Exploring the Application of Flood Scaling Property in Hydrological Model Calibration

YANCHEN ZHENG,^{a,b} JIANZHU LI,^a TING ZHANG,^a YOUTONG RONG,^c AND PING FENG^a

^a State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin, China

^b Department of Civil Engineering, University of Bristol, Bristol, United Kingdom

^c School of Geographical Sciences, University of Bristol, Bristol, United Kingdom

(Manuscript received 26 June 2021, in final form 20 September 2021)

ABSTRACT: Model calibration has always been one major challenge in the hydrological community. Flood scaling properties (FS) are often used to estimate the flood quantiles for data-scarce catchments based on the statistical relationship between flood peak and contributing areas. This paper investigates the potential of applying FS and multivariate flood scaling properties [multiple linear regression (MLR)] as constraints in model calibration. Based on the assumption that the scaling property of flood exists in four study catchments in northern China, eight calibration scenarios are designed with adopting different combinations of traditional indicators and FS or MLR as objective functions. The performance of the proposed method is verified by employing a distributed hydrological model, namely, the Soil and Water Assessment Tool (SWAT) model. The results indicate that reasonable performance could be obtained in FS with fewer requirements of observed streamflow data, exhibiting better simulation of flood peaks than the Nash–Sutcliffe efficiency coefficient calibration scenario. The observed streamflow data or regional flood information are required in the MLR calibration scenario to identify the dominant catchment descriptors, and MLR achieves better performance on catchment interior points, especially for the events with uneven distribution of rainfall. On account of the improved performance on hydrographs and flood frequency curve at the watershed outlet, adopting the statistical indicators and flood scaling property simultaneously as model constraints is suggested. The proposed methodology enhances the physical connection of flood peak among sub-basins and considers watershed actual conditions and climatic characteristics for each flood event, facilitating a new calibration approach for both gauged catchments and data-scarce catchments.

SIGNIFICANCE STATEMENT: This paper proposes a new hydrological model calibration strategy that explores the potential of applying flood scaling properties as constraints. The proposed method effectively captures flood peaks with fewer requirements of observed streamflow time series data, providing a new alternative method in hydrological model calibration for ungauged watersheds. For gauged watersheds, adopting flood scaling properties as model constraints could make the hydrological model calibration more physically based and improve the performance at catchment interior points. We encourage this novel method to be adopted in model calibration for both gauged and data-scarce watersheds.

KEYWORDS: Watersheds, Streamflow, Flood events, Hydrology, Satellite observations, Hydrologic models

1. Introduction

As an essential tool to represent the dynamic and complex hydrological processes in watersheds, hydrological models are widely used to support water resource management, planning, and decision-making (Niehoff et al. 2002; Bormann et al. 2007; Huisman et al. 2009; Daggupati et al. 2015). Generally, to provide reliable simulation results for floods with hydrological models, the model parameters should be calibrated carefully. However, the hydrological processes are a result of complex interactions of catchment physical processes, which are highly changeable in both space and time (Gupta 2004; Ayalew et al. 2014). Finding an optimal set of parameters to provide a reasonable streamflow prediction of catchment outlet or streamflow of any catchment interior locations is quite

difficult (Wanders et al. 2014), especially for data-scarce catchments. In addition, even for the gauged watershed, the model parameters that need to be calibrated for different model modules are also increasing with the wide application of distributed hydrological models, which further increases the difficulty of hydrological model calibration.

Beven (2006) believed that only a few parameter sets could meet the multiple calibration criteria simultaneously in model calibration, indicating that the model performance could be improved by adding more constraints. Hence, many studies tried to develop different calibration strategies to constrain the model performance based on all available information of the catchment. Calibrating the hydrological model with multivariable data is one of the common ways to improve the simulation performance among recent research (Chiang et al. 2014; White and Chaubey 2005). Numerous researchers calibrated streamflow and other hydrological variables simultaneously (Koren et al. 2008; Li et al. 2018; Pfannerstill et al. 2017; Rajib et al. 2016; Omani et al. 2017; Tuo et al. 2018; Lu et al. 2014; Wanders et al. 2014; Rajib et al. 2018). Multisite

Denotes content that is immediately available upon publication as open access.

Corresponding author: Jianzhu Li, lijianzhu@tju.edu.cn

DOI: 10.1175/JHM-D-21-0123.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

calibration method is also a widely used calibration strategy (Cao et al. 2006; Zhang et al. 2008, 2010; Niraula et al. 2012; Wi et al. 2015; Fenicia et al. 2016; Hughes et al. 2016).

a. Calibration strategies based on limited flow data for data-scarce catchments

However, sufficient observed measurements are required in multisite and multivariable calibration strategies. In this section, we introduce some other calibration strategies for data-scarce catchments.

Numerous studies are devoted to exploring the minimum streamflow time series length for achieving robust model calibration results. The recorded minimum streamflow time series length varied from 3 months to 8 years among different studies (Yapo et al. 1996; Brath et al. 2004; Perrin et al. 2007). Seibert and Beven (2009) indicated that a few runoff measurements can contain as much of the information as the continuous runoff time series. For instance, a single event or 10 observations during high flows could provide the same information as the continuous 3 months of data (Seibert and McDonnell 2013). They also proved that maximum flows series contain more information than minimum or mean flows. Similarly, Melsen et al. (2014) reported that the season (5 months) with the highest precipitation is sufficient to give a robust simulation of high flows over the full observation period. Singh and Bárdossy (2012) demonstrated that calibration on identified events representing 6%–7% of a 10-yr time series could obtain the reasonable results compared to calibration on the whole time series.

Some researchers further investigated the optimal sampling strategy to extract the most informative runoff measurements for constraining the model. Pool et al. (2017) compared 12 runoff measurements sampling strategy with selection ranging from simply monthly maximum runoff measurements to more complex observation times series data. They demonstrated that flow-duration curves were best estimated with strategies including low and mean flows, while strategies with high runoff magnitudes produced the best simulation of hydrographs.

Another type of calibration strategy is employing regional information or a priori information as model constraints to estimate the parameters (Hundecha and Bárdossy 2004; Göttinger and Bárdossy 2007). The goal of a regionalization approach is to transfer the information from gauged catchments to ungauged catchments based on the statistical relations between catchment characteristics and model parameters (Viviroli and Seibert 2015; Samaniego et al. 2017). This is the most common way to estimate the model parameters for ungauged catchments (Perrin et al. 2008; Rojas-Serna et al. 2016; De Lavenne et al. 2019). Additionally, combining the regionalization method and point flow measurement appears to be a promising method to estimate the model parameters for ungauged catchments. Some studies proved that employing a few runoff observations could improve the model calibration performance (Viviroli and Seibert 2015; Seibert 2010; Rojas-Serna et al. 2016; Pool et al. 2018).

b. Scaling property of floods

Mendoza et al. (2015) mentioned that integrating prior knowledge of hydrologic processes could improve the relatively poor performance of hydrological models. We attempted to consider how to add constraints for the model hydrologic processes based on limited data. In general, observations during wet periods, especially the event peak, contain valuable information for model parameter estimation (Yapo et al. 1996; Seibert and Beven 2009; Melsen et al. 2014; Pool et al. 2017). The peak discharge reflects various aspects of the rainfall rate, spatial–temporal variability, and watershed characteristics such as soil moisture, land use, and land cover (Ogden and David 2003). The statistical relationship of flood peaks with the size of contributing upstream area, which captures the interactions of the streamflow-generating process, forms the basis of many empirical approaches for the estimation of floods in data-scarce catchments (Ayalew et al. 2014; Farmer et al. 2015). A quantile-based flood scaling property has been applied in the study of design flood estimation along with flood risk assessment (Ishak et al. 2011; Basu and Srinivas 2015) and has been used to develop regional flood frequency (RFF) equations for many decades (Dawdy et al. 2012; Smith et al. 2015; Furey et al. 2016). However, Furey et al. (2016) pointed out that the quantile-based analysis allowed for event mixing, indicating that different flood events data were employed to determine a scaling relationship. Under these circumstances, the flood results could be underestimated. Another type of flood scaling property, namely, event-based flood scaling property, was taken as a more informative and robust method for establishing the relation between the flood peak and the drainage area in a watershed. A wide range of empirical and theoretical studies revealed that the scaling relationship at the flood event scale can be expressed in a power-law form $Q(A) = \alpha A^\theta$, where $Q(A)$ is peak discharge, A is the drainage area, α is the intercept, and θ is the exponent (Ogden and David 2003; Furey and Gupta 2005, 2007; Mandapaka et al. 2009; Ayalew et al. 2014, 2015).

Nevertheless, considering drainage area as the only explanatory variable of event flood peak is not sufficient in some cases. Jothityangkoon and Sivapalan (2001) illustrated that the spatial scaling of flood peaks is closely related to the space–time variability of rainfall–runoff processes through the underlying catchment water balance, including the effects of rainfall, evaporation, antecedent soil moisture storage, and the stream network geomorphology. Al-Rawas and Valeo (2010) explored the effect of various watershed characteristics on peak flood flow based on a large number of observed flood events. Ishak et al. (2011) employed a total of 10 explanatory variables to improve flood predictions. Hence, the so-called multivariate flood scaling property based on a multiple regression framework considering various watershed explanatory variables could help better grasp the rule of flood scaling and avoid estimation bias (Farmer et al. 2015).

c. Scope of this paper

This study aims to develop a novel calibration method that could be used in both gauged and data-scarce catchments for hydrological model calibration. The proposed method employs event-based scaling properties of flood as model constraints, which utilizes the statistical relationship between event flood peak and drainage area. Since estimation of the design flood in ungauged catchments is normally based on the scaling properties of flood, adopting the flood scaling property in model calibration exhibits the potential of applying the proposed method in both gauged catchments and data-scarce catchments. An innovation of this paper is to explore the potential of applying flood scaling properties as model constraints to calibrate a distributed hydrological model for the first time. Moreover, apart from the drainage area, the other catchment descriptors affecting flood peak are also considered as explanatory variables for describing the scaling properties

of flood. Thus, this paper aims to explore the following questions: 1) Does the flood scaling property have the potential to behave as a model constraint to calibrate a hydrological model? 2) How much can this new calibration strategy improve the model performance? 3) If more catchment descriptors are utilized as the explanatory variables, will the model calibration results outperform one that only considers drainage area?

2. Study area and data

a. Study area

Four watersheds, namely, Zijingguan watershed, Fuping watershed, Shifu watershed, and Kuancheng watershed, which are all located in the same typical temperate continental and semiarid climate region in north China, were selected

FIG. 1. Geographic location and detailed subbasin division map of the four study catchments: (a) Zijingguan watershed, (b) Shifu watershed, (c) Fuping watershed, and (d) Kuancheng watershed.

TABLE 1. Main catchment characteristics information on the study areas.

ID	Watershed name	Area (km ²)	No. of subbasins	Average annual catchment precipitation (mm)	Average annual discharge (m ³ s ⁻¹)
ZJG	Zijingguan watershed	1760	11	546	8.2
FP	Fuping watershed	2175	15	522	21.8
SF	Shifu watershed	1031	7	563	7.3
KC	Kuancheng watershed	1670	9	488	10.3

as the study catchments. The geographic locations of these four study catchments are shown in Fig. 1, and the main catchment characteristics are listed in Table 1. The drainage areas of these four study areas range from 1031 to 2175 km² with an average of 1659 km². Forest and grass are the dominant land cover type of these study areas. In these four catchments, short heavy rainfall is the main driving factor of the flood. Precipitation from July to August accounts for 65%–85% of the annual rainfall. Previous studies have proven that the event-based flood scaling property exists in this region (Li et al. 2013, 2019; Kang et al. 2019; Li et al. 2020).

b. Data

The observed daily streamflow data for the four watershed outlets during the flood seasons (June–September) were available. The corresponding periods of observed daily rainfall data were also collected. The number of rainfall gauges within the study catchment and the records period of both rainfall–runoff data were listed in Table 2. Additionally, daily precipitation data from Climate Forecast System Reanalysis (CFSR) datasets (Fuka et al. 2013; Dile and Srinivasan 2014) were adopted to integrate the missing values of the measured rainfall series for the dry season. Since the rainfall is mostly concentrated in the flood season in the four study catchments, flood events seldom occurred in the nonflood season. Furthermore, several researchers have proved that calibration with short, wet time periods of the observations is sufficient to provide the robust simulation compared to the calibration on the whole time series (Singh and Bárdossy 2012; Melsen et al. 2014; Seibert and McDonnell 2013; Seibert and Beven 2009). More importantly, our research goal is to compare the calibration results between different calibration strategies, rather than achieve the best performance. Hence, the available streamflow and rainfall records data are enough to obtain the reasonable calibration results.

1) SELECTION OF FLOOD EVENTS

Several flood events from the observed flood series were selected as key research objects to verify the flood scaling property in the study areas. Many previous studies indicated that a 2-yr return period flood, namely, natural bank-full discharge, could be considered as a threshold flow symbolizing the occurrence of flood events, which is often employed as a global minimum flood protection standard (Weeink 2010; Scussolini et al. 2016; Zheng et al. 2020). Thus, in this study, all flood events in the calibration periods with the flood peak larger than 2-yr return period flood were selected as significant flood events. The special number of identified flood events for four study catchments can be found in Table 2.

2) SWAT MODEL CONFIGURATION

The land use and land cover maps for the study areas were collected from the Institute of Geographic Sciences and Resources of the Chinese Academy of Sciences (<http://www.resdc.cn/>). The Harmonized World Soil Database (HWSD) v1.1 soil dataset (FAO/IIASA/ISRIC/ISS-CAS/JRC 2009) was utilized to extract the soil groups. The weather data spanning 1979–2014, including air temperature, wind speed, relative humidity, and solar data, provided by CFSR datasets (Fuka et al. 2013; Dile and Srinivasan 2014), were downloaded from the website <https://globalweather.tamu.edu/>.

Based on the homogeneity of land characteristics, the delineation of subbasins was performed by employing the ArcSWAT tool. Reasonable threshold values were adopted for soil, slope, and land cover to delineate the hydrological response units (HRUs) (Srinivasan et al. 2010; Sexton et al. 2010; Han et al. 2012; Her et al. 2015) for four study catchments. A total of 500 parameter sets were sampled from initial range by Latin hypercube sampling (LHS). The simulation results of these 500 parameter sets were utilized to conduct the global sensitivity analysis to screen out the

TABLE 2. Rainfall–runoff data records information for four study catchments.

ID	No. of rainfall gauges	Period of records	Warm-up periods	Calibration periods	Validation periods	No. of selected events in calibration periods
ZJG	8	1979–2012	1979–81	1982–99	2000–12	12
FP	9	1979–2000	1979–80	1981–95	1996–2000	8
SF	4	1992–2014	1992–93	1994–2008	2009–14	9
KC	6	1992–2014	1992–93	1994–2008	2009–14	10

TABLE 3. Potential parameters and their sensitivity ranking for the four study catchments. An asterisk indicates that this parameter is significantly sensitive in this catchment, and this parameter was selected to be calibrated in further analysis.

No.	Parameter name ^a	Description	Ranges	Sensitivity ranking			
				ZJG	FP	KC	SF
1	r_CN2.mgt	Curve number	−0.2 to 0.2	1*	1*	1*	1*
2	v_ALPHA_BF.gw	Base flow recession constant	0–1	7*	7*	9	11
3	v_GW_DELAY.gw	Estimated groundwater delay time (days)	0–500	12	11	10	8*
4	v_GWQMN.gw	Threshold shallow depth required for return flow to occur (mm)	0–5000	6*	6*	5*	4*
5	v_ESCO.hru	Soil evaporation compensation factor	0–1	8*	10	12	9*
6	v_CH_K2.rte	Effective hydraulic conductivity for the main channel	−0.01 to 150	2*	2*	3*	2*
7	v_ALPHA_BNK.rte	Base flow alpha factor for bank storage (days)	0–1	4*	4*	4*	7*
8	r_SOL_K.sol	Saturated hydraulic conductivity	−0.25 to 0.25	5*	5*	6*	6*
9	r_SOL_BD.sol	Bulk density of soil	−0.5 to 0.5	10	8*	7*	3*
10	v_SFTMP.bsn	Snowmelt base temperature (°C)	−5 to 5	9*	9	11	12
11	v_CH_N2.rte	Manning's <i>n</i> value for the main channel	0–0.3	3*	3*	2*	5*
12	v_CANMX.hru	Maximum canopy storage	0–100	11	12	8*	10

^a An “r_” means a relative change (existing or initial parameter value is multiplied by 1 + given value within the range) to the initial parameter values; “v_” means a replacement to the initial parameter values.

sensitive parameters. The initial parameter range was determined based on the previous research, in which the Soil and Water Assessment Tool (SWAT) model was also adopted to simulate the streamflow in the Zijianguan basin and its adjacent basin (Bu et al. 2018; Wang et al. 2021). The parameter range of some parameters, which are not used in the mentioned studies, are set to some widely used parameter range in SWAT-model-related research (Kouchi et al. 2017; Abbaspour et al. 2015; Chilkoti et al. 2018). Table 3 lists the potential parameters and their sensitivity ranking for the four study catchments. Only those significant sensitive parameters were selected to be calibrated in the comparison of different calibration scenarios, while the other parameters were fixed with the optimal value according to the first 500 runs model simulation results. The first 2 or 3 years were utilized as the warm-up period. Among the rest time periods of data, 70% of the data were allocated for calibration, and 30% for validation, as shown in Table 2. Due to the large proportion of missing data in Zijianguan watershed after 2000, the verification period for this catchment is divided for a relatively longer period to ensure that there are enough observed data for comparison.

3) COLLECTION OF CATCHMENT DESCRIPTORS

To better describe the flood peaks with employing multivariate flood scaling property, a total of 28 catchment descriptors were collected in this paper, which can be divided into two categories. The first category is the internal watershed characteristics, depending on the long-term controls such as the properties of the river network and other topographical factors. This type of catchment descriptor normally can be delineated and computed based on DEM and land use data. Since the watershed topography and stream network geomorphology exhibit limited changes over long periods, the internal catchment descriptor basically remains unchanged for the selected flood events.

The other type of catchment descriptor varies in time, depending on event characteristics such as external rainfall

characteristics, antecedent watershed soil moisture conditions, and evaporation. Consequently, rainfall characteristics (rainfall intensity, duration, etc.) were calculated based on the observed precipitation data for the selected flood events. Other catchment descriptors reflecting the antecedent watershed conditions, such as soil moisture content and evaporation data, were extracted from GLDAS-Noah datasets (Rodell et al. 2004; Beaudoin and Rodell 2015) due to the lack of observed data in the study areas. GLDAS-Noah v2.0 datasets, which are available at <https://giovanni.gsfc.nasa.gov/>, currently cover the period from January 1948 to December 2014 and contain 36 available variables, with a 3-h temporal interval and 0.25° spatial resolution. Five soil moisture content indexes and five evaporation indexes were selected from the GLDAS-Noah datasets. In this paper, the instant soil moisture content data at the closest time step before the date of each rainfall–runoff event were derived to reflect the antecedent wetness condition of the watershed. Both the mean values of soil moisture content data and evaporation data during each rainfall–runoff event were also collected. The area-weighted method was employed to calculate the average soil moisture content and evaporation value. As a result, a total of 28 catchment descriptors were adopted in this study, including 9 internal and 19 external catchment descriptors, as listed in Table 4.

3. Method

a. Identification of catchment descriptors

With drainage area being treated as the principal catchment descriptor in most studies (Meigh et al. 1997; Menabde and Sivapalan 2001; Jothityangkoon and Sivapalan 2001; Lima and Lall 2010; Lee and Huang 2016; Furey and Gupta 2007; Al-Rawas and Valeo 2010), the scaling property of flood generally takes the form expressed in Eq. (1). However, the flood peak is the result of a complex rainfall–runoff process under the influence of rainfall, watershed characteristics, and other

TABLE 4. List of the potential catchment descriptors influencing peak discharges.

Classification	No.	Factors	Abbreviation	Unit
Watershed and stream network characteristics	1	Drainage area	DA	km ²
	2	Drainage density	DD	km km ⁻²
	3	Basin slope	BS	‰
	4	Basin perimeter	BP	km
	5	Basin width	BW	km
	6	Stream length	SL	km
	7	Basin length	BL	km
	8	Shape factor	SF	— (dimensionless)
	9	Average forest area proportion	AFS	%
Rainfall characteristics	10	Average rainfall intensity	ARI	mm h ⁻¹
	11	Maximum rainfall intensity	MRI	mm h ⁻¹
	12	Antecedent precipitation index	API5	mm
	13	Rainfall duration	RD	h
Soil moisture data of a time step before the rainfall–runoff events	14	Soil moisture content of 0–10 cm	TSM0_10	kg m ⁻²
	15	Soil moisture content of 10–40 cm	TSM10_40	kg m ⁻²
	16	Soil moisture content of 40–100 cm	TSM40_100	kg m ⁻²
	17	Soil moisture content of 100–200 cm	TSM100_200	kg m ⁻²
	18	Root zone soil moisture	TRZSM	kg m ⁻²
Mean value of soil moisture data during the rainfall–runoff events	19	Soil moisture content of 0–10 cm	MSM0_10	kg m ⁻²
	20	Soil moisture content of 10–40 cm	MSM10_40	kg m ⁻²
	21	Soil moisture content of 40–100 cm	MSM40_100	kg m ⁻²
	22	Soil moisture content of 100–200 cm	MSM100_200	kg m ⁻²
	23	Root zone soil moisture	MRZSM	kg m ⁻²
Mean value of evaporation data during rainfall–runoff events	24	Evapotranspiration	EV	kg m ⁻² s ⁻¹
	25	Potential evaporation rate	PET	W m ⁻²
	26	Direct evaporation from bare soil	DEBS	W m ⁻²
	27	Canopy water evaporation	CWE	W m ⁻²
	28	Transpiration	TR	W m ⁻²

factors, so flood scaling property based on multiple regression analysis is also employed in this study as an objective function to avoid the bias caused by only considering the watershed area. The multiple regression equation is expressed in Eq. (2). Both forms of flood scaling property are considered as objective functions in model calibration to conduct further analysis. Employing the power-law equation to describe the relationship between the flood peak and the explanatory variables is the common and typical method of applying the flood scaling property (Ogden and David 2003; Furey and Gupta 2005, 2007). More importantly, previous studies have proven that the event-based flood scaling property in the form of power-law equation exists in the study area (Li et al. 2013, 2019; Kang et al. 2019; Li et al. 2020):

$$Q(A) = \alpha A^\theta, \quad (1)$$

$$Q = b_0 \times X_1^{b_1} \times X_2^{b_2} \times X_3^{b_3} \times \cdots \times X_q^{b_q}, \quad (2)$$

where Q is the flood peak discharge ($\text{m}^3 \text{s}^{-1}$), A is the drainage area, α is the coefficient, and θ is the exponent; b_0 is a constant, X_1 – X_q are the explanatory variables, namely, catchment descriptors, such as precipitation intensity, soil moisture content, and stream density, and b_1 – b_q are the regression coefficients.

The best subset regression method was performed on catchment descriptors with event flood peak to find the

optimal input combination of catchment descriptors. The goal of best subset selection is to choose the best subset of variables such that the resultant regression model has the best prediction accuracy (Zhu et al. 2020). Compared to the common variable selection method, best subset regression tests all possible combinations of the potential variables in a regression equation and identifies the best solution (Neter et al. 1989). This method is quite suitable for the feature selection problem with fewer candidate variables in total. In recent years, the best subset selection method has been frequently applied in hydrology community (Howard et al. 2010; Lacombe et al. 2014; Okcu et al. 2016).

For measuring the regression model performance, adjusted R^2 , Bayesian information criterion (BIC) (Schwarz 1978) and Mallows' coefficient (C_p) (Mallows 1973) were adopted as the evaluation metrics. As the modification version of R^2 , adjusted R^2 is more reasonable when evaluating model fit with many explanatory variables. The higher the adjusted R^2 , the better the model. To avoid information redundancy or multicollinearity problems among various catchments descriptors (Al-Rawas and Valeo 2010; Ishak et al. 2011), BIC and C_p criterion were employed to select the best performing model with the least number of input variables. A small BIC value implies a good selection of subset. The model with the lowest C_p value is preferred.

b. Objective functions and calibration criteria

When the flood scaling property is taken as the objective function, the flood peaks from each reach outlet for each selected flood events along with its corresponding contributing upstream area is employed to fit the power-law equation. The objective functions of flood scaling property are expressed in Eqs. (3) and (4). Figure 2 demonstrates the process of adopting flood scaling property in SWAT model calibration:

$$R_{fsj}^2 = \frac{1}{n} \sum_{i=1}^n R_{fsi}^2, \quad (3)$$

$$\max \{ R_{fsj_1}^2, R_{fsj_2}^2, R_{fsj_3}^2, \dots, R_{fsj_m}^2 \}$$

subject to

$$R_{fsjk}^2 > 0.6 \quad (k = 1, 2, 3, \dots, m)$$

where R_{fsi}^2 denotes the R^2 of each power-law equation for flood scaling property, and R_{fsj}^2 represents the mean value of R_{fsi}^2 ; n is the number of selected flood events, namely, the number of power-law equations; m is the simulation times of the SWAT model in each iteration.

For instance, 12 flood events were selected for the Zijingguan watershed. Consequently, a total of 12 power-law equations along with the R_{fsj}^2 of 12 flood events can be acquired for each model simulation. Li et al. (2020) have performed event-based flood scaling analysis in the Zijingguan watershed with the simulated subbasin peak discharge

using the Hydrologic Engineering Center Hydrologic Modeling System (HEC-HMS). They reported that R^2 varies for 14 different flood events in the range from 0.329 to 0.980 with an average of 0.735. In this study, the threshold of the mean value of 12 power-law equations R_{fsj}^2 was set as 0.6, which is considered as the basic requirement to verify whether this model simulation meets the flood scaling property. The larger the R_{fsj}^2 of 12 power-law equations, the better the model performance.

For multivariate flood scaling property, the flood peaks from each reach outlet of all selected flood events along with its corresponding catchment descriptors were extracted to fit the logarithmic form of the multiple regression equation. The multiple linear regression (MLR) equation and the objective functions are expressed in Eqs. (5) and (6):

$$\log Q = \log(b_0) + b_1 \log(X_1) + b_2 \log(X_2) + b_3 \log(X_3) + \dots + b_q \log(X_q), \quad (5)$$

$$\max \{ R_{mlrj_1}^2, R_{mlrj_2}^2, R_{mlrj_3}^2, \dots, R_{mlrj_m}^2 \}$$

subject to

$$R_{mlrjk}^2 > 0.6 \quad (k = 1, 2, 3, \dots, m)$$

$$p_{mlrjk} < 0.05 \quad (k = 1, 2, 3, \dots, m)$$

where Q is the peak discharge ($\text{m}^3 \text{s}^{-1}$), X_1 – X_q are the potential catchment descriptors of reach outlets, q is the number of catchment descriptors, and b_0 – b_q are the regression

FIG. 2. Schematic representation of employing flood scaling property in SWAT model calibration.

coefficients of the MLR equation. The term R_{mlrj}^2 denotes the R^2 of j th MLR equation, and p_{mlrj} is the p value of j th MLR equation.

Since the flood peaks from each reach outlet for all selected floods and its corresponding catchment descriptors were extracted to fit the MLR equation, only one MLR equation can be obtained during each SWAT model simulation. For each MLR equation, two basic requirements should be satisfied to achieve the best performance. First, the MLR equation should pass the F test to examine whether the fitting equation is significant. Thus, the p value of the MLR equation should be less than 0.05. Furthermore, a larger F statistic represents better regression fitting. If two MLR equations reach the same R_{mlrj}^2 , the F statistic can be utilized to determine the superiority of the two equations. Similarly, the goodness of fit for multivariate flood scaling property is also determined on the basis of R_{mlrj}^2 . The larger the R_{mlrj}^2 , the better the model performance; $R_{mlrj}^2 > 0.6$ is also utilized as the threshold criterion.

To compare the calibration performance under different objective functions, traditional statistical objective functions were also adopted. As the most widely adopted statistical index in model calibration (Tuo et al. 2018; Cao et al. 2018; Rajib et al. 2018; Odusanya et al. 2019; Lee et al. 2020), the Nash–Sutcliffe efficiency coefficient (NSE) was employed in this study. In addition, the Kling–Gupta efficiency (KGE) was also applied due to its three components structure allowing a better understanding of the model performance (Gupta et al. 2009; Knoben et al. 2019). The equations for NSE and KGE are presented in Eqs. (7) and (8):

$$\text{NSE} = \left[1 - \frac{\sum_{i=1}^n (E_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \right], \quad (7)$$

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \quad (8)$$

where E_i is the i th SWAT-simulated streamflow data; O_i is the corresponding observed streamflow data; $\bar{O}_i$ is the average of the observed streamflow data; $\bar{E}_i$ is the average of the simulated streamflow data; n is the total number of observed values; r is the Pearson correlation coefficient between observed and simulated streamflow data; α is the standard deviation of the simulated streamflow data over the standard deviation of observed streamflow data; and β is the ratio of the mean simulated streamflow data to observed streamflow data.

c. Calibration scenarios design and evaluation metric

To explore the model performance of adopting flood scaling property (FS) and MLR as objective functions, three groups of calibration scenarios were designed as follows.

- 1) As the comparison, traditional NSEs or KGEs were first employed as objective functions separately. These two objective functions are widely used in most of the

research. The observed streamflow data of watershed outlet is needed in these two calibration scenarios.

- 2) Then, the new FS or MLR objective functions were adopted to explore the potential of applying the scaling property of flood in model calibration. It is important to note that only the simulated event flood peaks during each model run need to be extracted to fit the power-law equations with corresponding catchment contributing areas and behaved as objective function in FS calibration scenarios. Observed streamflow data were employed to perform the flood frequency analysis and adopted as the criterion of selecting the significant flood events. For the calibration scenarios with MLR, the observed streamflow time series data are also not required during the calibration processes. But the event flood peaks of watershed outlets are still needed, as there are various catchment descriptors available to be chosen as explanatory variables. We need to figure out which catchment descriptors have an influential impact on event flood peaks. So, the identification of the most influential catchment descriptors on event flood peak should be performed first.
- 3) The combination of FS or MLR and traditional NSE or KGE were designed as another four calibration scenarios, namely, FS&NSE, FS&KGE, MLR&NSE, and MLR&KGE. In these four calibration scenarios, with considering FS or MLR as constraints, the best simulation was selected according to the NSE or KGE value. Under these circumstances, both the effect of flood scaling property and the simulation performance at the watershed outlet are guaranteed. Consequently, the observed streamflow data at the watershed outlet are also required in these calibration scenarios.

Hence, a total of eight calibration scenarios, including NSE, KGE, FS, MLR, FS&NSE, FS&KGE, MLR&NSE, and MLR&KGE, were designed.

For model evaluation criterion, apart from the performance of watershed outlet hydrograph represented by NSE or KGE value, flood frequency curve of annual maximum flood peak was also considered as an evaluation aspect in this study. As one of the widely used hydrological signatures, the flood frequency curve describes the relationship between the frequency and magnitude of the flood peak (Vogel and Fennessey 1994; McMillan 2021), which is also a major concern of the design flood calculation. Flood frequency curve has been adopted to behave as the objective function in model calibration in a growing number of research in recent years. (Westerberg et al. 2011; Pokhrel et al. 2012; Pfannerstill et al. 2014; Pool et al. 2018; Garcia et al. 2017). In the meanwhile, the flood frequency curve is also a critical descriptor of evaluating model performance (Ley et al. 2016; Chilkoti et al. 2018), especially in assessing the comparison results of several calibration strategies (Pool et al. 2017). As employing flood scaling properties as objective functions is a relatively new proposed calibration method, the influence caused by this method on both hydrograph and flood frequency curve would be of interest to us. So, we focus on the comparison results between different calibration scenarios including the

traditional objective functions and also the proposed ones. Since the simulated flood frequency curve could indicate the model performance, we prefer to employ flood frequency curve to behave as evaluation metric, helping us to find the optimal calibration scenario.

Since each flood frequency curve segment describes different hydrological processes (Yilmaz et al. 2008; Pokhrel et al. 2012; Pfannerstill et al. 2014), we partition the flood frequency curve into four segments and evaluate the performance on each segment by calculating the mean relative error (MRE) in this work. The four segments are 1) the very high-flow segment within a range of flow exceedance probability less than 5% (Q_5), 2) the high-flow segment within a range between Q_5 and Q_{20} , 3) the midflow segment within a range between Q_{20} and Q_{70} , and 4) the low-flow segment within a range of flow exceedance probability larger than 70%. The positive and negative of MRE value represents whether the simulated value overestimates or underestimates the observed value. The closer the absolute value of MRE is to 0, the better the model result is. Equation (9) is used to calculate the MRE for each flood frequency curve segment:

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^N \frac{E_i - O_i}{O_i}, \quad (9)$$

where E_i is the simulated data; O_i is the corresponding observed data; and N is the total number of observed values within each segment.

d. SWAT model calibration implementation process

The sequential uncertainty conformity (SUFI-2) optimization algorithm is one of the most widely used calibration approaches in the SWAT model. This algorithm is iterative and could achieve good calibration performance with the smallest number of simulations (Yang et al. 2008; Abbaspour et al. 2015, 2017). The specific schematic coupling of SWAT and SUFI-2 is presented in Fig. 3. More details about SUFI-2 algorithm can be found in Yang et al. (2008) and Abbaspour (2015).

Given that the objective of this paper is to apply flood scaling property in SWAT model calibration, the criterion metric is not the only traditional statistical index. Therefore, on the basis of the SUFI-2 algorithm procedure, the following two steps should be noted to employ the flood scaling property.

- 1) The first difference from the traditional calibration method is related to the extraction of the simulated value. Instead of only extracting the simulated value of the watershed outlet or a certain subbasin, the simulated flood peak of each reach outlet for the selected flood events must be extracted from the simulated subbasin streamflow. For instance, as presented in Fig. 1a, a total of 11 reach outlets were identified in the Zijianguan watershed. The simulated peak discharge at each reach outlet during the selected eight flood events must be extracted from the simulated subbasin streamflow. Moreover, the simulated value at the watershed outlet is also

FIG. 3. Schematic diagram showing the coupling process of SUFI-2 and SWAT and the implementation of flood scaling property during model calibration.

extracted to guarantee the calculation of traditional statistical indices.

- 2) The application of flood scaling property establishes the power-law fitting function between flood peak at each reach outlet and the corresponding contributing upstream area. For multivariate flood scaling property, a multiple linear regression equation is constructed with the flood peak and the identified catchment descriptors. New objective functions have been developed to employ the proposed calibration method, as only traditional statistical objective functions are available in SUFI2_goal_fn.exe in the SUFI-2 algorithm.

4. Results

a. Identification of catchment descriptors

Due to the lack of observed streamflow data of subbasins, the observed peak discharge of each flood event at watershed outlets from four study catchments and the corresponding potential catchment descriptors are employed to perform the best subset regression analysis. Table 5 lists the best performance regression models with adopting different number of catchment descriptors. The value of adjusted R^2 shows an upward tendency with an increase in the number of catchment descriptors, reaching 0.69 in model 5. However, BIC and Cp value obtain the minimum value in model 2 with a total number of four catchment descriptors, indicating that these four catchment descriptors could predict the reasonable results with an acceptable adjusted R^2 . Adding more explanatory variables in regression model may lead to information redundancy or overfitting problems. Interestingly, instead of commonly used drainage area, the shape factor (SF) is identified to be one of the significant catchment descriptors for flood peak. Since SF is computed by dividing the squared basin length by basin area, the results indicate that the shape of catchment has more effect on flood peak prediction than drainage area in the study areas. One possible reason is that the drainage area of these four study areas is close with a small standard deviation. To sum up, a total of four catchment descriptors, namely, SF, MSM10_40, TSM100_200, and EV, are selected as the explanatory variables in multivariate flood scaling calibration scenarios.

b. Different calibration scenarios result with the SUFI-2 algorithm

The three groups of calibration scenarios with a total number of eight are performed by SUFI-2 algorithm. For all eight

calibration scenarios, five iterations are executed with SWAT model running a total of 2500 times under each calibration scenario. The first 500 sets of parameters for each calibration scenario are obtained by LHS in the initial range of SWAT parameters, so the first 500 sets of iterative parameters adopted in the first group of each calibration scenario are identical. However, due to the different constraints of each calibration scenario, the range of new recommended parameters obtained by each iteration varies from different calibration scenarios, so finally, different streamflow simulation results are presented for the eight calibration scenarios. The simulated streamflow under eight calibration scenarios during the calibration period is shown in Fig. 4. Figure 5 presents the performance of two statistical indicators at the watershed outlet with the increase of model iterations.

1) CALIBRATION FOR NSE OR KGE AS OBJECTIVE FUNCTIONS

The mean NSE value for the four study watersheds at watershed outlet reaches 0.59 after five iterations with traditional NSE as the objective function. The KGE indicator achieves 0.65 when considering KGE as a single objective function. The results are not overly encouraging, which could be attributed to the complicated hydrological behaviors in subhumid and semiarid regions and limited observed data (Kannan et al. 2007; Jeong et al. 2010; Hapuarachchi et al. 2011; Ragettli et al. 2017). Thus, model calibration in these four study catchments is more difficult. Given that our research goal is to explore the potential of applying flood scaling property in model calibration, the high accurate model simulation result is not our main purpose. Thus, these two calibration scenarios results are utilized as the comparative experiments, representing the commonly adopted model calibration method. Additionally, as demonstrated in Fig. 4, it can be found that the shape of the hydrograph employing NSE as the objective function tends to be smoother than that employing KGE, while the simulated streamflow under KGE calibration scenarios presents better flood peak simulations.

2) CALIBRATION FOR FS OR MLR AS OBJECTIVE FUNCTIONS

When FS is considered as the single objective function and without constraining the watershed outlet streamflow, both the NSE and KGE indicators of streamflow simulation at four study watershed outlets are not high. The identical situation occurs in another calibration scenario, in which only MLR is considered as the objective function. However, under these

TABLE 5. Selection of regression models for different combination of input catchment descriptor. Bold indicates the best performance in each criterion.

Model	Catchment descriptors used	p value	Adjusted R^2	BIC	Cp
1	TSM100_200, EV	0.003	0.46	-15.1	-5.2
2	SF, MSM10_40, TSM100_200, EV	0.0001	0.64	-25.5	-11.3
3	SF, TSM100_200, TRZSM, MSM0_10, MSM10_40, EV	0.0015	0.66	-22.6	-9.2
4	BL, SF, MRI, TSM0_10, TSM10_40, TSM100_200, EV	0.0004	0.65	-19.4	-7.4
5	BL, SF, MRI, TSM0_10, TSM40_100, TSM100_200, MSM10_40, EV	0.0001	0.69	-22.3	-6.6

FIG. 4. SWAT model simulation by adopting the SUFI-2 algorithm under the eight calibration scenarios in (a) Zijiangguan (ZJG), (b) Fuping (FP), (c) Kuancheng (KC), and (d) Shifu (SF) catchments during calibration periods.

two calibration scenarios, the ability to capture and simulate the flood peak is better than the calibration scenarios with traditional objective functions. The reason may be that by employing FS or MLR as the objective function, the spatial scaling relationship between subbasins is enhanced during the calibration process. Consequently, the simulated flood peak value of each subbasin satisfies the statistical relationship with the basin area, meteorological factors, etc., which improves the simulation accuracy of the flood peak for each subbasin. Figure 6 displays the variation in the goodness of fit, namely, the mean R^2 of the power-law equation (FS calibration scenario) and R^2 of the multiple linear regression equation (MLR calibration scenario), for the four study catchments from the first model iteration to the fifth iteration.

Figure 6 illustrates that all R^2 values for each study catchment under these two calibration scenarios vary greatly in the first iteration. More specifically, as presented in Fig. 6a, except the mean value of R^2 in the first iteration is low in KC watershed with 0.35, the mean value of R^2 for the rest of the study watersheds could reach around 0.78 in the first iteration of FS calibration scenarios. This difference of R^2 among four study watersheds is more obvious in MLR calibration scenarios, and the average R^2 of the four watersheds fluctuates between

0.4 and 0.75 in the first model iteration. The results indicate that the simulated streamflow corresponding to many sets of parameters in the first iteration does not satisfy the flood scaling property. However, in the fifth iteration, the R^2 values for both the FS and MLR calibration scenario provide a better fit, exhibiting the potential of applying flood scaling property as model constraints.

3) CALIBRATION FOR THE COMBINATION OF FS OR MLR AND NSE OR KGE AS OBJECTIVE FUNCTIONS

As demonstrated in Fig. 5, in the KGE or NSE calibration scenarios could obtain the higher KGE or NSE evaluation value at watershed outlet than only FS or MLR employed as the single objective function. Since the FS and MLR calibration scenarios have constraints on event flood peak, facilitating the simulation of the event flood peaks is more consistent with the flood generation process. These calibration scenarios are more focused on the simulation about the event flood peak rather than the simulation results of hydrograph at watershed outlets. So basically, the results of calibration scenarios with single FS or MLR have poor NSE or KGE values of hydrograph for watershed outlets but high ability of

FIG. 5. Performance of two statistical indicators under eight different calibration scenarios with increasing model iterations.

capturing the event flood peaks. This finding illustrates that adding more constraints on watershed outlets with observed streamflow data could well improve the model performance compared to the calibration scenarios that only adopted FS or MLR.

Additionally, from Fig. 4, we can learn that the hydrograph of only FS or MLR adopted as model constraints tends to have more noise than that of the calibration scenarios with two objective functions. Since both the performance of flood peaks and hydrograph are taken into consideration in the calibration scenarios with two objective functions, the simulation results at watershed outlets have been improved under these calibration scenarios. The results further prove that adding the statistical NSE or KGE objective function on watershed outlet could produce better performance of the watershed simulated streamflow.

4) MODEL PERFORMANCE UNDER VALIDATION PERIODS

The model performance of eight calibration scenarios during the validation periods is presented in Fig. 7. The results present that the eight calibration scenarios can also achieve reasonable results in the validation periods, and each calibration scenario could capture the flood peak well. Similarly, the hydrographs in FS or MLR calibration scenarios contain more noises than those of the other calibration scenarios, which is consistent with the results in calibration periods.

As shown in Fig. 8, the assessment of flood frequency curve is also carried out with the observed and simulated flood peaks from both calibration and validation periods. The MRE values of each flood frequency curve segment under all designed calibration scenarios are presented in Table 6. The MRE values vary within each flood frequency curve segment,

FIG. 6. The variation in the goodness of fit of the fitting equation from the first iteration to the fifth iteration for the four study catchments under (a) FS and (b) MLR calibration scenarios.

FIG. 7. SWAT model simulation performance under the eight calibration scenarios in (a) ZJG, (b) FP, (c) KC, and (d) SF catchments during validation periods.

which indicates the various performance among different calibration scenarios. Usually, highest MRE values could be found in mid- or low-flow segments. The reason is that MRE is quite sensitive to the errors that occur in the lower magnitudes of the data (Dawson et al. 2007). From Fig. 5, we can learn that the NSE value is the highest in NSE calibration scenario, indicating the good calibration results in common sense. Yet, according to the results demonstrated by Fig. 8, the simulated streamflow results under the NSE calibration scenario obviously underestimate the flood peak at high-flow segments. Negative MRE values could be found in Q_5 and Q_{20} segments for all four study catchments. The KGE calibration scenario produces better simulation of flood peak than NSE calibration scenario with smaller absolute MRE values, yet still with the trends of underestimating the flood peak at high return periods in most cases.

In contrast, as demonstrated in Fig. 8, it is obvious that the flood frequency curve of FS or MLR calibration scenarios is relatively closer to the observed flood frequency curve at very high or high-flow segments, exhibiting better results than NSE or KGE calibration scenarios. In some cases, the FS or

MLR calibration scenario even overestimates the flood peak, which is consistent with the results of the simulated hydrograph. It is possible that applying flood scaling property as model constraints enhances the flood response process. Consequently, the ability of capturing the flood peak is improved, yet the simulation of flood peak is relatively larger than the observed data. As a result, the deviation of simulated flood frequency curve under FS or MLR calibration scenarios is relatively larger than observed data in the majorities of flow segments, especially in low-flow segment.

Furthermore, we can find that with adopting the NSE or KGE as one of the objective functions in FS or MLR calibration scenarios, the overestimation of flood frequency curve has been well improved with better accuracy. Compared to the calibration scenario with single FS or MLR objective function, improved MRE values could be found in these calibration scenarios at all flow segments. The evaluation results of flood frequency curve further illustrate the necessity of adopting the combination of statistical indicators and flood scaling property as objective functions in hydrological model calibration.

FIG. 8. Flood frequency analysis by employing the simulated and observed annual maximum peak discharge of both calibration and validation periods. (TOF denotes the theoretical frequency of the flood frequency curve, for instance, OBS TOF represents the theoretical frequency of the observed data.)

5. Discussion

a. Necessity of adopting flood scaling property as constraints

In some cases, only considering the performance at the catchment outlet is not sufficient, as the simulated flood processes at the interior point of the catchment may not coincide with the actual watershed properties. The application of flood scaling property as constraints could improve this situation. As demonstrated in Figs. 9g and 9h, two significant flood events, namely, the 1988 and 1996 flood events, in the Zijianguan watershed with different rainfall spatial distributions are selected to demonstrate the necessity of adopting flood scaling property as constraints. The simulated streamflow is extracted for a total of 10 catchment interior reach outlets and also the catchment outlet for these two flood events under NSE, FS, and MLR calibration scenarios, as shown in Figs. 9a–f. Due to the high rainfall amount at the Chajianlin rain gauge in the 1988 flood event, the flood generated at the south of the Zijianguan watershed tends to be larger.

Consequently, the flood peak from the hydrograph of the R1 reach outlet tends to be larger, even though the contributing area of the R1 reach outlet is quite small. As presented in Figs. 9a–c, the increase of flood peak for the R1 reach outlet could be captured under the three calibration scenarios (as delineated by the three boxes numbered 2, 3, and 4). However, under the NSE calibration scenario, the flood peak of the R10 reach outlet is even larger than that at the watershed outlet (as delineated by the box numbered 1), which is not consistent with the actual watershed situation, as the rainfall amount is largest at the watershed outlet. Obviously, this situation is well avoided in the FS and MLR calibration scenario, indicating the significance and validity of employing FS or MLR as an objective function. Thus, adopting the scaling property of flood has the potential to better constrain the performance of catchment interior points.

In addition, as shown in Figs. 9b and 9c, we can find that the hydrographs from different watershed interior points under MLR exhibit more reasonable performance than FS.

TABLE 6. The MRE values of each FDC segment under different calibration scenarios.

Catchments	MRE	Different calibration scenarios							
		NSE	KGE	FS	FS&NSE	FS&KGE	MLR	MLR&NSE	MLR&KGE
ZJG	Q_5	-0.359	-0.121	0.585	0.274	0.257	0.148	0.079	0.039
	Q_{20}	-0.722	-0.320	0.989	0.411	0.511	0.529	0.265	0.103
	Q_{mid}	-0.548	0.431	1.112	0.958	1.622	1.352	0.631	0.394
	Q_{70}	0.436	1.579	1.863	1.080	1.817	1.619	0.535	0.444
FP	Q_5	-0.266	0.189	0.225	0.108	0.055	0.597	-0.006	-0.029
	Q_{20}	-0.005	0.648	0.277	0.403	0.462	1.463	0.206	0.411
	Q_{mid}	0.428	1.036	0.107	0.579	0.863	2.212	0.398	1.072
	Q_{70}	0.511	-0.124	-0.248	-0.021	0.101	1.377	0.199	0.716
KC	Q_5	-0.470	-0.206	0.029	-0.414	-0.193	-0.105	-0.444	-0.484
	Q_{20}	-0.285	0.230	0.492	-0.312	0.052	-0.094	-0.328	-0.217
	Q_{mid}	0.233	1.364	1.510	0.110	1.145	-0.066	0.249	0.617
	Q_{70}	0.618	1.800	1.308	0.586	1.453	-0.050	0.964	1.322
SF	Q_5	-1.266	-0.759	-0.045	-0.491	-0.357	0.226	-0.935	-0.514
	Q_{20}	-1.012	-0.458	0.392	-0.269	-0.088	0.908	-0.510	-0.086
	Q_{mid}	-0.199	0.282	1.520	0.360	0.466	1.879	0.668	0.968
	Q_{70}	0.518	0.570	1.943	0.813	0.500	1.779	1.231	1.214

Specifically, in the 1988 flood events at the Zijianguan watershed, the magnitude of flood peak from watershed interior reach outlets tends to be determined by the size of contributing areas under FS, while MLR produces more reasonable results at R1 reach outlet with larger flood peak. For the 1996 flood event with a relatively uniform rainfall spatial distribution, the flood peak of each reach outlet basically exhibits similar tendency under both FS and MLR calibration scenarios. The results indicate that the uneven spatial distribution of rainfall would affect the formation of flood at catchment interior points. Hence, adopting MLR could produce better performance than FS, especially in the case of the uneven distribution of rainfall or other potential external factors.

b. Advantages of adopting flood scaling property as constraints

Compared to the commonly used NSE calibration scenario, the advantage of adopting flood scaling property as constraints is that the more accurate simulated flood peak can be obtained even with a small amount of the observed streamflow data. It is important to note that observed streamflow data are less required in the FS calibration scenarios. Only the simulated event flood peaks are extracted to fit the power-law equations and the goodness of fit behaves as the objective function. As for MLR calibration scenarios, the identification of catchment descriptors requires the observed streamflow data of this catchment or this region.

Additionally, as shown in Fig. 4, although the simulated streamflow results introduce noises in hydrograph under FS or MLR calibration scenarios, the simulation accuracy of flood peak is clearly improved compared to the NSE calibration scenario. It is obvious that the flood frequency curve provided by simulated annual peak discharge in NSE calibration scenario tends to underestimate the flood with more than 3-yr return period. In contrast, for flood frequency curve of FS or MLR calibration scenarios, the simulated flood frequency curve produces better results and even overestimates the

flood peak in some cases. However, the performance of flood frequency curve could be improved with adopting the combination of statistical indicators and flood scaling property as model constraints. To sum up, with only considering the flood scaling property as model constraints, reasonable results could be obtained without employing the observed streamflow time series data directly, which exhibits the potential of applying this method in ungauged watershed. Yet, if the observed streamflow data are available, employing the combination of statistical indicators and flood scaling property as objective functions could better constrain the model performance and improve the simulation of hydrograph at watershed outlet and the accuracy of the flood frequency curve.

c. Applications and future works

For employing flood scaling property as constraints in hydrological model calibration, it is necessary to judge whether the event-based flood scaling property exists in this river basin. In these study watersheds, previous studies have reported the existence of event-based flood scaling property (Li et al. 2013, 2019; Kang et al. 2019; Li et al. 2020). However, since the flood scaling property is an intrinsic law in river basins (Ogden and David 2003; Furey and Gupta 2005, 2007; Gupta et al. 2007; Ayalew et al. 2014, 2015; Furey et al. 2016; Lee and Huang 2016), the method proposed in this paper has good applicability.

In addition, as long as the hydrological model can output the streamflow process of subbasins or reaches, the flood scaling property can be employed as the objective function. In the future, more attempts should be undertaken to consider flood scaling property as constraints in various hydrological models, not just in the SWAT model. Moreover, since the physical conditions of different river basins are distinct, it is possible that the streamflow of the watershed outlet is well simulated by employing only flood scaling property in other river basins. Therefore, the proposed calibration method should be verified in more regions.

FIG. 9. Simulated streamflow process of reach outlets for typical flood events and the spatial distribution of rainfall. The reach outlet names in the legend are arranged according to the contributing area size. (a)–(c) The simulated streamflow of each reach outlet for the 1988 flood event under the NSE, FS, and MLR calibration scenarios, respectively. (d)–(f) The results for the 1996 flood event under the NSE, FS, and MLR calibration scenarios. (g),(h) The spatial distribution of rainfall for the 1988 and 1996 flood events, respectively.

6. Conclusions

To propose a new calibration method for hydrological models based on limited observed data, this paper investigates the potential of applying flood scaling property as one of the objective functions in hydrological model calibration. The main conclusions that were drawn are as follows.

Little information of observed streamflow time series data is needed when only FS is adopted as the single objective function in model calibration. Observed streamflow data were adopted to perform the flood frequency analysis and behaved

as the criterion of selecting the significant flood events. Yet, the FS calibration scenario produces a better simulation on flood peak than NSE calibration scenarios, especially for high return period floods, exhibiting the potential of applying this method in ungauged watersheds.

In contrast, MLR calibration scenario requires some streamflow data to identify the dominant catchment descriptors. FS and MLR calibration scenarios provide a similar hydrograph at the watershed outlet, while MLR achieves better performance at catchment interior points. If the rainfall spatial distribution of the

flood event is not uniform, adopting MLR as one of the model constraints in calibration is suggested.

Furthermore, it is important to note that better performance on hydrograph and flood frequency curve could be obtained with adopting the combination of statistical indicators and flood scaling property as objective functions. Under these calibration scenarios, not only could the accurate simulation of catchment outlet be guaranteed with high NSE or KGE values, but the performance on the catchment interior points is also better constrained.

In brief, for ungauged watersheds, although limited observed streamflow data are available, the simulated flood peak of the subbasin and its contributing drainage area could be extracted to fit the power-law equation and behave as an objective function, providing a new alternative method in hydrological model calibration. For gauged watersheds, adopting flood scaling property as model constraints could make the hydrological model calibration more physically based and improve the performance at catchment interior points. Therefore, the flood scaling property is encouraged to be utilized as an additional constraint in both gauged and ungauged watersheds for hydrological model calibration.

Acknowledgments. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported of this paper. This research is supported by the National Key Research and Development Program of China (2018YFC0407902). The authors thank the editors and reviewers for their constructive suggestions.

Data availability statement. Datasets for this research are available in the reference list. To be specific, land use and land cover maps and soil data are available from the Institute of Geographic Sciences and Resources of the Chinese Academy of Sciences and [FAO/IIASA/ISRIC/ISS-CAS/JRC \(2009\)](http://www.fao.org/fileadmin/templates/nr/documents/HWSD/) (at <http://www.resdc.cn/>, <http://www.fao.org/fileadmin/templates/nr/documents/HWSD/>). Weather data, including precipitation, air temperature, wind speed, relative humidity, and solar data, are available from Fuka et al. (2013) and Dile and Srinivasan (2014) at <https://globalweather.tamu.edu/>. GLDAS-Noah datasets are available from Rodell et al. (2004) and Beaudoin and Rodell (2015) at <https://giovanni.gsfc.nasa.gov/>.

REFERENCES

- Abbaspour, K. C., 2015: SWAT-CUP: SWAT calibration and uncertainty programs: A user manual. Eawag, Swiss Federal Institute of Aquatic Science and Technology, 100 pp., http://swat.tamu.edu/media/114860/usermanual_swatcup.pdf.
- , E. Rouholahnejad, S. Vaghefi, R. Srinivasan, H. Yang, and B. Kløve, 2015: A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *J. Hydrol.*, **524**, 733–752, <https://doi.org/10.1016/j.jhydrol.2015.03.027>.
- , S. A. Vaghefi, and R. Srinivasan, 2017: A guideline for successful calibration and uncertainty analysis for soil and water assessment: A review of papers from the 2016 international SWAT conference. *Water*, **10**, 6, <https://doi.org/10.3390/w10010006>.
- Al-Rawas, G. A., and C. Valeo, 2010: Relationship between wadi drainage characteristics and peak-flood flows in arid northern Oman. *Hydrol. Sci. J.*, **55**, 377–393, <https://doi.org/10.1080/02626661003718318>.
- Ayalew, T. B., W. F. Krajewski, R. Mantilla, and S. J. Small, 2014: Exploring the effects of hillslope-channel link dynamics and excess rainfall properties on the scaling structure of peak-discharge. *Adv. Water Resour.*, **64**, 9–20, <https://doi.org/10.1016/j.advwatres.2013.11.010>.
- , —, and —, 2015: Analyzing the effects of excess rainfall properties on the scaling structure of peak discharges: Insights from a mesoscale river basin. *Water Resour. Res.*, **51**, 3900–3921, <https://doi.org/10.1002/2014WR016258>.
- Basu, B., and V. V. Srinivas, 2015: A recursive multi-scaling approach to regional flood frequency analysis. *J. Hydrol.*, **529**, 373–383, <https://doi.org/10.1016/j.jhydrol.2015.07.037>.
- Beaudoin, H. K., and M. Rodell, 2015: GLDAS Noah land surface model L4 3 hourly 0.25 × 0.25 degree V2.0. Goddard Earth Sciences Data and Information Services Center, accessed 24 April 2019, <https://doi.org/10.5067/342OHQM9AK6Q>.
- Beven, K., 2006: A manifesto for the equifinality thesis. *J. Hydrol.*, **320**, 18–36, <https://doi.org/10.1016/j.jhydrol.2005.07.007>.
- Bormann, H., L. Breuer, T. Graff, and J. A. Huisman, 2007: Analysing the effects of soil properties changes associated with land use changes on the simulated water balance: A comparison of three hydrological catchment models for scenario analysis. *Ecol. Modell.*, **209**, 29–40, <https://doi.org/10.1016/j.ecolmodel.2007.07.004>.
- Brath, A., A. Montanari, and E. Toth, 2004: Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *J. Hydrol.*, **291**, 232–253, <https://doi.org/10.1016/j.jhydrol.2003.12.044>.
- Bu, J., C. Lu, J. Niu, and Y. Gao, 2018: Attribution of runoff reduction in the Juma River basin to climate variation, direct human intervention, and land use change. *Water*, **10**, 1775, <https://doi.org/10.3390/w10121775>.
- Cao, W., W. B. Bowden, T. Davie, and A. Fenemor, 2006: Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Hydrol. Processes*, **20**, 1057–1073, <https://doi.org/10.1002/hyp.5933>.
- Cao, Y., J. Zhang, M. Yang, X. Lei, B. Guo, L. Yang, Z. Zeng, and J. Qu, 2018: Application of SWAT model with CMADS data to estimate hydrological elements and parameter uncertainty based on SUFI-2 algorithm in the Lijiang River basin, China. *Water*, **10**, 742, <https://doi.org/10.3390/w10060742>.
- Chiang, L. C., Y. Yuan, M. Mehaffey, M. Jackson, and I. Chaubey, 2014: Assessing SWAT's performance in the Kaskaskia River watershed as influenced by the number of calibration stations used. *Hydrol. Processes*, **28**, 676–687, <https://doi.org/10.1002/hyp.9589>.
- Chilkoti, V., T. Bolisetti, and R. Balachandar, 2018: Multi-objective autocalibration of SWAT model for improved low flow performance for a small snowfed catchment. *Hydrol. Sci. J.*, **63**, 1482–1501, <https://doi.org/10.1080/02626667.2018.1505047>.
- Daggupati, P., and Coauthors, 2015: A recommended calibration and validation strategy for hydrologic and water quality models. *Trans. ASABE*, **58**, 1705–1719, <https://doi.org/10.13031/trans.58.10712>.
- Dawdy, D. R., V. W. Griffiths, and V. K. Gupta, 2012: Regional floodfrequency analysis: How we got here and where we are

- going. *J. Hydrol. Eng.*, **17**, 953–959, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000584](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000584).
- Dawson, C. W., R. J. Abraham, and L. M. See, 2007: HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ. Modell. Software*, **22**, 1034–1052, <https://doi.org/10.1016/j.envsoft.2006.06.008>.
- De Lavenne, A., V. Andréassian, G. Thirel, M. H. Ramos, and C. Perrin, 2019: A regularization approach to improve the sequential calibration of a semidistributed hydrological model. *Water Resour. Res.*, **55**, 8821–8839, <https://doi.org/10.1029/2018WR024266>.
- Dile, Y. T., and R. Srinivasan, 2014: Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: An application in the Blue Nile River basin. *J. Amer. Water Resour. Assoc.*, **50**, 1226–1241, <https://doi.org/10.1111/jawr.12182>.
- FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009: Harmonized World Soil Database (version 1.1). FAO, 43 pp.
- Farmer, W. H., T. M. Over, and R. M. Vogel, 2015: Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the Southeast United States. *Water Resour. Res.*, **51**, 1775–1796, <https://doi.org/10.1002/2014WR015924>.
- Fenicia, F., D. Kavetski, H. H. G. Savenije, and L. Pfister, 2016: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resour. Res.*, **52**, 954–989, <https://doi.org/10.1002/2015WR017398>.
- Fuka, D. R., C. A. MacAllister, A. T. Degaetano, and Z. M. Easton, 2014: Using the climate forecast system reanalysis dataset to improve weather input data for watershed models. *Hydrol. Processes*, **28**, 5613–5623, <https://doi.org/10.1002/hyp.10073>.
- Furey, P. R., and V. K. Gupta, 2005: Effects of excess rainfall on the temporal variability of observed peak-discharge power laws. *Adv. Water Resour.*, **28**, 1240–1253, <https://doi.org/10.1016/j.advwatres.2005.03.014>.
- , and —, 2007: Diagnosing peak-discharge power laws observed in rainfall–runoff events in Goodwin Creek experimental watershed. *Adv. Water Resour.*, **30**, 2387–2399, <https://doi.org/10.1016/j.advwatres.2007.05.014>.
- , B. M. Troutman, V. K. Gupta, and W. F. Krajewski, 2016: Connecting event-based scaling of flood peaks to regional flood frequency relationships. *J. Hydrol. Eng.*, **21**, 04016037, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001411](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001411).
- Garcia, F., N. Folton, and L. Oudin, 2017: Which objective function to calibrate rainfall–runoff models for low-flow index simulations? *Hydrol. Sci. J.*, **62**, 1149–1166, <https://doi.org/10.1080/02626667.2017.1308511>.
- Göttinger, J., and A. Bárdossy, 2007: Comparison of four regionalization methods for a distributed hydrological model. *J. Hydrol.*, **333**, 374–384, <https://doi.org/10.1016/j.jhydrol.2006.09.008>.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Gupta, V. K., 2004: Emergence of statistical scaling in floods on channel networks from complex runoff dynamics. *Chaos Solitons Fractals*, **19**, 357–365, [https://doi.org/10.1016/S0960-0779\(03\)00048-1](https://doi.org/10.1016/S0960-0779(03)00048-1).
- , B. M. Troutman, D. R. Dawdy, D. R. 2007: Towards a nonlinear geophysical theory of floods in river networks: An overview of 20 years of progress. *Nonlinear Dynamics in Geosciences*, Springer, 121–151.
- Han, E., V. Merwade, and G. C. Heathman, 2012: Implementation of surface soil moisture data assimilation with watershed-scale distributed hydrological model. *J. Hydrol.*, **416–417**, 98–117, <https://doi.org/10.1016/j.jhydrol.2011.11.039>.
- Hapuarachchi, H. A., Q. J. Wang, and T. C. Pagano, 2011: A review of advances in flash flood forecasting. *Hydrol. Processes*, **25**, 2771–2784, <https://doi.org/10.1002/hyp.8040>.
- Her, Y., J. Frankenberger, I. Chaubey, and R. Srinivasan, 2015: Threshold effects in HRU definition of the soil and water assessment tool. *Trans. ASABE*, **58**, 367–378, <https://doi.org/10.13031/trans.58.10805>.
- Howard, A. J., M. Bonell, D. Gilmour, and D. Cassells, 2010: Is rainfall intensity significant in the rainfall–runoff process within tropical rainforests of northeast Queensland? The Hewlett regression analyses revisited. *Hydrol. Processes*, **24**, 2520–2537, <https://doi.org/10.1002/hyp.7694>.
- Hughes, J. D., S. S. H. Kim, D. Dutta, and J. Vaze, 2016: Optimization of a multiple gauge, regulated river-system model. A system approach. *Hydrol. Processes*, **30**, 1955–1967, <https://doi.org/10.1002/hyp.10752>.
- Huisman, J. A., and Coauthors, 2009: Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) III: Scenario analysis. *Adv. Water Resour.*, **32**, 159–170, <https://doi.org/10.1016/j.advwatres.2008.06.009>.
- Hundecha, Y., and A. Bárdossy, 2004: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *J. Hydrol.*, **292**, 281–295, <https://doi.org/10.1016/j.jhydrol.2004.01.002>.
- Ishak, E., K. Haddad, M. Zaman, and A. Rahman, 2011: Scaling property of regional floods in New South Wales Australia. *Nat. Hazards*, **58**, 1155–1167, <https://doi.org/10.1007/s11069-011-9719-6>.
- Jeong, J., N. Kannan, J. Arnold, R. Glick, L. Gosselink, and R. Srinivasan, 2010: Development and integration of sub-hourly rainfall–runoff modeling capability within a watershed model. *Water Resour. Manage.*, **24**, 4505–4527, <https://doi.org/10.1007/s11269-010-9670-4>.
- Jothityangkoon, C., and M. Sivapalan, 2001: Temporal scales of rainfall–runoff processes and spatial scaling of flood peaks: Space–time connection through catchment water balance. *Adv. Water Resour.*, **24**, 1015–1036, [https://doi.org/10.1016/S0309-1708\(01\)00044-6](https://doi.org/10.1016/S0309-1708(01)00044-6).
- Kang, Y. F., J. Z. Li, and Q. S. Ma, 2019: Using HEC-HMS model to simulate flooding in Zijianguan watershed. *J. Irrig. Drain.*, **38**, 108–115, <https://doi.org/10.13522/j.cnki.gggs.20180520>.
- Kannan, N., S. M. White, F. Worrall, and M. J. Whelan, 2007: Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000. *J. Hydrol.*, **332**, 456–466, <https://doi.org/10.1016/j.jhydrol.2006.08.001>.
- Knoben, W. J. M., J. E. Freer, and R. A. Woods, 2019: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.*, **23**, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>.
- Koren, V., F. Morel, and M. Smith, 2008: Use of soil moisture observations to improve parameter consistency in watershed calibration. *Phys. Chem. Earth*, **33**, 1068–1080, <https://doi.org/10.1016/j.pce.2008.01.003>.
- Kouchi, D. H., K. Esmaili, A. Faridhosseini, S. H. Sanaeinejad, D. Khalili, and K. C. Abbaspour, 2017: Sensitivity of calibrated parameters and water resource estimates on different objective functions and optimization algorithms. *Water*, **9**, 384, <https://doi.org/10.3390/w9060384>.
- Lacombe, G., S. Douangsavanh, R. M. Vogel, M. McCartney, Y. Chemin, L. M. Rebelo, and T. Sotoukee, 2014: Multivariate

- power-law models for streamflow prediction in the Mekong Basin. *J. Hydrol. Reg. Stud.*, **2**, 35–48, <https://doi.org/10.1016/j.ejrh.2014.08.002>.
- Lee, K. T., and J. K. Huang, 2016: Influence of storm magnitude and watershed size on runoff nonlinearity. *J. Earth Syst. Sci.*, **125**, 777–794, <https://doi.org/10.1007/s12040-016-0700-3>.
- Lee, S., G. W. McCarty, G. E. Moglen, X. Li, and C. W. Wallace, 2020: Assessing the effectiveness of riparian buffers for reducing organic nitrogen loads in the Coastal Plain of the Chesapeake Bay watershed using a watershed model. *J. Hydrol.*, **585**, 124779, <https://doi.org/10.1016/j.jhydrol.2020.124779>.
- Ley, R., H. Hellebrand, C. M. Casper, and F. Fenicia, 2016: Comparing classical performance measures with signature indices derived from flow duration curves to assess model structures as tools for catchment classification. *Hydrol. Res.*, **47**, 1–14, <https://doi.org/10.2166/nh.2015.221>.
- Li, J. Z., P. Feng, and Z. Wei, 2013: Incorporating the data of different watersheds to estimate the effects of land use change on flood peak and volume using multi-linear regression. *Mitig. Adapt. Strategies Global Change*, **18**, 1183–1196, <https://doi.org/10.1007/s11027-012-9416-0>.
- , Q. S. Ma, Y. Tian, Y. M. Lei, T. Zhang, and P. Feng, 2019: Flood scaling under nonstationarity in Daqinghe River basin, China. *Nat. Hazards*, **98**, 675–696, <https://doi.org/10.1007/s11069-019-03724-y>.
- , K. Lei, T. Zhang, W. Zhong, A. Q. Kang, Q. S. Ma, and P. Feng, 2020: A framework for event-based flood scaling analysis by hydrological modeling in data-scarce regions. *Hydrol. Res.*, **51**, 1091–1103, <https://doi.org/10.2166/nh.2020.042>.
- Li, Y., S. Grimaldi, V. R. N. Pauwels, and J. P. Walker, 2018: Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: The impact on predictions at gauged and ungauged locations. *J. Hydrol.*, **557**, 897–909, <https://doi.org/10.1016/j.jhydrol.2018.01.013>.
- Lima, C. H. R., and U. Lall, 2010: Spatial scaling in a changing climate: A hierarchical bayesian model for non-stationary multi-site annual maximum and monthly streamflow. *J. Hydrol.*, **383**, 307–318, <https://doi.org/10.1016/j.jhydrol.2009.12.045>.
- Lu, S., N. Kayastha, H. Thodsen, A. Van Griensven, and H. E. Andersen, 2014: Multiobjective calibration for comparing channel sediment routing models in the soil and water assessment tool. *J. Environ. Qual.*, **43**, 110–120, <https://doi.org/10.2134/jeq2011.0364>.
- Mallows, C. L., 1973: Some comments on *Cp*. *Technometrics*, **15**, 661–675, <https://doi.org/10.1080/00401706.1973.10489103>.
- Mandapaka, V., W. F. Krajewski, R. Mantilla, and V. K. Gupta, 2009: Dissecting the effect of rainfall variability on the statistical structure of peak flows. *Adv. Water Resour.*, **32**, 1508–1525, <https://doi.org/10.1016/j.advwatres.2009.07.005>.
- McMillan, H. K., 2021: A review of hydrologic signatures and their applications. *Wiley Interdiscip. Rev.: Water*, **8**, e1499, <https://doi.org/10.1002/wat2.1499>.
- Meigh, J. R., F. A. K. Farquharson, and J. V. Sutcliffe, 1997: A worldwide comparison of regional flood estimation methods and climate. *Hydrol. Sci. J.*, **42**, 225–244, <https://doi.org/10.1080/02626669709492022>.
- Melsen, L. A., A. J. Teuling, S. W. Berkum, P. J. J. F. Torfs, and R. Uijlenhoet, 2014: Catchments as simple dynamical systems: A case study on methods and data requirements for parameter identification. *Water Resour. Res.*, **50**, 5577–5596, <https://doi.org/10.1002/2013WR014720>.
- Menabde, M., and M. Sivapalan, 2011: Linking space–time variability of river runoff and rainfall fields: A dynamic approach. *Adv. Water Resour.*, **24**, 1001–1014, [https://doi.org/10.1016/S0309-1708\(01\)00038-0](https://doi.org/10.1016/S0309-1708(01)00038-0).
- Mendoza, P. A., M. P. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. Gupta, 2015: Are we unnecessarily constraining the agility of complex process-based models? *Water Resour. Res.*, **51**, 716–728, <https://doi.org/10.1002/2014WR015820>.
- Neter, J., W. Wasserman, and M. H. Kutner, 1989: *Applied Linear Regression Models*. 2nd ed. Richard D. Irwin, Inc., 1127 pp.
- Niehoff, D., U. Fritsch, and A. Bronstert, 2002: Land-use impacts on storm-runoff generation: Scenario land-use change and simulation of hydrological response in a mesoscale catchment in SW-Germany. *J. Hydrol.*, **267**, 80–93, [https://doi.org/10.1016/S0022-1694\(02\)00142-7](https://doi.org/10.1016/S0022-1694(02)00142-7).
- Niraula, R., L. M. Norman, T. Meixner, and J. B. Callegary, 2012: Multi-gauge calibration for modeling the semi-arid Santa Cruz watershed in Arizona-Mexico border area using SWAT. *Air Soil Water Res.*, **5**, 41–57, <https://doi.org/10.4137/ASWR.S9410>.
- Odusanya, A. E., B. Mehdi, C. Schürz, A. O. Oke, O. S. Awokola, J. A. Awomeso, J. O. Adejwon, and K. Schulz, 2019: Multi-site calibration and validation of SWAT with satellite-based evapotranspiration in a data sparse catchment in southwestern Nigeria. *Hydrol. Earth Syst. Sci.*, **23**, 1113–1144, <https://doi.org/10.5194/hess-23-1113-2019>.
- Ogden, F. L., and D. R. David, 2003: Peak discharge scaling in small Hortonian watershed. *J. Hydrol. Eng.*, **8**, 64–73, [https://doi.org/10.1061/\(ASCE\)1084-0699\(2003\)8:2\(64\)](https://doi.org/10.1061/(ASCE)1084-0699(2003)8:2(64)).
- Okcu, D., A. O. Pektas, and A. Uyumaz, 2016: Creating a non-linear total sediment load formula using polynomial best subset regression model. *J. Hydrol.*, **539**, 662–673, <https://doi.org/10.1016/j.jhydrol.2016.04.069>.
- Omani, N., R. Srinivasan, K. Smith, and R. Karthikeyan, 2017: Glacier mass balance simulation using SWAT distributed snow algorithm. *Hydrol. Sci. J.*, **62**, 546–560, <https://doi.org/10.1080/02626667.2016.1162907>.
- Perrin, C., L. Oudin, V. Andréassian, C. Rojas-Serna, C. Michel, and T. Mathevet, 2007: Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models. *Hydrol. Sci. J.*, **52**, 131–151, <https://doi.org/10.1623/hysj.52.1.131>.
- , V. Andréassian, C. R. Serna, T. Mathevet, and N. Le Moine, 2008: Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments. *Water Resour. Res.*, **44**, W08447, <https://doi.org/10.1029/2007WR006579>.
- Pfannerstill, M., B. Guse, and N. Fohrer, 2014: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *J. Hydrol.*, **510**, 447–458, <https://doi.org/10.1016/j.jhydrol.2013.12.044>.
- , K. Bieger, B. Guse, D. D. Bosch, N. Fohrer, and J. G. Arnold, 2017: How to constrain multi-objective calibrations of the SWAT model using water balance components. *J. Amer. Water Resour. Assoc.*, **53**, 532–546, <https://doi.org/10.1111/1752-1688.12524>.
- Pokhrel, P., K. K. Yilmaz, and H. V. Gupta, 2012: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *J. Hydrol.*, **418**–**419**, 49–60, <https://doi.org/10.1016/j.jhydrol.2008.12.004>.
- Pool, S., D. Viviroli, and J. Seibert, 2017: Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? *J. Hydrol.*, **554**, 613–622, <https://doi.org/10.1016/j.jhydrol.2017.09.037>.

- , M. Vis, and J. Seibert, 2018: Evaluating model performance: Towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrol. Sci. J.*, **63**, 1941–1953, <https://doi.org/10.1080/02626667.2018.1552002>.
- Ragetti, S., J. Zhou, H. Wang, C. Liu, and L. Guo, 2017: Modeling flash floods in ungauged mountain catchments of China: A decision tree learning approach for parameter regionalization. *J. Hydrol.*, **555**, 330–346, <https://doi.org/10.1016/j.jhydrol.2017.10.031>.
- Rajib, M. A., V. Merwade, and Z. Yu, 2016: Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *J. Hydrol.*, **536**, 192–207, <https://doi.org/10.1016/j.jhydrol.2016.02.037>.
- , —, and —, 2018: Rationale and efficacy of assimilating remotely sensed potential evapotranspiration for reduced uncertainty of hydrologic models. *Water Resour. Res.*, **54**, 4615–4637, <https://doi.org/10.1029/2017WR021147>.
- Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- Rojas-Serna, C., L. Lebecherel, C. Perrin, V. Andréassian, and L. Oudin, 2016: How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments. *Water Resour. Res.*, **52**, 4765–4784, <https://doi.org/10.1002/2015WR018549>.
- Samaniego, L., and Coauthors, 2017: Toward seamless hydrologic predictions across spatial scales. *Hydrol. Earth Syst. Sci.*, **21**, 4323–4346, <https://doi.org/10.5194/hess-21-4323-2017>.
- Schwarz, G., 1978: Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464, <https://doi.org/10.1214/aos/1176344136>.
- Scussolini, P., J. C. J. H. Aerts, B. Jongman, L. M. Bouwer, H. C. Winsemius, H. De Moel, and J. Ward, 2016: FLOPROS: An evolving global database of flood protection standards. *Nat. Hazards Earth Syst. Sci.*, **16**, 1049–1061, <https://doi.org/10.5194/nhess-16-1049-2016>.
- Seibert, J., 2010: Gauging the ungauged catchment: A combination of regionalization and limited streamflow observations. *Geophysical Research Abstracts*, Vol. 12, Abstract 12465, <https://meetingorganizer.copernicus.org/EGU2010/EGU2010-12465.pdf>.
- , and K. J. Beven, 2009: Gauging the ungauged basin: How many discharge measurements are needed? *Hydrol. Earth Syst. Sci.*, **13**, 883–892, <https://doi.org/10.5194/hess-13-883-2009>.
- , and J. J. McDonnell, 2013: Gauging the ungauged basin: Relative value of soft and hard data. *J. Hydrol. Eng.*, **20**, A4014004, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000861](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000861).
- Sexton, A. M., A. M. Sadeghi, X. Zhang, R. Srinivasan, and A. Shirmohammadi, 2010: Using NEXRAD and rain gauge precipitation data for hydrologic calibration of SWAT in a northeastern watershed. *Trans. ASABE*, **53**, 1501–1510, <https://doi.org/10.13031/2013.34900>.
- Singh, S. K., and A. Bárdossy, 2012: Calibration of hydrological models on hydrologically unusual events. *Adv. Water Resour.*, **38**, 81–91, <https://doi.org/10.1016/j.advwatres.2011.12.006>.
- Smith, A., C. Sampson, and P. Bates, 2015: Regional flood frequency analysis at the global scale. *Water Resour. Res.*, **51**, 539–553, <https://doi.org/10.1002/2014WR015814>.
- Srinivasan, R., X. Zhang, and J. Arnold, 2010: SWAT ungauged: Hydrological budget and crop yield predictions in the upper Mississippi River basin. *Trans. ASABE*, **53**, 1533–1546, <https://doi.org/10.13031/2013.34903>.
- Tuo, Y., G. Marcolini, M. Disse, and G. Chiogna, 2018: A multi-objective approach to improve swat model calibration in alpine catchments. *J. Hydrol.*, **559**, 347–360, <https://doi.org/10.1016/j.jhydrol.2018.02.055>.
- Viviroli, D., and J. Seibert, 2015: Can a regionalized model parameterisation be improved with a limited number of runoff measurements? *J. Hydrol.*, **529**, 49–61, <https://doi.org/10.1016/j.jhydrol.2015.07.009>.
- Vogel, R. M., and N. M. Fennessey, 1994: Flow-duration curves. I: New interpretation and confidence intervals. *J. Water Resour. Plann. Manage.*, **120**, 485–504, [https://doi.org/10.1061/\(ASCE\)0733-9496\(1994\)120:4\(485\)](https://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(485)).
- Wanders, N., M. F. Bierkens, S. M. de Jong, A. de Roo, and D. Karssenbergh, 2014: The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resour. Res.*, **50**, 6874–6891, <https://doi.org/10.1002/2013WR014639>.
- Wang, Z., J. Cao, and H. Yang, 2021: Multi-time scale evaluation of forest water conservation function in the semiarid mountains area. *Forests*, **12**, 116, <https://doi.org/10.3390/f12020116>.
- Weeink, W., 2010: Thresholds for flood forecasting and warning. M.S. thesis, Civil Engineering and Management, University of Twente, 89 pp.
- Westerberg, I. K., J.-L. Guerrero, P. M. Younger, K. J. Beven, J., Seibert, S. Halldin, J. E. Freer, and C.-Y. Xu, 2011: Calibration of hydrological models using flow-duration curves. *Hydrol. Earth Syst. Sci.*, **15**, 2205–2227, <https://doi.org/10.5194/hess-15-2205-2011>.
- White, K. L., and I. Chaubey, 2005: Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. *J. Amer. Water Resour. Assoc.*, **41**, 1077–1089, <https://doi.org/10.1111/j.1752-1688.2005.tb03786.x>.
- Wi, S., Y. C. E. Yang, S. Steinschneider, A. Khalil, and C. M. Brown, 2015: Calibration approaches for distributed hydrologic models in poorly gaged basins: Implication for streamflow projections under climate change. *Hydrol. Earth Syst. Sci.*, **19**, 857–876, <https://doi.org/10.5194/hess-19-857-2015>.
- Yang, J., P. Reichert, K. C. Abbaspour, J. Xia, and H. Yang, 2008: Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *J. Hydrol.*, **358**, 1–23, <https://doi.org/10.1016/j.jhydrol.2008.05.012>.
- Yapo, P. O., H. V. Gupta, and S. Sorooshian, 1996: Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. *J. Hydrol.*, **181**, 23–48, [https://doi.org/10.1016/0022-1694\(95\)02918-4](https://doi.org/10.1016/0022-1694(95)02918-4).
- Yilmaz, K. K., H. V. Gupta, and T. Wagener, 2008: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resour. Res.*, **44**, W09417, <https://doi.org/10.1029/2007WR006716>.
- Zhang, X., R. Srinivasan, and M. Van Liew, 2008: Multi-site calibration of the SWAT model for hydrologic modeling. *Trans. ASABE*, **51**, 2039–2049, <https://doi.org/10.13031/2013.25407>.
- , —, and —, 2010: On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. *Hydrol. Processes*, **24**, 955–969, <https://doi.org/10.1002/hyp.7528>.
- Zheng, Y., J. Li, L. Dong, Y. Rong, A. Kang, and P. Feng, 2020: Estimation of initial abstraction for hydrological modeling based on global land data assimilation system-simulated datasets. *J. Hydrometeorol.*, **21**, 1051–1072, <https://doi.org/10.1175/JHM-D-19-0202.1>.
- Zhu, J., C. Wen, Z. Jin, H. Zhang, and X. Wang, 2020: A polynomial algorithm for best-subset selection problem. *Proc. Natl. Acad. Sci. USA*, **117**, 33 117–33 123, <https://doi.org/10.1073/pnas.2014241117>.