Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

Natural Hazards and Earth System Sciences
Open Access
Discussions

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

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

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

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

Abstract. Flood forecasting and simulation in semiarid regions are always poor, and a single criterion assessment provides limited information for decision making. Here, we propose a multicriterion assessment framework combining the absolute relative error, the flow partitioning and the confidence interval estimated by the Hydrologic Uncertainty Processor (HUP) to assess the most striking feature of an event-based flood—the peak flow. The physically based model MIKE SHE and three conceptual models (two models with a single runoff generation mechanism, the Xianjiang model (XAJ) and the Shanbei model (SBM), and one model with the mixed runoff generation mechanism, the vertically mixed runoff model (VMM)) are compared in terms of flood modeling performance in four semiarid catchments (Qiushui River, Qingjian River, Tuwei River and Kuye River) in the middle Yellow River. Our results show that VMM has a better flood estimation performance than the other models, and under the multicriterion assessment framework, the average acceptance of flood events accounts for 58%, but when absolute relative error 20% is used as the performance criterion, its figure is only 41% in four semiarid catchments.

1 Introduction

Arid and semiarid regions account for approximately one-third of the global land surface and half of China. A trend towards a warmer climate has increased global incidences of intense precipitation events. Arid and semiarid regions, i.e., areas where the annual rain is less than 250 and 250–500 mm/a, respectively, are particularly vulnerable to this change in climate (Khomsi et al., 2016; Yatheendradas et al., 2008). More than 50% of flood-related casualties occur in these regions worldwide (Brito and Evers, 2016).

The hydrological model plays an important role in flood simulation and forecasting (Devia et al., 2015). Many studies focus on the improvement and estimation of hydrologic models in humid catchments, and similar work for semiarid catchments is relatively few (Jiang et al., 2015). The runoff generation mechanism of semiarid catchments is complex and may be dominated by infiltration excess and saturation excess mechanisms simultaneously (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 network of rain gauges is generally sparse. Rainfall data are important inputs for hydrologic models, and the high temporal-spatial rainfall variability combined with sparse rain gauges makes 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, but the low spatial and temporal resolutions of the products limit their applicability to subdaily rainstorms (Dinku et al., 2007). Weather radar has high spatial resolution (1 km) and temporal resolution (15 min). However, the radar costs are too high to be used on a large scale in semiarid areas (Young et al., 1999).

Literature on the subdaily modeling of rainfall-runoff is limited in semiarid catchments. Due to quick 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 to compare three hydrologic models (the lumped SCS model, simple distributed SCS model, and distributed 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 IHACRES model, distributed IHACRES model, and a 2-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, it is urgent to search for useful information based on the limited accuracy of modeling results to serve as flood warnings and to improve decision making.

In this study, a multicriterion assessment framework combining the absolute relative error, partitioning flow zones and the confidence interval estimated by Hydrologic Uncertainty Processor (HUP) is proposed to provide information for engineers' decision making. Four hydrologic models (the vertically mixed runoff model (VMM), MIKE SHE, Xinanjiang model (XAJ) and Shanbei model (SBM)) are compared on the basis of the performance of modeling results in four catchments in the middle Yellow River. The global sensitive analysis (GSA) method PAWN is used to analyze the parametric sensitivity of VMM. The remainder of the paper is organized as follows. The section below presents a description of the study area and the data set used. The VMM model, model calibration, initial conditions of the VMM model, parametric analysis, multicriterion assessment framework, comparison of models and model validation are introduced in the methodology section. The results of the model comparison, sensitivity analysis and analysis of the multicriterion assessment framework for VMM are described in the results and discussion section, with the final section presenting the conclusions of the study.

2 Study area and data

The 4 selected study catchments are all key tributaries located in the middle Yellow River, China (Fig. 1). The maximum and minimum areas of catchments are 1989 km^2 and 8706 km^2, respectively. The average annual temperature is 6–14°C. The average annual precipitation is 1010–1150 mm, of which 65–80% is concentrated in summer. The rainfall is generally characterized by high intensity and short duration. The average annual evaporation is 1010–1150 mm. All selected catchments are semiarid due to an aridity index between 2.31 and 2.78 (UNEP, 1992). More information about the catchments is listed in Table 1.

The study catchments have poor vegetation coverage and serious soil erosion. Some hydrologists have studied daily and monthly rainfall runoff, and few attempts have been applied to model hourly flood flows (Cheng et al., 2009; Wang et al., 2006). With the rapid increase in population and economic development, flood disasters have received increasing attention. Hence, modeling floods and providing a useful method for decision makers in charge of flood defense are essential and urgent.

Streamflow and rainfall data are from 1983 to 2009. Hourly streamflow data came from hydrological stations. Nine rainfall gauge stations in the Qiushui River catchment, 15 rainfall gauge stations in the Qingjian River catchment, 12 rainfall gauge stations in the Tuwei River and 41 rainfall gauge stations in the Kuye River were selected. Thiessen polygon methods were used to interpolate the rainfall data.

3 Methodology

3.1 Vertically mixed runoff model

The VMM is a conceptual hydrologic model developed by Bao and Wang (1997), and has been used in many areas in China, especially in semiarid and subhumid catchments (Bao and Zhao, 2014; Li, 2018; Li et al., 2018a; Wen and Cai, 2015). Compared with other conceptual models, such as the XAJ model (Zhao, 1992), Sacramento Soil Moisture Accounting Model (SSMA) (Burnash et al. 1973), etc., VMM is able to simulate the saturation excess and infiltration excess runoff generation mechanisms simultaneously. As shown in Fig. 2, 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 VMM. *FA* is regulated by the tension water storage capacity curve, part of which supplements the tension water storage (*W*), and the rest of which forms the below-ground runoff (*RB*) (including subsurface runoff and ground runoff). Here, the calculation of runoff generation is described briefly. For more detailed information about VMM, you can refer to Bao and Zhao (2014).

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

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

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

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 by Eq. (2).

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

where

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

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

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

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

RB can be calculated by Eq. (5).

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

where

$$W^* = WMM\left[1 - \left(1 - \frac{W}{WM}\right)\right]^{\left(\frac{1}{B}+1\right)} \tag{6}$$

$$WMM = WM(1 + B) \tag{7}$$

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

The outlet runoff R can be calculated as follows.

$$R = RS + RB \tag{8}$$

3.2 Initial condition of event-based VMM

The initial condition has important effects in modeling flood events. Therefore, initial tension water storage ($W0$) and initial free water storage ($S0$) should be determined at the beginning of each flood event calculation. We use daily rainfall data over the period of 1983–2009 to simulate the daily streamflow with daily based VMM in each catchment's outlet. The simulation results are accepted when achieving water balance compared with observed streamflow. Hence, tension water storage W and free water storage S per day during 1983–2009 can be achieved, and the initial conditions ($W0$ and $S0$) of each flood event are determined based on daily values of W and S.

3.3 Model calibration

To consider the spatial variation in rainfall, the subcatchments are divided, and the lumped model VMM is applied to each

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

subcatchment. Because only one streamflow gauge station is present in each catchment, the spatial variation in model parameters cannot be recognized by calibration. Thus, the parameters are set uniformly in all the subcatchments. The fourteen parameters (Table 2) of VMM are calibrated by the global optimization algorithm SCE-UA (Duan et al., 1993). The ranges of parameters are determined based on previous literature and prior knowledge (Bao and Zhao, 2014; Li et al., 2018).

In semiarid catchments, due to the rapid rise and fall of floods (usually less than 24 hours), accurate simulation of the full hydrograph is not needed and cannot be realized. 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 there is no need for full fitness of simulated and observed streamflows. (McIntyre and Al-Qurashi, 2009; SHARMA and MURTHY, 1998) used absolute relative error to evaluate model outputs (flow peak and flow volume) for semiarid areas, and the calibrated results indicated that the flow peak results are more accurate than suggested based on NSE. Thus, the simulated hydrograph is reasonable for the majority of flood events. The equations are as follows:

$$E_p = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Q_p^i - Q_{p'}^i\right|}{Q_{p'}^i} \tag{9}$$

$$E_v = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Q_v^i - Q_{v'}^i\right|}{Q_{v'}^i} \tag{10}$$

where E_p and E_v are 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 $Q_{p'}$ are the simulated and measured values of peak flow per event, respectively; and Q_v and $Q_{v'}$ are the simulated and measured values of flow volume per event, respectively.

Simultaneously, constraining the model output with peak flows and flow volumes can be expressed as:

$$E_{pv} = \frac{E_p + E_v}{2} \tag{11}$$

where E_{pv} is the objective value. The closer E_{pv} is to 0, the better the model outputs are. We set the number of iterations to 2000 in the model calibration step.

3.4 Model comparison

To achieve a better performance in event-based rainstorm flood simulations, three hydrologic models, including two conceptual models, XAJ and SBM, and one distributed model, MIKE SHE, are used for comparison with the VMM model. XAJ was developed by (Zhao, 1992) and has a single saturation excess runoff generation mechanism. XAJ has been successfully applied in humid and subhumid catchments (Lü et al., 2013; Wei-jian et al., 2016). We test the performance of XAJ in the semiarid catchments of the middle Yellow River. SBM was developed by Zhao (1983) and has a single infiltration excess runoff generation mechanism. SBM is generally used in semiarid or arid catchments (Bao et al., 2017). MIKE SHE is one of the most widely used physically based distributed hydrologic models (Li et al., 2018b; Refsgaard, 1995; Rujner et al., 2018; Samaras et al., 2016).

3.5 A multicriterion assessment framework of flood events

Due to strong spatially variability of rainfall, complexity of landscape characterizes and lack of enough rain gauges, event-based floods simulation and forecasting in semiarid catchments are very difficult. Although some hydrologists improve flood simulation and forecasting by improving hydrologic models, the improvements are always limited or only are suitable to some specific regions (Collier, 2007). Flood peak is the most significant flood 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 other of that is acceptable (Li et al., 2014; McIntyre and Al-Qurashi, 2009). To provide more information for the decision maker of flood defense, generalized likelihood uncertainty estimation (GLUE) and Bayesian method are used to provided probabilistic forecasting like 95% confidence interval (Christiaens and Feyen, 2002; Li et al., 2017), but the way may not lead to clear decision (Beven, 2007).

In this study, to acquire a utility method for decision maker, we propose a mulit-criterion assessment framework for flood forecasting in the catchments of middle Yellow Rivers. This framework can be described as follows:

(C1) the absolute relative error of peak flow should be less than 20%.

(C2) modeling and observation of peak flows should be in the same flow zone: the observed peak flows Q_p of all flood events in a catchment are 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 respectively: if each $Q_p \leqslant Q_{p25}$, 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 rest flow peaks belongs to medium flow zone. Both of the 25th percentile and 75th percentile is commonly used to distinguish zones.

(C3) the modeling peak flows should fall within one standard deviation (σ) of mean (about 68.3% confidence interval) of peak flows estimated by hydrologic uncertainty processor (HUP), one component of Bayesian approach (detailed can be found in (Krzysztofowicz, 1999; Biondi et al., 2010)).

The key of framework is C2, and C1 is used to avoid errors caused by flow zone boundaries. For example, when $\mathrm{Q_{p75}} = 200$ m^3/s, modelling peak flow equals 198 m^3/s and observed peak flow equals 201 m^3/s, only using the condition C2 may lead the modeling result not to be accepted. So, adding C1 can deal with the problem. C3 is used to test the confidence level of modeling peak flows. A modeling peak flow that can be accepted should satisfy condition C1 or condition C2 and then condition C3.

3.6 Parameter sensitivity analysis

To assess the effects of inputs on model output, sensitivity analysis (SA) was proposed (Saltelli et al., 1989). SA can be classified into GSA and local sensitivity analysis (LSA). Compared with LSA, GSA is able to analyze 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 based on cumulative density function. PAWN has advantages over the parameter ranking and time-consuming nature of other GSA methods (Khorashadi Zadeh et al., 2017). In this study, we use the PAWN method to perform GSA on the VMM model.

Considering $x_{i,j}$ ($i, j = 1, 2, \cdots$, 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 $F_{y_i}(y_i)$:

$$KS(x_{i,j}) = \max_{1 \le j \le n} \left| F_{y_i}(y_i) - F_{(y_i|x_{i,j})}(y_i) \right| \tag{12}$$

As *KS* varies with $x_{i,j}$, the maximum of all possible *KS* is seen as the PAWN index P_i:

$$P_i = \max_{1 \le j \le n} KS(x_{i,j}) \tag{13}$$

P_i ranges from 0 to 1. The closer P_i is to 1, the more sensitive x_i is. A P_i 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, as Pianosi and Wagener (2018) suggested, the number of evaluations is set to 500.

3.7 Model validation

The modeling time step was hourly, and the modeling period was between 1983 and 2009. In the Qiushui River, 20 flood events were selected, and the first 15 events were used for calibration, and the remaining 5 events were used for validation. Similarly, in the Qingjian River, 29 flood events were selected, of which 24 events were used for calibration, and the remaining 5 events were used for validation. In the Tuwei River, 23 flood events were selected, of which 18 events were used for calibration, and the remaining 5 events were used for validation. Finally, in the Kuye River, 28 flood events were selected, of which 23 events were used for calibration, and the remaining 5 events were used for validation.

4 Results and discussion

4.1 Comparison of model results

For flow peaks, the absolute relative error of the peak flow for each model in the four catchments is shown in Fig. 3. Except

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

for the validation period in the Kuye River catchment, it is obvious that VMM performs better than the other models, with lower absolute relative errors for both median and average peak flows. In addition, we can see that VMM has a relatively small range of absolute relative error values for peak flows, which is similar to MIKE SHE. MIKE SHE has a moderate performance in terms of average and median peak flows. Although SBM performs as well as VMM in the Tuwei River catchment, it performs as poorly as XAJ in other catchments, with a large range and a large value of absolute relative error for peak flows.

Table 3 and Table 4 show the average performance in terms of absolute relative error for flow volume E_v and the average performance for lag time for the four models in each catchment, respectively. VMM has the minimum averages of E_v and lag time, with values of 39.01% and 3.05 h, respectively (Table 3 and Table 4). In contrast, XAJ has the maximum averages of E_v and lag time, with values of 58.93% and 4.51 h, respectively. MIKE SHE and SBM have similar performances in terms of average E_v and lag time.

Overall, VMM has the best performance in event-based flood modeling in the four studied catchments in the middle Yellow River, and XAJ has the worst performance. MIKE SHE is slightly superior to SBM. Although MIKE SHE is a distributed hydrologic model with more complex structures and more explicit physical meaning than the conceptual model VMM, it does not necessarily achieve better results than conceptual models, which is consistent with other studies, because distributed models lack sufficiently high-resolution data (Beven, 2002, 2011; Michaud and Sorooshian, 1994; Seyfried and Wilcox, 1995). Both infiltration excess and saturation excess can be simulated in VMM, which may be 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 VMM

The GSA method PAWN is applied to estimate the influence of parameter uncertainty on the model output results. Fig. 4 (a) and Fig. 4 (b) show the SA results of all study catchments for the objective Eq. (9) and Eq. (11), respectively. The most sensitive parameters – *CS*, *IM* and *KE* – are not affected by different objective functions. The rankings of other parameters are slightly influenced by different objective functions, such as *CG* expect for WM. *WM* ranks sixth when Eq. (11) is the objective function and 12th when Eq. (9) is the objective function. *WM* controls the tension water content in the soil, which determines the amount of rainfall stored in the soil and the generation of runoff. Therefore, it is reasonable to conclude that when the weight of the flow volume is added in Eq. (9), which can be expressed as Eq. (11), the ranking of *WM* increases, in this case to sixth place.

4.3 Multicriterion assessment framework of VMM

The multicriterion assessment framework we propose is applied to assess the ability of VMM to model flood peaks in four catchments. The framework requires that an accepted flood event must meet the requirements of C1 or C2, and C3 must be satisfied. Flood events conforming to conditions for C1, C2 or C3 can be obtained from Fig. 5. We find that the peak flows of

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

most flood events fall between the 15.85th percentile and the 81.45th percentile (68.3% confidence interval) estimated by HUP, which means that the VMM modeling results satisfy C3 fairly well. Under the premise of satisfying C3, the number of modeling events satisfying C2 is slightly more than that satisfying C1. Under the multicriterion assessment framework, 15 of the 20 (75.0%) flood events in the Qiushui River catchment, 12 of the 29 (41.4%) flood events in the Qingjian River catchment, 15 of the 23 (65.2%) in the Tuwei River catchment, and 16 of the 28 (57.1%) in the Kuye River catchment can be accepted. In this case, the average acceptance rate for the four catchments is 58%, which is greater than the acceptance rate of 41% for C1.

The multicriterion assessment framework can provide more reasonable information for decision makers. Taking the 13th flood event of the Kuye River catchment as an example, the observed and modeled peak flows are 1230 m^3/s and 1510 m^3/s, respectively. As shown in Fig. 5, the absolute relative error for peak flow is greater than 20%, and the peak flows do not fall in the 68.3% confidence interval, but they are in the same zone, i.e., the medium flow zone. For the Kuye River catchment, it is reasonable to believe that the peak flows 1230 m^3/s and 1510 m^3/s correspond to the same level according to the known flood peak data, which is the role played by C2. Although the dividing flow zone method of C2 is coarse, it is convenient and is beneficial for flood defense.

5 Conclusion

In this study, a multicriterion assessment framework of flood peaks is proposed with the VMM model in four catchments in the middle Yellow River. The main conclusions are as follows:

(1) Compared with MIKE SHE, XAJ and SBM, VMM has better performance in terms of modeling event-based floods in semiarid catchments in the middle Yellow River.

(2) In the four catchments, by PAWN analysis of VMM, *CS*, *IM*, and *KE* are the most sensitive parameters and are not affected by the choice of objective functions, whereas *WM* is the most sensitive parameter.

(3) The multicriterion assessment framework can provide more reasonable information than single criteria (such as absolute relative error of peak flows) when engineers need to make decisions regarding semiarid catchments.

The condition C2 will be affected by the number of known peak flows if data availability is limited. The framework is suitable for semiarid regions with poor modeling results and can enrich people’s decision making.

Code availability

We have shared the MATLAB code of VMM model at https://doi.org/10.4211/hs.c5232287d5c04bfb8cac5ce4e391ea0f.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (grant nos. 2016YFC0402706 and 2016YFC0402709) and the Major Program of the National Natural Science Foundation of China (grant no. 41730750). We would like to thank Francesca Pianosi (University of Bristol) for providing the program code of PAWN at https://www.safetoolbox.info/pawn-method/. We also thank the anonymous reviewers as their comments have largely improved this work.

References:

Andersen, F. H.: Hydrological modeling in a semi-arid area using remote sensing data. University of Copenhagen: Copenhagen, Denmark, 2008.

Burnash, R. J. C., Ferral, R. L., McGuire, R. A. A.: Generalized streamflow simulation system, conceptual modeling for digital computers, 1973.

Bao, W.: Improvement and application of the Green-Ampt infiltration curve, Yellow River, 09, 1–3, 1993. (In Chinese)

Bao, W. and Wang, C.: Application of the Vertically mixed runoff model, Journal of China Hydrology, 03, 19-22, 1997. (In Chinese)

Bao, W. and Zhao, L.: Application of Linearized Calibration Method for Vertically Mixed Runoff Model Parameters, J. Hydrol. Eng., 33(4), 85–91, 2014.

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(7), 284–293, 2017.

Beven, K.: Surface water hydrology—runoff generation and basin structure, Rev. Geophys., 21(3), 721–730, 1983.

Beven, K. and Freer, J.: A dynamic TOPMODEL, Hydrol. Process, 15(10), 1993–2011, 2001.

Beven, K.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, Hydrol. Process, 16(2), 189–206, 2002.

Beven. K.: Environmental modelling: An uncertain future?, CRC press, 2007.

Beven, K.: Rainfall-runoff modelling: the primer, John Wiley & Sons, 2011.

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(1–4), 38–54, 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(4), 1019–1033, 2016.

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(8), 3873–3878, 1973.

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(2), 373–391, 2002.

Cheng, L., Xu, Z., Luo, R., and Mi, Y.: SWAT application in arid and semi-arid region: a case study in the Kuye River Basin, Geogr. Res., 01, 65–73, 2009. (In Chinese)

Collier, CG.: Flash flood forecasting: What are the limits of predictability? ， Q. J. Roy. Meteor. Soc., 133(622), 3–23, 2007.

Devia, G. K., Ganasri, B. P., and Dwarakish, G. S.: A review on hydrological models. Aquatic Procedia, 4, 1001–1007, 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(7), 1503–1526, 2007.

Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, J. Optimiz. Theory App., 76(3), 501–521, 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(7), 279, 2018.

Jiang, Y., Liu, C., Li, X., Liu, L., and Wang, H.: Rainfall-runoff modeling, parameter estimation and sensitivity analysis in a semiarid catchment, Environ. Modell Softw., 67, 72–88, 2015.

Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resour. Res., 35(9), 2739-2750, 1999.

Khomsi, K., Mahe, G., Tramblay, 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(5), 1079–1090, 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. Modell Softw., 91, 210–222, 2017.

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, 2013.

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(5), 1026-1035, 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. Env. Res. Risk A., 31(5), 1045–1059, 2017.

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, 2017.

Li, D.: Hydrologic model: the vertically mixed runoff model (vmm), HydroShare, https://doi.org/10.4211/hs.c5232287d5c04bfb8cac5ce4e391ea0f, 2018.

Li, D., Liang, Z., Hou, B., Li, B., and Wang, J.: Application of vertical hybrid model on flood forecast and early-warning of Qiushui River, Yellow River, 06, 24–28, 2018a. (In Chinese)

Li, D., Liang, Z., Li, B., Lei X., and Zhou Y.: Multi-objective calibration of MIKE SHE with SMAP soil moisture datasets, Hydrology Research, https://doi.org/10.2166/nh.2018.110, 2018b.

McIntyre, N., and Al-Qurashi, A.: Performance of ten rainfall – runoff models applied to an arid catchment in Oman, Environ. Modell Softw., 24, 726–738, 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, 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, 1994.

Morris, M.D.: Factorial sampling plans for preliminary computational experiments, Technometrics, 33, 161–174, 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, 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, 1970.

Pianosi, F., and Wagener, T.: A simple and efficient method for global sensitivity analysis based on cumulative distribution functions, Environ. Modell Softw., 67, 1–11, 2015

Pianosi, F., and Wagener, T.: Distribution-based sensitivity analysis from a generic input-output sample, Environ. Modell Softw., 108, 197–207, 2018.

Pilgrim, D. H., Chapman, T. G. and Doran, D.G.: Problems of rainfall-runoff modelling in arid and semiarid regions, Hydrological Sciences Journal, 33, 379–400, 1988.

Refsgaard, C. J.: Mike she, Computer models of catchment hydrology, 809–846, 1995.

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, 2018.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto M.: Sensitivity Analysis in Practice. J. Am. Stat. Assoc., 101, 398–399, 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, 1–32, 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, 1995.

United Nations Environment Programme (UNEP), World Atlas of Desertification, 1992.

Wen, S., and Cai, T.: Application of improved vertical mixed runoff model in Dalinghe river basin, Water Resources & Hydropower of Northeast, 33(4), 45–48, 2015. (In Chinese)

Wang, G., Wang, J., Jing, X., and Li, H.: Application of SIMHYD model in the Qinjiang River Basin, Yellow River, 03, 29–30, 2006. (In Chinese)

Sharma, K. D., and Murthy, J. S. R.: A practical approach to rainfall-runoff modelling in arid zone drainage basins, Hydrological Sciences Journal, 43(3), 331–348, 1998.

Simard, R., and Ecuyer, P. L.: Computing the two-sided Kolmogorov-Smirnov distribution, J Stat. Softw., 39, 1–18, 2011.

Sobol, I. M.: Sensitivity estimates for nonlinear mathematical models, Mathematical modelling and computational experiments, 1, 407–414, 1993.

Guo, W., Wang, C. Ma, T., Zeng, X. and Yang, H.: A distributed Grid-Xinanjiang model with integration of subgrid variability of soil storage capacity. Water Science and Engineering, 9, 97–105, 2016.

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(5), 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, 1999.

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

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

Table 1. Characteristics of the four catchments

Catchment	Area (km^2)	Outlet station	Area* (km^2)	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 a catchment controlled by the outlet station in the table.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

Natural Hazards and Earth System Sciences
Open Access
Discussions

Table 2. Calibrated parameters of VMM

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

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

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

Table 3. The performance (in terms of absolute relative error) for peak flow E_v in each catchment in the four models

Unit: %

	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	45.2	43.15

*The average E_v of the four catchments for each model

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

Table 4 The lag time of peak flow in the four catchments in the four models Unit: h

	Qiushui 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

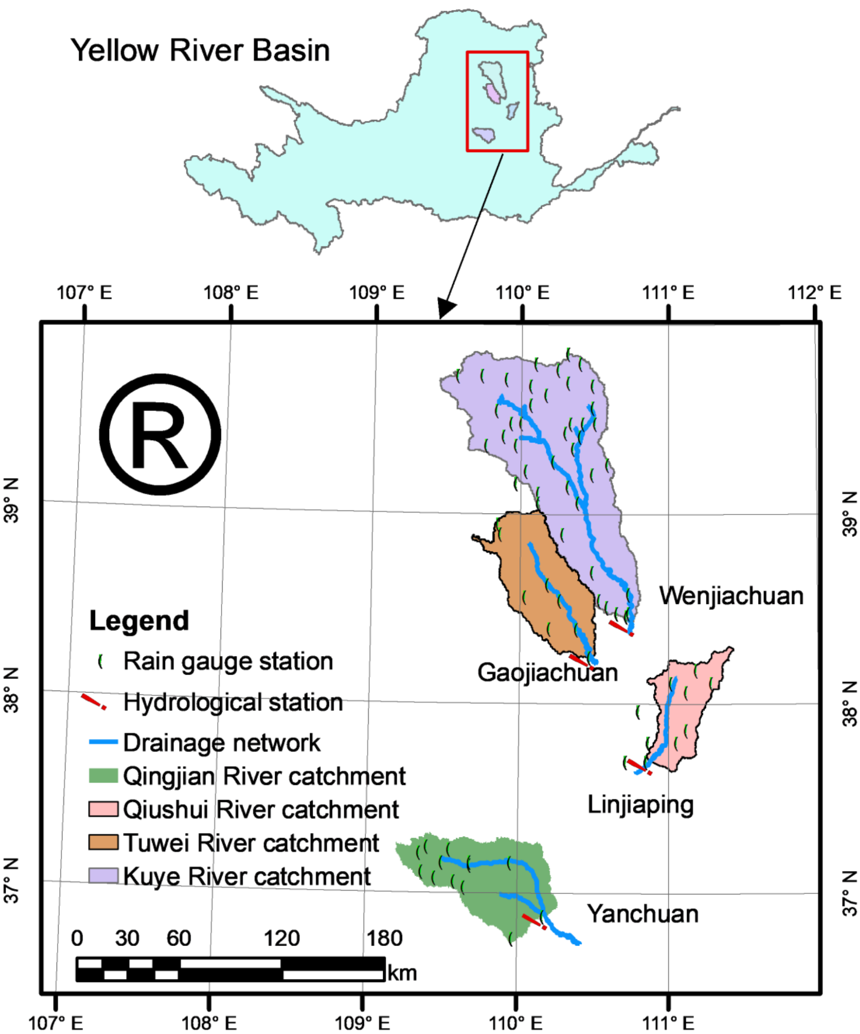

Figure 1: Location of Qingjian River catchment, Qiushui River catchment, Tuwei River catchment and Kuye River catchment.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

Natural Hazards and Earth System Sciences
Open Access
Discussions

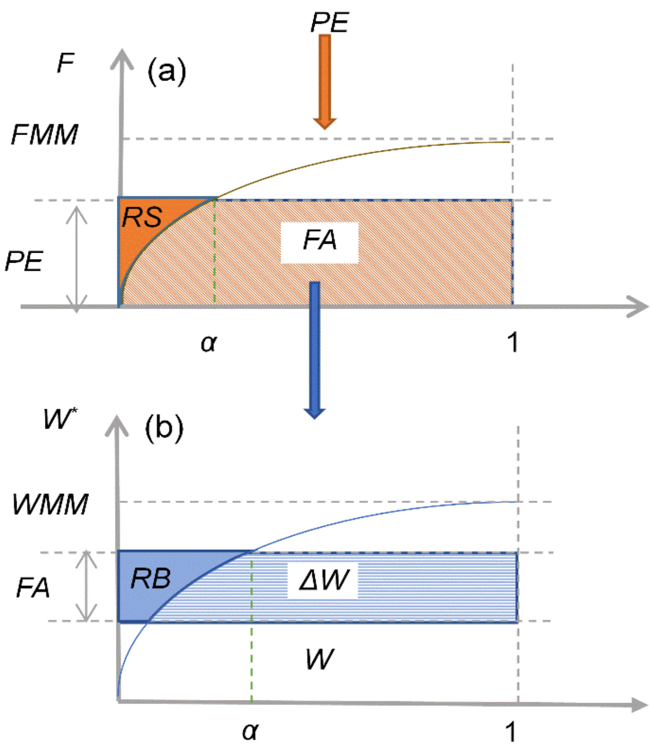

Figure 2: Runoff generation module in VMM. (a) The infiltration capacity curve; (b) the tension water content storage capacity curve. α is the fracture area that is saturated, and *F* represents the point infiltration capacity.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

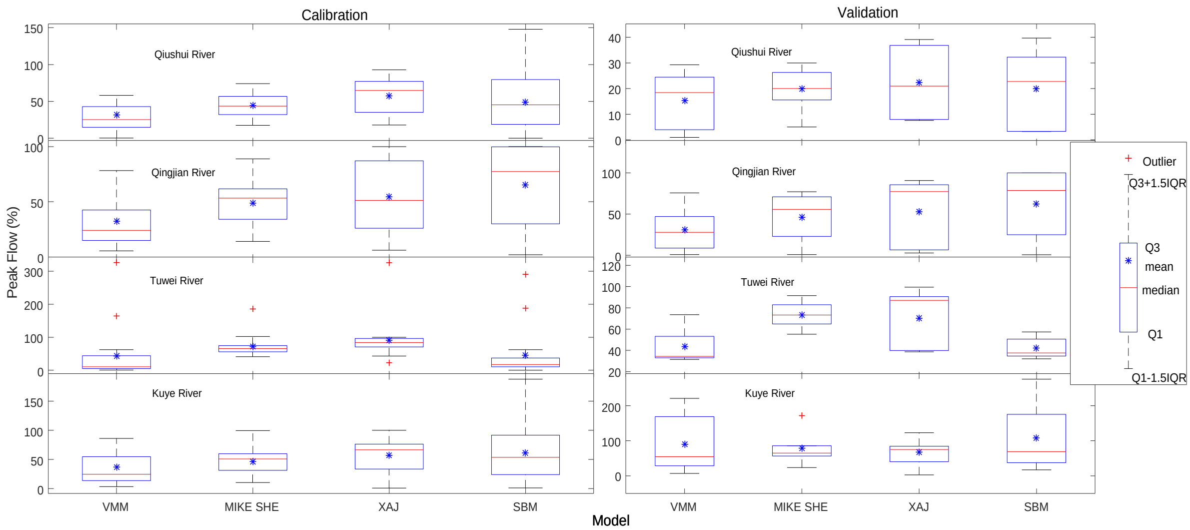

Figure 3: Boxplot of peak flows in the four catchments; Q1 and Q3 mean the first quantile and third quantile, respectively; interquartile range (IQR) = Q3 − Q1; an outlier is defined as an extreme value that exceeds the IQR.

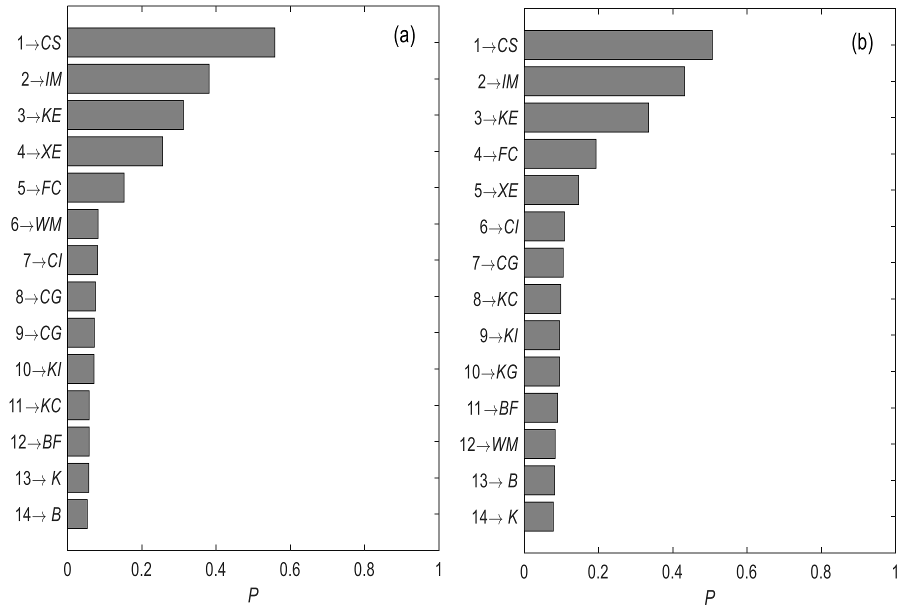

Figure 4: Sensitivity rankings of VMM parameters based on PAWN value *P* for different objective functions; (a) E_{pv} as the objective function; (b) E_p as the objective function. The numbers on the ordinate represent the sensitivity ranking, where the larger the value, the greater the sensitivity.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-402
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 8 February 2019

Natural Hazards and Earth System Sciences
Open Access
Discussions

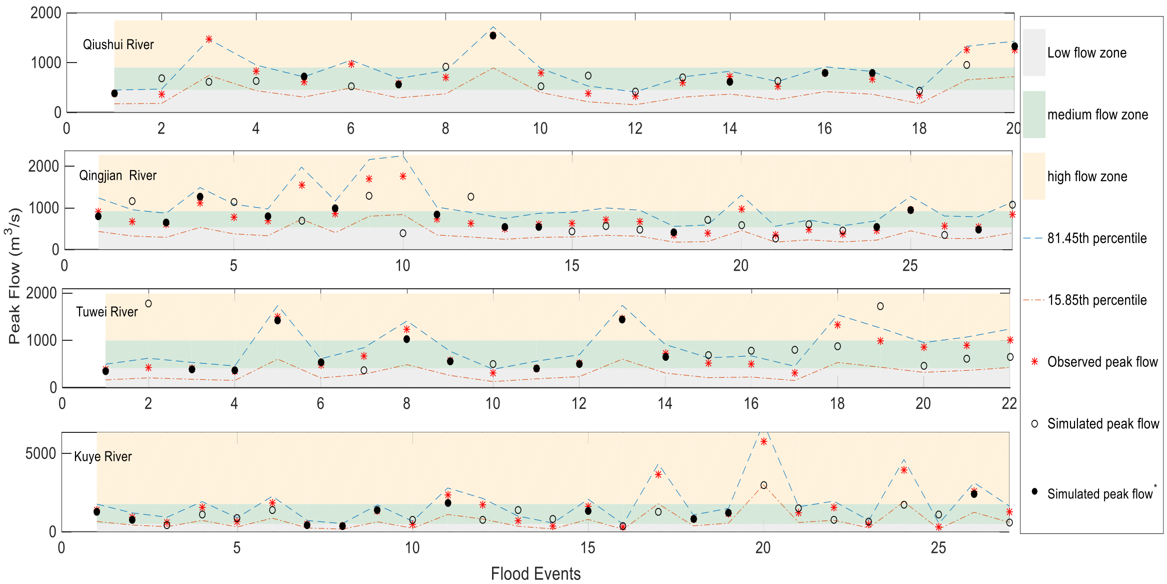

Figure 5: Multicriterion assessments of event-based floods with VMM in the four catchments. Simulated peak flow[*] means the value of the modeled peak flow meets the condition C1.