jin et al. 2010) are calculated:

$$P - 95CI = \frac{NQ_{in}}{n} \times 100\% \tag{3}$$

Table 2 Parameter ranges used in the Monte Carlo simulations for HBV

Parameter	Definition	Range	Optimal value
TT(°C)	Threshold temperature	0–1	0.03
CFMAX(mm/°C)	Degree-day factor	1.5~4	2.94
FC(mm)	Maximum soil moisture storage	90–200	99
LP	Threshold for evaporation reduction	0.7–1	0.81
BETA	Parameter that determines the relative contribution to runoff from rain or snowmelt	1–4	2.29
PERC(mm/day)	Maximum percolation from the upper to the lower groundwater box	0–1.5	0.63
UZL(mm)	Threshold parameter	0–100	60.43
K_0	Recession coefficient of upper zone	0.05–0.9	0.83
K_1	Recession coefficient of lower zone	0.01–0.3	0.05
K_2	Recession coefficient of deep zone	0.001–0.1	0.03
MAXBAS	Transformation function parameter	1–10	1.30

where n is the number of time steps; NQ_{in} is the number of observations which are contained in the calculated confidence intervals.

The Average Relative Length (ARIL) is proposed by Jin et al. (2010) as follows:

$$ARIL = \frac{1}{n}\sum \frac{Limit_{Upper,t} - Limit_{Lower,t}}{R_{obs,t}} \tag{4}$$

$Limit_{Lower,t}$ and $Limit_{Upper,t}$ are the lower and upper boundary values of 95 % confidence intervals, and $R_{obs,t}$ is the observed.

The average asymmetry degree (AAD) of the prediction bounds (Xiong et al. 2009) with respect to the corresponding observed discharge is simulated:

$$AAD = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Limit_{Upper,t} - R_{obs,t}}{Limit_{Upper,t} - Limit_{Lower,t}} - 0.5\right| \tag{5}$$

An AAD value less than 0.5 indicate that, on average, the observed discharge lie within the uncertainty bands. Whereas the greater the value of AAD, the more asymmetrical the uncertainty bands are around the observed water levels (Xiong et al. 2009).

The Average deviation amplitude (ADA) is used by Xiong et al. (2009) as follows:

$$ADA = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{2}\left(Limit_{Upper,t} + Limit_{Lower,t}\right) - R_{obs,t}\right| \tag{6}$$

Table 3 Parameter ranges used in the Monte Carlo simulations for TOPMODEL

Parameter	Definition	Range	Optimal value
SR_{MAX} (m)	The soil profile storage available for transpiration	0.01–0.02	0.013
SR_0 (m)	The initial storage deficit in the root zone	0–0.1	0.0026
SZM(m)	The parameter of the exponential transmissivity function or recession curve	0.002–0.09	0.07
T_0 (m^2/h)	Effective lateral saturated transmissivity	0.1–5	0.25
Rv(*m*/*h*)	The velocity of overland flow concentration	2000–2700	2273

Table 4 Parameter ranges used in the Monte Carlo simulations for XAJ model

Parameter	Definition	Range	Optimal value
Kc	Ratio of potential evapotranspiration to pan evaporation	0.2–0.4	0.35
UM (mm)	Averaged soil moisture storage capacity of the upper layer	10–40	33
LM (mm)	Averaged soil moisture storage capacity of the lower layer	60–90	80
C	Coefficient of the deep layer, that depends on the proportion of the basin area covered by vegetation with deep roots	0.05–0.2	0.06
WM (mm)	Average soil moisture storage capacity of the whole layer	120–300	225
B	Exponential parameter with a single parabolic curve, which represents the non-uniformity of the spatial	0.3–0.7	0.65
I_m	Percentage of impervious and saturated areas in the catchment	0.005–0.02	0.01
SM (mm)	Areal mean free water capacity of the surface soil layer, which represents the maximum possible deficit of free water storage	5–50	24
EX	Exponent of the free water capacity curve influencing the development of the saturated area	1–1.65	1.58
KG	Outflow coefficients of the free water storage to groundwater relationships	0.05–0.7	0.36
KI	Outflow coefficients of the free water storage to interflow relationships	*KG*+*KI*=0.7	0.34
CI	Recession constants of the lower interflow storage	0.5–0.9	0.87
CG	Recession constants of the groundwater storage	0.95–0.998	0.99
CS(*UH*)	Recession constants in the lag and route method for routing through the channel system within each sub-basin	0.5–0.8	0.78
L(h)	Lag in time (routing period)	0–3	1
KE(h)	Parameter of the Muskingum method	$KE=\Delta t$	24
XE	Parameter of the Muskingum method	0–0.5	0.12

Both ARIL and P-95CI were recently applied in uncertainty quantification of hydrological modelling (Jin et al. 2010; Li et al. 2010, 2011). They are different criterion to measure different uncertainty in hydrological simulations. Generally, P-95CI can be used to quantify the random and systematic errors in simulations. From the statistical perspective, high values of P-95CI (e.g. >50 %) stand for more random errors in simulations. Whereas, low P-95CI values (e.g. <50 %) suggest systematic errors are major sources of uncertainties in simulations. However, ARIL stands for the systematic error in most simulations. In most cases, it can be used to assess the systematic error in hydrological simulations.

4 Results and Discussion

4.1 Model Calibration and Verification

This section presents an inter-comparison between the observed and simulated daily streamflows (1960–2005) by three hydrological models. The statistics presented here are RMSE, correlation coefficient (R^2) and Nash-Sutcliffe (NS) coefficient of efficiency (Eq. 20). Table 5 summarizes the performance scores for three hydrological models in the headwater catchment of Yellow River. It suggests that three models can reproduce historical streamflows well. In both calibration and validation, HBV model has the highest skills (lowest RMSE, highest R^2 and NS) followed by XAJ and TOPMODEL.

Table 5 Skill scores for three hydrological models in calibration (1961–1986) and validation (1987–2011)

Calibration and validation		XAJ	TOP	HBV
Calibration (1960–1986)	RMSE	272.00	271.95	256.68
	Correlation	0.90	0.90	0.93
	NS	0.80	0.80	0.82
Validation (1987–2011)	RMSE	260.56	248.14	223.86
	Correlation	0.85	0.86	0.91
	NS	0.70	0.73	0.78

4.2 Uncertainty Analysis

4.2.1 Daily Streamflow

Intercomparison of the uncertainty measures among the three hydrological models In this section, different uncertainty measures of the three hydrological models in simulating daily runoff are quantified and compared using a threshold value of 0.7 in the GLUE estimates. Here, the percentage of observations that are contained in the calculated confidence intervals (P-95CI) is used here as an uncertainty measure. In this study, XAJ model produces the best simulation (P-95CI=71.57 %).

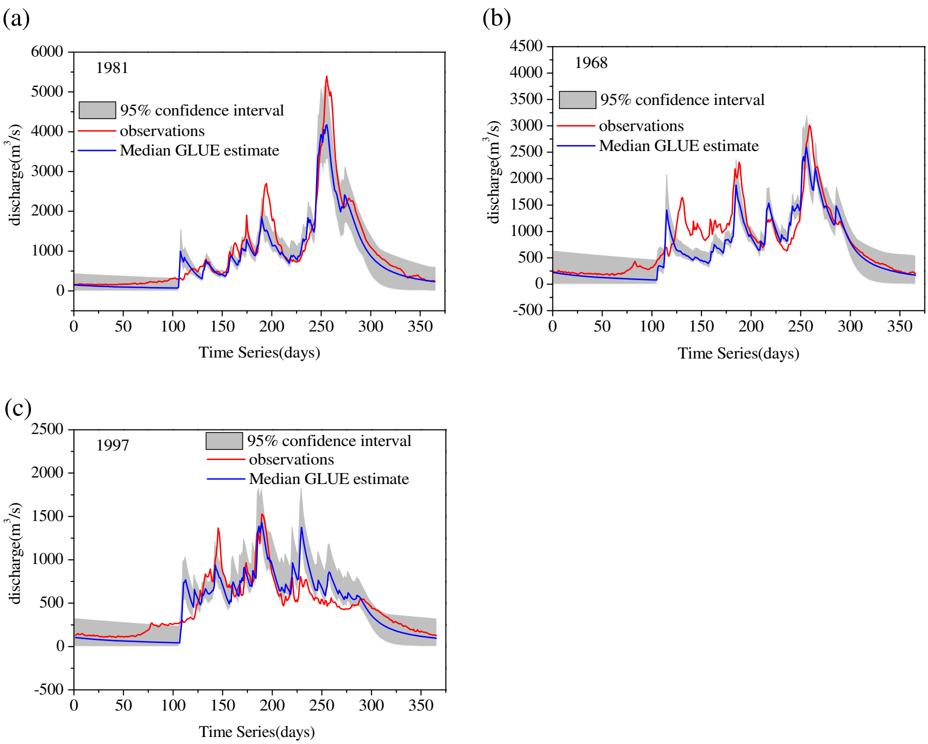

Fig. 2 Comparison of 95 % confidence interval of runoff with threshold of 0.7 using Xinanjiang model. (**a**) in 1981 flood year; (**b**) in 1968 normal year; (**c**) in 1997 dry year

For sake of a clear illustration, 3 typical years are presented here (i.e. 1981 as a typical flood year, 1968 as a normal year, and 1997 as a dry year). As can be seen from Figs. 2a-4a, XAJ model performed best in simulating daily flow in the flood year with the uncertainty bounds include a large percentage of the observations, about 82.19 % (Table 6). The GLUE estimate results of TOPMODEL (Fig. 3a) and HBV (Fig. 4a) are not as satisfactory as those obtained from XAJ, mainly due to overestimation or underestimation of the hydrograph in dry period. Similarly in the normal year, the best performance is also obtained by XAJ model (Fig. 2b).

Figures 2c, 3c, 4c show the 95 % confidence intervals of daily discharge in the dry year for the three models (Threshold=0.7). The similar performance is shown for the XAJ model (P-95CI=64.66 %) and TOPMODEL (P-95CI=65.48 %). However, the P-95CI obtained from HBV is less than 50 % (about 48.30 %), because the flow during the dry (wet) period is always under (over) estimated. Whereas, the simulated flood peak time during spring matches well with observations.

In summary, the XAJ model generally reproduces satisfied daily runoff processes when threshold value equals to 0.7, particularly the observations in dry period (December-February) are completely within the confidence intervals (P-95CI=98.46 %). TOPMODEL works better in wet period (June-August), however, its performance during dry period is poor (P-95CI= 9.34 %). Meanwhile, the GLUE uncertainty bounds produced by TOPMODEL (HBV model) are above (below) the observations during dry period.

Table 6 Intercomparison of uncertainty measures in simulating daily streamflow by three hydrological models under different threshold values (0.5, 0.6, 0.7)

Measures	Time Period	0.5			0.6			0.7		
		XAJ	TOP	HBV	XAJ	TOP	HBV	XAJ	TOP	HBV
P-95CI (%)	1960–2011	83.25	83.35	72.36	79.53	70.74	64.94	71.57	59.11	61.75
	1981	89.86	97.53	71.51	87.67	81.92	66.30	82.19	67.4	63.72
	1968	77.32	81.69	68.85	76.23	74.04	63.93	69.67	54.1	60.26
	1997	92.60	98.9	73.42	85.48	78.9	58.90	64.66	65.48	48.30
	Dry period	98.84	75.24	51.62	98.79	32.74	42.98	98.46	9.34	31.24
ARIL	1960–2011	1.45	1.37	1.07	1.34	1.21	0.91	1.13	0.98	0.73
	1981	1.46	1.49	1.06	1.32	1.29	0.88	1.19	1.02	0.71
	1968	1.43	1.27	1.02	1.36	1.11	0.87	1.11	0.87	0.70
	1997	1.49	1.72	1.24	1.36	1.52	1.04	1.14	1.23	0.84
	Dry period	2.69	2.05	1.05	2.6	1.78	0.93	2.35	1.41	0.72
ADA (m^3/s)	1960–2011	178.98	183.48	182.49	176.36	187.51	182.46	166.08	186.20	178.33
	1981	213.16	180.92	271.01	214.38	182.86	268.58	198.42	192.52	231.48
	1968	221.31	247.24	247.03	222.73	248.93	240.03	203.85	243.62	231.66
	1997	116.22	150.05	197.91	117.62	156.24	205.36	135.82	161.64	221.97
	Dry period	69.39	159.63	84.16	61.60	175.41	107.21	44.42	175.92	93.58
AAD	1960–2011	0.29	0.37	0.46	0.33	0.43	0.53	0.43	0.54	0.71
	1981	0.24	0.22	0.50	0.28	0.26	0.77	0.31	0.36	1.01
	1968	0.36	0.37	0.64	0.41	0.42	1.03	0.55	0.52	1.04
	1997	0.23	0.26	0.81	0.26	0.31	0.99	0.38	0.39	1.28
	Dry period	0.14	0.64	0.68	0.16	0.78	0.73	0.17	0.99	0.92

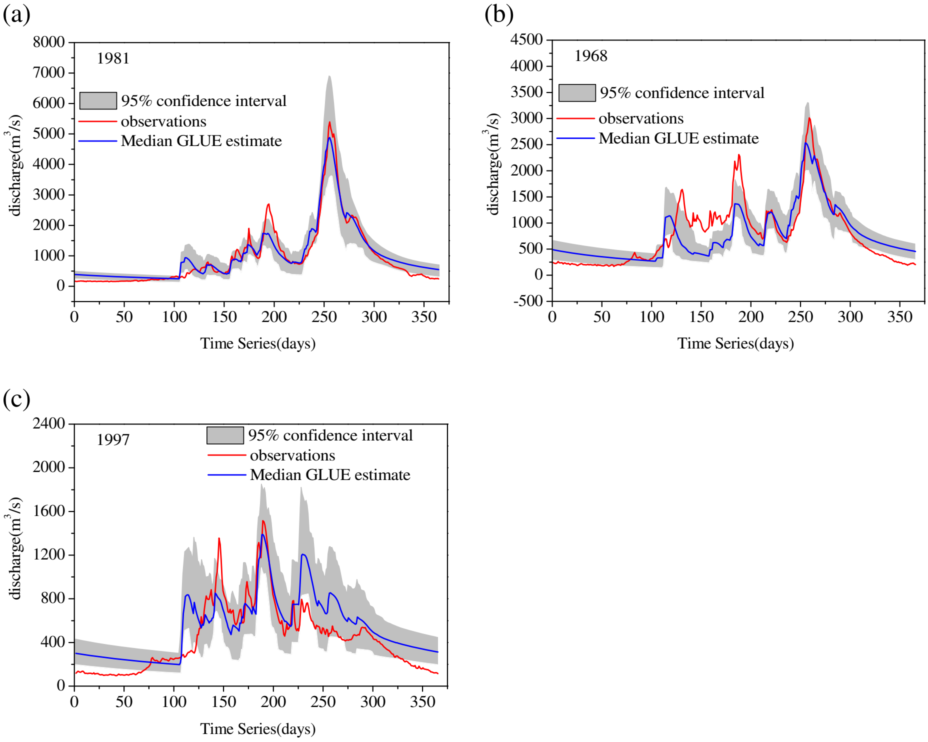

Fig. 3 Comparison of 95 % confidence interval of runoff with threshold of 0.7 using TOPMODEL. (**a**) in 1981 flood year; (**b**) in 1968 normal year; (**c**) in 1997 dry year

Sensitivity analysis of GLUE estimates to the threshold values in the likelihood function This section aims to address the uncertainty behaviors by a series of hydrological models (ranging from conceptual to distributed models) constructed with different assumptions, structures, running with different inputs. A sensitivity analysis is therefore conducted among three typical hydrological models to illustrate their distinct performances in response to different threshold values using the GLUE method. The illustration is done by calculating model uncertainty with threshold values of 0.7, 0.6 and 0.5. The ARIL, P-95CI, AAD and ADA for three hydrological models with different threshold values are compared in Table 6.

Results indicate that: (1) Decreasing ARIL, P-95CI and increasing AAD are found, when the threshold value increases for all the three models. This means GLUE estimates are sensitive to the choice of the threshold values based on three model simulations. As an example (the daily runoff simulation by TOPMODEL in 1968 normal year), P-95CI, ARIL and AAD are 81.69 %, 1.27 and 0.37 when the threshold value is 0.5. While they are 54.1 %, 0.87 and 0.52 corresponding to the threshold value of 0.7. (2) The best hydrological model differs to different threshold values. For instance, the GLUE estimate of the daily runoff in the normal year indicates the best performance (the highest P-95CI) is generated by TOPMODEL and XAJ model, with threshold value of 0.5 and 0.7. In addition, a high P-95CI value, narrow band-width, high degree of symmetry and small deviation amplitude, cannot be obtained simultaneously. It is similar with the results presented by Xiong et al. (2009). (3) There are no identical changes of ADA when threshold value increases (or P-95CI decreases). This is similar with the results (Alvisi and Franchini 2011) in the application of

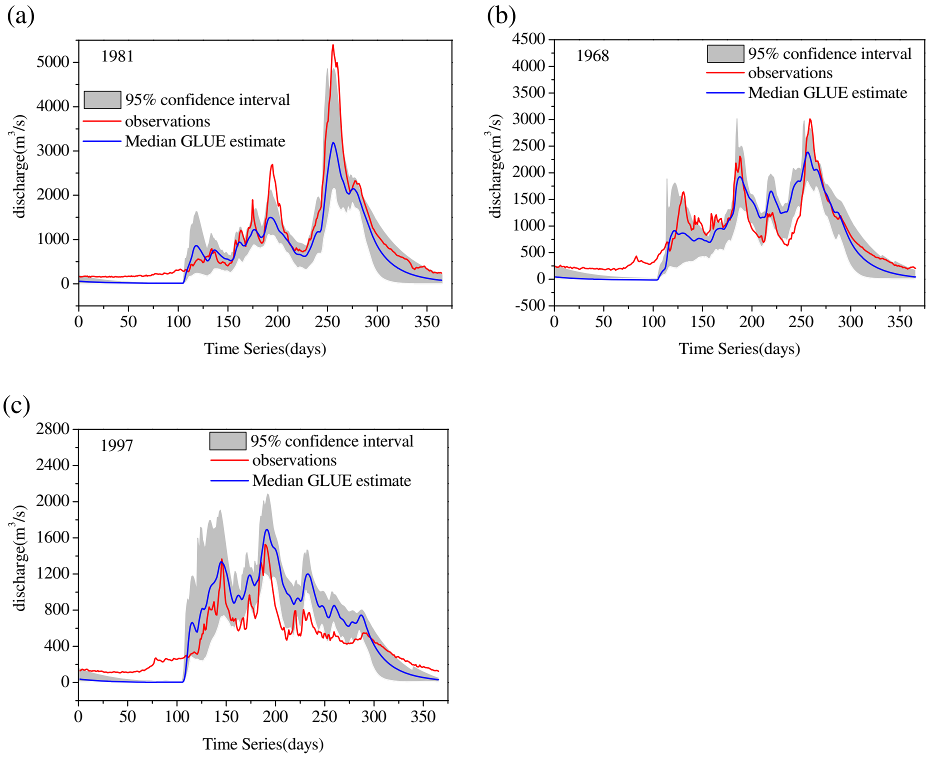

Fig. 4 Comparison of 95 % confidence interval of runoff with threshold of 0.7 using HBV model. (**a**) in 1981 flood year; (**b**) in 1968 normal year; (**c**) in 1997 dry year

the same index to evaluate the uncertainty bands obtained through Bayesian neural network and fuzzy neural network.

4.2.2 Extreme Streamflows

Intercomparison for the uncertainty measures among the three hydrological models In this section, the GLUE estimates of extreme streamflow (Q10 and Q25,) are herein used to examine the uncertainty behaviors of three hydrological models in simulating streamflow extremes (Q10 and Q25) in the headwater region of the Yellow River. Q10 (Q25) represents the flow corresponding to 10 % (25 %) exceedence probability during 1960–2011.

To perform the evaluation in extreme flow simulation, a lower threshold value of 0.4 is used. Fig. 5a and b show the 95 % confidence interval of Q10 and Q25 simulations by XAJ model using the GLUE method. Fig. 5a demonstrates a poor performance of Q10 GLUE estimates, where the observations fall outside the GLUE uncertainty bounds, especially in the calibration period (1960–1986). Meanwhile, the P-95CI is below 50 %, suggesting the other sources of uncertainty such as model structure are perhaps more important than parameter uncertainty alone (Thorndahl et al. 2008). Whereas, a better simulation of Q25 is produced (P-95CI =71.74 %). The P-95CIs for GLU estimate of high-flows (Q10 and Q25) show the TOPMODEL performs best among the three hydrological models, where the uncertainty bounds include a large percentage of observations with 80.43 % (Q10, Fig. 5c)

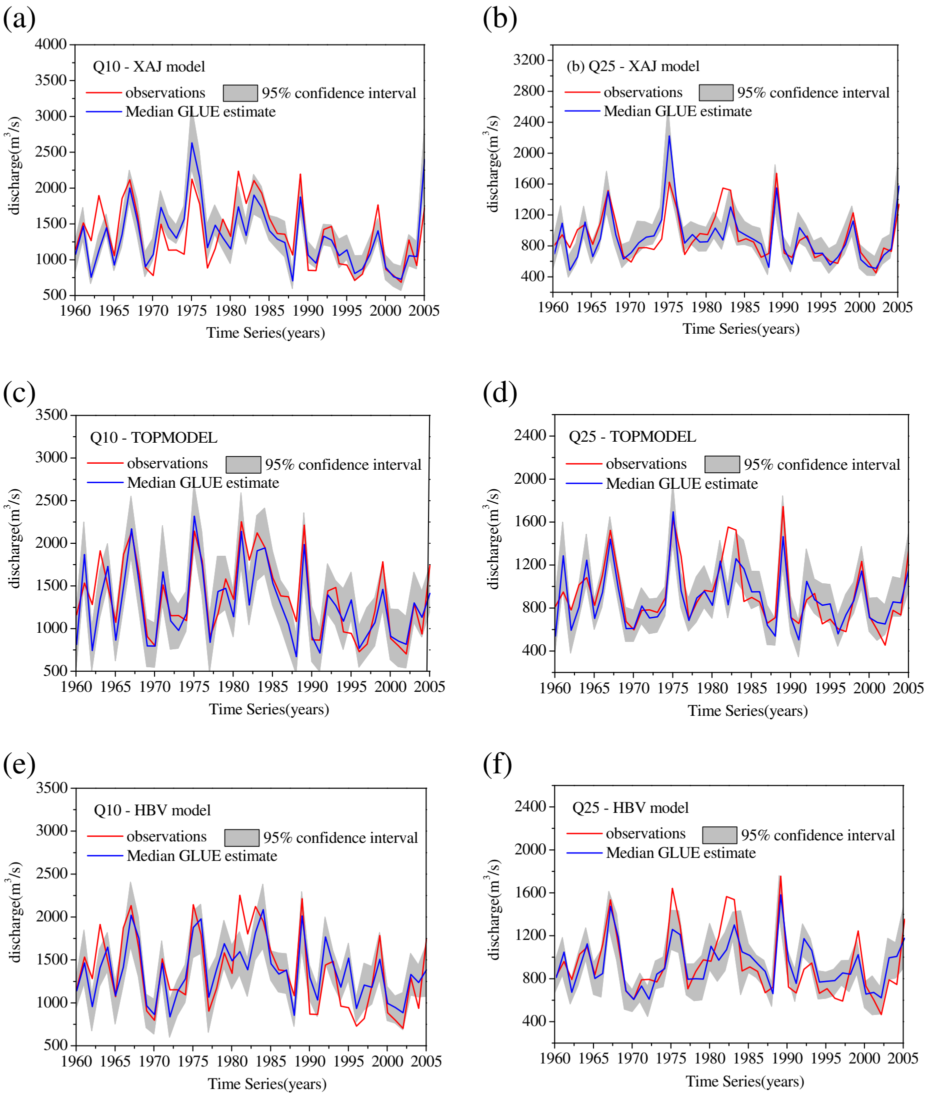

Fig. 5 Comparison of 95 % confidence interval of runoff with threshold of 0.4. (**a**) hydrological extreme (Q10) with XAJ model simulation; (**b**) hydrological extreme (Q25) with XAJ model simulation; (**c**) hydrological extreme (Q10) with TOPMODEL simulation; (**d**) hydrological extreme (Q25) with TOPMODEL simulation; (**e**) hydrological extreme (Q10) with HBV model simulation; (**f**) hydrological extreme (Q25) with HBV model simulation

and 76.09 % (Q25, Fig. 5d). The imperfect simulations by HBV model (Fig. 5e and f) suggest that the observations for Q10 and Q25 are often outside the uncertainty bounds during the validation period (1987–2011).

Sensitivity analysis of GLUE estimates to the threshold values in the likelihood function
This section presents a sensitivity analysis of GLUE estimates for high-flows (Q10 and Q25) to the threshold values (0.2, 0.3, and 0.4, Table 7). Table 7 shows that: (1) A high P-95CI

Table 7 Intercomparison of uncertainty measures in simulating extreme high-flows (Q10 and Q25) by three hydrological models under different threshold values (0.2, 0.3, 0.4)

Measures	High-flows	0.2			0.3			0.4		
		XAJ	TOP	HBV	XAJ	TOP	HBV	XAJ	TOP	HBV
P-95CI (%)	Q10	69.57	86.96	71.74	60.87	86.90	67.39	49.97	80.43	61.13
	Q25	80.43	80.43	76.09	78.00	78.26	71.74	71.74	76.09	69.25
ARIL	Q10	0.41	0.55	0.49	0.36	0.52	0.47	0.30	0.49	0.42
	Q25	0.46	0.49	0.47	0.43	0.45	0.45	0.38	0.41	0.40
ADA (m^3/s)	Q10	239.03	183.48	203.86	237.37	181.41	203.80	235.72	183.12	199.48
	Q25	140.36	126.98	139.72	139.01	127.99	136.69	134.55	131.11	136.41
AAD	Q10	0.46	0.28	0.34	0.53	0.29	0.36	0.64	0.31	0.38
	Q25	0.34	0.30	0.33	0.37	0.33	0.35	0.40	0.36	0.37

value, narrow band-width, high degree of symmetry and small deviation amplitude, cannot be obtained collectively. ARIL and P-95CI decrease and AAD increases, while the threshold value increases for the three models. However, it is difficult to quantify the link between the threshold values and ADA. This is similar to the GLUE estimates for daily runoff. (2) There exists significant disparity between daily flow series and extreme streamflows in sensitivity degree to the threshold value in the likelihood function. In detail, the change range for Q10 is up to 6.53 % when the threshold value changes is 20 %, while the range for daily runoff is about 24.24 % as a result of rather large range for flow during dry period (about 65.90 %, Table 6). Consequently, the GLUE estimates for high flow are less sensitive to the choice of the threshold values than the daily runoff in TOPMODEL simulations. However, the GLUE estimates for high flow are more sensitive to the choice of the threshold values than the daily runoff in XAJ model simulations. Therefore, the sensitivity of GLUE estimates for extreme flow to the threshold values depends on the major uncertainty sources from parameter, model structure and inputs. (3) High P-95CIs of GLUE estimate for high-flows (Q10 and Q25) show TOPMODEL generally performs best in different threshold values. While XAJ model produces the smallest ARIL under different threshold values.

5 Conclusion

In the work, uncertainty behaviors of three typical lumped and semi-distributed hydrological models (HBV, XAJ and TOPMODEL) in simulating runoff processes and extremes using the GLUE methodology are identified in the headwater region of the Yellow River, China. The major points are summarized as follows:

(1) In general, the results show that three models can reproduce historical daily runoff series with an acceptable accuracy. When the threshold value in GLUE estimate equals to 0.7, it is found that XAJ model has good strength in simulating daily runoff. In particular, the observations in dry period are completely within the confidence intervals, suggesting the flow during dry period can be grasped well by XAJ model. Nevertheless, there are always drawbacks for different models. The daily streamflow in wet period are either underestimated or overestimated by XAJ model. Moreover, the imperfect performance during dry period of TOPMODEL and

HBV model is dissimilar. The flow during dry period is always under (over) estimated for HBV model (TOPMODEL), and both the PCI-95 values are below 50 %, indicating non-aleatory errors in input data and model structure remain important. Hereby, it is discussed as following: Firstly, it is difficult to describe spatial distribution of the runoff generation mechanism with lumped or semi-distributed hydrological models. In addition, Xiong and Guo (2004) indicated the reason why the moderate or even bad performance of TOPMODEL is shown. TOPMODEL paid more attention to the runoff generation processes rather than the water budget accounting, as demonstrated by the very simple linear relationships used to estimate both the evaporation loss in the root storage zone and the vertical water movement from the unsaturated zone to the saturated zone. As for the HBV model, the runoff concentration is simplified by means of a triangular weighting function. This will inevitably introduce uncertainties to a certain degree. Therefore, the performance of TOPMODEL and HBV model in simulating low-flows is less satisfactory.

TOPMODEL performs best in high-flows (Q10 and Q25) among the three hydrological models. The uncertainty bounds include a large percentage of observations, about 80.43 % (Q10) and 76.09 % (Q25). However, the P-95CI in the GLUE estimate by XAJ model is below 50 %. In addition to the effect of the model structure, the prior distribution of parameter may influence the result of GLUE estimate, in this study just the uniform distribution is selected, a further investigation is needed to identify how some other prior distributions of parameter will influence the model simulation.

(2) Results suggest that a high P-95CI value, narrow band-width, high degree of symmetry and small deviation amplitude with respect to uncertainty interval, cannot be obtained simultaneously. ARIL as well as P-95CI decrease and AAD increases as the threshold value increase for the GLUE estimate from three models, while it is difficult to quantify the relationship between the threshold values and ADA. The performance statistics obtained using a variety of uncertainty measures for three models confirm this point. However, there is significant disparity between daily runoff and extreme streamflows in sensitivity degree to the threshold value of the likelihood function. Jin et al. (2010) indicated the drawback of GLUE is related, to some extent, to the subjectivity in choosing the likelihood function. Indeed, the GLUE estimates for daily runoff indicate that the P-95CI, ARIL and AAD are sensitive to the choice of the threshold values. Meanwhile, the selected hydrological model which performs best in term of P-95CI (ARIL) varies with the threshold value. While regarding to the high-flows (Q10 and Q25), whether the GLUE estimates for it are more sensitive to the choice of the threshold values than the daily runoff or not depends on which uncertainty source is more important, including parameter uncertainty, model structure and input uncertainty. In detail, the best P-95CI in the GLUE estimates of high-flow (Q10) is always produced by TOPMODEL in different threshold values. The probable reason is that the parameter uncertainty is more important than the other sources of uncertainty such as model structure.

Acknowledgements The work was jointly supported by grants from the National Natural Science Foundation of China (40901016, 40830639, 40830640), a grant from the State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering (2009586612, 2009585512), grants of Special Public Sector Research Program of Ministry of Water Resources (201201025), and the Fundamental Research Funds for the Central Universities (2010B00714).

References

Abebe N, Ogden F, Pradhan N (2010) Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. Journal of Hydrology 389(3–4):301–310

Alvisi S, Franchini M (2011) Fuzzy neural networks for water level and discharge forecasting with uncertainty. Environmental Modelling and Software 26:523–537

Aronica G, Hankin B, Beven K (1998) Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. Advances in Water Resources 22(4):349–365

Arora VK (2002) The use of the aridity index to assess climate change effect on annual runoff. J Hydrol 265:164–177

Bahremand A, Smedt FD (2010) Predictive analysis and simulation uncertainty of a distributed hydrological model. Water Resour Manage 2010(24):2869–2880

Bates BC, Campbell EP (2001) A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall–runoff modeling. Water Resources Research 37:937–947

Benke KK, Lowell KE, Hamilton AJ (2007) Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. Mathematical and Computer Modeling 47(11–12):1134–1149

Bergström S (1976) Development and application of a conceptual runoff model for Scandinavian catchments. Ph.D. Thesis. SMHI Reports RHO No. 7, Norrköping, Sweden

Bergström S (1995) The HBV model. In: Singh VP (ed) Computer models of watershed hydrology. Water Resources Publications, Highlands Ranch, CO., pp 443–476

Beven K (2001) How far can we go in distributed hydrological modeling? Hydrology and Earth System Sciences 5(1):1–12

Beven KJ, Binley A (1992) The future of distributed models, model calibration and uncertainty prediction. Hydrological Processes 6:279–298

Beven KJ (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16:41–51

Beven KJ, Freer J (2001a) A dynamic TOPMODEL. Hydrological Processes 15:1993–2011

Beven KJ, Freer J (2001b) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249:11–29

Beven KJ, Kirby MJ (1979) A physically based variable contributing area model of basin hydrology. Hydrological Science Bulletin 24:43–69

Beven KJ (1997) Distributed hydrological modelling: Applications of the TOPMODEL concept. John Wiley and Sons, Chichester

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. Journal of Hydrology 292:153–172

Carpenter TM, Georgakakos KP (2004) Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrological model. Journal of Hydrology 298:202–221

Cameron D, Beven KJ, Tawn J, Blazkova S, Naden P (1999) Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). Journal of Hydrology 219:169–187

Chowdhury S, Sharma A (2007) Mitigating parameter bias in hydrological modelling due to uncertainty in covariates. Journal of Hydrology 340:197–204

Davies A (1997) Monthly snowmelt modelling for large-scale climate change studies using the degree day approach. Ecological Modelling 101(2–3):303–323

Diaz-Ramirez JN, McAnally WH, Martin JL (2012) Sensitivity of simulating hydrologic processes to gauge and radar rainfall data in subtropical coastal catchments. Water Resour Manage 26:3515–3538

Engeland K, Xu C-Y, Gottschalk L (2005) Assessing uncertainties in a conceptual water balance model using Bayesian methodology. Hydrological Sciences Journal 50(1):45–63

Freer J, Beven K, Ambroise B (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research 32(7):2161–2173

Georgakakos KP, Seo DJ, Gupta H et al (2004) Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. Journal of Hydrology 298:222–241

Huang B, Jiang B (2002) AVTOP: A full integration of TOPMODEL into GIS. Environmental Modelling and Software 17(3):261–268

Iorgulescu I, Beven KJ, Musy A (2005) Data-based modelling of runoff and chemical tracer concentrations in the haute-menthue (Switzerland) research catchment. Hydrological Processes 19:2557–2574

Jiang T, Chen Y, Xu C, Chen X, Singh V (2007) Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. Journal of Hydrology 336(3–4):316–333

`Jin X, Xu C, Zhang Q, Singh VP (2010) Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. Journal of Hydrology 383:147–155

Ju Q, Yu Z, Hao Z, Ou G, Zhao J, Liu D (2009) Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. Neurocomputing 72(13–15):2873–2883

Kavetski D, Franks S, Kuczera G (2002) Confronting input uncertainty in environmental modelling. In: Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds) Calibration of watershed models. AGU Water Science and Applications Series, Duan, pp 49–68

Kavetski D, Kuczera G, Franks SW (2006a) Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory Water Resources Research 42:W03407

Kavetski D, Kuczera G, Franks SW (2006b) Bayesian analysis of input uncertainty in hydrological modeling: 2. Application Water Resources Research 42(581):W03408

Kuczera G, Parent E (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The metropolis algorithm. Journal of Hydrology 211:69–85

Lamb R, Beven KJ, Myrabø S (1998) Use of spatially distributed water table observations to constrain uncertainty in a rainfallrunoff model. Advances in Water Resources 22(4):305–317

Li L, Xia J, Xu C-Y, Singh VP (2010) Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. Journal of Hydrology 390:210–221

Li L, Xu C-Y, Xia J, Engeland K, Reggiani P (2011) Uncertainty estimates by Bayesian method with likelihood of AR (1) &normal model and AR (1) &multi-normal model in different time-scales hydrological models. Journal of Hydrology 406:54–65

Lin K, Guo S, Zhang W, Liu P (2007) A new baseflow separation method based on analytical solutions of the Horton infiltration capacity curve. Hydrological Processes 21(13):1719–1736

Marshall L, Nott D, Sharma A (2007) Towards dynamic catchment modelling: A Bayesian hierarchical modelling framework. Hydrological Processes 21:847–861

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models- Part 1-a. discussion of principles. Journal of Hydrology 10(3):282–290

Quinn P, Beven KJ, Lamb R (1995) The ln(α/tan β) index: how to calculate it and how to use it in the TOPMODEL framework. Hydrol Process 9:161–185

Rango A (1992) World-wide testing of the Snowmelt runoff model with applications for predicting the effects of climate change. Nordic Hydrology 23:155–172

Refsgaard JC, Storm B (1996) Construction calibration and validation of hydrological models. In: Abbott MB, Refsgaard JC (eds) Distributed hydrological modelling, water science and technology library, vol 22, Kluwer Academic Publishers. Dordrecht, The Netherlands, pp 41–54

Seibert J (2003) Reliability of model predictions outside calibration conditions. Nordic Hydrology 34:477–492

Stedinger JR, Vogel RM, Lee SU, Batchelder R (2008) Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research 44:W00B06. doi:10.1029/2008WR006822

Talebizadeh M, Morid S, Ayyoubzadeh SA et al (2010) Uncertainty analysis in sediment load modeling using ANN and SWAT model. Water Resour Manage 24:1747–1761

Thorndahl S, Beven KJ, Jensen JB (2008) Event based uncertainty assessment in urban drainage modeling, applying the GLUE methodology. Journal of Hydrology 357:421–437

Vazquez RF, Beven K, Feyen J (2009) GLUE based assessment on the overall predictions of a MIKE SHE application. Water Resour Manage 23:1325–1349

Wu S, Li J, Huang GH (2008) Characterization and evaluation of elevation data uncertainty in water resources modeling with GIS. Water Resour Manage 22:959–972

Xiong LH, Guo SL (2004) Effects of the catchment runoff coefficient on the performance of TOPMODEL in rainfall-runoff modeling. Hydrological Processes 18:1823–1836

Xiong LH, Wan M, Wei XJ (2009) Indices for assessing the prediction bounds of hydrological models and application by generalized likelihood uncertainty estimation. Hydrological Sciences Journal 54:852–871

Xu C-Y, Tunemar L, Chen YD, Singh VP (2006) Evaluation of seasonal and spatial variations of conceptual hydrological model sensitivity to precipitation data errors. Journal of Hydrology 324:80–93

Yang J, Reichert P, Abbaspour KC, Xia J, Yang H (2008) Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology 358:1–23

Yang T, Shao Q, Hao Z, Chen X, Zhang Z, Xu C, Sun L (2010) Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in Pearl River Basin, Southern China. Journal of Hydrology 380(3–4):386–405

Zhao RJ, Zhuang YL, Fang LR, Liu XR, Zhang QS (1980) The Xinanjiang model. Hydrological Forecasting Proceedings Oxford Symposium, 129. IAHS publication, pp. 351–356.