Multicriteria 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 multicriteria assessment framework that combines the absolute relative error, flow partitioning and confidence interval estimated by the Hydrologic Uncertainty Processor (HUP) to assess the most striking feature of an event-based flood: the peak flow. The vertically mixed runoff model (VMM) is compared with three models, one physical-based model, the MIKE SHE and two conceptual models, the Xinanjiang model (XAJ) and the Shanbei model (SBM). The 100 flood events in the four catchments of the Yellow River are modeled over the period of 1983–2009. Our results show that the VMM has a better flood estimation performance than the other models, and under the multicriteria assessment framework, the average acceptance of flood events is improved. In addition, the framework may provide reasonable flood early warning information for decision-makers.

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 the global incidence 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).

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 similar work for semiarid catchments is relatively few (Jiang et al., 2015). The runoff generation mechanism of semiarid catchments is 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 a 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, although 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 for large-scale 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 that produced the largest peak flows from 1957–1977 to compare three hydrologic models, i.e., 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, determining how to use modeling results with limited accuracy to provide guidance for flood early warning is important.

In this study, a multicriteria assessment framework that combines the absolute relative error, flow partitioning and the confidence interval estimated by Hydrologic Uncertainty Processor (HUP) is proposed to provide information for engineers' decision making. Four hydrological models: the vertically mixed runoff model (VMM), the MIKE SHE model, Xinanjiang model (XAJ) and Shanbei model (SBM), are compared based on the performance of the 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 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 model methodology, initial condition 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 model. The final section presents 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 ranges from 6–14°C. The average annual precipitation ranges from 1010–1150 mm, and 65 to 80% is concentrated in summer (Li et al., 2019; Li and Huang, 2017). The rainfall is generally characterized by high intensity and short duration. The average annual evaporation ranges from 1010–1150 mm. All selected catchments are semiarid due to 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 serious soil erosion, and the average sediment concentration reaches 126 kg m^{-3} 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 modeling is from 1983 to 2009. Continuous streamflow and rainfall data are 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 are 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 and 41 rainfall gauging stations in the Kuye River were selected. 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), Sacramento Soil Moisture Accounting Model (SSMA) (Burnash et al. 1973), among others, the VMM is able to simulate 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 rest forming the 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 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):

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

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 the VMM

The initial condition has important effects in modeling flood events. The VMM model was run continuously from 1983 to 2009 for each catchment. Two initial values are the initial tension water storage (*W0*) and the initial free water storage (*S0*) should be determined. Both of them represent the moisture content of the soil and were assumed to be zero due to the dry conditions at 00:00:00 on January 1, 1983. 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 simulation was daily between flood events and hourly within flood events.

3.3 Model calibration

To consider the spatial variation in rainfall, the subcatchments are divided and the VMM model is applied to each subcatchment. Due to the fact that only one streamflow gauging station is available for each catchment, the spatial variation in each catchment's parameters cannot be determined by calibration. Thus, the parameters are set uniformly in all subcatchments. The fourteen parameters (Table 2) of the 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 h), accurate simulations of the full hydrograph are not needed and cannot be realized. 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 Murphy (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 flow peak results are more accurate than suggested based on the NSE. Thus, the simulated hydrograph is reasonable for the majority of flood events. The equations are as follows:

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

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

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:

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

where E_{pv} is the objective value. The closer E_{pv} is to 0, the better the model outputs. 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, 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 (Cheng et al., 2006; Lü et al., 2013;). SBM was developed by Zhao (1983) and has a single infiltration excess runoff generation mechanism. 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, MIKE SHE originated from the Système Hydrologique Européen (SHE) program, and it is a deterministic, physically based distributed hydrologic model that can simulate surface water flow, unsaturated flow and saturated flow (Jayatilaka et al., 1998). MIKE SHE 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 for flood events

Due to strong spatially variability of rainfall, complex landscape characteristics, insufficient rain gauges, and dispersion, flood event simulation and forecasting in semiarid catchments are very difficult. Although some hydrologists improve flood simulations and forecasting by improving hydrologic models, the improvements are always limited or are suitable for only certain 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 Bayesian framework with the Markov Chain Monte Carlo sampling are used to provide probabilistic forecasting, such as the 95% confidence 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 useful method for decision maker, we propose a multicriteria assessment framework for flood forecasting in the catchments of middle Yellow River. This framework can be described as follows:

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

(C2) modeled and observed peak flows should be in the same flow zone: the observed peak flow 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: if $Q_p \leq Q_{p25}$, then the peak flow Q_p belongs to the low flow zone; and 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 the 25th percentile and 75th percentile are commonly used to distinguish zones.

(C3) The modeled peak flows should fall within one standard deviation (σ) of the mean (approximately 68.3% confidence interval) peak flow estimated by the HUP, one component of the Bayesian forecasting system detailed in Krzysztofowicz (1999)

and Biondi et al. (2010).

The key of the framework is C2, and C1 is used to avoid errors caused by flow zone boundaries. For example, when $Q_{p75} = 200$ m^3/s, the modeled peak flow equals 198 m^3/s, and the observed peak flow equals 201 m^3/s. However, using only condition C2 may lead to inappropriate model results; adding C1 can help address the problem. C3 is used to test the confidence level of modeled peak flows. A modeled 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 the model output, a sensitivity analysis (SA) was proposed (Saltelli et al., 1989). The SA can be classified into a GSA and local sensitivity analysis (LSA). Compared with the LSA, the 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 (derived from the authors' names), 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 use the PAWN method to perform a 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 included in 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, 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.

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, except for the validation period in the Kuye River catchment shown in Fig. 3 (h), Figures 3 (a)–(g) reveal that the VMM has lowest values for both calibration and validation; in most cases, MIKE SHE performs better than the XAJ and SBM, i.e., Figures 3 (a), (b), (c), (d), (g), (h). Except for the good performance in the Tuwei River catchment, the SBM is as poor as the XAJ in other catchments. In terms of ranges of the absolute relative errors for peak flows, the VMM and MIKE SHE have relatively small ranges (Figs. 3(a), (c), (d), (g)); and the SBM and XAJ have large ranges in most cases (Figs. 3(a), (b), (c), (d), (g)).

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. 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 has the maximum average 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.

The analysis of Fig. 3, Table 3 and Table 4 above indicates that the VMM has the best performance for flood modeling in the four studied catchments of the middle Yellow River and the XAJ has the worst performance. In addition, MIKE SHE is slightly superior to the 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 because distributed models lack sufficiently high-resolution data, and this finding 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, 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 the VMM

The GSA method PAWN is applied to estimate the influence of parameter uncertainty on the model output results. Figure 4 (a) and Figure 4 (b) show the SA results of all study catchments for the objective function E_p (Eq. 9) or E_{pv} (Eq. 11), respectively. The higher the ranking is, the more sensitive the parameters. Parameters *CS*, *IM* and *KE* have the highest rankings whether the

objective function of the VMM model 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. *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 E_p, which can be expressed as E_{pv}, the ranking of *WM* increases, in this case to sixth place.

4.3 Multicriteria assessment framework of the VMM

The multicriteria assessment framework we propose is applied to assess the ability of the VMM to model flood peaks in four catchments. The framework requires that an accepted flood event should meet one of the requirements of C1 and C2; in addition, C3 needs to be satisfied simultaneously. The observed peak flows and the modeled peak flows under the conditions C1, C2 or C3 are shown in Fig. 5. We find that the majority of peak flows for the observations and modeling 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 greater than that satisfying C1. Under the multicriteria assessment framework, 15 of the 20 (75.0%) modeled flood events in the Qiushui River catchment, 12 of the 29 (41.4%) of that in the Qingjian River catchment, 15 of the 23 (65.2%) of that in the Tuwei River catchment, and 16 of the 28 (57.1%) of that in the Kuye River catchment can be accepted. The average acceptance rate for the four catchments is 58%, which is greater than the acceptance rate of 41% under the single criterion C1, a common assessment for peak flows.

The multicriteria assessment framework may provide more reasonable and reliable flood early warning 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 (d), the absolute relative error for peak flow is greater than 20%, and the peak flow does not fall in the 68.3% confidence interval; however, these parameters 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 beneficial for flood defense.

5 Conclusions

In this study, a multicriteria assessment framework of flood peaks is proposed with the vertically mixed runoff model(VMM) in four catchments in the middle Yellow River. The main conclusions are as follows.

(1) Compared with the distributed model MIKE SHE and the two conceptual models Xinanjiang (XAJ) and Shanbei (SBM), the VMM has better performance for modeling flood events in semiarid catchments of the middle Yellow River.

(2) In the four catchments, the parameters confluence coefficient of surface flow (*CS*), impermeable area (*IM*), and residence

time of Muskingum (*KE*) are the most sensitive based on an analysis by the global sensitivity method PAWN; in addition, the sensitivity ranking of the parameter *WM* related with the soil moisture capacity is the most affected by the objective functions.

(3) The multicriteria assessment framework may provide more reliable flood early warning information than single criterion (such as absolute relative error of peak flows) when engineers need to make decisions in semiarid catchments.

The condition C2, which divides peak flows into three flow zones, will be affected by the number of known peak flows when data availability is limited. The framework is suitable for semiarid regions with poor modeling results and can provide guidance for decision making.

Code availability

We have shared the MATLAB code of the 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 no. 2016YFC0402706) and National Natural Science Foundation of China (grant nos. 41730750, 41877147). 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, whose comments have largely improved this work.

References:

Andersen, F. H.: Hydrological modeling in a semi-arid area using remote sensing data, Ph.D. 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(7), 284–293, 2017.

Bao, W.: Improvement and application of the Green-Ampt infiltration curve, Yellow River, 9, 1–3, 1993. (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.

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

Beven, K. J.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, Hydrol. Process., 16(2), 189–206, 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., 2011.Beven, K. J., and Freer, J.: A dynamic TOPMODEL, Hydrol. Process., 15(10), 1993–2011, 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(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.

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, 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(2), 373–391, 2002.

Collier, C. G.: Flash flood forecasting: What are the limits of predictability?, Q. J. Roy. Meteor. Soc., 133(622), 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(8), 3873–3878, 1973.

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.

Jayatilaka, C., Storm, B., and Mudgway, L.: Simulation of water flow on irrigation bay scale with MIKE-SHE, J. Hydrol., 208, 108-130, 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. Modell. Softw., 67, 72–88, 2015.

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.

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

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, 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., https://doi.org/10.1002/ldr.3274, 2019.

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, D.: Hydrologic model: the vertically mixed runoff model (vmm), HydroShare, https://doi.org/10.4211/hs.c523228 7d5c04bfb8cac5ce4e391ea 0f, 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., 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, 2017.

Li, Z. J., and Zhang, K.: Comparison of three GIS-based hydrological models, J. Hydrol. Eng., 13, 364–370, 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, 2013.

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, Hydrolog. Sci. J., 33, 379–400, 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, 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(6), 1499–1518, 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.

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

Simard, R., and Ecuyer, P. L.: Computing the two-sided Kolmogorov-Smirnov distribution, J. Stat. Softw., 39(11), 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.

United Nations Environment Programme (UNEP): 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, Wuhan, China, 4–5 July, 583–588, 2009.

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), 61–74, 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, 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. (In Chinese)

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

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 the catchment controlled by the outlet station in the table.

Table 2. Calibrated parameters of the VMM

Symbol	Definition	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.

Table 3. 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

Table 4 Lag time of the 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

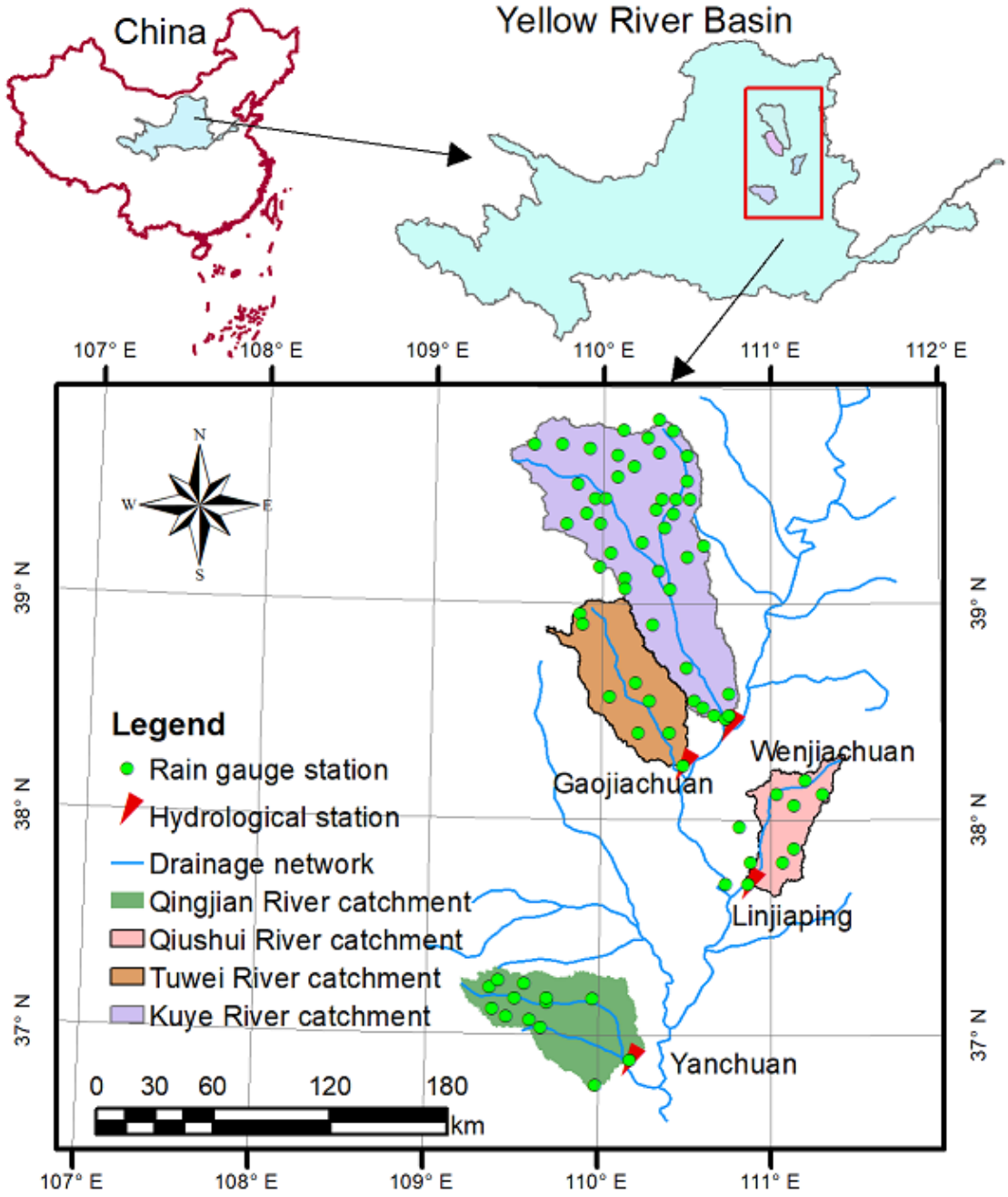

Figure 1: Location of the Qingjian River catchment, Qiushui River catchment, Tuwei River catchment and Kuye River catchment.

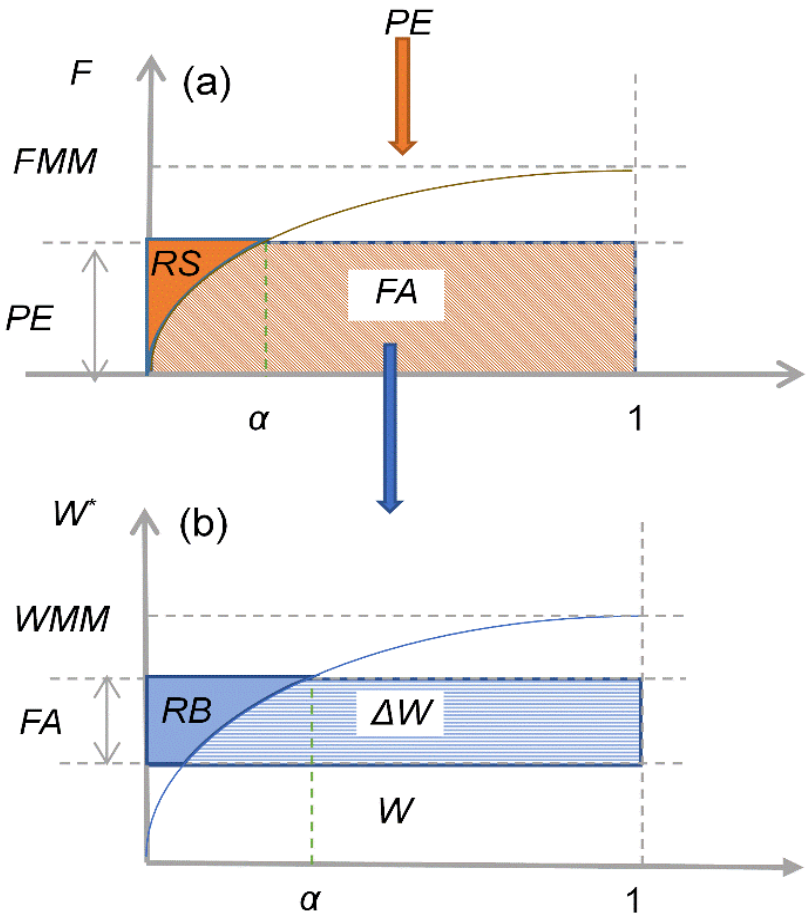

Figure 2: Runoff generation module in the VMM. (a) Infiltration capacity curve; and (b) tension water content storage capacity curve. *α* is the fracture area that is saturated, and *F* represents the point infiltration capacity.

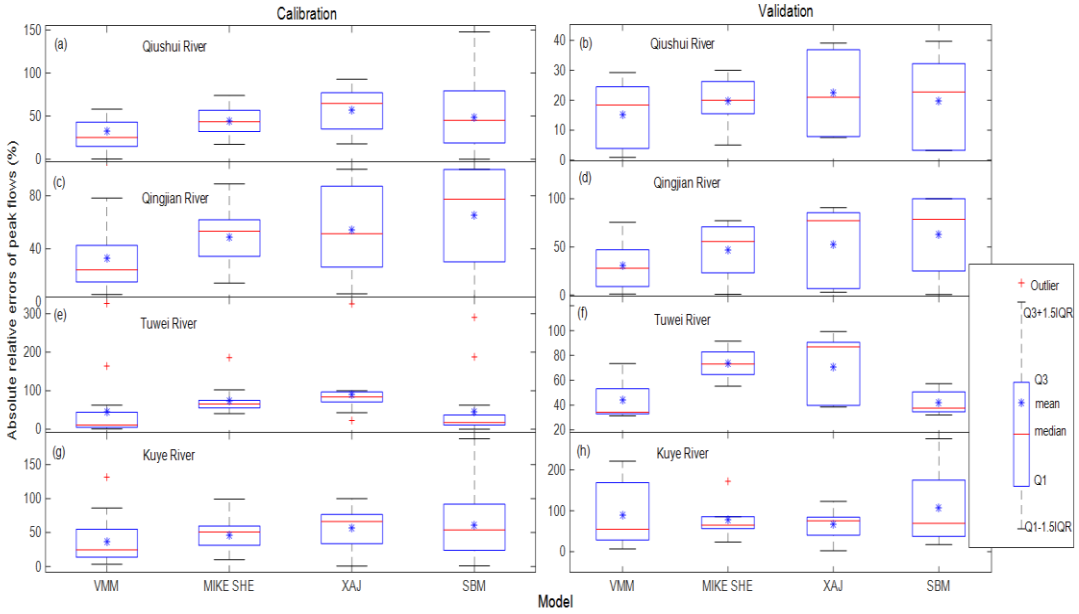

Figure 3: Boxplot of absolute relative errors of peak flows in the four catchments; Q1 and Q3 mean the first quantile and third quantile, respectively; interquartile range (IQR) = Q3 − Q1; and an outlier is defined as an extreme value that exceeds the IQR.

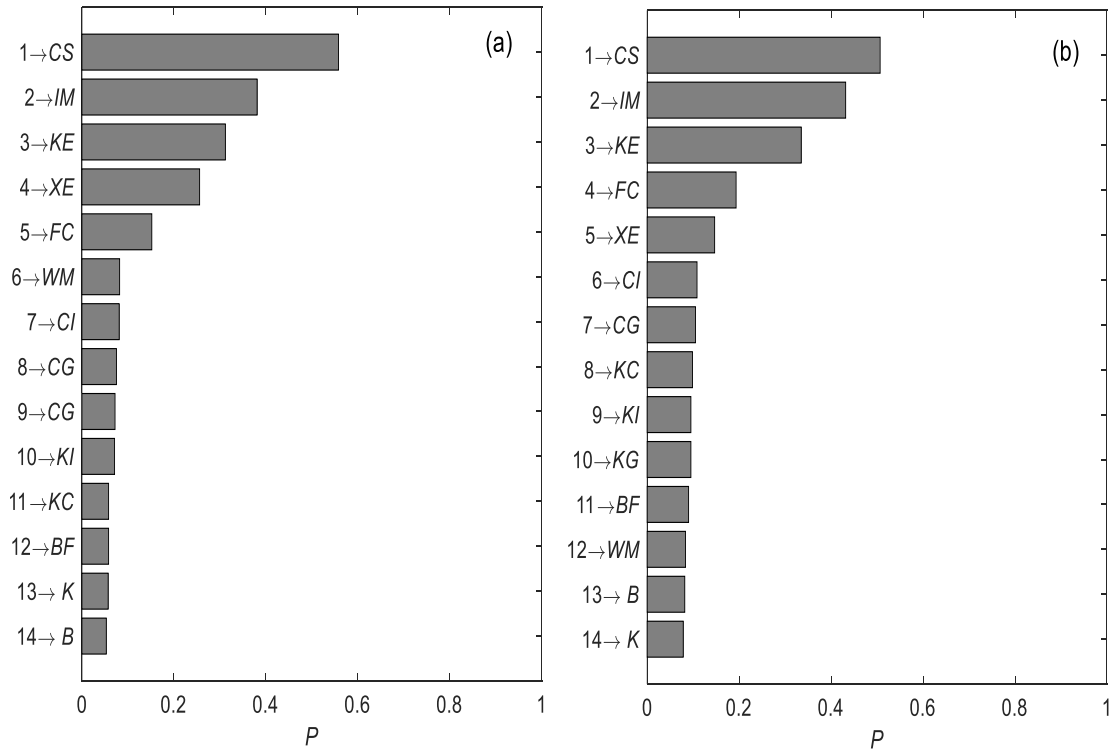

Figure 4: Sensitivity rankings of the VMM parameters based on the global sensitivity analysis method PAWN for different objective functions: (a) E_{pv} as the objective function, and (b) E_p as the objective function. The value P is used to assess the sensitivity degree of the parameter with PAWN method,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

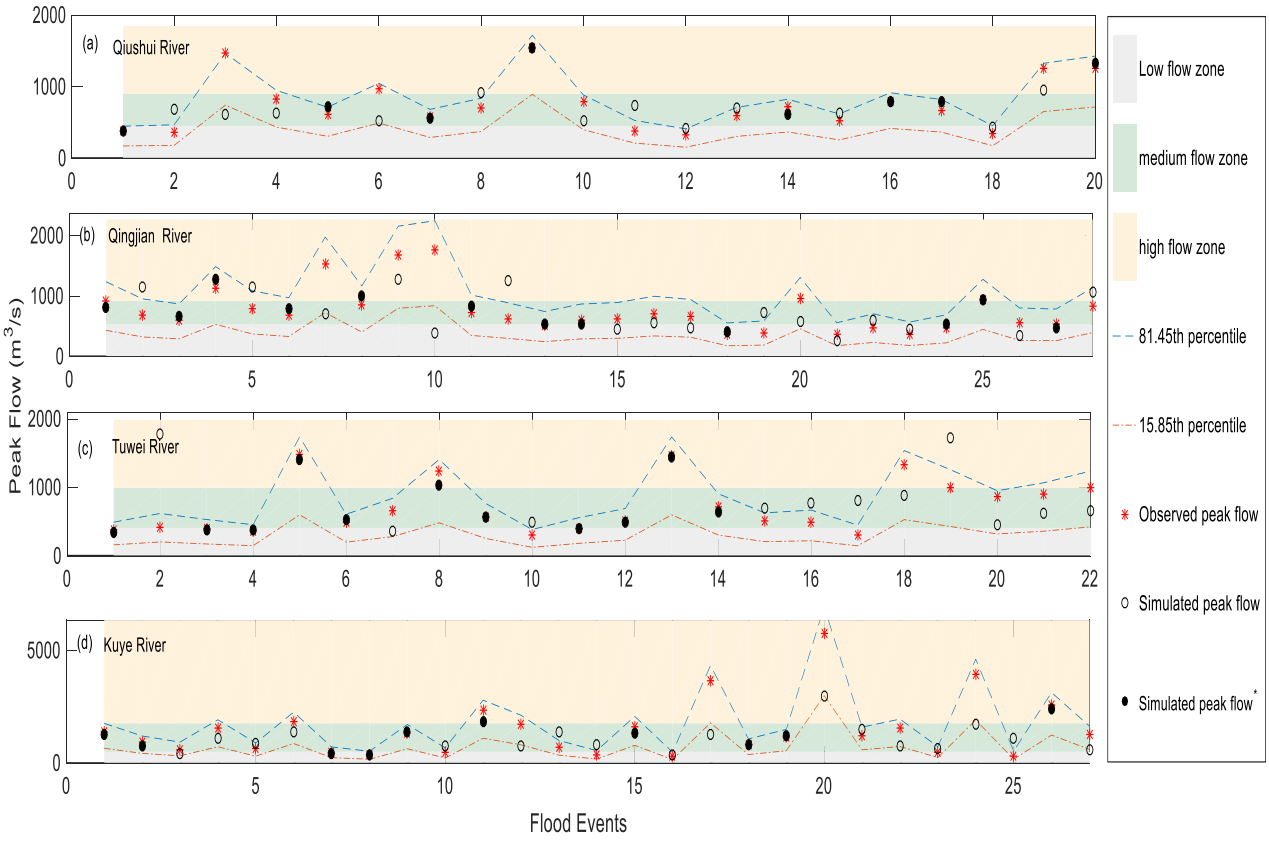

Figure 5: Observed peak flows (red asterisk) and simulated peak flows (solid ball and circle) with the VMM for each catchment under the conditions C1, C2 and C3. Flood peaks conforming to the condition C1 is represented by a solid ball; the three flow zones (low, medium and high flow zones) divided by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