



# Multicriteria assessment framework of flood events simulated with vertically mixed runoff model in semiarid catchments in the middle Yellow River

Dayang Li, Zhongmin Liang, Yan Zhou, Binquan Li, and Yupeng Fu

College of Hydrology and Water Resources, Hohai University, Nanjing, 210098, China

**Correspondence:** Binquan Li (libinquan@hhu.edu.cn)

Received: 26 December 2018 – Discussion started: 8 February 2019

Revised: 6 July 2019 – Accepted: 28 July 2019 – Published: 13 September 2019

**Abstract.** Flood forecasting in semiarid regions is always poor, and a single-criterion assessment provides limited information for decision making. Here, we propose a multicriteria assessment framework called flood classification–reliability assessment (FCRA) that combines the absolute relative error, flow classification and uncertainty interval estimated by the hydrologic uncertainty processor (HUP) to assess the most striking feature of an event-based flood: the peak flow. A total of 100 flood events in four catchments of the middle reaches of the Yellow River are modeled with hydrological models over the period of 1983–2009. The vertically mixed runoff model (VMM) is compared with one physically based model, the MIKE SHE model (originating from the *Système Hydrologique Européen* program), and two conceptual models, the Xinanjiang model (XAJ) and the Shanbei model (SBM). Our results show that the VMM has a better flood estimation performance than the other models, and the FCRA framework can provide reasonable flood classification and reliability assessment information, which may help decision makers improve their diagnostic abilities in the early flood warning process.

## 1 Introduction

Arid and semiarid regions account for approximately one-third of the global land surface and half of China's land surface. A trend towards a warmer climate has increased the global incidence of intense precipitation events. Arid and semiarid regions, i.e., areas where the annual rainfall is less than 250 and 250–500 mm a<sup>-1</sup>, respectively, are particularly

vulnerable to this change in climate (Khomsli et al., 2016; Yatheendradas et al., 2008). More than 50 % of flood-related casualties occur in these regions worldwide (Brito and Evers, 2016).

Hydrological models play an important role in flood simulation and forecasting (Devia et al., 2015). Many studies have focused on the improvement and estimation of hydrologic models in humid catchments, although there are fewer similar studies for semiarid catchments (Jiang et al., 2015). The runoff generation mechanisms for semiarid catchments are complex and may be simultaneously dominated by infiltration excess and saturation excess mechanisms (Beven, 1983; Beven and Freer, 2001).

Modeling semiarid catchments is a difficult task due to the strong spatial variability in rainfall and complexity of landscape characteristics (vegetation, soil, etc.) (Pilgrim et al., 1988). Compared with humid catchments, the rainfall of semiarid catchments is characterized by high intensity and short duration (Andersen, 2008). In certain areas with developing economies and small populations, the rain gauge networks are generally sparse. Rainfall data are important inputs for hydrologic models, and the high temporal–spatial rainfall variability combined with sparse rain gauges make modeling runoff more difficult (Hao et al., 2018; Li and Huang, 2017; Mwakalila et al., 2001).

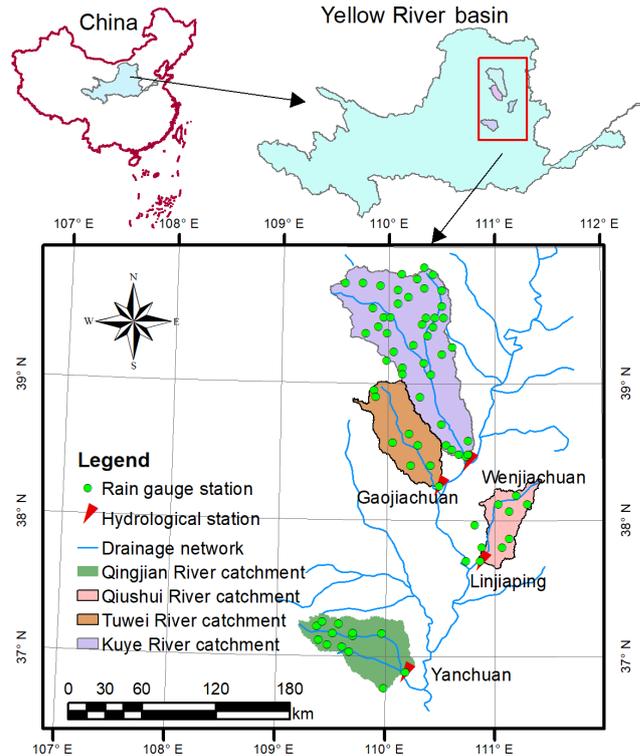
Satellite technology has the possibility to solve the issue of low rain gauge densities, although the low spatial and temporal resolutions of the products limit their applicability to sub-daily rainstorms (Dinku et al., 2007). Weather radar has high spatial resolution (1 m) and temporal resolution (15 min). However, the radar costs are too high to be used for large-scale semiarid areas (Young et al., 1999).

Literature on the subdaily modeling of rainfall runoff is limited in semiarid catchments. Due to rapid times to peak and scarce rainfall data, capturing rainstorm flood responses is more difficult than estimating daily, monthly or annual runoff (Andersen, 2008; McMichael et al., 2006). Flood simulation results in semiarid catchments are often poor. Michaud and Sorooshian (1994) used 24 severe rainstorms that produced the largest peak flows from 1957 to 1977 to compare three hydrologic models, i.e., the lumped Soil Conservation Service (SCS) model, the simple distributed SCS model, and the distributed kinematic runoff and erosion (KINEROS) model, in the Walnut Gulch catchment, and none of them were able to accurately simulate flood events. McIntyre and Al-Qurashi (2009) analyzed 27 flood events with three hydrologic models, the lumped Identification of Hydrographs and Components from Rainfall, Evaporation and Streamflow (IHACRES) model, the distributed IHACRES model, and a two-parameter regression model in a catchment in Oman. The average absolute relative errors in the flow peak and flow volume were 53 % and 36 %, respectively, for the best performing models. Under current technical conditions, it seems difficult to achieve an acceptable simulation/forecasting result for flood events in semiarid catchments. Therefore, determining how to use modeling results with limited accuracy to provide guidance for early flood warnings is important.

In this study, a multicriteria assessment framework that combines the absolute relative error, flow classification and uncertainty interval estimated by the hydrologic uncertainty processor (HUP) is proposed to provide more information for engineers' decision making. Four hydrological models, the vertically mixed runoff model (VMM), the MIKE SHE model (originated from the *Système Hydrologique Européen* program), the Xinanjiang model (XAJ) and the Shanbei model (SBM), are compared based on the performance of the modeling results in four catchments in the middle reaches of the Yellow River. The global sensitivity analysis (GSA) method PAWN (derived from the authors' names) is used to analyze the parametric sensitivity of the VMM. The rest of the paper is organized as follows. Section 2 describes the study area and the data set used. Section 3 presents the VMM methodology, model set, model calibration and validation, multicriteria assessment framework, and parameter sensitivity analysis. Section 4 describes the results and discussion of model comparison, sensitivity analysis and analysis of the multicriteria assessment framework for the VMM. The final section presents the conclusions of the study.

## 2 Study area and data

The four selected study catchments are all key tributaries located in the middle reaches of the Yellow River, China (Fig. 1). The maximum and minimum areas of the catchments are 1989 and 8706 km<sup>2</sup>, respectively. The average an-



**Figure 1.** Locations and rain gauging nets of the Qingjian River catchment, Qiushui River catchment, Tuwei River catchment and Kuye River catchment.

nual temperature ranges from 6 to 14 °C. The average annual precipitation ranges from 1010 to 1150 mm, and 65 % to 80 % is concentrated in summer (Li et al., 2019; Li and Huang, 2017; Xiao et al., 2019). The rainfall is generally characterized by high intensity and short duration. The average annual evaporation ranges from 1010 to 1150 mm. All selected catchments are semiarid based on an aridity index between 2.31 and 2.78 (UNEP, 1992). This catchment information is listed in Table 1.

The lack of vegetation in these catchments leads to severe soil erosion, and the average sediment concentration reaches 126 kg m<sup>-3</sup> according to Li et al. (2019). Some hydrologists have studied daily and monthly rainfall runoff, although few studies have modeled hourly floods. With the rapid increase in population and economic development, flood disasters have received increasing attention. Hence, it is important for decision makers to know how to evaluate the flood risk when a flood is approaching.

The period used in the model was from 1983 to 2009. Continuous streamflow and rainfall data were collected from streamflow gauging stations and rain gauging stations at a daily time step, respectively; streamflow and rainfall data for each of the flood events were collected at an hourly time step. Nine rainfall gauging stations in the Qiushui River catchment, 15 rainfall gauging stations in the Qingjian River catchment, 12 rainfall gauging stations in the Tuwei River

Table 1. Characteristics of the four catchments.

Catchment	Area (km <sup>2</sup> )	Outlet station	Area* (km <sup>2</sup> )	Mean annual precipitation (mm)	Mean evaporation (mm)	Aridity index
Qiushui River	1989	Linjiaping	1873	499	1150	2.31
Qingjian River	4080	Yanchuan	3468	451	1080	2.4
Tuwei River	3294	Gaojiachuan	2095	377	1050	2.78
Kuye River	8706	Wenjiachuan	8645	410	1010	2.46

\* The area of the catchment controlled by the outlet station indicated in the table.

and 41 rainfall gauging stations in the Kuye River were selected. The Thiessen polygon method was used to interpolate the rainfall data for each catchment.

### 3 Methodology

#### 3.1 Vertically mixed runoff model

The VMM is a lumped, continuous hydrologic model and has been used in many areas in China, especially in semiarid and subhumid catchments (Bao and Zhao, 2014; Li, 2018; Wang and Ren, 2009). Compared with other conceptual models, such as the XAJ model (Zhao, 1992) and the Sacramento Soil Moisture Accounting (SSMA) model (Burnash et al., 1973), among others, the VMM is capable of simulating the saturation excess and infiltration excess runoff generation mechanisms simultaneously. As shown in Fig. 2, the VMM combines the infiltration capacity curve and tension water content storage capacity curve in the vertical direction. Net rainfall (observed rainfall after removal of evaporation, *PE*) is partitioned into surface runoff (*RS*) and infiltration flow (*FA*) by the infiltration capacity curve in the VMM. *FA* is regulated by the tension water storage capacity curve, part of which supplements the tension water storage (*W*), with the remainder of the rainfall forming groundwater flow (*RB*) (including unsaturated flow and saturated flow). Here, the calculation of runoff generation is described briefly. More detailed information about the VMM is contained in Bao and Zhao (2014).

The improved Green–Ampt infiltration curve (Bao, 1993) is applied in the VMM as the infiltration capacity curve, and the equation is as follows:

$$FM = FC \left( 1 + K \frac{WM - W}{WM} \right), \tag{1}$$

where *FM* is the average point infiltration capacity of the catchment, and the descriptions of *WM*, *K*, and *FC* are shown in Table 2.

*FA* is calculated using Eq. (2):

$$FA = \begin{cases} FM - FM \left( 1 - \frac{PE}{(FMM)^{1+BF}} \right) & PE < FMM \\ FM & PE \geq FMM \end{cases}, \tag{2}$$

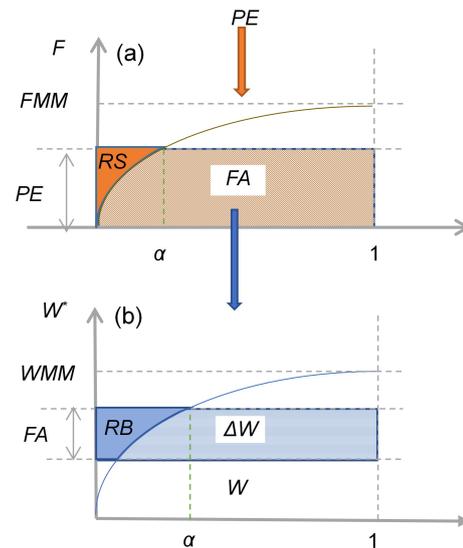


Figure 2. Runoff generation module in the VMM. (a) Infiltration capacity curve and (b) tension water content storage capacity curve.  $\alpha$  is the fracture area that is saturated and *F* represents the point infiltration capacity.

where

$$FMM = FM(1 + BF), \tag{3}$$

in which *FMM* is the maximum point infiltration capacity of the catchment and *BF* is defined in Table 2.

The part that exceeds the average point infiltration capacity of the catchment *FM* forms *RS*. *RS* can be calculated with Eq. (4).

$$RS = PE - FA \tag{4}$$

*RB* can be calculated using Eq. (5):

$$RB = \begin{cases} FA - WM + W + WM \left( 1 - \frac{W^* + FA}{WMM} \right)^{B+1} & FA + W^* < WMM \\ FM - WM + W & FA + W^* \geq WMM \end{cases}, \tag{5}$$

**Table 2.** Parameter values of the VMM.

Symbol	Definition	Range*
<i>KC</i>	Ratio of potential evapotranspiration to pan evaporation	[0.5, 1.5]
<i>WM</i>	Mean areal maximum possible soil moisture, mm	[50, 200]
<i>FC</i>	Stable infiltration capacity, mm h <sup>-1</sup>	[5, 100]
<i>K</i>	Infiltration index related to soil permeability	[0.05, 1]
<i>BF</i>	Index of the watershed infiltration capacity curve	[0, 0.5]
<i>B</i>	Index of the watershed water storage capacity curve	[1, 2]
<i>KI</i>	Outflow coefficient of interflow, d	[0.1, 0.5]
<i>KG</i>	Outflow coefficient of groundwater, d	[0.5, 2]
<i>CS</i>	Confluence coefficient of surface flow	[0.05, 0.9]
<i>CI</i>	Recession coefficient of interflow, d	[0.5, 0.95]
<i>CG</i>	Recession coefficient of ground flow, d	[0.90, 0.99]
<i>KE</i>	Residence time of Muskingum, h	[0.5, 5]
<i>XE</i>	Muskingum coefficient	[0.01, 0.49]
<i>IM</i>	Impermeable area	[0, 1]

\* In  $[a, b]$ ,  $a$  and  $b$  represent the lower and upper bounds of the parameters, respectively.

where

$$W^* = WMM \left[ 1 - \left( 1 - \frac{W}{WM} \right) \right]^{\left( \frac{1}{B} + 1 \right)}$$

$$WMM = WM(1 + B), \quad (6)$$

in which  $WMM$  is the maximum point tension water storage capacity of the catchment,  $W^*$  is the ordinate of Fig. 2b, which represents the point tension water content capacity in the catchment, and  $B$  is defined in Table 2.

The outlet runoff  $R$  can be calculated as follows:

$$R = RS + RB. \quad (7)$$

### 3.2 Model set of the VMM

The VMM was run continuously from 1983 to 2009 for each catchment. Rainfall data were available only at an hourly time step over the periods of flood events, and for other periods, they were available at a daily time step. Hence, the time step of simulations was set as daily between flood events and hourly within flood events for each catchment. To consider the spatial variation in rainfall, the subcatchments were divided according to the stream networks, and each subcatchment contained at least one rainfall gauging station. The areal mean rainfall of each subcatchment was calculated using the Thiessen polygon method. Because streamflow data were only available in the outlet streamflow gauging station for each catchment, the spatial variation in each catchment's parameters could not be determined by calibration. Thus, the parameters (Table 2) were set uniformly in all subcatchments. Two initial values, the initial tension water storage ( $W_0$ ) and the initial free water storage ( $S_0$ ), were used to describe the initial catchment moisture condition. The initial values are smaller for drier catchments, and the minimum values are zero. In this study, the initial values were assumed

to be zero uniformly due to the dry conditions at 00:00:00 LT on 1 January 1983 for each catchment. It should be noted that continuous simulations for each catchment eliminate the need to set the initial values for each flood event in a catchment.

### 3.3 Model calibration

The 14 parameters (Table 2) of the VMM were calibrated using the Shuffled Complex Evolution (SCE-UA) global optimization algorithm (Duan et al., 1993). The ranges of parameters were determined based on previous literature and prior knowledge (Bao and Zhao, 2014; Li et al., 2018). Due to the rapid rise and fall of floods (usually less than 24 h) in semi-arid catchments, accurate simulations of the full hydrograph are not needed and cannot be achieved. The Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) is widely used as an objective function of calibration in humid catchments; however, it may not be suitable for semiarid catchments because a good fit is not required between the simulated and observed streamflows. McIntyre and Al-Qurashi (2009) and Sharma and Murthy (1998) used the absolute relative error to evaluate model outputs (flow peak and flow volume) for semiarid areas, and the calibrated results indicated that the peak flow results are more accurate than the suggested results based on the NSE. Thus, the simulated hydrograph is reasonable for the majority of flood events. The equations are as follows:

$$E_p = \frac{1}{n} \sum_{i=1}^n \frac{|Q_p^i - Q_{pm}^i|}{Q_{pm}^i}, \quad (8)$$

$$E_v = \frac{1}{n} \sum_{i=1}^n \frac{|Q_v^i - Q_{vm}^i|}{Q_{vm}^i}, \quad (9)$$

where  $E_p$  and  $E_v$  are the average performances (in terms of absolute relative error) for peak flows and flow volumes in each catchment, respectively;  $n$  is the number of events; the index  $i$  denotes each event;  $Q_p$  and  $Q_{pm}$  are the simulated and measured values of peak flow per event, respectively; and  $Q_v$  and  $Q_{vm}$  are the simulated and measured values of flow volume per event, respectively.

Constraining the model output with peak flows and flow volumes can be expressed as follows:

$$E_{pv} = \frac{E_p + E_v}{2}, \quad (10)$$

where  $E_{pv}$  is the objective value. The model outputs become better as the value of  $E_{pv}$  approaches 0. The number of iterations was set to 2000 in the calibration process.

### 3.4 Model comparison

To achieve a better performance in rainstorm flood simulations, three hydrologic models, including two conceptual models, XAJ and SBM, and one distributed model, MIKE SHE, were used for comparison with the VMM. The XAJ model was developed by Zhao (1992) and has a single saturation excess runoff generation mechanism. The XAJ model has been successfully applied in humid and subhumid catchments (Cheng et al., 2006; Lü et al., 2013). The SBM was developed by Zhao (1983) and has a single infiltration excess runoff generation mechanism. The SBM is generally used in semiarid or arid catchments in China (Bao et al., 2017; Li and Zhang, 2008; Zhao et al., 2013). In addition, the MIKE SHE model is a deterministic, physically based distributed hydrologic model that can simulate surface water flow, unsaturated flow and saturated flow (Jayatilaka et al., 1998). The MIKE SHE model has been used to solve water resources and environment problems at different spatiotemporal scales (Li et al., 2018; Rujner et al., 2018; Samaras et al., 2016).

### 3.5 Multicriteria assessment framework: flood classification–reliability assessment for flood events

Flood simulations and forecasting in semiarid catchments are very difficult due to strong spatial variability of rainfall, complex landscape characteristics, etc. Although some hydrologists improve flood simulations and forecasting by improving hydrologic models, the improvements are always limited or are suitable for only specific regions (Collier, 2007). The flood peak is the most significant feature in semiarid regions. Determining the extent to which the calculation of flood peaks can be accepted is crucial. Generally, the absolute relative error is used to measure the calculation of flood peak accuracy; for example, 20%, 30% or similar values are acceptable (Li et al., 2014; McIntyre and Al-Qurashi, 2009). To provide more information for flood defense management, the generalized likelihood uncertainty estimation (GLUE) and the Bayesian framework with Markov

chain Monte Carlo sampling are used to provide probabilistic forecasting, such as the 95% uncertainty interval (Christiaens and Feyen, 2002; Li et al., 2017), although these methods may not lead to clear decisions (Beven, 2007).

In this study, to obtain a better diagnostic and discriminatory method for the decision maker, we propose a multicriteria assessment framework called the flood classification–reliability assessment (FCRA) in the catchments of the middle reaches of the Yellow River. The FCRA framework consists of two parts: (i) flood classification and (ii) flood reliability assessment. The first part represents floods that are classified with percentiles and the absolute relative error; the other represents the reliability of flood modeling that is evaluated with the Bayesian method. Peak flows, as the most prominent features of flood events, are assessed with the FCRA framework. Detailed descriptions can be found as follows.

- C1. The absolute relative error of peak flow should be less than 20%.
- C2. The modeled and observed peak flows should be in the same flow zone: the observed peak flow  $Q_p$  for all flood events in a catchment is divided into three zones (low-flow zone, medium-flow zone, high-flow zone), with 25th percentiles  $Q_{p25}$  and 75th percentiles  $Q_{p75}$  as the boundary points; if  $Q_p \leq Q_{p25}$ , then the peak flow  $Q_p$  belongs to the low-flow zone; if  $Q_p \geq Q_{p75}$ , then the peak flow  $Q_p$  belongs to the high-flow zone; the remaining flow peaks belong to the medium-flow zone. Both the 25th and 75th percentiles are commonly used to distinguish zones.
- C3. The observed peak flows should fall within 1 standard deviation ( $\sigma$ ) of the mean (approximately 68.3% uncertainty interval) peak flow estimated by the hydrologic uncertainty processor (HUP), one component of the Bayesian forecasting system detailed in Krzysztofowicz (1999) and Biondi et al. (2010).

Conditions C1 and C2 are flood classification criteria. If the observed and modeled peak flows meet one of the two conditions, it is believed that they are the same types of floods. The key of the FCRA framework is condition C2, and condition C1 is used to avoid errors caused by flow zone boundaries. For example, when  $Q_{p75} = 200 \text{ m}^3 \text{ s}^{-1}$ , the modeled peak flow equals  $198 \text{ m}^3 \text{ s}^{-1}$  and the observed peak flow equals  $201 \text{ m}^3 \text{ s}^{-1}$ . However, using only condition C2 may lead to inappropriate model results; adding condition C1 can help address the problem. Condition C3 is used to assess the reliability of peak flow modeling; a small uncertainty interval (68.3%) is used that has narrow upper and lower limits. This interval may reduce the numbers of observed peak flows that fall within the confidence level. A modeled peak flow that can be accepted should satisfy condition C1 or condition C2 and then condition C3.

### 3.6 Parameter sensitivity analysis

A sensitivity analysis (SA) (Saltelli et al., 1989) was proposed to assess the effects of inputs on the model output. The SA can be classified into a GSA and local sensitivity analysis (LSA). Compared with the LSA, the GSA is capable of analyzing the effects of inputs within the entire input domain. The Fourier amplitude sensitivity test (Cukier et al., 1973), Sobol method (Sobol, 1993) and Morris screening method (Morris, 1991) are the most widely used GSA methods in the assessment of parameter sensitivity in hydrologic models. Pianosi and Wagener (2015) proposed the novel GSA method PAWN, which is based on the cumulative density function. PAWN has advantages over the parameter ranking and time-consuming nature of other GSA methods (Khorashadi et al., 2017). In this study, we used the PAWN method to perform a GSA on the VMM.

Considering  $x_{i,j}$  ( $ij = 1, 2, \dots$ , where  $i$  and  $j$  represent the  $i$ th input parameters and the  $j$ th sampling, respectively) as sensitivity inputs, then the sensitivity of  $x_{i,j}$  can be measured by the distance between  $F_{(y_i|x_{i,j})}(y_i)$  (the cumulative probability distribution function of  $y_i$  when  $x_{i,j}$  changes between the upper bound and lower bound) and  $F_{y_i}(y_i)$  (the cumulative probability distribution function of  $y_i$  when  $x_i = \frac{1}{n} \sum_{j=1}^n x_{i,j}$ , where  $n$  is the number of samplings per input parameter). The Kolmogorov–Smirnov statistic (Simard and Ecuyer, 2011) is used to measure the distance between  $F_{(y_i|x_i)}(y_i)$  and  $F_{y_i}(y_i)$ :

$$KS(x_{i,j}) = \left| F_{y_i}(y_i) - F_{(y_i|x_{i,j})}(y_i) \right|. \quad (11)$$

As  $KS$  varies with  $x_{i,j}$ , the maximum of all possible  $KS$  values is included in the PAWN index  $P_i$ :

$$P_i = KS(x_{i,j}). \quad (12)$$

$P_i$  ranges from 0 to 1 and  $x_i$  becomes more sensitive as  $P_i$  approaches 1. A  $P_i$  value equal to 1 indicates that  $x_i$  has no effect on the model. For more information about PAWN, please refer to Pianosi and Wagener (2015). In this study, the number of evaluations was set to 500, as suggested by Pianosi and Wagener (2018).

### 3.7 Model validation

The modeling period was between 1983 and 2009. In the Qiushui River, 20 flood events were selected, with the first 15 events used for calibration and the remaining five events used for validation. Similarly, in the Qingjian River, 29 flood events were selected, with 24 events used for calibration and the remaining five events used for validation. In the Tuwei River, 23 flood events were selected, with 18 events used for calibration and the remaining five events used for validation. Finally, in the Kuye River, 28 flood events were selected, with 23 events used for calibration and the remaining five events used for validation.

**Table 3.** Performance (in terms of absolute relative error) for peak flow  $E_v$  in each catchment in the four models. Values are given as a percentage.

	Qiushui River	Qingjian River	Tuwei River	Kuye River	Average*
VMM	26.52	58.50	40.20	30.80	39.01
MIKE SHE	40.50	60.70	45.30	38.20	46.18
XAJ	56.60	66.61	60.20	52.30	58.93
SBM	38.14	55.82	35.50	45.20	43.15

\* The average  $E_v$  of the four catchments for each model.

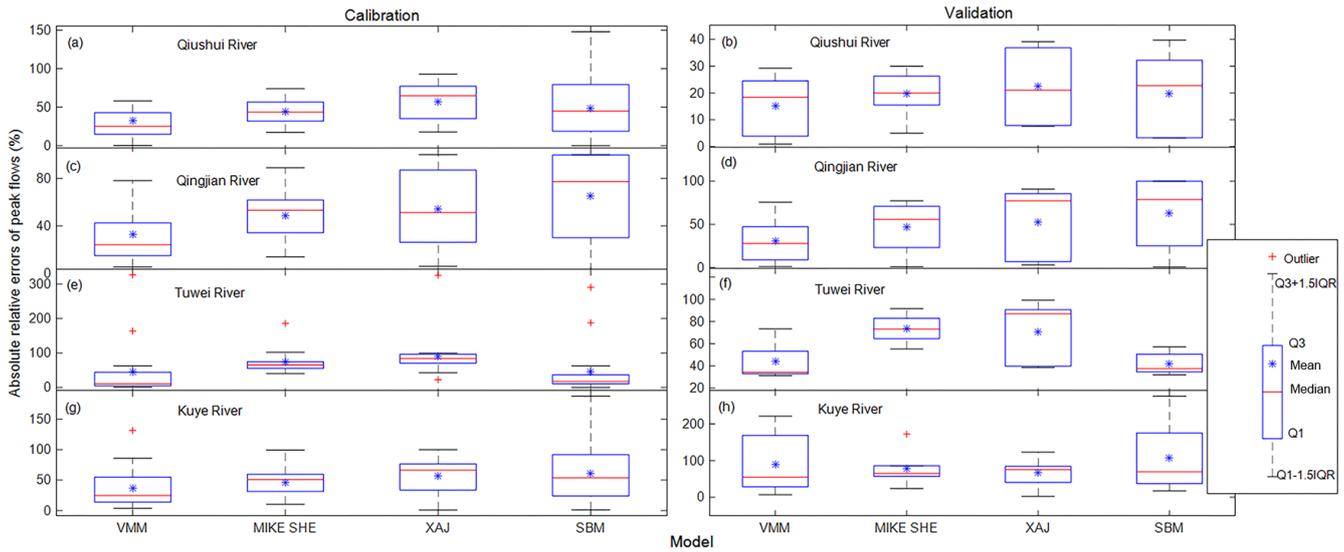
## 4 Results and discussion

### 4.1 Comparison of model results

Boxplots of the absolute relative errors of the peak flows for each model in the four catchments are shown in Fig. 3. In terms of the median and average of the absolute relative errors for peak flows, the VMM has the lowest values for both calibration and validation in Fig. 3a–g except for the validation period in the Kuye River catchment in Fig. 3h; in most cases, the MIKE SHE model has lower median and average values than the XAJ and SBM, i.e., Fig. 3a–d, g and h. Low median and average values indicate that more modeled flood events have good performance in a catchment. Except for the good performance in the Tuwei River catchment, the results using the SBM are as poor as those using the XAJ model in the other catchments. In terms of interquartile ranges (IQRs) of the absolute relative errors for peak flows, the VMM and MIKE SHE models have relatively small ranges (Fig. 3a, c, d and g) and the SBM and XAJ models have large ranges in most cases (Fig. 3a–d and g). This indicates that the VMM and MIKE SHE models are more robust to reproduce the peak flows in the middle reaches of the Yellow River.

Tables 3 and 4 show the average performance in terms of the absolute relative error for flow volume  $E_v$  and the lag time for the four models in each catchment, respectively. Low  $E_v$  and lag time values indicate that the model is highly capable of reproducing the flow volumes and time-to-peak values. The VMM has the minimum average  $E_v$  and lag time, with values of 39.01 % and 3.05 h, respectively (Tables 3 and 4). In contrast, the XAJ model has the maximum average  $E_v$  and lag time, with values of 58.93 % and 4.51 h, respectively. The MIKE SHE and SBM have similar performances in terms of average  $E_v$  and lag time.

The analysis of Fig. 3 and Tables 3 and 4 above indicates that the VMM has the best performance to reproduce the peak flows, flow volumes and lag times in the four studied catchments of the middle reaches of the Yellow River and the XAJ model has the worst performance. In addition, the MIKE SHE model is superior for reproducing the peak flows but exhibits similar performance compared with the SBM for reproducing the flow volume and lag time. Al-



**Figure 3.** Boxplot of absolute relative errors of peak flows in the four catchments.  $Q_1$  and  $Q_3$  represent the 1st quantile and 3rd quantile, respectively; interquartile range  $IQR = Q_3 - Q_1$ . An outlier is defined as an extreme value that exceeds the range of  $Q_1 - 1.5 IQR$  and  $Q_3 + 1.5 IQR$ .

**Table 4.** Lag time of the peak flow in the four catchments in the four models. Values are in hours.

	Qishui River	Qingjian River	Tuwei River	Kuye River	Average*
VMM	2.20	3.02	3.46	3.50	3.05
MIKE SHE	2.50	3.50	4.20	3.90	3.53
XAJ	4.10	3.81	5.62	4.50	4.51
SBM	4.00	2.95	3.46	4.20	3.65

\* The average lag time in the four catchments for each model.

though the MIKE SHE model is a distributed hydrologic model with more complex structures and more explicit physical meaning than the conceptual VMM, it does not achieve better results than the conceptual VMM due to a lack of sufficiently high-resolution data, and this is consistent with other studies (Beven, 2002, 2011; Michaud and Sorooshian, 1994; Seyfried and Wilcox, 1995). Both infiltration excess and saturation excess can be simulated via the VMM; it may be the reason why it performs better than the other two conceptual models (XAJ and SBM), which have single runoff generation mechanisms (saturation excess and infiltration excess, respectively).

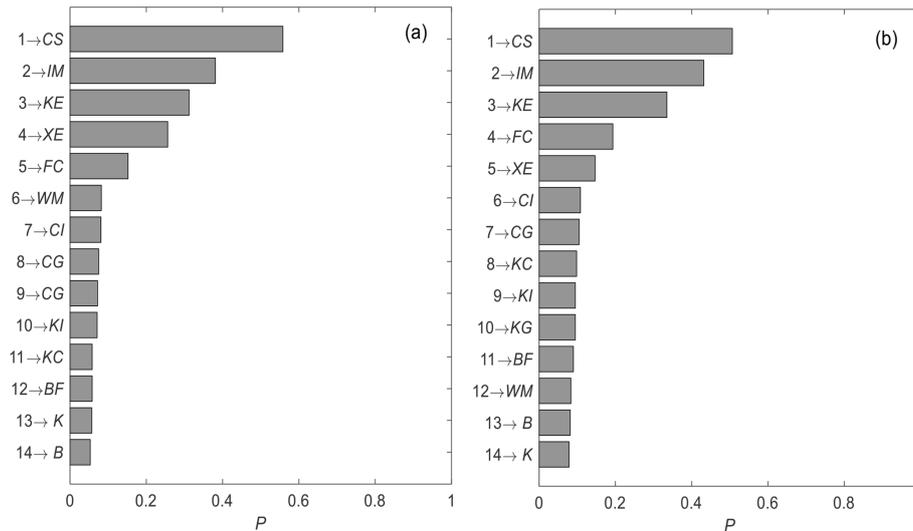
#### 4.2 Sensitivity analysis of the VMM

The GSA method PAWN is applied to estimate the influence of parameter uncertainty on the model output results. Figure 4a and b show the average SA results of all study catchments for the objective functions  $E_p$  (Eq. 9) and  $E_{pv}$  (Eq. 11), respectively. The parameters become more sensitive as the

ranking becomes higher. Parameters  $CS$ ,  $IM$  and  $KE$  have the highest rankings whether the objective function of the VMM is  $E_p$  or  $E_{pv}$ . The rankings of other parameters are influenced slightly by different objective functions, such as  $CG$ , except for  $WM$ .  $WM$  ranks sixth when  $E_{pv}$  is the objective function and 12th when  $E_p$  is the objective function. This ranking is because  $WM$  controls the tension water content in the soil, which determines the amount of rainfall stored in the soil and the generation of runoff. There may be a strong relationship between flow volume and  $WM$ . Therefore,  $WM$  has a higher ranking when the objective function considers the effect of flow volume.

#### 4.3 Flood classification–reliability assessment of the VMM

The FCRA framework we propose is applied to assess the ability of the VMM to model flood events in the four catchments. FCRA requires that the accepted modeled peak flows have the same flood types (high flow, medium flow or low flow) as the observed peak flows; in addition, the modeled peak flows should be reliable. Similar types of peak flows that represent the modeled peak flows should meet one of the requirements of conditions C1 and C2; the modeled peak flows that are reliably represented need to meet condition C3. The observed peak flows and the modeled peak flows under condition C1, C2 or C3 are shown in Fig. 5. The percentages of modeled peak flows that meet the conditions are presented in Table 5. Although the percentages of the modeled peak flows that meet condition C1 are less than 50% (Table 5), they reduce the boundary effects of flood classification. Taking the 13th flood event of the Kuye River catchment as an



**Figure 4.** Sensitivity rankings of the VMM parameters based on the global sensitivity analysis method PAWN for different objective functions. (a)  $E_{pv}$  as the objective function and (b)  $E_p$  as the objective function. The value  $P$  is used to assess the sensitivity degree of the parameter with the PAWN method, and a larger value corresponds to greater sensitivity. The numbers on the ordinate represent the sensitivity rankings.

**Table 5.** The percentage of modeled peak flows that meet various conditions for the VMM. Values are given as a percentage.

	C1	C2	C3	C*	C**
Qiushui River	40.00	75.00	95.00	75.00	75.00
Qingjian River	41.38	44.83	100.00	58.62	58.62
Tuwei River	47.83	69.57	100.00	69.57	69.57
Kuye River	35.71	64.29	100.00	67.86	67.86

C\* represents the modeled peak flows that meet condition C1 or C2; this means modeled and observed peak flows are the same type; C\*\* represents the modeled peak flows that meet conditions C\* and C3.

example, the observed and modeled peak flows are  $1230$  and  $1510 \text{ m}^3 \text{ s}^{-1}$ , respectively. As shown in Fig. 5d, the absolute relative error for peak flow is greater than 20%. In addition, for the Kuye River catchment, it is reasonable to believe that the peak flows  $1230$  and  $1510 \text{ m}^3 \text{ s}^{-1}$  may have the same risk according to the known flood peak data, which can be classified as the same flood type (medium flow) according to condition C2.

From Table 5, we find that 95% or more of modeled peak flows meet condition C3; this indicates that almost all modeled peak flows have less uncertainty and more reliability in the selected catchments. Figure 5 shows more directly that the majority of peak flows for the observations and modeling fall between the 15.85th percentile and the 81.45th percentile (68.3% uncertainty interval) estimated by HUP, which is consistent with Table 5. In addition, the percentages of modeled flood events and observed peak flows that are the same flood types (shown in Table 5 with C\*) equal the acceptance rate (shown in Table 5 with C\*\*) for each catchment due to

the high reliability of modeled peak flows. Under the FCRA framework, the acceptance rates ( $C^{**}$ ) for the catchments are more than 65% except for the Qingjian River catchment. This indicates that the FCRA framework may have the diagnostic capability to assess the modeled flood events in the four semiarid catchments.

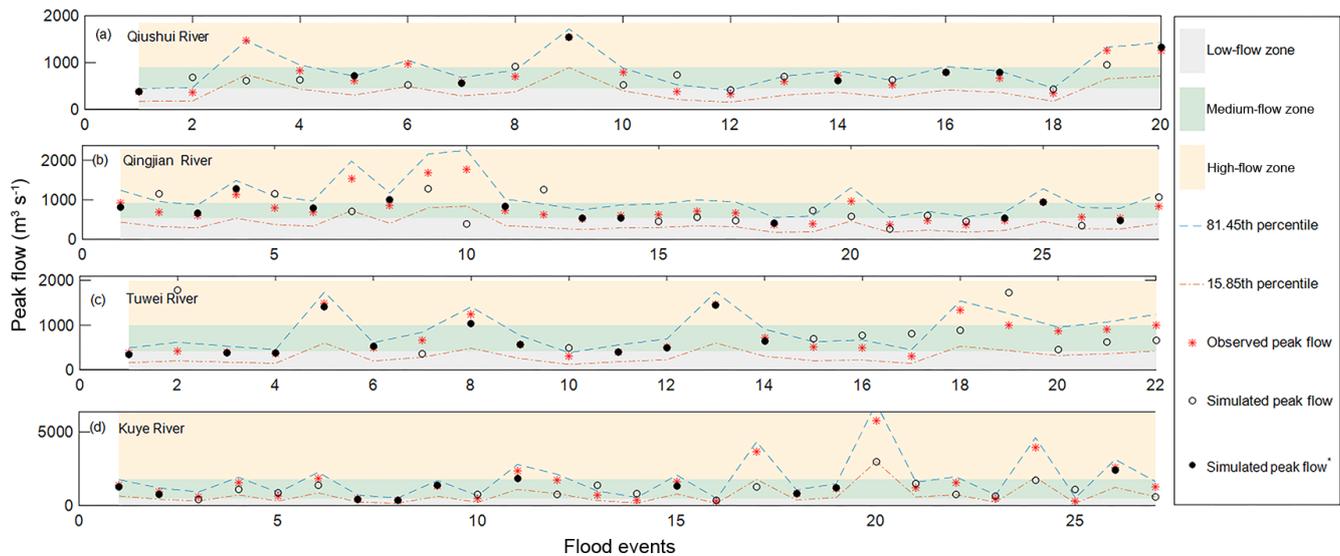
Under the FCRA framework, a modeled flood event could be assessed to determine what flood type (high flow, medium flow, low flow) it is and how reliable it is. This information is meaningful in the early flood warning process in semiarid catchments. Although FCRA is simple and even coarse, it is convenient and beneficial in helping engineers make decisions when a flood is approaching.

## 5 Conclusions

In this study, a multicriteria assessment framework of flood events called the flood classification–reliability assessment (FCRA) is proposed with the VMM in four semiarid catchments of the middle reaches of the Yellow River. The main conclusions are as follows.

Compared with the distributed model MIKE SHE and the two conceptual models, XAJ and SBM, the VMM has a better performance for modeling flood events in the middle reaches of the Yellow River.

In the four catchments, the parameter confluence coefficient of surface flow (CS), impermeable area (IM) and residence time of Muskingum (KE) in the VMM are the most sensitive parameters based on an analysis using the global sensitivity method PAWN; in addition, the sensitivity ranking



**Figure 5.** Observed peak flows (red asterisk) and simulated peak flows (circles) with the VMM for each catchment under conditions C1–C3. Flood peaks conforming to condition C1 are represented by solid circles and the others are empty; the three flow zones (low, medium and high flow zones) classified by condition C2 are shown in gray, green and off-white, respectively; the 68.3 % uncertainty interval of peak flows estimated by condition C3 is shown between the blue dashed line (81.45th percentile) and the red dashed–dotted line (15.85th percentile).

of the parameter  $WM$  related with the soil moisture capacity is the most affected by the objective functions.

The FCRA framework combining flood classification and reliability assessment may have the reliable diagnostic capability to assess flood events in the early flood warning process. It should be noted that condition C2, which divides peak flows into three flow zones, will be affected by the number of observed peak flows when data availability is limited. The framework is suitable for semiarid regions with poor modeling results and provides guidance for decision making.

**Code availability.** We have shared the MATLAB code for the VMM at <https://doi.org/10.4211/hs.c5232287d5c04bfb8cac5ce4e391ea0f> (Li, 2018).

**Author contributions.** DL wrote the text and developed the MATLAB code for the VMM. DL, ZL, YZ and BL conceived the study. All co-authors jointly worked on improving the text and responded to the editor’s and the reviewers’ suggestions.

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** We would like to thank Francesca Pianosi (University of Bristol) for providing the program code of PAWN at <https://www.safetoolbox.info/pawn-method/> (last access: 20 August 2019). We also thank the editor and the anonymous reviewers, whose comments have largely improved this work.

**Financial support.** This research has been supported by the National Key Research and Development Program of China (grant no. 2016YFC0402706), the National Natural Science Foundation of China (grant nos. 41730750, 41877147), and the Special Scientific Research Fund of Public Welfare Industry of Ministry of Water Resources, China (grant no. 201501004), sponsored by the Qing Lan Project.

**Review statement.** This paper was edited by Joaquim G. Pinto and reviewed by two anonymous referees.

## References

- Andersen, F. H.: Hydrological modeling in a semi-arid area using remote sensing data, PhD thesis, University of Copenhagen, Copenhagen, Denmark, 2008.
- Bao, H., Wang L., Zhang, K., and Li, Z.: Application of a developed distributed hydrological model based on the mixed runoff generation model and 2D kinematic wave flow routing model for better flood forecasting, *Atmos. Sci. Lett.*, 18, 284–293, <https://doi.org/10.1002/asl.754>, 2017.
- Bao, W.: Improvement and application of the Green-Ampt infiltration curve, *Yellow River*, 9, 1–3, 1993.
- Bao, W. and Zhao, L.: Application of Linearized Calibration Method for Vertically Mixed Runoff Model Parameters, *J. Hydrol. Eng.*, 33, 85–91, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000984](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000984), 2014.
- Beven, K. J.: Surface water hydrology – runoff generation and basin structure, *Rev. Geophys.*, 21, 721–730, <https://doi.org/10.1029/RG021i003p00721>, 1983.

- Beven, K. J.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, *Hydrol. Process.*, 16, 189–206, <https://doi.org/10.1002/hyp.343>, 2002.
- Beven, K. J.: *Environmental modelling: An uncertain future?*, CRC Press, London, UK, 328 pp., 2007.
- Beven, K. J.: *Rainfall-runoff modelling: the primer*, John Wiley & Sons, UK, 488 pp., <https://doi.org/10.1002/9781119951001>, 2011.
- Beven, K. J. and Freer, J.: A dynamic TOPMODEL, *Hydrol. Process.*, 15, 1993–2011, <https://doi.org/10.1002/hyp.252>, 2001.
- Biondi, D., Versace, P., and Sirangelo, B.: Uncertainty assessment through a precipitation dependent hydrologic uncertainty processor: An application to a small catchment in southern Italy, *J. Hydrol.*, 386, 38–54, <https://doi.org/10.1016/j.jhydrol.2010.03.004>, 2010.
- Brito, M. and Evers, M.: Multi-criteria decision-making for flood risk management: A survey of the current state of the art, *Nat. Hazards Earth Syst. Sci.*, 16, 1019–1033, <https://doi.org/10.5194/nhess-16-1019-2016>, 2016.
- Burnash, R. J., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system, conceptual modeling for digital computers, Report by the Joliet Federal State River Forecasts Center, Sacramento, CA, 204 pp., 1973.
- Cheng, C. T., Zhao, M. Y., Chau, K., and Wu, X. Y.: Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure, *J. Hydrol.*, 316, 129–140, <https://doi.org/10.1016/j.jhydrol.2005.04.022>, 2006.
- Christiaens, K. and Feyen, J.: Constraining soil hydraulic parameter and output uncertainty of the distributed hydrological MIKE SHE model using the GLUE framework, *Hydrol. Process.*, 16, 373–391, <https://doi.org/10.1002/hyp.335>, 2002.
- Collier, C. G.: Flash flood forecasting: What are the limits of predictability?, *Q. J. Roy. Meteorol. Soc.*, 133, 3–23, <https://doi.org/10.1002/qj.29>, 2007.
- Cukier, R., Fortuin, C., Shuler, K., and Petschek, A., and Schaibly, J.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory, *J. Chem. Phys.*, 59, 3873–3878, <https://doi.org/10.1063/1.1680571>, 1973.
- Devia, G. K., Ganasri, B. P., and Dwarakish, G. S.: A review on hydrological models, *Aquat. Proced.*, 4, 1001–1007, <https://doi.org/10.1016/j.aqpro.2015.02.126>, 2015.
- Dinku, T., Ceccato, P., Kopec, E. G., Lemma, M., Connor, S. J., and Ropelewski, C. F.: Validation of satellite rainfall products over East Africa's complex topography, *Int. J. Remote Sens.*, 28, 1503–1526, <https://doi.org/10.1080/01431160600954688>, 2007.
- Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, *J. Optimiz. Theory Appl.*, 76, 501–521, <https://doi.org/10.1007/bf00939380>, 1993.
- Hao, G., Li, J., Song, L., Li, H., and Li, Z.: Comparison between the TOPMODEL and the Xin'anjiang model and their application to rainfall runoff simulation in semi-humid regions, *Environ. Earth Sci.*, 77, 279, <https://doi.org/10.1007/s12665-018-7477-4>, 2018.
- Jayatilaka, C., Storm, B., and Mudgway, L.: Simulation of water flow on irrigation bay scale with MIKE-SHE, *J. Hydrol.*, 208, 108–130, [https://doi.org/10.1016/s0022-1694\(98\)00151-6](https://doi.org/10.1016/s0022-1694(98)00151-6), 1998.
- Jiang, Y., Liu, C., Li, X., Liu, L., and Wang, H.: Rainfall-runoff modeling, parameter estimation and sensitivity analysis in a semiarid catchment, *Environ. Model. Softw.*, 67, 72–88, <https://doi.org/10.1016/j.envsoft.2015.01.008>, 2015.
- Khomsi, K., Mahe, G., Trambly, Y., Sinan, M., and Snoussi, M.: Regional impacts of global change: seasonal trends in extreme rainfall, run-off and temperature in two contrasting regions of Morocco, *Nat. Hazards Earth Syst. Sci.*, 16, 1079–1090, <https://doi.org/10.5194/nhess-16-1079-2016>, 2016.
- Khorashadi, Z. F., Nossent, J., Sarrazin F., Pianosi, F., Griensven, V. A., Wagener, T., and Bauwens, W.: Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model, *Environ. Model. Softw.*, 91, 210–222, <https://doi.org/10.1016/j.envsoft.2017.02.001>, 2017.
- Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resour. Res.*, 35, 2739–2750, <https://doi.org/10.1029/1999wr900099>, 1999.
- Li, B., Yu, Z., Liang, Z., Song, K., Li, H., Wang, Y., Zhang, W., and Acharya, K.: Effects of Climate Variations and Human Activities on Runoff in the Zoige Alpine Wetland in the Eastern Edge of the Tibetan Plateau, *J. Hydrol. Eng.*, 19, 1026–1035, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000868](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000868), 2014.
- Li, B., Liang, Z., He, Y., Hu, L., Zhao, W., and Acharya, K.: Comparison of parameter uncertainty analysis techniques for a TOPMODEL application, *Stoch. Environ. Res. Risk A.*, 31, 1045–1059, <https://doi.org/10.1007/s00477-016-1319-2>, 2017.
- Li, B., Liang, Z., Bao, Z., Wang, J., and Hu, Y.: Changes in streamflow and sediment for a planned large reservoir in the middle Yellow River, *Land Degrad. Dev.*, 30, 878–893, <https://doi.org/10.1002/ldr.3274>, 2019.
- Li, D.: Hydrologic model: the vertically mixed runoff model (vmm), *HydroShare*, <https://doi.org/10.4211/hs.c5232287d5c04bfb8cac5ce4e391ea0f>, 2018.
- Li, D., Liang, Z., Li, B., Lei, X., and Zhou, Y.: Multi-objective calibration of MIKE SHE with SMAP soil moisture datasets, *Hydrol. Res.*, 50, 644–654, <https://doi.org/10.2166/nh.2018.110>, 2018.
- Li, X. and Huang, C. C.: Holocene palaeoflood events recorded by slackwater deposits along the Jin-shan Gorges of the middle Yellow River, China, *Quatern. Int.*, 453, 85–95, <https://doi.org/10.1002/jqs.2536>, 2017.
- Li, Z. J. and Zhang, K.: Comparison of three GIS-based hydrological models, *J. Hydrol. Eng.*, 13, 364–370, [https://doi.org/10.1061/\(ASCE\)1084-0699\(2008\)13:5\(364\)](https://doi.org/10.1061/(ASCE)1084-0699(2008)13:5(364)), 2008.
- Lü, H., Hou, T., Horton, R., Zhu, Y., Chen, X., Jia, Y., Wang, W., and Fu, X.: The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method, *J. Hydrol.*, 480, 102–114, <https://doi.org/10.1016/j.jhydrol.2012.12.011>, 2013.
- McIntyre, N. and Al-Qurashi, A.: Performance of ten rainfall-runoff models applied to an arid catchment in Oman, *Environ. Model. Softw.*, 24, 726–738, <https://doi.org/10.1016/j.envsoft.2008.11.001>, 2009.
- McMichael, C. E., Hope, A. S., and Loaiciga, H. A.: Distributed hydrological modelling in California semi-arid shrublands: MIKE SHE model calibration and

- uncertainty estimation, *J. Hydrol.*, 317, 307–324, <https://doi.org/10.1016/j.jhydrol.2005.05.023>, 2006.
- Michaud, J. and Sorooshian, S.: Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed, *Water Resour. Res.*, 30, 593–605, <https://doi.org/10.1029/93wr03218>, 1994.
- Morris, M. D.: Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161–174, <https://doi.org/10.2307/1269043>, 1991.
- Mwakalila, S., Campling, P., Feyen, J., Wyseure, G., and Beven, K.: Application of a data-based mechanistic modelling (DBM) approach for predicting runoff generation in semi-arid regions, *Hydrol. Process.*, 15, 2281–2295, <https://doi.org/10.1002/hyp.257>, 2001.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Pianosi, F. and Wagener, T.: A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, *Environ. Model. Softw.*, 67, 1–11, <https://doi.org/10.1016/j.envsoft.2015.01.004>, 2015.
- Pianosi, F. and Wagener, T.: Distribution-based sensitivity analysis from a generic input-output sample, *Environ. Model. Softw.*, 108, 197–207, <https://doi.org/10.1016/j.envsoft.2018.07.019>, 2018.
- Pilgrim, D. H., Chapman, T. G., and Doran, D. G.: Problems of rainfall-runoff modelling in arid and semiarid regions, *Hydrolog. Sci. J.*, 33, 379–400, <https://doi.org/10.1080/02626668809491261>, 1988.
- Rujner, H., Uuml, G., Leonhardt, N., Marsalek, J., and Viklander, M.: High-resolution modelling of the grass swale response to runoff inflows with Mike SHE, *J. Hydrol.*, 562, 411–422, <https://doi.org/10.1016/j.jhydrol.2018.05.024>, 2018.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M.: Sensitivity Analysis in Practice, *J. Am. Stat. Assoc.*, 101, 398–399, <https://doi.org/10.1198/jasa.2006.s80>, 1989.
- Samaras, A. G., Gaeta, M. G., Moreno, M. A., and Archetti R.: High-resolution wave and hydrodynamics modelling in coastal areas: operational applications for coastal planning, decision support and assessment, *Nat. Hazards Earth Syst. Sci.*, 16, 1499–1518, <https://doi.org/10.5194/nhess-16-1499-2016>, 2016.
- Seyfried, M. S. and Wilcox, B. P.: Scale and the nature of spatial variability: Field examples having implications for hydrologic modeling, *Water Resour. Res.*, 31, 173–184, <https://doi.org/10.1029/94wr02025>, 1995.
- Sharma, K. D. and Murthy, J. S. R.: A practical approach to rainfall-runoff modelling in arid zone drainage basins, *Hydrolog. Sci. J.*, 43, 331–348, <https://doi.org/10.1080/02626669809492130>, 1998.
- Simard, R. and Ecuyer, P. L.: Computing the two-sided Kolmogorov-Smirnov distribution, *J. Stat. Softw.*, 39, 1–18, <https://doi.org/10.18637/jss.v039.i11>, 2011.
- Sobol, I. M.: Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comput. Exp.*, 1, 407–414, 1993.
- UNEP – United Nations Environment Programme: World Atlas of Desertification, Edward Arnold, London, 69 pp., 1992.
- Wang, G. and Ren, L.: A Contrastive Study of Simulation Results between GWSC-VMR and Hybrid Runoff Model in Dianzi Basin, in: International Conference on Environmental Science and Information Application Technology, 4–5 July 2009, Wuhan, China, 583–588, 2009.
- Xiao, Z., Liang, Z., Li, B., Hou, B., Hu, Y., and Wang, J.: New flood early warning and forecasting method based on similarity theory, *J. Hydrol. Eng.*, 24, 04019023, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001811](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001811), 2019.
- Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., and Stewart, A.: Understanding uncertainty in distributed flash flood forecasting for semiarid regions, *Water Resour. Res.*, 44, 61–74, <https://doi.org/10.1029/2007WR005940>, 2008.
- Young, C. B., Nelson, B. R., Bradley, A. A., Smith, J. A., Peters-Lidard, C. D., Kruger, A., and Baeck, M. L.: An evaluation of NEXRAD precipitation estimates in complex terrain, *J. Geophys. Res.-Atmos.*, 104, 19691–19703, <https://doi.org/10.1029/1999jd900123>, 1999.
- Zhao, L., Xia, J., Xu, C. Y., Wang, Z., Sobkowiak, L., and Long, C.: Evapotranspiration estimation methods in hydrological models, *J. Geogr. Sci.*, 23, 359–369, <https://doi.org/10.1007/s11442-013-1015-9>, 2013.
- Zhao, R. J.: Watershed Hydrological Model: Xin'anjiang Model and Shanbei Model, Water and Power Press, Beijing, China, 1983.
- Zhao, R. J.: The Xinanjiang model applied in China, *J. Hydrol.*, 135, 371–381, [https://doi.org/10.1016/0022-1694\(92\)90096-E](https://doi.org/10.1016/0022-1694(92)90096-E), 1992.