Impacts of spatio-temporal resolutions of precipitation on flood events simulation based on multi-model structures — A case study over Xiang River Basin in China

Qian Zhu[1, *], Xiaodong Qin[1], Dongyang Zhou[1], Tiantian Yang[2], Xinyi Song[3]

[1]School of Civil Engineering, Southeast University, Nanjing 211189, China; zhuqian@seu.edu.cn

[2]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, 73019 , USA; Tiantian.Yang@ou.edu

[3]School of Hydraulic and Environmental Engineering, Changsha University of Science & Technology, Changsha 410114, China; songxy@csust.edu.cn

Correspondence to: Qian Zhu(zhuqian@seu.edu.cn)

Key words: satellite-based precipitation; spatio-temporal resolutions; hydrological modelling; Long Short-Term Memory (LSTM); calibration strategies; flood events

Abstract. Accurate flood events simulation and prediction, enabled by effective models and reliable data, are critical for mitigating the potential risk of flood disaster. This study aims to investigate the impacts of spatio-temporal resolutions of precipitation on flood events simulation in a large-scale catchment of China. We use the high spatio-temporal resolutions Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG) products and a gauge-based product as precipitation forcing for hydrologic simulation. Three hydrological models (HBV, SWAT, and DHSVM) and a data-driven model (Long Short-Term Memory (LSTM) network) are utilized for flood events simulation. Two calibration strategies are carried out, one of which targets at matching the flood events with peak discharge exceeding 8600 m^3/s between January 2015 and December 2017, and the other one is the conventional strategy to match the entire streamflow time series. The results indicate that the event-based calibration strategy improves the performance of flood events simulation, compared with conventional calibration strategy, except for DHSVM. Both hydrological models and LSTM yield better flood events simulation at finer temporal resolution, especially in flood peaks simulation. Furthermore, SWAT and DHSVM are less sensitive to the spatial resolutions of IMERG, while the performance of LSTM obtains improvement when degrading the spatial resolution of IMERG-L. Generally, the LSTM outperforms the hydrological models in most flood events, which implies the usefulness of the deep learning algorithms for flood event simulation.

1 Introduction

The global climate change increases the risk of floods, which brings heavy casualties and losses of property (Hirabayashi et al., 2013). In China, the flood events seem to become more frequent over the mid to lower reaches of the Yangtze River due to the increasing intensity and frequency of rainfall extremes (Piao et al., 2010). In June 2017, large-scale flood events induced by heavy rainfall in Hunan province, located in the southern China, affected more than ten million people and caused economic losses of more than 40 billion Chinese Yuan. Reliable flood events simulation and prediction are the key to minimize the losses and impacts caused by flood events.

Numerous models are applied to simulate the flood events, most of which are conceptual/physically based hydrologic models (Dutta et al., 2000; Koutroulis and Tsanis, 2010; Nikolopoulos et al., 2013; Wu

et al., 2014; Mei et al., 2016; Yang et al., 2017; Yu et al., 2018; Grimaldi et al., 2019), and others are based on artificial neural networks (Shrestha et al., 2005; Badrzadeh et al., 2015). Owing to the continuous development of artificial neural networks, deep learning (DL) is emerged as a dominant tool, which impacts various scientific disciplines in recent years (Akbari Asanjan et al., 2018; Shen, 2018; Shen et al., 2018; Zhang et al., 2018). Among various DL methods, the Long Short-Term Memory (LSTM) network is appropriate for capturing the relationship between rainfall and runoff, because of its ability to learn long-term dependencies and delays between the input and output, and shows extraordinary potential in hydrological simulation (Hu et al., 2018; Liao et al., 2019; Fan et al., 2020; Kao et al., 2020; Ni et al., 2020; Zhu et al., 2020b). Both hydrological models and deep-learning based models require multi-sources inputs, particularly precipitation, which is the key forcing variable in hydrological process to simulate/predict flood events.

Traditionally, in-situ precipitation is utilized for hydrological simulation. However, because of the uneven distribution of in-situ observations and its unavailability in less developed regions, satellite-based precipitation products have been widely used as an alternative precipitation source, and further applied for flood events simulation (Maggioni and Massari, 2018). Among them, the Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG) (Huffman et al., 2015) is a high spatio-temporal resolution satellite-based precipitation product released by National Aeronautics and Space Administration (Rafieeinasab et al.), whose accuracy and hydrological utility have been evaluated from multiple aspects, such as with different temporal scales (e.g., daily and sub-daily) (Tang et al., 2016; Yuan et al., 2018; Su et al., 2020), and on basins with different climate conditions (O et al., 2017; Wang et al., 2017; Zubieta et al., 2017; Fang et al., 2019; Jiang and Bauer-Gottwein, 2019). Many studies show that the performance of IMERG varies across different climate regions and terrain. In addition, most of the IMERG-related studies are conducted to assess its performance at a specific spatio-temporal resolution, a few of which consider the impacts of different spatio-temporal resolutions on its accuracy. Among limited studies, Tang et al. (2016) evaluated the IMERG products at hourly, 3-hourly and daily scales, and they revealed that the statistical indices of IMERG increase with coarser temporal resolutions. Su et al. (2020) assessed the IMERG products at multiple spatial and temporal resolutions by upscaling, and they summarized that degrading the spatio-temporal resolution improves the accuracy of IMERG products. However, these two studies just evaluated the accuracy of IMERG products at multiple spatio-temporal scales, rather than the effects of spatio-temporal resolutions of IMERG products on their hydrological applications (e.g., flood simulation).

As proved by Huang et al. (2019), the spatio-temporal resolutions affect the accuracy of precipitation estimates, and the effects can be propagated to the flood events simulation through the hydrological processes. However, the impact of precipitation with different spatio-temporal resolutions on hydrological simulation has not yet been determined, which is related to many different factors, such as the structure of hydrological models (Arnaud et al., 2011; Yu et al., 2014), the scale of catchment and event characteristics (Lobligeois et al., 2014; Ficchì et al., 2016). Most studies investigated the sensitivity of hydrological models to spatio-temporal resolution based on one model structure with in-situ precipitation, and they concluded that the accuracy of hydrological simulation is not always higher with shorter time steps or higher spatial resolutions (Liang et al., 2004; Arnaud et al., 2011; Lobligeois et al., 2014; Yu et al., 2014; Rafieeinasab et al., 2015; Ficchì et al., 2016; Melsen et al., 2016; Buitink et al., 2019; Huang et al., 2019). For instance, some studies present better hydrological simulation forced by

in-situ precipitation with lower spatio-temporal resolutions to some extent (Liu et al., 2012; Apip et al., 2012; Lobligeois et al., 2014; Ficchì et al., 2016). As we all know, high spatio-temporal resolution is one of the advantages of satellite-based precipitation products, however, there are also studies pointing out that degrading the spatio-temporal resolution can improve the accuracy of precipitation (Su et al., 2020). But rare studies have been conducted to probe the effects of spatio-temporal of satellite-based precipitation on flood simulation, not to mention its impact on flood simulation with models based on DL methods (*e.g.*, LSTM). What's more important, to our best knowledge, the sensitivity of models with different structures, such as lumped hydrological model, semi-distributed/distributed hydrological model, and data-driven model, to the spatio-temporal resolutions of precipitation has not been investigated. Therefore, three widely used and typical conceptual/physically based models (lumped HBV model, semi-distributed SWAT model, and distributed DHSVM model), and one data-driven model (LSTM) which shows good performance in hydrological simulation, are employed to probe the impacts of spatio-temporal resolutions of precipitation on flood events simulation.

Apart from the factors mentioned above, the rationality of calibration is another important factor to affect the accuracy of hydrological simulation. Many studies investigate the influences of the choice of objective function and calibration method on hydrological simulation, but most of them use the calibration strategy based on entire streamflow time series instead of flood events (Moussa and Chahinian, 2009; Noilhan et al., 2010; Nikolopoulos et al., 2013; Badrzadeh et al., 2015; Yoshimoto and Amarnath, 2017; Spellman et al., 2018). However, some studies prove that the event-based calibration can improve the performance of streamflow simulation. For instance, Yu et al. (2018) developed the sub-daily SWAT-EVENT model for event-based flood simulation, which particularly improved the performance of flood events simulation, especially the accuracies of the flood peaks. And Xie et al. (2019) compared the continuous modelling and event-based modelling based on the generalized likelihood uncertainty estimation (GLUE), and found the event-based simulation showed better overall performance. However, studies about event-based calibration are still quite limited, particularly for LSTM. Therefore, in this study, we conduct different calibration strategies aimed at obtaining the best possible flood events simulation.

The main objectives of this study are: (1) to investigate the impact of spatio-temporal resolutions of satellite-based precipitation estimates derived from IMERG on streamflow simulation, particularly flood events simulation, over a watershed of 82,375 km^2; (2) to explore and compare the performance of hydrological models with different structures and LSTM on flood events simulation based on gauge-based and satellite-based precipitation products; (3) to study the potential benefits of the calibration strategy based on flood events. The remaining sections of the paper are organized as follows: the descriptions of the study area and data are presented in Section 2; the methodology is introduced in Section 3; Section 4 provides the results; the discussion is stated in Section 5; and conclusions are summarized in Section 6.

2 Study area and data

2.1 Study area

The Xiang River basin is a humid region, located in the middle reach of the Yangtze River, within 110.50°E-114.25°E, 24.50°N-28.25°N in the southern China, which covers an area of about 82,375 km^2 above the Xiangtan hydrological station (Fig. 1). Together with the impact of diverse topographic types and a dominant subtropical monsoon climate, the precipitation is characterized with strong temporal and

spatial variability (Zhu et al., 2017). The average annual temperature of the basin is around 17℃, and the mean annual total precipitation is around 1,400-1,700 mm, most of which falls from April to September.

Concentrated storm events during the flood season cause frequent floods throughout the basin. Since the Xiang River basin is the most densely populated and economically developed area in Hunan Province (Zhu et al. 2020a), it is critical to accurately simulate and predict flood events in the region for effective flood risk management.

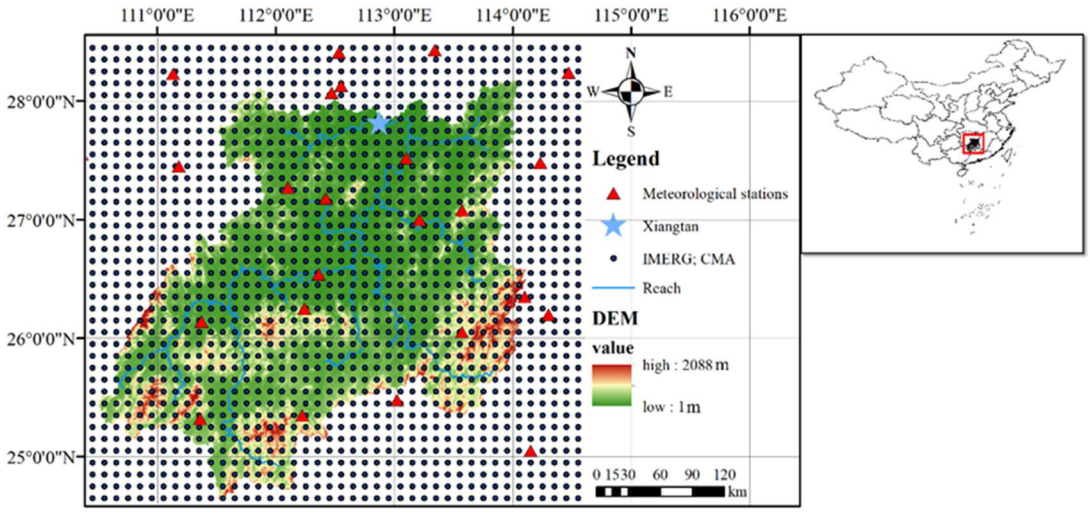


Fig. 1. The spatial distribution of meteorological stations, the outlet of the study area, and precipitation from IMERG and CMA.

2.2 Data description

IMERG V05B is a widely used satellite-based precipitation product with a spatio-temporal resolution of 0.1° and 30-min released by NASA, which consists of multiple rainfall retrieval algorithms and combines various precipitation-relevant remote sensing data sources obtained from the GPM sensors (Huffman et al., 2015). The IMERG system is firstly run twice to produce IMERG Early Run and IMERG Late Run (hereafter IMERG-E and IMERG-L) with a latency of 4 hours and 12 hours in near real-time (NRT). And then through the bias adjustment with monthly Global Precipitation Climatology Centre (GPCC) gauge observations, IMERG Final Run (hereafter IMERG-F) is generated with 2.5 months latency.

A precipitation product released by China Meteorological Administration (hereafter CMA), which merges rain gauge data from more than 30,000 automatic weather stations (AWSs) in China with the Climate Prediction Center morphing technique (CMORPH) precipitation product by an improved probability density function-optimal interpolation method (PDF-OI), is used as the reference precipitation dataset in this study (Shen et al., 2014). CMA provides precipitation estimates in a spatial resolution of 0.1° and a temporal resolution of 1-hour, which is proved to be a reliable precipitation product as a result of high density of the AWSs and rigorous quality control of the source data. Therefore, CMA has already been applied as the benchmark in some studies (Wang et al., 2017; Tang et al., 2017; Su et al., 2020).

Daily gauge meteorological variables (maximum and minimum temperature, relative humidity, wind speed and solar radiation) at 27 meteorological stations over the Xiang River basin are obtained from CMA. The available hourly streamflow observation at Xiangtan Station is provided by Hunan

Hydrological Bureau of China. Fig. 2 shows the time series of the hourly streamflow and corresponding gauge-based precipitation between 2015 and 2017, where eleven historical flood events are selected with in this study. The flood events are the streamflow time series with one-month span whose peak flow exceeded 8600 m^3/s, corresponding to 97^{th} approximately the quantile level(Zhu et al., 2020a). The period of the time series containing the selected flood events is from April 2014 to December 2017. The DEM (digital elevation model) with 90m resolution is derived from NASA Shuttle Radar Topographic Mission (SRTM) (Farr et al., 2007). Land cover and soil data with a resolution of 1km are obtained from Global Land Cover 2000 and Environmental and Ecological Science Data Center for West China, National Natural Science Foundation of China, respectively.

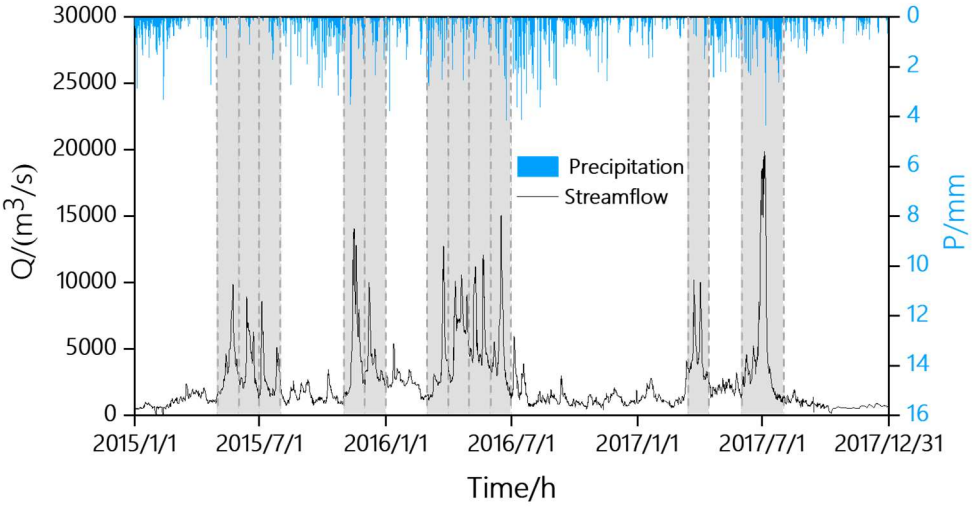


Fig. 2. Time series of observed hourly streamflow in Xiangtan station and basin-average precipitation from CMA, with eleven selected flood events covered by shaded areas.

3. Methodology

In this study, the IMERG precipitation products (IMERG-E, IMERG-L and IMERG-F) are assessed against the reference precipitation, namely CMA, under different spatio-temporal resolutions. As mentioned above, three widely used and typical conceptual/physically based models (lumped HBV model, semi-distributed SWAT model, and distributed DHSVM model), and one data-driven model (LSTM), are employed to probe the impacts of spatio-temporal resolutions of precipitation on flood events simulation. To investigate the impacts of spatial resolutions of precipitation on flood simulation, precipitation estimates with different spatial resolutions, which is obtained by inverse distance interpolation (Franke, 1982), are used to force the selected models, which are SWAT and DHSVM, as well as LSTM. To study the influence of temporal resolutions of precipitation on flood simulation, HBV, DHSVM and LSTM are utilized, which are forced by precipitation with different temporal resolutions. These four models are calibrated with two calibration strategies to investigate the potential benefits of the calibration strategy based on flood events. Finally, the performance of flood events simulation under different scenarios are compared and discussed. The designed framework for this study is shown in a flowchart in Fig. 3.

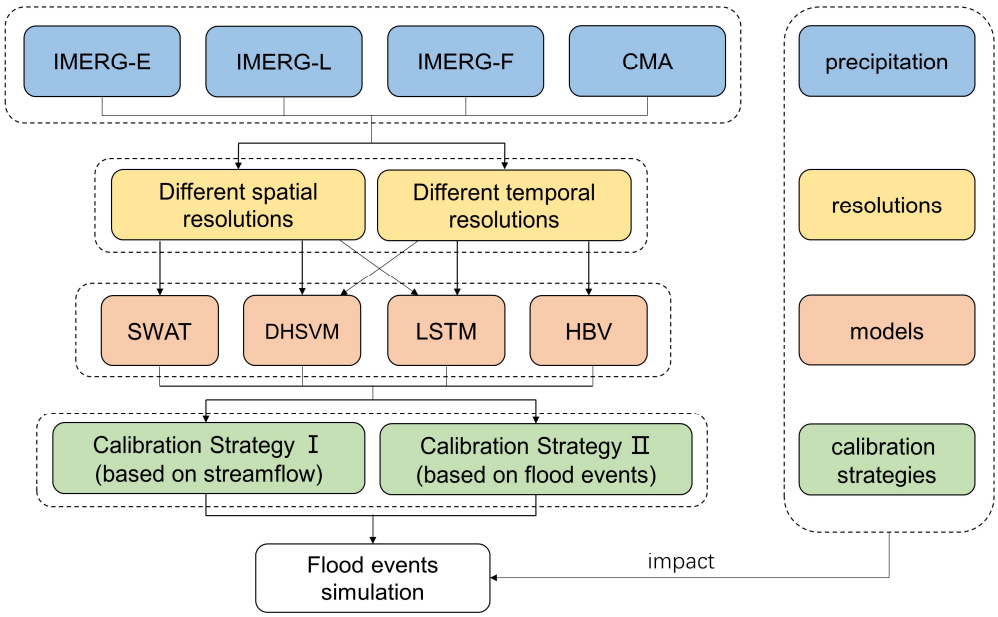


Fig. 3. Methodological flowchart adopted in this study.

3.1 Hydrological models and LSTM

3.1.1 The HBV model

The conceptual HBV model is originally developed by the Swedish Meteorological and Hydrological Institute (SMHI) (Bergström and Forsman, 1973). Various versions of the HBV model have been developed and widely used in hydrological simulation and flood forecasting due to its simplicity and effectivity (Alfredsen and Hailegeorgis, 2015; Grimaldi et al., 2019; Huang et al., 2019). A lumped version of HBV model (Aghakouchak et al., 2013) is used in this study, which is operated at hourly and daily time steps with the inputs of precipitation, temperature and potential evapotranspiration. The potential evapotranspiration is calculated with the Penman-Monteith equation (Beven, 1979) based on gauge meteorological data, and all the inputs are averaged over the basin with the Thiessen polygon method. Three main modules (soil moisture routine, response routine, and transformation routine) are contained in HBV model, while the module of snow routine is not included in this case because of the temperature above 0℃ perennially over the Xiang River basin.

3.1.2 The SWAT model

The SWAT model is a semi-distributed hydrological model developed by the Agricultural Research Center of the United States Department of Agriculture (USDA) (Arnold et al., 1998). The SWAT 2012 is used in this study, which is operated on daily time step with the inputs of geographical data (DEM, land use and soil) (Figure 4), precipitation and other meteorological variables mentioned above. The SWAT model divides the watershed into sub-basins according to DEM, and then segregates them into multiple hydrological response units (HRUs) as the basic computational unit based on different types of soil, land use, and slope. The Xiang River Basin is divided into 25 sub-basins and 495 HRUs in this study. Forest-evergreen is the dominant land cover category with coverage of 62%, and Ferralsols is the main soil type with coverage of 58%, as shown in Fig. 4. The hydrologic cycle simulated by SWAT is based on the water balance equation, which mainly includes surface runoff, evapotranspiration, soil moisture and groundwater.

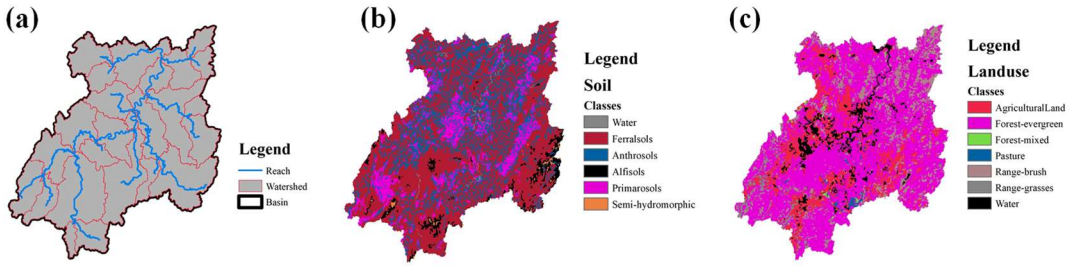


Fig. 4. The (a) sub-basin divisions, (b) soil types, and (c) land use of Xiang River Basin used in SWAT model.

3.1.3 The DHSVM model

The DHSVM model is a fully distributed, physics-based hydrological model developed by the Pacific Northwest National Laboratory (PNNL) and the University of Washington (Wigmosta et al., 1994). DHSVM uses near-surface meteorology including air temperature, wind speed, humidity, precipitation, as well as incoming short- and long-wave radiation as hydro-climate inputs to solve energy and water balance. The model represents a dynamic watershed process at specific spatial scales considered the effect of topography, soil, and vegetation. The DHSVM model mainly consists of seven modules, including evapotranspiration module, surface snow melting module, canopy snow melting module, unsaturated soil moisture module, saturated soil flow module, surface runoff module and flow routing module. The version used in this study is DHSVM 3.1.2 with the grid resolution of 3,000 m. Six soil types and eight vegetation classes are derived, and the spatial distributions of them are shown in Fig.5.

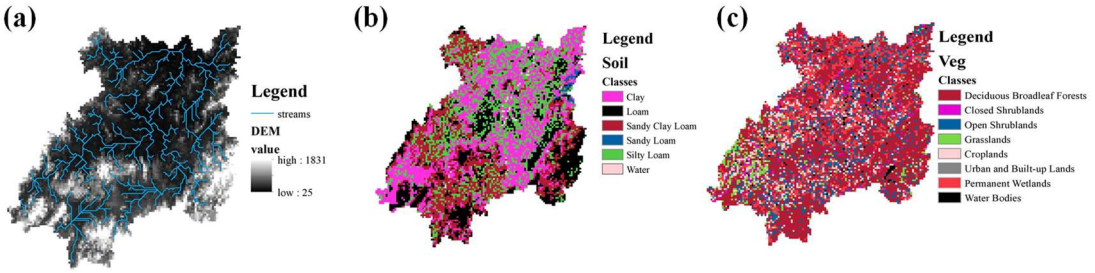


Fig. 5. The (a) river network divisions, (b) soil types, and (c) vegetation types of Xiang River Basin used in DHSVM model.

3.1.4 The Long Short-Term Memory network

The LSTM is a type of recurrent neural network (RNN), which is first proposed by Hochreiter and Schmidhuber (1997). LSTM is designed to overcome the error backflow problems with exploding and vanishing gradients by introducing three gates, namely forget, input, and output gates, into the repeating modules of neural network. The forget gate decides the information removed from the previous cell state. The input gate determines information updated to the present cell state, and the output gate controls which part of the cell state output to the new hidden state. Therefore, LSTM can learn long-term dependencies between input and output features, which makes it appropriate for rainfall-runoff modelling. In this study, LSTM is developed using the deep learning framework PyTorch (Paszke et al., 2019), which has 100 hidden states and a single fully connected layer with a dropout rate of 0.5 (Srivastava et al., 2014). Precipitation and temperature are selected as the inputs of LSTM, and the output of LSTM is streamflow. The inputs for the complete sequence $x = [x_1, \ldots, x_n]$, where x_t is a vector containing the input features of time t, and the dimension of the x_t corresponds to the number of grids of the

precipitation data. The outputs for the complete sequence $y = [y_1, \ldots, y_n]$, where y_t is the streamflow of time *t*.

3.2 Two strategies for parameter calibration

3.2.1 Calibration Strategy I

As stated above, almost all the parameter calibration for hydrological modes is based on entire streamflow time series, which is defined as Calibration Strategy I in this study. It is a conventional calibration method to optimize the parameters of hydrological model. For the HBV model, the whole period is divided into three periods: warm-up period (April 2014 to December 2014), calibration period (January 2015 to December 2016) and validation period (January 2017 to December 2017). The calibration is conducted by maximizing the Nash–Sutcliffe efficiency coefficient (NSE) of the streamflow simulated during calibration period via the SCE-UA algorithm (Duan et al., 1994).

For the SWAT model, the whole period is also divided into three periods, and they are the same as HBV. The calibration is accomplished with a separate tool named SWAT Calibration and Uncertainty Program (SWAT-CUP) (Abbaspour et al., 2007). Parallel Sequential Uncertainty Fitting Version 2 (SUFI-2) is stable and always converging, and it is well appropriate for global optimization (Abbaspour et al. 2007), which is the reason why it is adopted in this study for parameter calibration. The objective function is also to reach the maximum value of NSE for the streamflow simulated in the calibration period.

The warm-up, calibration, and validation periods of DHSVM are the same as HBV and SWAT, as well as the objective function. The parameter calibration of DHSVM is executed by an auto-calibration module based on ε-dominance non-dominated sorted genetic algorithm II (ε-NSGAII) (Pan et al., 2018). Parallel computing with a message passing interface (MPI) program is applied in this study.

Regarding the training of LSTM, the learnable parameters of the network are updated depending on a given loss function. Same as the selected hydrological models, the NSE is chosen as the objective criterion for the LSTM (Kratzert et al., 2019), and adaptive moment estimation (Adam) (Kingma and Ba, 2014) with the learning rate of 0.0001 is used as the optimization algorithm. The data set is divided into three parts generally, namely training, validation, and test data. The first two parts are used to determine the parameters of the networks, and the last one is used to evaluate the performance of actual application. In this study, the whole data set is divided into training set (October 2015 to December 2017) and validation set (April 2014 to September 2015). The absence of test set is because of the limited available period of the data, while the selection of training period will be discussed in detail in section 5.3. Each LSTM network is trained with three different random initial seeds for 1,500 epochs to account for the stochasticity in the network initialization. Among total 4,500 trained models, the best model is selected through comprehensive consideration of both calibration and validation NSE of the streamflow simulation.

3.2.2 Calibration Strategy II

Calibration Strategy II is designed in this study particularly for flood events, which conducts the calibration based on the performance of flood event simulation. Eleven historical flood events occurred between January 2015 and December 2017 are selected to conduct the flood events simulation (Fig. 2). The calibration is conducted by maximizing the mean NSE of the flood events simulated during calibration period for the HBV model. For the SWAT and DHSVM models, numerous sets of parameters (the number is 1,000 in this study) are obtained through optimization algorithm, and the best fitted parameters set is selected with the largest NSE for the flood events simulation. Considering the LSTM,

among total 4,500 trained models, the best model is also selected by maximizing the mean NSE of the flood events simulation (4 flood events during calibration and 4 flood events during validation).

3.3 Diagnostic statistics

To quantitatively evaluate the performance of streamflow and flood events simulation, three evaluation indices are selected in this study, namely NSE, BIAS-P and KGE. The formulas of these indices are listed as follows:

$$NSE = 1 - \frac{\sum_{t=1}^{T}\left(Q_o^t - Q_s^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - \overline{Q_o}\right)^2} \quad (1)$$

$$BIAS - P = \left|\frac{Q_s^p - Q_o^p}{Q_o^p} \times 100\right| \quad (2)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (3)$$

Where Q_o^t and Q_s^t are the values of the observed and simulated flood events at time t; Q_o^p and Q_s^p are the observed and simulated flood peaks of the flood events; r is the linear correlation between observations and simulations, α a measure of the flow variability error, and β a bias term.

4. Results

4.1 The performance of flood events simulation based on two different calibration strategies

Fig. 6 shows the distributions of NSE and BIAS-P values, which are used to evaluate the performance of four precipitation sources on flood events with two different calibration strategies at daily scale.

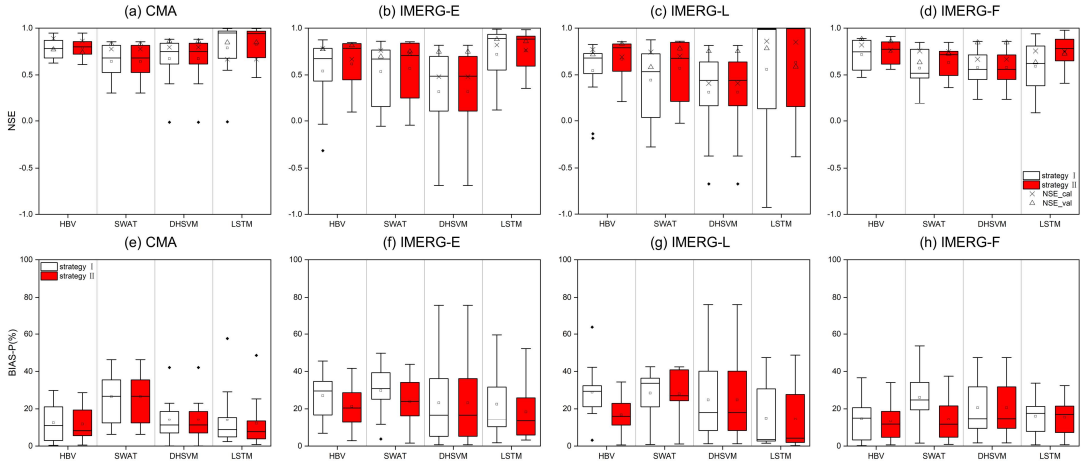


Fig. 6. The NSE and BIAS-P of flood events simulation forced by (a, e) CMA, (b, f) IMERG-E, (c, g) IMERG-L and (d, h) IMERG-F using two calibration strategies (White box is based on calibration strategy I; red box is based on calibration strategy II). The box plots show the 25th, 50th, and 75th percentiles, and the mean value is given and shown by a square. The cross represents the NSE of simulated streamflow during calibration, and the triangle represents the NSE of simulated streamflow during validation.

For the performance of HBV, it can be seen that flood events simulation with calibration strategy II shows better performance, for the mean NSE values of CMA, IMERG-E, IMERG-L, IMERG-F increase from 0.78, 0.54, 0.54, 0.72 with calibration strategy I to 0.79, 0.62, 0.67, 0.75 with calibration strategy II, respectively (Fig. 6). And the corresponding mean BIAS-P values decrease from 12.0%, 27.0%, 29.0%, 14.6% to 11.4%, 21.2%, 16.7%, 13.1%. Meanwhile, the uncertainty of NSE and BIAS-P values of flood events simulation is reduced, with less occurrences of poor flood events simulation. The flood events simulated by the CMA have the highest NSE among all precipitation sources, ranging from 0.61 to 0.95,

and its averaged value is 0.79. It proves the capability of HBV in flood events simulation. When comparing the performance of IMERG precipitation estimates, the IMERG-F performs the best with both calibration strategies.

In terms of the streamflow and flood events simulation based on SWAT, Fig. 6 shows that the performance of the two calibration strategies with CMA is comparable, while for IMERG precipitation estimates, the strategy II outperforms the other one. Specifically, for streamflow simulation, the NSE values in the validation period of IMERG-E, IMERG-L, IMERG-F show a significant increase from 0.70, 0.58, 0.63 with the strategy I to 0.75, 0.78, 0.73 with the strategy II, respectively. For flood events simulation, the mean NSE values based on the strategy II are 0.57, 0.58, 0.63 forced with IMERG-E, IMERG-L, IMERG-F, which are 0.53, 0.44, 0.57 based on the strategy I. The corresponding mean BIAS-P values are reduced from 29.8%, 28.4%, 26.1% to 23.9%, 28.0%, 13.2%. Compared to HBV and SWAT, the two calibration strategies present little difference in streamflow and flood events simulation based on DHSVM, which indicates the performance of DHSVM is stable when using different calibration strategies.

For the LSTM, the NSE values of flood events simulation also show higher mean values and smaller uncertainty based on the strategy II for all precipitation products. The flood events simulation based on IMERG-L shows the most significant improvement with the mean NSE value increasing from 0.62 with the strategy I to 0.77 with the strategy II. The flood events simulation based on CMA and IMERG-E show lightly lower medium NSE values of 0.94, 0.88 with the strategy II than 0.95, 0.99 with strategy I. But they show higher 25th NSE with strategy II, especially LSTM driven by IMERG-E, which increase from 0.58 with strategy I to 0.66 with strategy II. Therefore, although strategy II has a lower median performance than strategy I in individual cases, it still significantly improves the performance of LSTM, particularly in terms of uncertainty.

According to the above results, it can be concluded that the calibration strategy II outperforms than the strategy I. Therefore, the following parts are based on the calibration strategy II.

4.2 Impact of spatial resolutions of precipitation on flood events simulation

To investigate the impact of spatial resolutions of precipitation on flood events simulation, the IMERG-E, IMERG-L, IMERG-F, and CMA are adopted to force the SWAT model, the DHSVM model and the LSTM model under 0.1°, 0.25° and 0.5°, respectively.

Fig. 7 shows the distributions of statistical indices, namely NSE, BIAS-P and KGE, which are used to evaluate the performance of different precipitation sources with different spatial resolutions on flood events simulation. From the BIAS-P of flood events simulated with SWAT, it can be seen that spatial resolution significantly affects the performance of precipitation on flood events simulation. For instance, CMA performs the best at 0.25° with the mean BIAS-P of 26.5%, while IMERG-E, IMERG-L and IMERG-F display the best performance at 0.5° with the mean BIAS-P of 23.7%, 22.9% and 13.8%, respectively. Similar to its performance in BIAS-P, in terms of mean NSE, CMA also performs the best under 0.25° with the mean NSE of 0.66. IMERG-E presents little difference at different spatial resolutions, while IMERG-L performs slightly better at 0.5° with the mean NSE of 0.61and the middle NSE of 0.76. The performance of IMERG-F gets worse as the resolution is coarser, regardless of the NSE or BIAS-P values. According to the KGE values, the performances based on CMA, IMERG-E and IMERG-L show improvement at coarser spatial resolutions. Except for IMERG-F, whose KGE values are stable at 0.71.

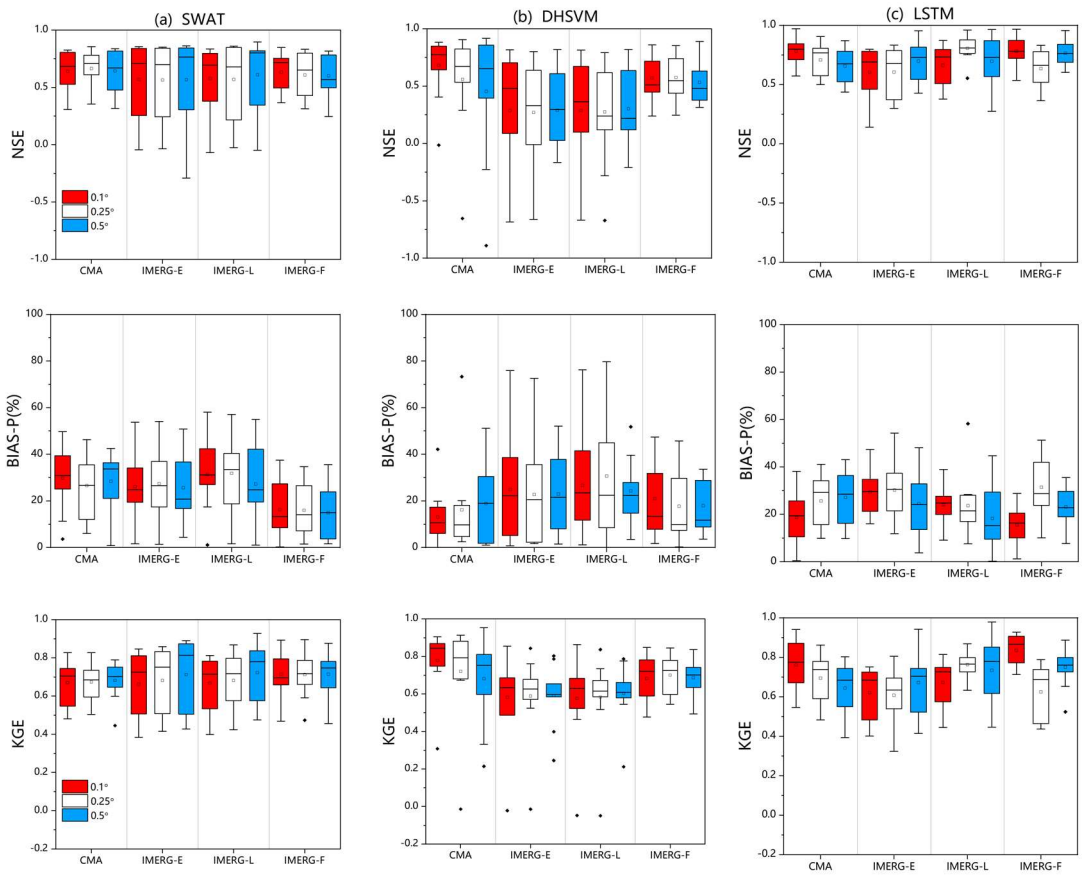


Fig. 7. The performance of flood events simulation based on (a) SWAT, (b) DHSVM, and (c) LSTM forced by precipitation with different spatial resolutions. The box plots show the 25th, 50th, and 75th percentiles, and the mean value is given and shown by a square.

Compared to SWAT, DHSVM shows different performance forced by precipitation with different spatial resolutions. The mean NSE of flood events simulated with CMA declines from 0.68 to 0.45 when the spatial resolution of precipitation changes from 0.1° to 0.5°, the mean KGE declines from 0.77 to 0.66, and the mean BIAS-P increases from 13.9% to 19.3%. By contrast, the difference of flood events simulated with IMERG forcing at different spatial resolutions is smaller, for instance, the mean NSE values decrease from 0.29 to 0.27 for IMERG-E, 0.30 to 0.29 for IMERG-L, and 0.58 to 0.53 for IMERG-F. However, the uncertainty of NSE, KGE and BIAS-P values of flood events simulated with IMERG decreases as the spatial resolution is finer. Among 11 flood events simulation, the performance of 4 flood events simulation get better as the spatial resolution gets coarser. And the difference among the three IMERG precipitation estimates is illustrated clearly in Fig. 7(b): the distribution of NSE, KGE and BIAS-P of simulated flood events forced with IMERG-E is more scattered than the others, while the uncertainty of IMGER-F is the smallest.

Similar to DHSVM, LSTM shows different performance forced by precipitation with different spatial resolutions. CMA and IMERG-F performs the best at 0.1° with the mean BIAS-P of 18.64%, 15.55% and mean NSE of 0.78,0.78. The 25th NSE of flood events simulated with CMA increases from 0.52 to 0.72, the 75th NSE increases from 0.78 to 0.83 while the spatial resolution is finer. The KGE show the same pattern to NSE for CMA and IMERG-F. By contrast, IMERG-E performs the best at 0.5° with the mean NSE of 0.69 and medium NSE of 0.68 while IMERG-L performs the best at 0.25° with the mean

NSE of 0.80 and the medium NSE of 0.81. In the light of BIAS, IMERG-E and IMERG-L achieve the best performance on flood events simulation at 0.5°, the mean values of which are 24.55%, and 18.27%, 0.77. In contrast of BIAS-P, LSTM driven by IMERG-L shows the best KGE at 0.25° with the mean KGE of 0.76 and the smallest uncertainty, which is the same as NSE.

Compared with the SWAT and DHSVM, the LSTM shows better performance on flood events simulation. The mean NSEs of LSTM are higher than 0.7 in most cases, while the mean NSEs of SWAT is around 0.6, and the largest mean NSE of DHSVM is 0.68. The 25th NSE of LSTM are higher than 0.5 in most cases, while the 25th NSE of DHSVM is around 0.15. The smallest 75th NSE of LSTM is 0.78, while the 75th NSE of DHSVM are around 0.6. The mean KGEs of SWAT and LSTM are similarly around 0.7, which are around 0.6 for DHSVM. In addition, LSTM also shows a relatively lower BIAS-P (the mean values less than 25%).

4.3 Impact of temporal resolutions of precipitation on flood events simulation

To investigate the impact of temporal resolutions on flood events simulation, HBV, DHSVM and LSTM are adopted to be forced by the selected four precipitation sources at hourly and daily time scales. To compare the influences of temporal resolutions, the flood events simulated at hourly scale are aggregated into daily time series.

The performance of different precipitation datasets with different temporal resolutions on flood events simulation is shown in Fig. 8. In HBV-based simulation, the mean NSE of the flood events simulation at the hourly scale is about 0.03 higher than that at the daily scale for all precipitation products. The mean KGE of the flood events simulation at the hourly scale is also higher than that at the daily scale for IMERG forcing, while the mean KGE of flood events simulated with CMA shows a decrease of about 0.03 at the hourly scale. In terms of BIAS-P, compared with the small difference between the performance of flood events simulated with CMA at hourly and daily scales, the performance of IMERG-E, IMERG-L and IMERG-F on flood events simulation at hourly scale is much better than that at daily scale (mean BIAS-P values of 15.1% vs 21.2%, 13.7% vs 16.7% and 11.1% vs 13.1% for IMERG-E, IMERG-L and IMERG-F, respectively).

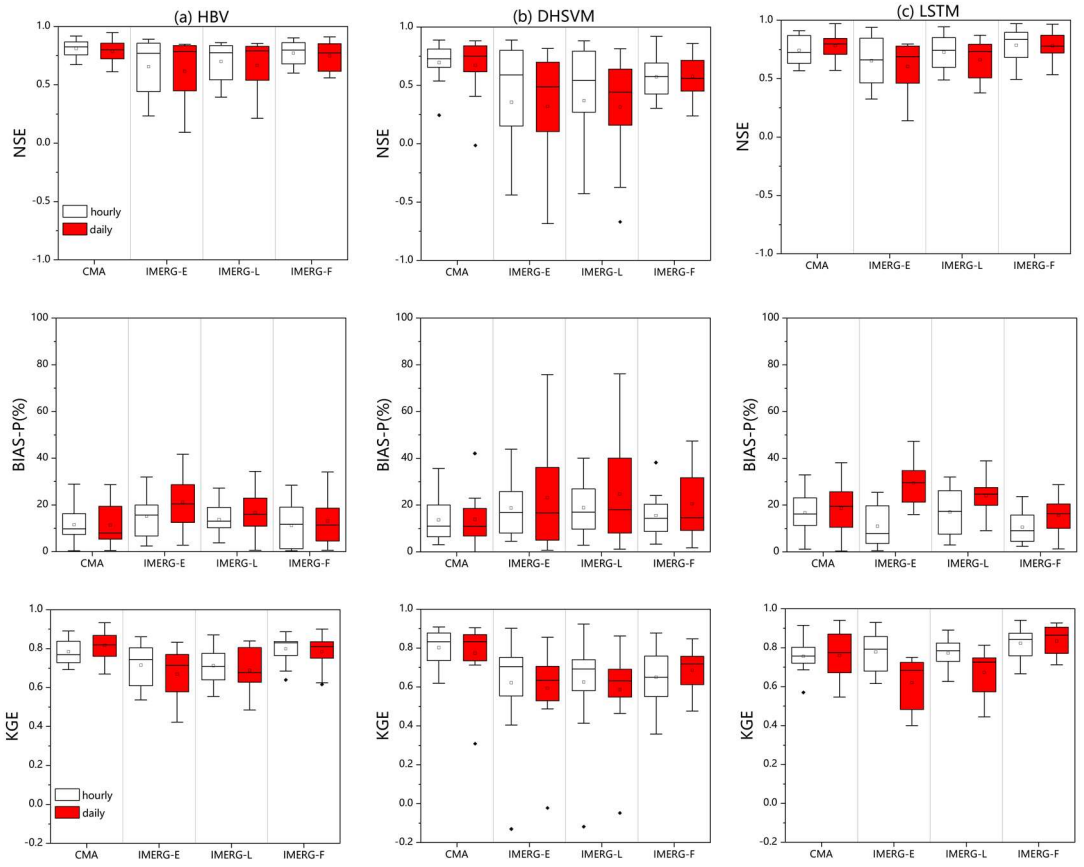


Fig. 8 The performance of flood events simulation based on (a) HBV, (b) DHSVM, and (c) LSTM forced by precipitation with temporal resolution. The box plots show the 25th, 50th, and 75th percentiles, and the mean value is given and shown by a square.

Similar performance is also presented in DHSVM-based simulation. According to the NSE and KGE values, the performances based on all precipitation products show improvement at the hourly scale. More obvious improvement is shown in terms of BIAS-P, which is dropped by 5% at the hourly scale for IMERG products.

The performance based on LSTM is also shown in Fig. 8. Consistent with the results obtained by the HBV and DHSVM, all precipitation sources also have relatively better performance at the hourly scale. For example, the mean BIAS-P of CMA is reduced from 18.64% at the daily scale to 16.7% at the hourly scale. And IMERG-E, IMERG-L and IMERG-F obtain better performance at the hourly scale with the mean NSE of 0.65, 0.73 and 0.78, the mean KGE of 0.78, 0.77 and 0.82, respectively.

Compared with the HBV and DHSVM, the LSTM shows higher mean NSE values of flood events simulation, except for the simulation based on IMERG-L, while the HBV forced by CMA and IMERG-F presents smaller uncertainty. In terms of BIAS-P, two models show comparable performance with the mean values around 15%. The performance on flood events simulation of HBV is more stable but slightly poorer than LSTM in general.

5. Discussion

5.1 Comparison of two different calibration strategies

Two different calibration strategies are used to simulate flood events in this study. Compared with the conventional method choosing the fit parameter set based on entire streamflow time series (Calibration

Strategy I), selecting the parameter set that results in the best flood events simulation (Calibration Strategy II) shows better performance on flood event simulation (Fig. 6). However, the CMA shows the similar results under two different calibration strategies in SWAT-based flood events simulation, so as the DHSVM-based simulation. Furthermore, the CMA shows little difference with other precipitation forcing. Although we targeted in difference between the two strategies in flood events simulation, their performances in the whole streamflow simulation time series are also compared, which is presented in Table 1 (The mean value is the average NSE of the four precipitation products with the same calibration strategy). According to the mean NSE values, calibration strategy II outperforms calibration strategy I. To be specific, for HBV, SWAT, DHSVM and LSTM models, among the four precipitation products, there are two, three, three and three NSE values larger with calibration strategy II than that with calibration strategy I. These findings indicate that both precipitation accuracy and calibration strategy used in hydrological models are important uncertainty sources for flood simulation. From lumped model to distributed model, precipitation accuracy becomes the major source of uncertainty to streamflow/flood events simulation instead of hydrological model, the reason of which is that hydrological models describe the hydrological process more and more comprehensively. In the application of LSTM for flood events simulation, a large number of equivalent simulations with different parameter sets are generated, which is similar to the parameter equifinality in hydrological simulation. When comparing the two calibration strategies, the calibration strategy II is an effective way to train the LSTM model to obtain the best flood events simulation.

Table. 1 The NSE values of the whole streamflow simulation time series forced by CMA, IMERG-E, IMERG-L, IMERG-F

Model	Strategies	CMA	IMERG-E	IMERG-L	IMERG-F	Mean
HBV	Strategies I	0.77	0.77	0.72	0.88	0.79
	Strategies II	0.73	0.81	0.82	0.86	0.80
SWAT	Strategies I	0.83	0.75	0.76	0.73	0.77
	Strategies II	0.83	0.84	0.82	0.70	0.80
DHSVM	Strategies I	0.86	0.75	0.75	0.85	0.80
	Strategies II	0.82	0.87	0.86	0.87	0.86
LSTM	Strategies I	0.92	0.89	0.87	0.85	0.88
	Strategies II	0.93	0.91	0.86	0.85	0.89

5.2 Comparison of the performance of precipitation products on flood events simulation at different spatio-temporal resolutions

As illustrated in Fig.7 and Fig.8, the performance of precipitation products on flood events simulation is affected by both the spatial and temporal resolutions. Impacts of spatial resolution on flood events simulation behave differently among different models and precipitation sources. For the study area, under 0.25° spatial resolution, the CMA obtains the best flood events simulation based on SWAT. The impact of spatial resolution on the capture of precipitation variability during flood event periods can propagate to the flood events simulation. The best results are obtained under 0.25° spatial resolution, the possible reason can be that finer spatial resolution (0.1°) increases the uncertainty of precipitation sets, nevertheless coarser spatial resolution (0.5°) decreases the sufficiency of datasets. For SWAT driven by CMA, it shows the best 75th NSE and the worst 25th NSE at 0.5° while the DHSVM driven by CMA

shows the same pattern at 0.5°, which proves that coarser spatial resolution decreases the sufficiency of datasets. But the DHSVM driven by CMA shows the best performance at 0.1°, which proves that the effects of increasing and decreasing spatial resolution are simultaneous and affect different models differently. It indicates the choice of the dataset is influenced by the resolution range, which must be adapted to the model definition, for the proper spatial resolution is essential to both minimize the uncertainty and assure the sufficiency(Grusson et al., 2017).

The SWAT and DHSVM model driven by IMERG perform similarly under different spatial resolutions, which is consistent with previous research (Lobligeois et al., 2014; Huang et al., 2019), where insignificant improvement was reported with higher spatial resolution of observed rainfall in a large catchment area. It probably dues to the large catchment area and only the outlet station is used for calibration. Liang et al. (2004) found a critical resolution (1/8° for the VIC model) for a watershed with 1,233 km^2, beyond which the spatial resolution shows limited impact on model performance. For our study area (82,375 km^2), when the spatial resolution of precipitation changes from 0.1° to 0.5°, a small variation is shown in the performance of flood events simulation, which indicates the critical resolution may be larger for a large watershed. For data-driven model, CMA and IMERG-F show better performance under 0.1° spatial resolution in the LSTM-based simulation, which indicates that a higher spatial resolution, namely a larger dataset, can improve the performance of flood events simulation. Similar conclusion is drawn from previous study conducted by Sun et al. (2017), which also found that a deep learning model performs better with larger datasets. In addition, the simulation with IMERG-L and IMERG-E at 0.1° spatial resolution is not satisfactory, which may be related to the choice of hyperparameters and the limited data. However, after upscaling, the performance of LSTM in flood events simulation is greatly improved when the IMERG-L data is applied with 0.25° spatial resolution, which implies that scale transformation can be regarded as an approach of data enhancement in hydrological simulation based on deep learning.

In order to compare the performance of different models on flood events simulation in the same spatial resolutions, some results presented in Fig.7 are illustrated in Fig.9. Overall, the LSTM shows better performance in most cases, for instance, in Fig. 9 (a) and Fig. 9 (c), LSTM is better than other models with the largest mean NSE and the smallest range between 25^{th} and 75^{th} percentile. There is also exception, for example, in Fig. 9 (b), the range of NSE between 25^{th} and 75^{th} percentile of SWAT with CMA is smaller than that of LSTM, but its mean and medium values of NSE are lower. Therefore, it can be summarized that the performance of LSTM has a higher likelihood of success than the other models. For KGE at 0.1° (Fig.9 (d)), LSTM also show better performance than the other models expect that simulated with CMA, with which DHSVM is better than LSTM, and they show similar results with 0.5° (Fig. 9 (e)).

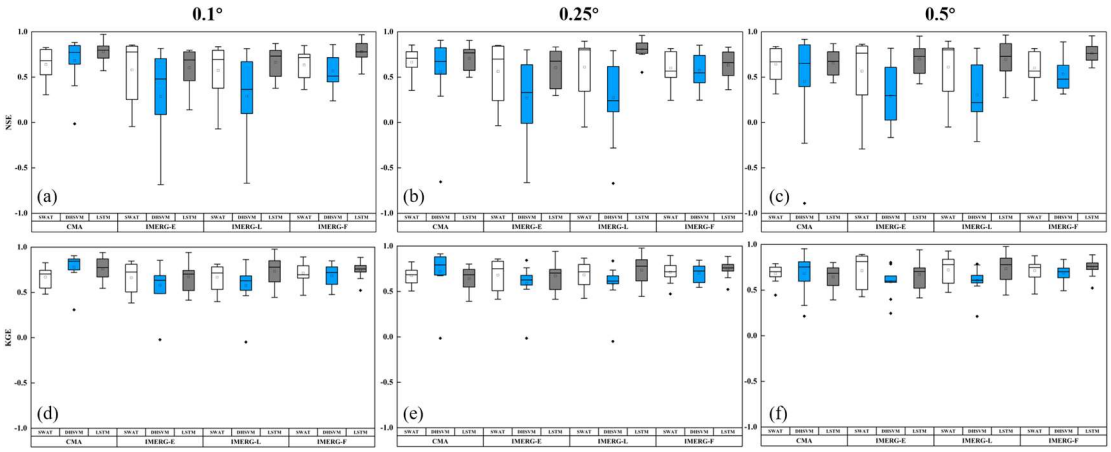


Fig. 9. The NSE and KGE of flood events simulation forced by CMA, IMERG-E, IMERG-L and IMERG-F with different spatial resolutions. The box plots show the 25th, 50th, and 75th percentiles, and the mean value is given and shown by a square.

The influence of spatio-temporal resolution on flood events simulation is affected by model structure. For instance, based on NSE, the SWAT shows the best performance at 0.25° with CMA forcing, but the LSTM shows the best performance at 0.1°. Similarly, based on KGE, the SWAT performs the best at 0.5° with CMA forcing, but the LSTM has the best performance at 0.1°. On one hand, the difference in performance between NSE and KGE is due to their different statistical focus, with NSE giving larger weights to high values, especially flood peaks, which leads to different performance with different statistical metrics. On the other hand, the difference between SWAT and LSTM is due to their model structure. The SWAT operates as a physically driven model, where the impact of the spatial resolution of the precipitation dataset will propagate during hydrological process, which makes finer spatial resolution does not necessarily lead to the improved performance as indicated by studies such as Huang et al., 2019. This is probably exemplified by the SWAT performs better at 0.25° with CMA forcing based on NSE, while it performs better at 0.5° based on KGE. Regarding LSTM, as a deep learning model, some studies have highlighted significant performance enhancements when applied to larger, reliable datasets (Sun et al., 2017). Consequently, when forced by CMA and IMERG-F, LSTM shows the best performance across all statistical metrics at 0.1°, rather than at 0.25° or 0.5°. The deviations observed in IMERG-E and IMERG-L deviate from this pattern are likely attributable to inherent errors within the precipitation product itself. We previously evaluated the applicability of the IMERG dataset in the Xiangjiang River Basin, and found that IMERG-E and IMERG-L have larger uncertainties and errors than IMERG-F (Zhu et al., 2020a). The CMA has been confirmed by several studies to be a more reliable precipitation product in the Xiangjiang River Basin and always used as a reference precipitation product (Wang et al., 2017; Tang et al., 2017; Su et al., 2020). This probably makes IMERG-E and IMERG-L do not bring enough performance improvement to LSTM when the spatial resolution is finer.

Considering the impacts of temporal resolutions on flood events simulation, for HBV and DHSVM, the flood events simulation at hourly scale outperforms than that at daily scale in general, which indicates that a higher temporal resolution can improve the performance of hydrological models. Meanwhile, hourly precipitation sources also show better performance of flood events simulation with LSTM, especially for the simulation of flood peaks.

5.3 Comparison of different models on flood events simulation

In this study, a lumped hydrological model (HBV), a semi-distributed hydrological model (SWAT), a fully distributed hydrological model (DHSVM) and a data-driven model (LSTM) are utilized to simulate flood events. In order to compare the performance of different models on flood events simulation more clearly, some results presented in Fig.6 are illustrated in Fig.10. As shown in Fig. 10, HBV and SWAT forced by CMA show comparable runoff simulation performance, while HBV shows better performance than SWAT in flood events simulation. The inability of the SWAT model to capture the flood events is also proved in previous studies (Zhu et al., 2016; Yu et al., 2018). Furthermore, when driven by IMERG, HBV outperforms SWAT and DHSVM, especially by IMERG-E and IMERG-L. It is because the hydrological model with a simpler structure can reduce the impact of errors in radar rainfall estimation, which is better constrained during its propagation in the hydrological process.(Zhu et al., 2013).

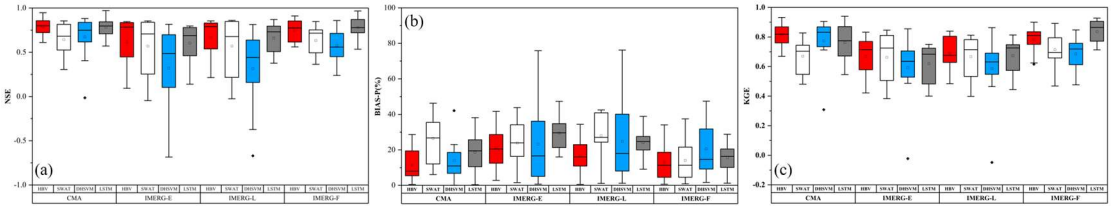
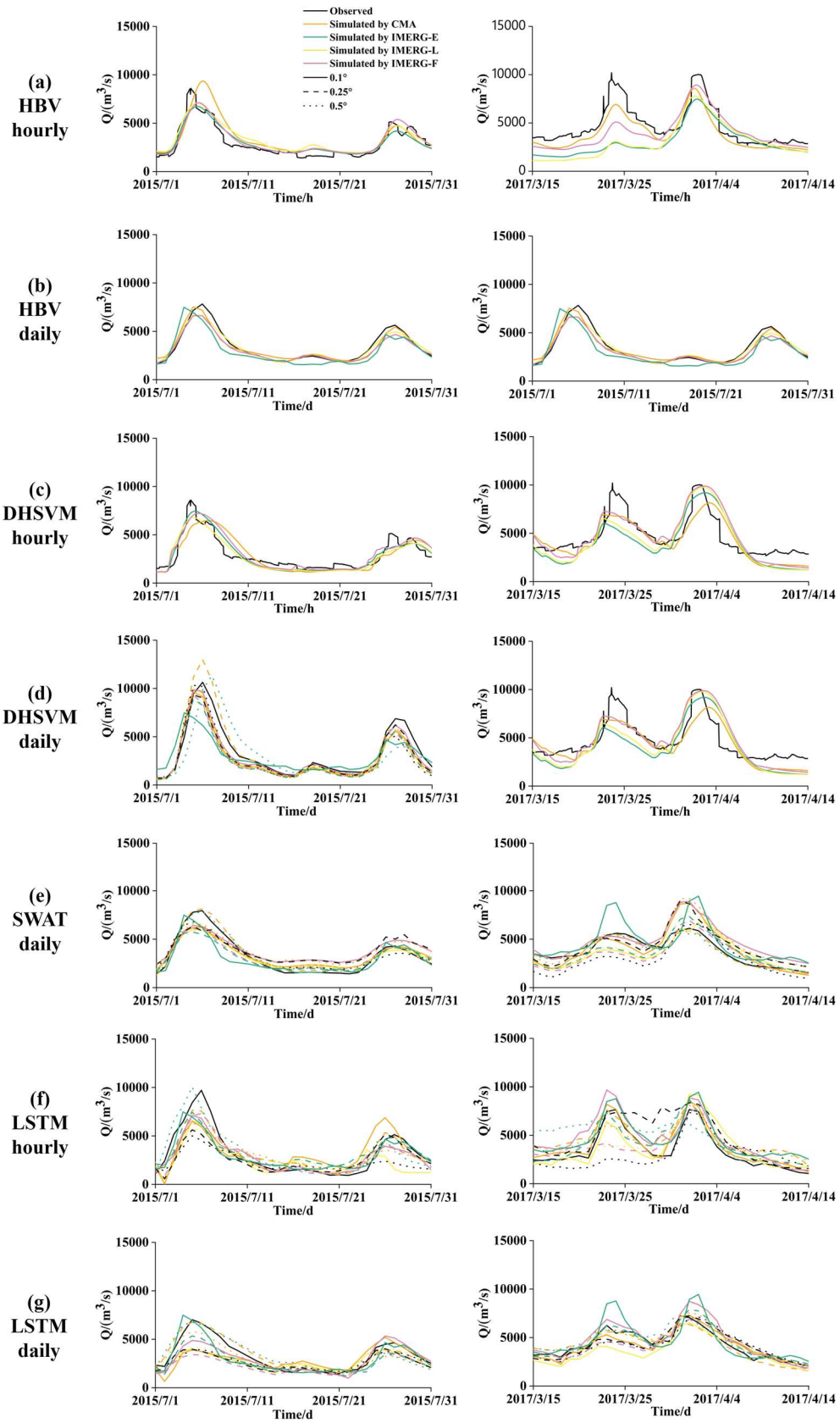


Fig. 10. The (a) NSE, (b) BIAS-P and (c) KGE of flood events simulation forced by CMA, IMERG-E, IMERG-L and IMERG-F using calibration strategies II. The box plots show the 25th, 50th, and 75th percentiles, and the mean value is given and shown by a square.

The comparisons of SWAT, DHSVM and LSTM at different spatial resolutions are also illustrated.As a data-driven approach, LSTM shows better performance than SWAT and DHSVM in terms of flood events simulation and shows reduced uncertainty and a higher likelihood of success than HBV, which is considered an appropriate model in this case. Among IMERG products, IMERG-F outperforms IMERG-E and IMERG-L in flood events simulation based on hydrological model, while IMERG-E and IMERG-L show comparable and even better performance than IMERG-F based on LSTM. This phenomenon shows that LSTM can deal with the error of precipitation products during the learning process. In many previous studies, LSTM is forced by large data sets, such as the CAMELS data set, the lower bound of data requirements used for calibration is considered as the daily time series of 15 years (Kratzert et al. 2018, 2019). In this study, as mentioned above, the calibration (October 2015 to December 2017) and validation (April 2014 to September 2015) of LSTM are different from those of hydrological models. For hydrological models, the calibration period is from January 2015 to December 2016, and the calibration period is January 2017 to December 2017. We tried to use the same calibration data in LSTM as the hydrological model, but the results about flood events simulation is not satisfactory, where its NSE of validation period is less than 0.5. The reason is that two major flood events are not included in the calibration period used in hydrological model. As a result, LSTM failed to learn the input–output relationship during the periods of flood events. Containing the characteristics of inputs as many as possible is critical for data-driven model, for instance LSTM, to capture the accurate relationship between the inputs and the output. Therefore, we use the data in the latter part for calibration, through which the performance of LSTM is significantly improved. It should be notable that reliance on data may still be a potential barrier for LSTM in the data-sparse areas. In addition to obtaining more data for the input, such as the remote sensing data, how to make good use of limited data should also be considered in the future

studies. What's more, based on the same computer specification (Intel i5-9300H CPU, 8 GB Memory), the running time of one simulation based on HBV, SWAT, DHSVM, and LSTM are 0.2 seconds, 1 minute, 54 minutes and 1.2 seconds, respectively. Results obtained from this case show that LSTM can provide reasonable accuracy in flood events simulation whilst it is also competitive in computational efficiency.

In order to compare the performance of flood events simulation with different scenarios, two randomly selected flood events simulation from July 1st, 2015, to July 31st, 2015, and from March 15th, 2017 to April 14th, 2017 are shown in the Fig. 11. The first flood event is the typical one with single peak occurred during the calibration period of HBV, SWAT and DHSVM models, and the latter one is with twin peaks occurred during the validation period of HBV, SWAT and DHSVM. While for LSTM, the occurrence times of the two selected flood events are in its validation and calibration period, respectively. From the figure, it can be seen that hydrological models generally show good capability to capture the first flood event. However, for the second flood event from March 15th, 2017 to April 14th, 2017, an obvious underestimation of the first peak exists in the flood simulation, which is primarily caused by the bias of precipitation products, which are comprehensively evaluated in our previous study (Zhu et al., 2020a). The underestimation of the second flood peak is reduced in LSTM-based simulations, which implies the ability of LSTM to correct the propagation of influence from the bias of precipitation. Since the hydrological models may smooth the short-term variability of input, the flood events simulated with hydrological models show relatively smooth runoff processes, compared with LSTM. Meanwhile, the performance of LSTM is not stable under different spatial resolutions, compared with the SWAT and DHSVM. Compared with spatio-temporal resolutions of precipitation and simulation models, precipitation source is the primary uncertainty source for flood events simulation, which indicates the importance of choosing appropriate precipitation source for ungagged regions.



580 **Fig. 11. Comparison of HBV-, SWAT-, DHSVM-, and LSTM- based flood events simulation from July 1st, 2015 to July 31st, 2015, and from March 15th, 2017 to April 14th, 2017 forced by CMA, IMERG-E, IMERG-L, and IMERG-F with different spatio-temporal resolutions.**

6. Conclusion

In this study, we investigated the impacts of temporal and spatial resolutions of precipitation on flood events simulation over a large-scale catchment. We accomplished the study with the application of HBV, SWAT, DHSVM and LSTM forced by high spatio-temporal resolution gauge-based and satellite-based precipitation products. The main conclusions of this study are summarized as follows:

(1) According to the comparison of two calibration strategies, event-based calibration strategy leads to better performance of flood event simulation based on lumped HBV model and semi-distributed SWAT model. However, there is little difference between two calibration strategies application on distributed DHSVM model. For the data-driven model, LSTM, the event-based strategy also leads to better results.

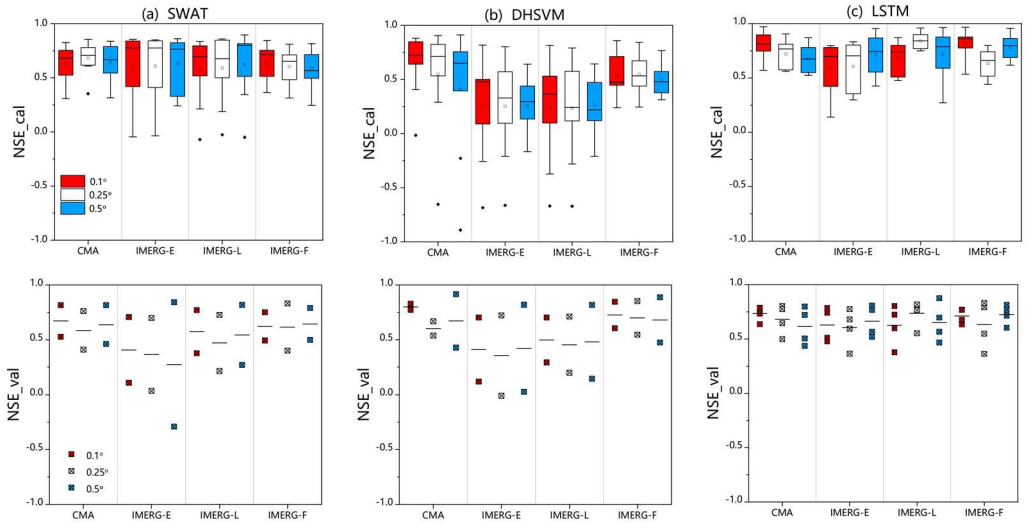
(2) Considering the impact of temporal resolution, both hydrological models and LSTM perform better at hourly scale on flood events simulation than at daily scale, especially in flood peaks. However, the influence of spatial resolution on flood events simulation has no significant pattern in this case, which varies with models and precipitation sources.

(3) Three hydrological models and LSTM are used to simulate the flood events forced by gauge-based and satellite-based precipitation products in this study. The hydrological models and LSTM forced by IMERG precipitation estimates can achieve acceptable flood events simulation in most cases. In some cases, the LSTM outperforms the hydrological models. However, it should be notable that the performance of LSTM largely depends on the input data and settings such as the choice of hyperparameters, which may be unstable in some other cases.

Acknowledgement

This study is financially supported by the National Natural Science Foundation of China (52009020) and the Natural Science Foundation of Jiangsu Province (BK20180403). This study is also financially supported by the High-level innovation and entrepreneurship talents plan of Jiangsu Province “Coupling remote sensing datasets to investigate impacts of hydrological key variables on flood extremes”, and partially supported by the U.S. Department of Energy (DOE Prime Award # DE-IA0000018).

Appendix A



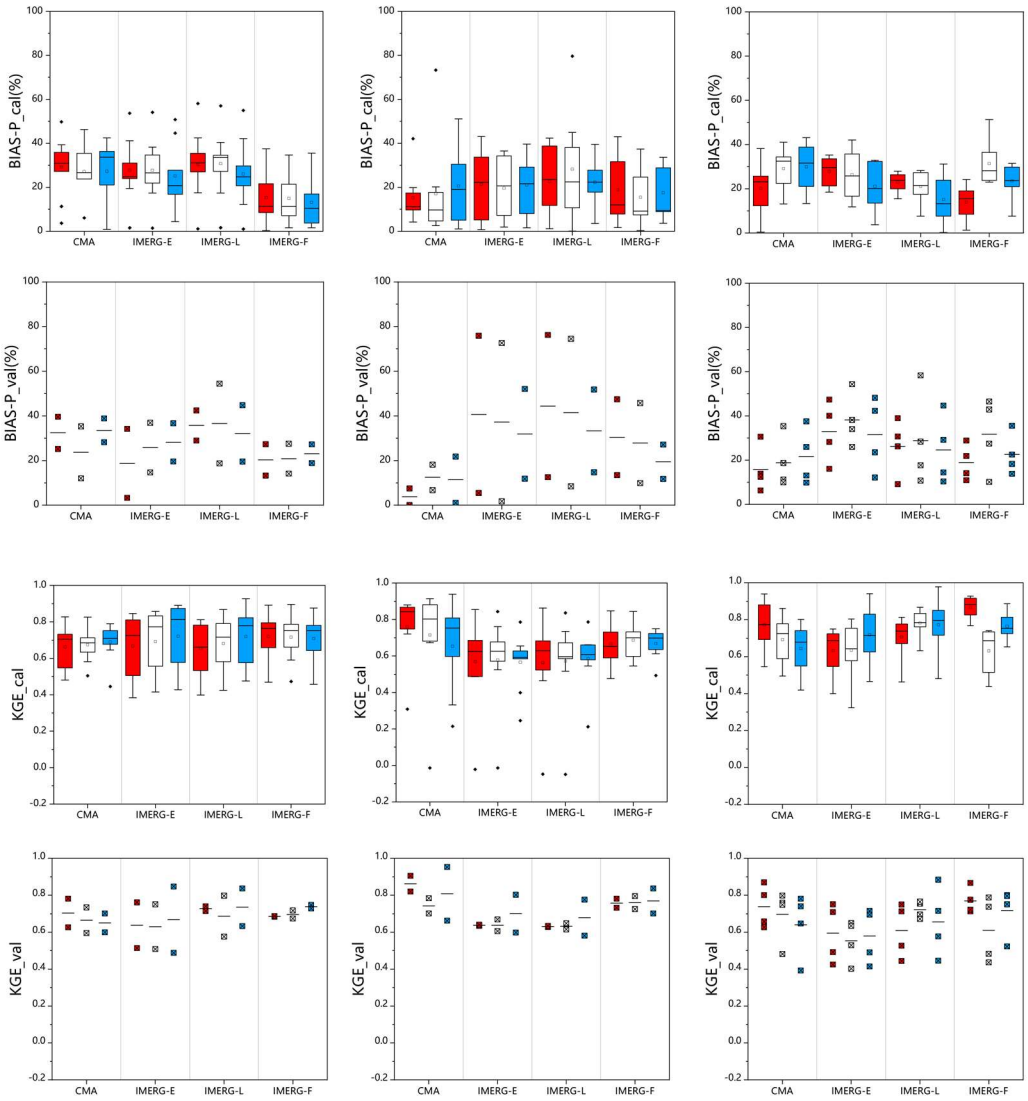
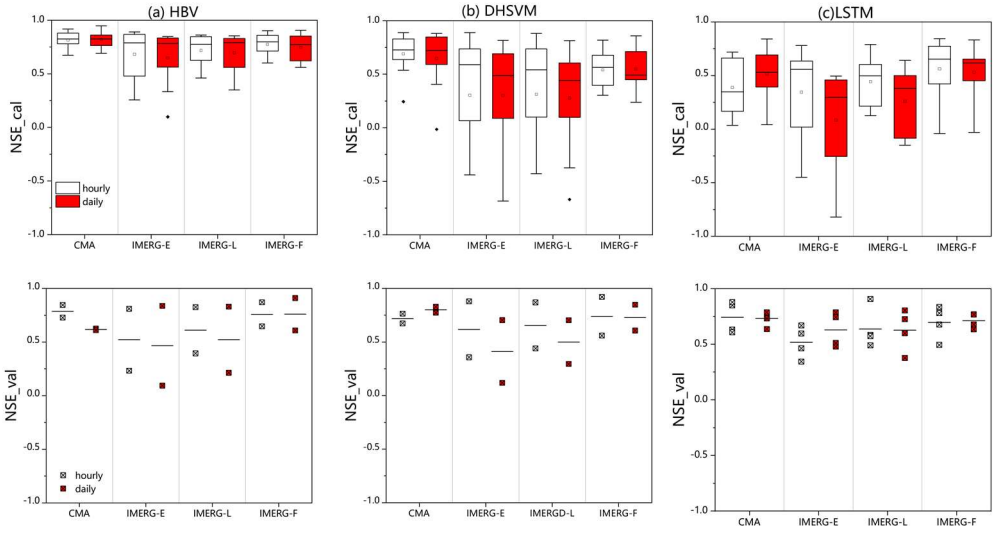


Fig. A1. Same as Fig. 7, but the results in calibration and validation periods are separated.

Appendix B



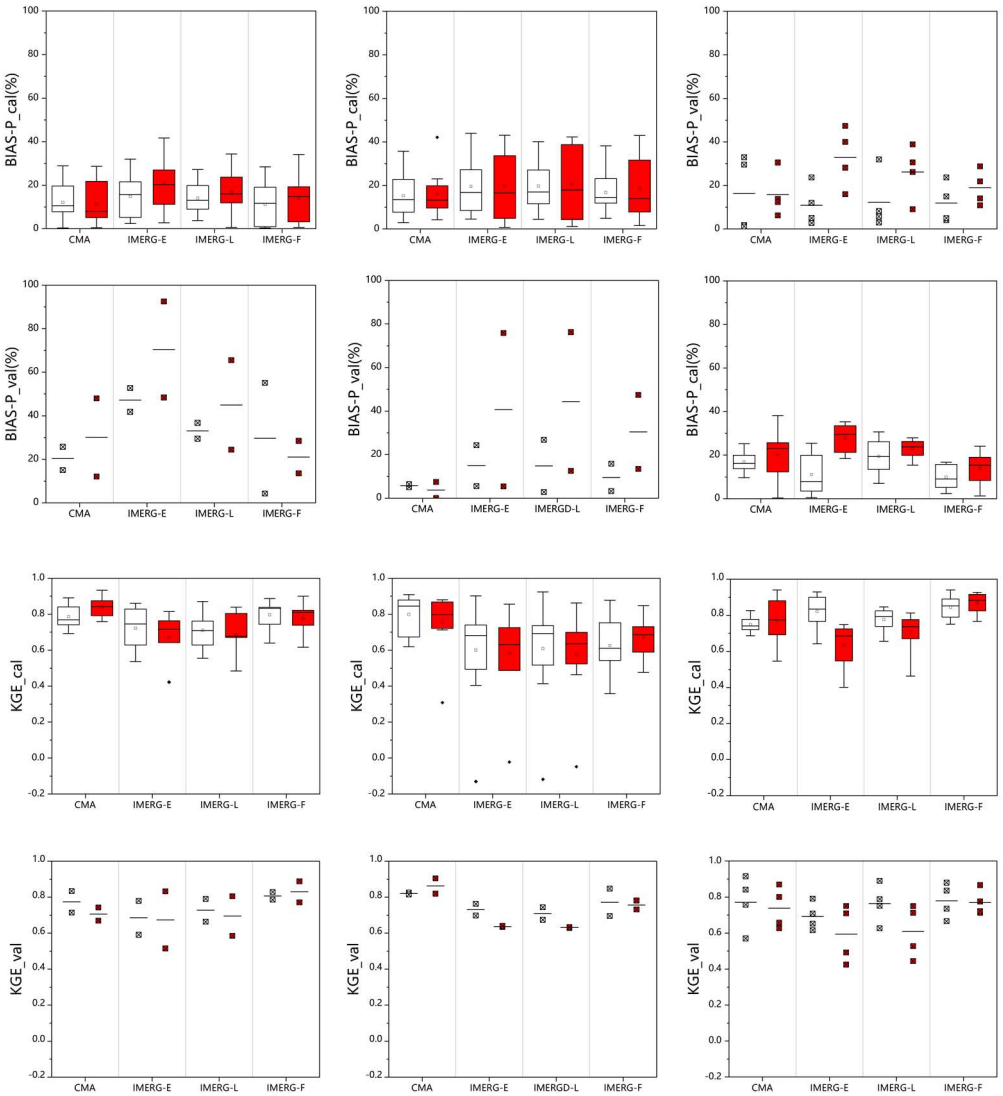


Fig. B1. Same as Fig. 8, but the results in calibration and validation periods are separated.

Appendix C

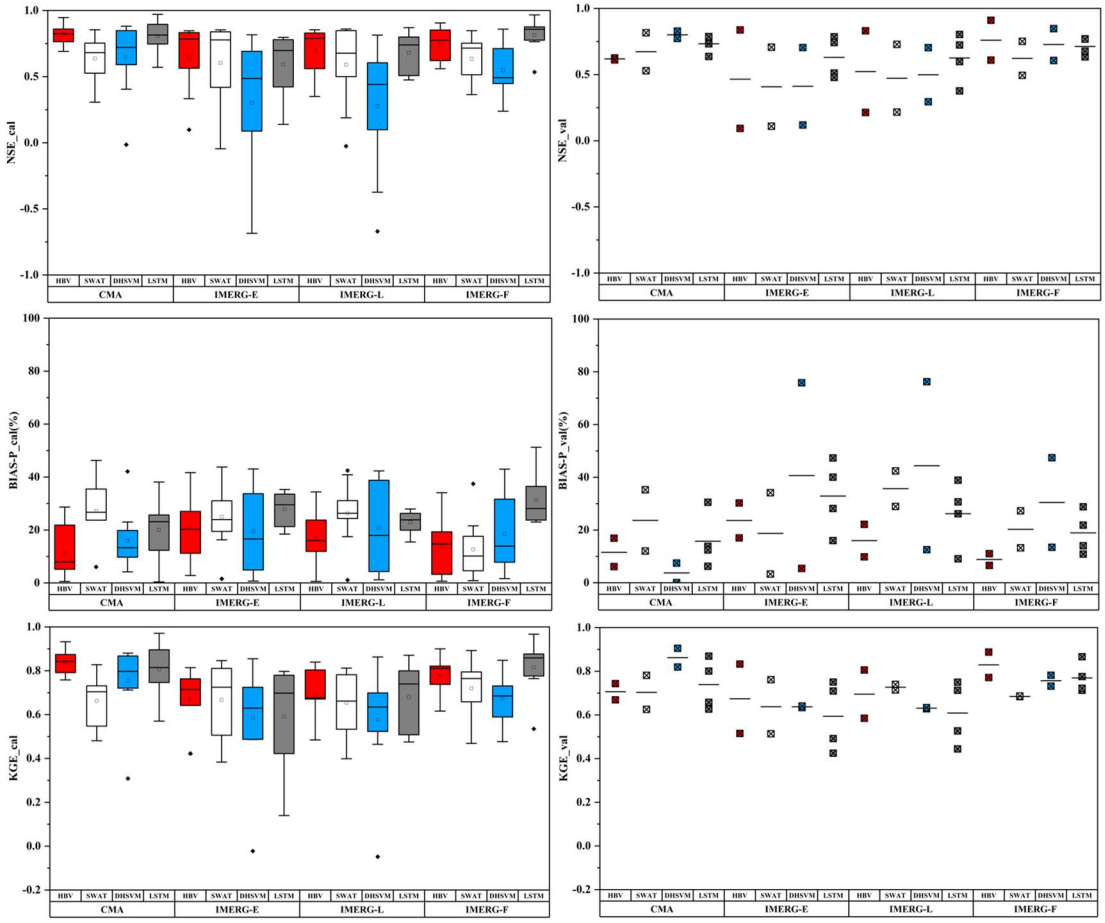


Fig. C1. Same as Fig. 10, but the results in calibration and validation periods are separated.

Reference

Abbaspour, K. C., Vejdani, M., and Haghighat, S.: SWAT-CUP Calibration and Uncertainty Programs for SWAT, Modsim 2007: International Congress on Modelling and Simulation, 1603-1609, 2007.

AghaKouchak, A., Nakhjiri, N., and Habib, E.: An educational model for ensemble streamflow simulation and uncertainty analysis, Hydrology and Earth System Sciences, 17, 445-452, 10.5194/hess-17-445-2013, 2013.

Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., and Peng, Q.: Short-Term Precipitation Forecast Based on the PERSIANN System and LSTM Recurrent Neural Networks, Journal of Geophysical Research: Atmospheres, 123, 10.1029/2018jd028375, 2018.

Alfredsen, K. and Hailegeorgis, T. T.: Comparative evaluation of performances of different conceptualisations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments, Hydrology Research, 46, 607-628, 10.2166/nh.2014.051, 2015.

Apip, Sayama, T., Tachikawa, Y., and Takara, K.: Spatial lumping of a distributed rainfall-sediment-runoff model and its effective lumping scale, Hydrol Process, 26, 855-871, 10.1002/hyp.8300, 2012.

Arnaud, P., Lavabre, J., Fouchier, C., Diss, S., and Javelle, P.: Sensitivity of hydrological models to uncertainty in rainfall input, Hydrological Sciences Journal, 56, 397-410, 10.1080/02626667.2011.563742, 2011.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modeling and assessment - Part 1: Model development, J Am Water Resour As, 34, 73-89, 1998.

Badrzadeh, H., Sarukkalige, R., and Jayawardena, A. W.: Hourly runoff forecasting for flood risk management: Application of various computational intelligence models, Journal of Hydrology, 529, 1633-1643, 10.1016/j.jhydrol.2015.07.057, 2015.

Bergström, S. and Forsman, A.: Development of a conceptual deterministic rainfall-runoff mode, Nord. Hydrol, 4, 240-253, 1973.

Beven, K.: A sensitivity analysis of the Penman-Monteith actual evapotranspiration estimates, Journal of Hydrology, 44, 169-190, 10.1016/0022-1694(79)90130-6, 1979.

Buitink, J., Uijlenhoet, R., and Teuling, A. J.: Evaluating seasonal hydrological extremes in mesoscale (pre-)Alpine basins at coarse 0.5° and fine hyperresolution, Hydrology and Earth System Sciences, 23, 1593-1609, 10.5194/hess-23-1593-2019, 2019.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, Journal of Hydrology, 158, 265-284, 10.1016/0022-1694(94)90057-4, 1994.

Dutta, D., Herath, S., and Musiake, K.: Flood inundation simulation in a river basin using a physically based distributed hydrologic model, Hydrol Process, 14, 497-519, 2000.

Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., and Jiang, J.: Comparison of Long Short Term Memory Networks and the Hydrological Model in Runoff Simulation, Water, 12, 10.3390/w12010175, 2020.

Fang, J., Yang, W., Luan, Y., Du, J., Lin, A., and Zhao, L.: Evaluation of the TRMM 3B42 and GPM IMERG products for extreme precipitation analysis over China, Atmospheric Research, 223, 24-38, 10.1016/j.atmosres.2019.03.001, 2019.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, Reviews of Geophysics, 45, 10.1029/2005rg000183, 2007.

Ficchì, A., Perrin, C., and Andréassian, V.: Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events, Journal of Hydrology, 538, 454-470, 10.1016/j.jhydrol.2016.04.016, 2016.

Franke, R.: Scattered Data Interpolation - Tests of Some Methods, Math Comput, 38, 181-200, 1982.

Grimaldi, S., Schumann, G. J. P., Shokri, A., Walker, J. P., and Pauwels, V. R. N.: Challenges, Opportunities, and Pitfalls for Global Coupled Hydrologic-Hydraulic Modeling of Floods, Water Resources Research, 55, 5277-5300, 10.1029/2018wr024289, 2019.

Grusson, Y., Anctil, F., Sauvage, S., and Sánchez Pérez, J.: Testing the SWAT Model with Gridded Weather Data of Different Spatial Resolutions, Water, 9, 10.3390/w9010054, 2017.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nat Clim Change, 3, 816-821, 2013.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Comput, 9, 1735-1780, 1997.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., and Lou, Z.: Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation, Water, 10, 10.3390/w10111543, 2018.

Huang, Y., Bárdossy, A., and Zhang, K.: Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data, Hydrology and Earth System Sciences, 23, 2647-2663, 10.5194/hess-23-2647-2019, 2019.

Huffman, G. J., Bolvin, D. T., and Nelkin, E. J.: Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation, NASA/GSFC Code, 612, 2019, 2015.

Jiang, L. and Bauer-Gottwein, P.: How do GPM IMERG precipitation estimates perform as hydrological model forcing? Evaluation for 300 catchments across Mainland China, Journal of Hydrology, 572, 486-500, 10.1016/j.jhydrol.2019.03.042, 2019.

Kao, I. F., Zhou, Y., Chang, L.-C., and Chang, F.-J.: Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting, Journal of Hydrology, 583, 10.1016/j.jhydrol.2020.124631, 2020.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Koutroulis, A. G. and Tsanis, I. K.: A method for estimating flash flood peak discharge in a poorly gauged basin: Case study for the 13–14 January 1994 flood, Giofiros basin, Crete, Greece, Journal of Hydrology, 385, 150-164, 10.1016/j.jhydrol.2010.02.012, 2010.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrology and Earth System Sciences, 23, 5089-5110, 10.5194/hess-23-5089-2019, 2019.

Liang, X., Guo, J., and Leung, L. R.: Assessment of the effects of spatial resolutions on daily water flux simulations, Journal of Hydrology, 298, 287-310, 10.1016/j.jhydrol.2003.07.007, 2004.

Liao, W., Yin, Z., Wang, R., and Lei, X.: Rainfall-Runoff Modelling Based on Long Short-Term Memory (LSTM), 38th IAHR World Congress - "Water: Connecting the World", 10.3850/38wc092019-1488, 2019.

Liu, J., Chen, X., Wu, J., Zhang, X., Feng, D., and Xu, C.-Y.: Grid parameterization of a conceptual distributed hydrological model through integration of a sub-grid topographic index: necessity and practicability, Hydrological Sciences Journal, 57, 282-297, 10.1080/02626667.2011.645823, 2012.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., and Loumagne, C.: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, Hydrology and Earth System Sciences, 18, 575-594, 10.5194/hess-18-575-2014, 2014.

Maggioni, V. and Massari, C.: On the performance of satellite precipitation products in riverine flood modeling: A review, Journal of Hydrology, 558, 214-224, 10.1016/j.jhydrol.2018.01.039, 2018.

Mei, Y., Nikolopoulos, E., Anagnostou, E., Zoccatelli, D., and Borga, M.: Error Analysis of Satellite Precipitation-Driven Modeling of Flood Events in Complex Alpine Terrain, Remote Sensing, 8, 10.3390/rs8040293, 2016.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., and Uijlenhoet, R.: Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin, Hydrology and Earth System Sciences, 20, 2207-2226, 10.5194/hess-20-2207-2016, 2016.

Moussa, R. and Chahinian, N.: Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events, Hydrology and Earth System Sciences, 13, 519-535, 2009.

Ni, L., Wang, D., Singh, V. P., Wu, J., Wang, Y., Tao, Y., and Zhang, J.: Streamflow and rainfall forecasting by two long short-term memory-based models, Journal of Hydrology, 583, 10.1016/j.jhydrol.2019.124296, 2020.

Nikolopoulos, E. I., Anagnostou, E. N., and Borga, M.: Using High-Resolution Satellite Rainfall Products to Simulate a Major Flash Flood Event in Northern Italy, Journal of Hydrometeorology, 14, 171-185, 10.1175/jhm-d-12-09.1, 2013.

Noilhan, J., Martin, E., Anquetin, S., Saulnier, G.-M., Habets, F., Ducrocq, V., Vincendon, B., Chancibault, K., and Bouilloud, L.: Coupling the ISBA Land Surface Model and the TOPMODEL Hydrological Model for Mediterranean Flash-Flood Forecasting: Description, Calibration, and Validation, Journal of Hydrometeorology, 11, 315-333, 10.1175/2009jhm1163.1, 2010.

O, S., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., and Petersen, W. A.: Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria, Hydrology and Earth System Sciences, 21, 6559-6572, 10.5194/hess-21-6559-2017, 2017.

Pan, S., Liu, L., Bai, Z., and Xu, Y.-P.: Integration of Remote Sensing Evapotranspiration into Multi-Objective Calibration of Distributed Hydrology–Soil–Vegetation Model (DHSVM) in a Humid Region of China, Water, 10, 10.3390/w10121841, 2018.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z. M., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S.,

Steiner, B., Fang, L., Bai, J. J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, Adv Neur In, 32, 2019.

Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., Ding, Y., Friedlingstein, P., Liu, C., Tan, K., Yu, Y., Zhang, T., and Fang, J.: The impacts of climate change on water resources and agriculture in China, Nature, 467, 43-51, 10.1038/nature09364, 2010.

Rafieeinasab, A., Norouzi, A., Kim, S., Habibi, H., Nazari, B., Seo, D.-J., Lee, H., Cosgrove, B., and Cui, Z.: Toward high-resolution flash flood prediction in large urban areas – Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling, Journal of Hydrology, 531, 370-388, 10.1016/j.jhydrol.2015.08.045, 2015.

Shen, C.: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resources Research, 54, 8558-8593, 10.1029/2018wr022643, 2018.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., and Tsai, W.-P.: HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community, Hydrology and Earth System Sciences, 22, 5639-5656, 10.5194/hess-22-5639-2018, 2018.

Shen, Y., Zhao, P., Pan, Y., and Yu, J.: A high spatiotemporal gauge-satellite merged precipitation analysis over China, Journal of Geophysical Research: Atmospheres, 119, 3063-3075, 10.1002/2013jd020686, 2014.

Shrestha, R. R., Theobald, S., and Nestmann, F.: Simulation of flood flow in a river system using artificial neural networks, Hydrology and Earth System Sciences, 9, 313-321, 2005.

Spellman, P., Webster, V., and Watkins, D.: Bias correcting instantaneous peak flows generated using a continuous, semi-distributed hydrologic model, Journal of Flood Risk Management, 11, 10.1111/jfr3.12342, 2018.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, J Mach Learn Res, 15, 1929-1958, 2014.

Su, J., Lü, H., Crow, W. T., Zhu, Y., and Cui, Y.: The Effect of Spatiotemporal Resolution Degradation on the Accuracy of IMERG Products over the Huai River Basin, Journal of Hydrometeorology, 21, 1073-1088, 10.1175/jhm-d-19-0158.1, 2020.

Sun, C., Shrivastava, A., Singh, S., and Gupta, A.: Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, Ieee I Conf Comp Vis, 843-852, 10.1109/Iccv.2017.97, 2017.

Tang, G., Ma, Y., Long, D., Zhong, L., and Hong, Y.: Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales, Journal of Hydrology, 533, 152-167, 10.1016/j.jhydrol.2015.12.008, 2016.

Tang, G., Zeng, Z., Ma, M., Liu, R., Wen, Y., and Hong, Y.: Can Near-Real-Time Satellite Precipitation Products Capture Rainstorms and Guide Flood Warning for the 2016 Summer in South China?, IEEE Geoscience and Remote Sensing Letters, 14, 1208-1212, 10.1109/lgrs.2017.2702137, 2017.

Wang, Z., Zhong, R., Lai, C., and Chen, J.: Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility, Atmospheric Research, 196, 151-163, 10.1016/j.atmosres.2017.06.020, 2017.

Wigmosta, M. S., Vail, L. W., and Lettenmaier, D. P.: A Distributed Hydrology-Vegetation Model for Complex Terrain, Water Resources Research, 30, 1665-1679, 1994.

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., and Wang, J.: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model, Water Resources Research, 50, 2693-2717, 10.1002/2013wr014710, 2014.

Xie, H., Shen, Z., Chen, L., Lai, X., Qiu, J., Wei, G., Dong, J., Peng, Y., and Chen, X.: Parameter Estimation and Uncertainty Analysis: A Comparison between Continuous and Event-Based Modeling of Streamflow Based on the Hydrological Simulation Program–Fortran (HSPF) Model, Water, 11, 10.3390/w11010171, 2019.

Yang, Y., Du, J., Cheng, L., and Xu, W.: Applicability of TRMM satellite precipitation in driving hydrological model for identifying flood events: a case study in the Xiangjiang River Basin, China, Natural Hazards, 87, 1489-1505, 10.1007/s11069-017-2836-0, 2017.

Yoshimoto, S. and Amarnath, G.: Applications of Satellite-Based Rainfall Estimates in Flood Inundation Modeling—A Case Study in Mundeni Aru River Basin, Sri Lanka, Remote Sensing, 9, 10.3390/rs9100998, 2017.

Yu, D., Xie, P., Dong, X., Hu, X., Liu, J., Li, Y., Peng, T., Ma, H., Wang, K., and Xu, S.: Improvement of the SWAT model for event-based flood simulation on a sub-daily timescale, Hydrology and Earth System Sciences, 22, 5001-5019, 10.5194/hess-22-5001-2018, 2018.

Yu, Z., Lu, Q., Zhu, J., Yang, C., Ju, Q., Yang, T., Chen, X., and Sudicky, E. A.: Spatial and Temporal Scale Effect in Simulating Hydrologic Processes in a Watershed, Journal of Hydrologic Engineering, 19, 99-107, doi:10.1061/(ASCE)HE.1943-5584.0000762, 2014.

Yuan, F., Wang, B., Shi, C., Cui, W., Zhao, C., Liu, Y., Ren, L., Zhang, L., Zhu, Y., Chen, T., Jiang, S., and Yang, X.: Evaluation of hydrological utility of IMERG Final run V05 and TMPA 3B42V7 satellite precipitation products in the Yellow River source region, China, Journal of Hydrology, 567, 696-711, 10.1016/j.jhydrol.2018.06.045, 2018.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., and Zhuang, J.: Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm, Journal of Hydrology, 565, 720-736, 10.1016/j.jhydrol.2018.08.050, 2018.

Zhu, D., Peng, D. Z., and Cluckie, I. D.: Statistical analysis of error propagation from radar rainfall to hydrological models, Hydrology and Earth System Sciences, 17, 1445-1453, 10.5194/hess-17-1445-2013, 2013.

Zhu, Q., Hsu, K.-l., Xu, Y.-P., and Yang, T.: Evaluation of a new satellite-based precipitation data set for climate studies in the Xiang River basin, southern China, International Journal of Climatology, 37, 4561-4575, 10.1002/joc.5105, 2017.

Zhu, Q., Xuan, W. D., Liu, L., and Xu, Y. P.: Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China, Hydrol Process, 30, 3061-3083, 10.1002/hyp.10846, 2016.

Zhu, Q., Zhou, D., Luo, Y., Xu, Y.-P., Wang, G., and Gao, X.: Suitability of high-temporal satellite-based precipitation products in flood simulation over a humid region of China, Hydrological Sciences Journal, 66, 104-117, 10.1080/02626667.2020.1844206, 2020a.

Zhu, S., Luo, X., Yuan, X., and Xu, Z.: An improved long short-term memory network for streamflow forecasting in the upper Yangtze River, Stochastic Environmental Research and Risk Assessment, 10.1007/s00477-020-01766-4, 2020b.

Zubieta, R., Getirana, A., Espinoza, J. C., Lavado-Casimiro, W., and Aragon, L.: Hydrological modeling of the Peruvian–Ecuadorian Amazon Basin using GPM-IMERG satellite-based precipitation dataset, Hydrology and Earth System Sciences, 21, 3543-3555, 10.5194/hess-21-3543-2017, 2017.