



Comprehensive evaluation of hydrological models for climate change impact assessment in the Upper Yangtze River Basin, China

Shanshan Wen, et al. *[full author details at the end of the article]*

Received: 15 January 2020 / Accepted: 29 October 2020 / Published online: 24 November 2020

© Springer Nature B.V. 2020

Abstract

Climate change has substantial impacts on regional hydrology in the major river basins. To figure out such latent hydrological impacts of changing climate, more reliable hydrological simulations are imperative. In this study, we evaluated the impacts of climate change on hydrological regime in the Upper Yangtze River Basin based on four downscaled and bias-corrected Global Climate Model outputs from Coupled Model Intercomparison Project Phase 5 under four Representative Concentration Pathways (RCP2.6, RCP4.5, RCP6.0, and RCP8.5) driving three hydrological models. Two model evaluation approaches were applied: simple and comprehensive. The comprehensive approach was used to evaluate models in the historical period, optimizing objective function at four gauges, and hydrological models were weighted for impact assessment based on their performance. In such a way, projected streamflow time series are obtained under different emission scenarios. Results show that the annual average discharge is projected to increase by 4.1–10.5% under the RCP scenarios at the end of twenty-first century relative to the reference period (1970–1999). Moreover, the high flow is projected to increase and the low flow to decrease indicating a higher probability of flood and drought occurrence in the basin. The severity of floods and droughts may increase. In comparison with the simple one-site model evaluation approach, the comprehensive method reveals that the anticipated extreme flow events would be less severe, and annual mean discharge slightly lower. The projected results imply that application of the comprehensive model evaluation approach could narrow the simulated spreads of projections significantly, and might provide more credible results.

Keywords Hydrological modeling · Discharge · Comprehensive model evaluation · Model weighting · Upper Yangtze River Basin

Shanshan Wen, Buda Su and Tong Jiang contributed equally to this work.

This article is part of a Special Issue on “How evaluation of hydrological models influences results of climate impact assessment”, edited by Valentina Krysanova, Fred Hattermann, and Zbigniew Kundzewicz

Shanshan Wen and Buda Su are first authors

1 Introduction

Climate change is a present reality that alters our physical environment and perturbs the societies around the globe. The earth's temperature equilibrium has been destabilized due to enhanced global warming, resulting in changes of the radiation balance of ground surface and atmosphere circulation. The global warming speeds up the circulation of water vapor, influences the spatiotemporal distribution and intensity of precipitation, directly influences the hydrological features, such as evaporation, runoff, and soil water, which could lead to raise of the number of extreme climate events (Beaulieu et al. 2012; IPCC 2013). Hydrological cycle is recognized as a complex process; the climate change impact on the water cycle is determined as one of the most challenging issues of widespread concern, especially under the background of the continuous fast development of the economy (Tomer and Schilling, 2009).

There are large uncertainties involved in the impact analysis (Bastola et al. 2011, Teng et al. 2012). Several studies revealing discharge trends have been published to highlight the alteration of global streamflow in response to climate change (Ejder et al. 2016; Su et al. 2017). The global and regional climate models and the hydrological models with their structures and parameterizations could obviously influence the uncertainty in climate change impacts on water resources (Bosshard et al. 2013; Vetter et al. 2017). Therefore, enormous research efforts have focused on climate change impact assessment by using combinations of multiple GCMs, emission scenarios, and hydrological models (HMs) (Braud et al. 2010; Vetter et al. 2015; Su et al. 2017). The multi-model ensembles of climate models and sets of hydrological impact models allow to obtain more robust results (Knutti and Sedláček 2013; Kundzewicz et al. 2018).

HMs are usually used as important and convenient tools for impact assessment. One of the important features in using HMs for climate change impact studies is its scale. Typically, regional HMs are calibrated and validated specifically for the location of interest, whereas global HMs are usually applied for impact studies with a default global parameterization without calibration. Several model intercomparison studies (Gosling et al. 2017; Hattermann et al. 2017; Kundzewicz et al. 2018) suggested that the calibration and validation of HMs are highly recommended before applying them for climate change impact assessment, especially at the regional/basin scale.

The simple and commonly used model calibration/validation (or evaluation) approach mainly focuses on a single variable (i.e., streamflow) in a single point (catchment outlet). Though this method is still being widely used in climate change impact assessment nowadays (Hattermann et al. 2017; Huang et al. 2017; Gao et al. 2020), significant efforts have been done for using multi-site multi-variable and multi-objective approaches for evaluation of models in advance of climate change impact assessment (Chiang et al. 2014; Franco and Bonumá 2017; Eghdamirad et al. 2019; Puertes et al. 2019). The results of such studies have shown that a better streamflow performance can be achieved for the whole catchment, compared to the single-variable (streamflow only) and single-site calibration strategy. Furthermore, temporal and spatial variability in the basin could be represented through different parameter values; hence, an effective HM evaluation procedure should include testing the model performance for several relevant variables considering their spatial distribution, rather than in a particular point (i.e., catchment outlet) only (Krysanova et al. 2016). Besides, it is important to test models in a variety of climate conditions (including the conditions of the selected study area) (Kundzewicz et al. 2018; Motavita et al. 2019).

When multi-site calibration is ready, a multi-model combination (MMC) approach could be used to better characterize the uncertainty in simulated results. The simplest and common way to generate a MMC is to use the arithmetic mean or median of the selected models. However, although many previous studies have demonstrated that the ensemble mean/median result often shows a better performance than the individual models (Cloke and Pappenberger 2009; Arsenaault et al. 2015), it is still considered problematic solution because of ignoring the limitations of the model applicability, and assigning the same weights to all models regardless their performance. Thus, another approach employing a weighting scheme was recommended by some researchers in the field of climate change impact research, in which different weighting coefficients are applied to each model to reflect their relative strengths or weaknesses in performance (Shamseldin et al. 2007). Though all the aforementioned points could play important roles in hydrological impact assessment, seldom all of them are taken into consideration.

A comprehensive model evaluation framework has been proposed by Krysanova et al. (2018), with a view to generating the model-based projections of hydrological variables, and it is recommended by the regional water sector group in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP, www.isimip.org). It is assumed that after the comprehensive model evaluation, the uncertainties related to HMs would be notably reduced in climate change impact assessment, and the models will provide more credible results. Therefore, it is imperative to test how this framework will work at both the global and regional scales. In this paper, we selected the Upper Yangtze River Basin (UYR) as the study area to study the influence of the HM evaluation methods on climate change impacts.

Numerous studies have been conducted to evaluate streamflow variations and to simulate hydrological projections in the context of climate change in the UYR (Sun et al. 2013; Chen et al. 2014; Liu and Du 2017; Wang et al. 2019; Gao et al. 2020). Both climate and socioeconomic factors and variabilities play crucial roles in the change of hydrological regime in this basin (Chai et al. 2019). A notable decrease in annual runoff was observed in the Upper Yangtze River since the second half of the twentieth century, global climate change is increasing the risk of disastrous extreme hydrological events, and the booming economy inevitably is increasing water demands. Research on the influence of future climate changes on hydrological processes in the basin would be useful for the development of water resources and for the regional economy. Whereas many previous studies reported that the changes in streamflow may occur in the future, the projected impacts usually incorporate significant uncertainty.

Therefore, the main purpose of this study is to investigate whether and how the impact model evaluation influences the climate change impact assessment results using three semi-distributed HMs driven by climate projections from four global climate models (GCMs), which have been prepared within the ISIMIP2b simulation round. The aforementioned comprehensive model performance evaluation approach will be used first, and further the projected discharge will be compared with simulations based on the simple HM calibration/validation approach to explore whether the comprehensive evaluation will provide different impact results than the simple one, and whether the uncertainty related to hydrological models could be reduced. For this purpose, differences in annual average discharge, high/low flows, and mean seasonal dynamics of river flow will be analyzed. It is believed that the methods and results of this study on climate change impacts for water sector at the regional scale will contribute to further research, and will be helpful in policy making to ensure sustainable water management and climate change adaptation issues.

2 Materials and methods

2.1 Study area

This study was conducted for the Upper Yangtze River Basin which is located upstream of the Cuntan hydrological gauging station (29° 37' N, 106° 36' E) in China. The area extends from longitude 90° 30' E to 108° 10' E and from latitude 25°N to 35° 45' N. It covers a drainage area of about 8.6×10^5 km², and spreads over six provinces, providing more than 80% of the river flow to the Three Gorges Project (Fig. 1). Three other hydrological stations, Zhimenda, Shigu, and Pingshan, are located in the mainstream of the Yangtze River upstream of Cuntan station and their drainage areas are 1.4×10^5 km², 2.1×10^5 km², and 4.5×10^5 km², respectively.

The UYR has a complex terrain, with an altitude level ranging from about 150 to 6500 m above sea level. The mountains and plateaus lie in the western part of the UYR, and account for more than 90% of the total basin area, whereas the rest is plain area located in the eastern region. Forest coverage is around 26% of the entire basin. Furthermore, average annual mean temperature in the UYR is 12.3 °C, and average annual precipitation is about 1018 mm (based on data for 1961–2010 period). The spatial distributions are highly uneven. The mean annual temperature and precipitation range from below 0 °C and 500 mm in the west to over 17 °C and 1200 mm in the east, respectively (Guan et al. 2015; Su et al. 2017). It is considered a sensitive area under climate change conditions. The climate of this region is significantly influenced by the southeast monsoon, southwest monsoon, and the Tibetan Plateau. Notably, the flood season normally occurs in May to September, and the amount of precipitation in this period accounts for about 78% of the total annual precipitation, whereas the peak flood season is from July to August (Su et al. 2017). The flood features (i.e., high peaks, flash floods) in the UYR are controlled by rainfall distribution and topography. The UYR region faces acute water shortages in the dry season, whereas frequent floods occur in the wet season. Such diverse effects of climate create many difficulties in the development and utilization of water resources in the region.

In terms of demographic distribution, the population in this region is about 180 million (2010 census), which is 35% of the population in the entire Yangtze Basin, or 14% of the total

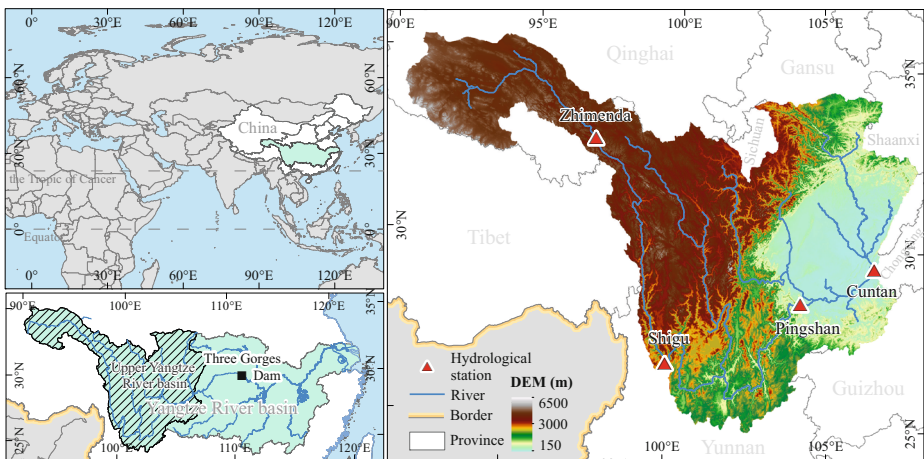


Fig. 1 Location of the Upper Yangtze River Basin and its topography, as well as four hydrological stations used for hydrological model evaluation

population in China (National Bureau of Statistics 2011). This region has significant contribution to augment sustainable development in the Yangtze River Basin, and even in the whole nation.

2.2 Materials

2.2.1 Observation and reanalysis data

The digital elevation model constructed from the Shuttle Radar Topography Missions with 90 m resolution (Jarvis et al. 2008) was used to generate river network, delineate sub-basins, and estimate slope parameters as the basis for HM simulations. In this study, soil features like soil depth, texture, and bulk density were derived from the Harmonized World Soil Database (FAO et al. 2009), and other parameters like field capacity, available water capacity, and saturated conductivity were derived using pedo-transfer functions and tables (hoc AG-Boden 2006). In addition, information on land use was taken from Global Land Cover data for the year of 2000 (Bartholomé and Belward 2005).

To calibrate and validate the three HMs, observed daily river discharge data at the Cuntan, Pingshan, Shigu, and Zhimenda stations in the UYR was obtained from the China Hydrological Yearbook, Yangtze for the 1979–2013 period. Meteorological variables including daily air temperature, air humidity, precipitation, solar radiation, and wind speed on a $0.5^\circ \times 0.5^\circ$ spatial resolution for the period of 1979–2013 were from the global meteorological reanalysis data source EWEMBI (abbreviation for Earth2Observe, WFDEI, and ERA-Interim data Merged, and Bias-corrected for ISIMIP, Lange 2018) downloaded from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). This reanalysis dataset was developed to support the bias correction of climate model outputs in phase 2b of ISIMIP (Frieler et al. 2017). Moreover, to assess the quality of the EWEMBI dataset for the simulation period, gridded observed daily meteorological datasets at 0.25° resolution were derived using quality-controlled gauge observations from more than 2400 meteorological stations over China (Wu and Gao 2013). The National Meteorological Information Center of the China Meteorological Administration is responsible for processing and maintaining these station based variables (Ren et al. 2010).

In this paper, daily evapotranspiration data is also used for cross-checking the model performance by comparing with evapotranspiration simulated by HMs. For that, proxy-observed evapotranspiration data was downloaded from the third version of Global Land Evaporation Amsterdam Model (GLEAM v3.0) (Miralles et al. 2011; Martens et al., 2016). The GLEAM is a set of algorithms dedicated for the estimation of terrestrial evaporation and root-zone soil moisture from satellite data, rather than deriving evapotranspiration directly from satellite imageries. It offers a long-term daily dataset spanning since 1980 with the higher spatial resolution of $0.25^\circ \times 0.25^\circ$ than other evapotranspiration products, widely used in hydrology and climate research, provide (Miralles et al. 2014).

2.2.2 Climate model data

In this study, we used four single-realizations of the GCM outputs from the fifth phase of the Coupled Model Intercomparison Project (CMIP5) for climate change impact assessment: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, and MIROC5 (Taylor et al. 2012), hereinafter abbreviated as GFDL, Had, IPSL, and MIROC, respectively. These four GCMs provide

the full set of output variables for multi-sectoral simulations, and span the space of global mean temperature change and relative precipitation change as best as possible compared to all other CMIP5 GCMs (Warszawski et al. 2014); therefore, they were selected by the ISIMIP group and used in this study. The atmospheric component resolutions and host institution names are presented in Table S1.

The daily climate variables including precipitation (Pre), near-surface daily average/maximum/minimum air temperatures ($T_{ave}/T_{max}/T_{min}$), wind speed, relative humidity, and surface downwelling shortwave radiation from four GCM outputs were used. All raw data were interpolated to 0.5° horizontal resolution by using a first-order conservative remapping scheme (Jones 1999), and bias-corrected by using a trend-preserving statistical bias correction algorithm with EWEMBI reanalysis data at the daily time step (Hempel et al. 2013). Climate simulations including the historical and future periods were selected under four representative concentration pathways (RCPs: RCP2.6, RCP4.5, RCP6.0, and RCP8.5). The simulations cover the period 1970 to 2099. The interval 1986–2005 was selected to evaluate the quality of the climatic data, and the 30-year period of 1970–1999 was used to evaluate the performance of the hydrological models and as a reference period for estimation of impacts. The period 2006–2099 was used for projections under different scenarios.

2.3 Methods

2.3.1 Hydrological models and model performance evaluation

Three semi-distributed HMs were used in this study to generate daily discharge (Table S2). They are the following: (i) the SWAT (Soil and Water Assessment Tool) model, a process based continuous-time hydrological model which describes major hydrological processes at the catchment scale (Arnold et al. 1998); (ii) the VIC (Variable Infiltration Capacity) model, a grid-based hydrological simulator for large-scale applications (Liang et al. 1994); and (iii) a modified version of the conceptual rainfall-runoff model HBV (*Hydrologiska Byråns Vattenavdelning*, Bergström, and Forsman 1973) named HBV-D, which was developed by Krysanova et al. (1999). These three models differ in their levels of complexity, mathematical process formulation, and spatial resolution. All three models have been successfully applied in many different hydro-climates for a broad range of hydrological problems and climate impact studies (Huang et al. 2017; Krysanova and Hattermann 2017; Su et al. 2017). The HBV-D model requires precipitation and temperature as input, and the other two models need all meteorological data listed in Section 2.2. More detailed description of these three HMs can be found in Vetter et al. (2015) and Gao et al. (2020).

The spatial resolution of the VIC model in this study is $0.5^\circ \times 0.5^\circ$, in agreement with the resolution of EWEMBI climate data, and totally 308 grid cells completely cover the UYR. In SWAT model simulation, UYR was subdivided into 36 sub-basins which was further subdivided into over 310 Hydrological Response Units (HRUs) after overlaying the sub-basins with the spatial distribution of land use, soil, etc.

2.3.2 Comprehensive model evaluation approach

The model performance has been tested via calibration and validation of relevant variables for the historical period (1970–1999) by applying a comprehensive model evaluation approach (Krysanova et al. 2018) for the whole UYR. In addition, the simple method based on the one-

site calibration procedure was applied in this study: only for discharge at the outlet station (Cuntan). The five requirements for the comprehensive approach were slightly modified for our study, and are as follows:

- (1) To evaluate the quality of reanalysis data collected from EWEMBI against the station-based observational climate data for the period 1986–2005, as the former is used as input data for HM simulations.
- (2) Check the model performance for the historical period or sub-periods with varying climate conditions which may be climatically similar to the projected future climate. This is to ensure that the models can work realistically under the future climate conditions. In this study, the interval from 1979 to 1996 is a mixed period combining wetter and drier years, and it was selected as the calibration period. To validate the HM performance under different climate conditions (to assure more robust impact simulation results), the validation period was divided into a relatively wet (1997–2004) period and a relatively dry spell (2005–2013) period. See supplementary material for further detailed explanation.
- (3) Calibrate and validate model performance at multiple gauges within the basin and for multiple variables (runoff, evapotranspiration, etc.) to ensure internal consistency of the simulated processes. In this study, data from the outlet station Cuntan and three intermediate gauges (Zhimenda, Shigu, and Pingshan) were used for calibration and validation. In addition, evapotranspiration was also used to validate the model performance.
- (4) Validation for specific hydrological indicators. At this step, we validated the models for the annual high-flow (Q10) and low-flow (Q90) percentiles which were calculated from the daily mean streamflow, as they are important for impact assessment.
- (5) Test the observed trend directions and slopes. To ensure a good model performance, trend comparison is required, and the observed trends (or lack of trends) should be reproduced by the models. If both the observed and simulated streamflows show the same trend directions at the selected confidence level (95 or 99%), or both do not show statistically significant trends, this step can be considered passed satisfactorily in the aspect of trend detection.

If a HM performs well in calibration and validation on the basis of the above mentioned requirements, it can be considered ready for impact assessment. In the case of an ensemble approach, the models could be weighted based on their performance results (Krysanova et al. 2018; Kundzewicz et al. 2018). For that, a simple multiple linear regression (MLR) method could be used for weighting the multi-HMs, to get an MMC result (Uyanık and Güler, 2013). More detailed information about the MLR method and the weighting scheme are given in supplementary material.

In the simple evaluation method, a single-objective function is used to pursue the minimum of the absolute error between the observed and simulated streamflows at one station. Applying this approach, an unweighted multi-model mean is used, disregarding the different performances.

During the course of model evaluation, several goodness-of-fit criteria were used: Nash-Sutcliffe efficiency (NSE), Kling-Gupta efficiency (KGE), coefficient of determination (R^2), and ratio of root mean square error to standard deviation of observation (RSR). These all are commonly used metrics in hydrological and climate impact studies. The optimum value

metrics is 1 for the first three, and 0 for the fourth, indicating perfect correspondence between simulations and observations. For monthly data, the values of $NSE \geq 0.7$, $KGE \geq 0.7$, $R^2 \geq 0.8$, and $RSR \leq 0.6$ are considered ranges of the good model performance suggested by previous studies (Moriassi et al. 2015; Huang et al. 2017).

Besides, SUFI (Sequential Uncertainty Fitting)-2 algorithm has been used to estimate the sensitivity and uncertainty of hydrological models (Abbaspour et al., 2004). It is operated with successive iterations, the calibrated parameter value intervals are reduced at each iteration. In this study, 3–4 iterations with 800 model runs were conducted, and the best performing parameter set was derived from the final iteration for each hydrological model.

2.3.3 Trend analysis

The Sen's slope method was used to estimate the magnitude of trend in a data series, and the Mann–Kendall test was used for identifying the significance of trend at different confidence levels (Mann 1945; Sen 1968; Kendall 1975; Hipel and McLeod 1994). In this study, the trend tests were not only used in analyzing long-term variations in streamflow, precipitation, etc., but were also applied for checking the HM performance in terms of trend detection.

3 Results

3.1 Quality of reanalysis data

Precipitation (Pre) and temperature indices (T_{ave} , T_{max} , T_{min}) are indispensable input data for all HMs applied in this study. The inter-annual variations of these variables derived from meteorological stations were compared with those from the EWEMBI reanalysis dataset for the UYR (Figs. S1–S3, here and later “S” means that table or figure is in supplementary). Though obvious discrepancies between the annual average temperature and precipitation time series exist, very consistent annual fluctuations and trends could be found. On average, both the EWEMBI and station-based temperatures are significantly increasing by 0.2–0.3 °C/10a in the basin. The annual fluctuations in temperature and precipitation of these two data sources were quite reasonable ensuring correlation coefficients above 0.89 (significant at 0.05 level) (Fig. S1). The probability density functions (PDFs) of regional daily T_{ave} , T_{max} , T_{min} , and Pre derived from these two datasets also showed similar patterns, and their spatial correlation coefficients were higher than 0.94 (Fig. S2, S3). In regard to the spatial and temporal distributions, trends and relative variations in the UYR are acceptable using EWEMBI as proxy for observed data. Therefore, it could be used for further hydrological analysis. More details about the comparison between the reanalysis data and the observed climate data are provided in supplementary material.

3.2 Hydrological model evaluation

The results of comprehensive calibration and validation in terms of fit metrics (NSE, R^2 , KGE, RSR) for the monthly data from three models and four hydrological gauges are presented in Table 1. It can be seen that the dynamics are simulated satisfactorily by all three HMs at

Table 1 Criteria of fitting for three hydrological models in the calibration and validation periods in the Upper Yangtze River Basin, based on the multi-gauge comprehensive calibration/validation procedure; good criteria values are displayed in italics

Criteria	Station	Calibration (1979–1996)			Validation (1997–2004)			Validation (2005–2013)		
		HBV-D	VIC	SWAT	HBV-D	VIC	SWAT	HBV-D	VIC	SWAT
NSE	Cuntan	<i>0.88</i>	<i>0.90</i>	<i>0.96</i>	<i>0.88</i>	<i>0.88</i>	<i>0.96</i>	<i>0.90</i>	<i>0.86</i>	<i>0.95</i>
	Pingshan	<i>0.85</i>	<i>0.89</i>	<i>0.95</i>	<i>0.84</i>	<i>0.89</i>	<i>0.96</i>	<i>0.84</i>	<i>0.92</i>	<i>0.96</i>
	Shigu	<i>0.82</i>	<i>0.86</i>	<i>0.73</i>	<i>0.85</i>	<i>0.93</i>	<i>0.84</i>	<i>0.85</i>	<i>0.92</i>	<i>0.85</i>
	Zhimenda	<i>0.75</i>	<i>0.35</i>	<i>0.84</i>	<i>0.60</i>	<i>0.15</i>	<i>0.70</i>	<i>0.62</i>	<i>0.41</i>	<i>0.67</i>
R ²	Cuntan	<i>0.97</i>	<i>0.99</i>	<i>0.98</i>	<i>0.98</i>	<i>0.99</i>	<i>0.98</i>	<i>0.96</i>	<i>0.97</i>	<i>0.98</i>
	Pingshan	<i>0.93</i>	<i>0.97</i>	<i>0.97</i>	<i>0.95</i>	<i>0.97</i>	<i>0.99</i>	<i>0.92</i>	<i>0.96</i>	<i>0.98</i>
	Shigu	<i>0.92</i>	<i>0.93</i>	<i>0.95</i>	<i>0.94</i>	<i>0.97</i>	<i>0.96</i>	<i>0.92</i>	<i>0.96</i>	<i>0.96</i>
	Zhimenda	<i>0.89</i>	<i>0.88</i>	<i>0.93</i>	<i>0.89</i>	<i>0.91</i>	<i>0.94</i>	<i>0.88</i>	<i>0.92</i>	<i>0.95</i>
KGE	Cuntan	<i>0.83</i>	<i>0.70</i>	<i>0.95</i>	<i>0.81</i>	<i>0.69</i>	<i>0.94</i>	<i>0.86</i>	<i>0.69</i>	<i>0.96</i>
	Pingshan	<i>0.88</i>	<i>0.72</i>	<i>0.97</i>	<i>0.81</i>	<i>0.67</i>	<i>0.93</i>	<i>0.89</i>	<i>0.96</i>	<i>0.96</i>
	Shigu	<i>0.85</i>	<i>0.90</i>	<i>0.57</i>	<i>0.85</i>	<i>0.88</i>	<i>0.67</i>	<i>0.91</i>	<i>0.95</i>	<i>0.75</i>
	Zhimenda	<i>0.59</i>	<i>0.35</i>	<i>0.82</i>	<i>0.51</i>	<i>0.30</i>	<i>0.69</i>	<i>0.57</i>	<i>0.39</i>	<i>0.55</i>
RSR	Cuntan	<i>0.35</i>	<i>0.32</i>	<i>0.19</i>	<i>0.35</i>	<i>0.35</i>	<i>0.20</i>	<i>0.31</i>	<i>0.37</i>	<i>0.22</i>
	Pingshan	<i>0.39</i>	<i>0.33</i>	<i>0.23</i>	<i>0.40</i>	<i>0.34</i>	<i>0.20</i>	<i>0.40</i>	<i>0.29</i>	<i>0.20</i>
	Shigu	<i>0.42</i>	<i>0.38</i>	<i>0.52</i>	<i>0.39</i>	<i>0.27</i>	<i>0.40</i>	<i>0.38</i>	<i>0.28</i>	<i>0.39</i>
	Zhimenda	<i>0.50</i>	<i>0.81</i>	<i>0.39</i>	<i>0.63</i>	<i>0.92</i>	<i>0.55</i>	<i>0.61</i>	<i>0.77</i>	<i>0.58</i>

Cuntan, Pingshan, and Shigu gauges, with NSE and KGE ranging from 0.7 to 0.96, R² exceeding 0.9 and RSRs lower than 0.5 in both calibration and validation periods.

In the calibration period, most of the fit metrics of the HBV-D and SWAT models are within the thresholds at Zhimenda gauging stations. In the validation period, HBV-D and VIC perform poorly. SWAT is the only model that performs good or satisfactory in both periods. The Zhimenda-controlled area is the source region of the Yangtze River, with a high altitude, low temperature, presence of frozen soil, and a permanent snow cover. The relatively poor performance at this station could be also explained by an inaccuracy of precipitation data in the high alpine area of the catchment above this station. However, taking into account the fact that the Zhimenda-controlled area only accounts for roughly 16% of drainage area of the whole Cuntan-controlled basin, we can conclude that discharge results simulated by the HMs are still satisfactory for the whole catchment. The differences between three HMs are quite obvious for the Zhimenda. The SWAT-simulated results are better than of the other two, and this will be taken into consideration in model weighting before the multi-model ensemble will be used for impact assessment.

Table 2 shows the model performance for the monthly discharge corresponding to the one-site simple calibration procedure, using only the outlet station (Cuntan) to calibrate/validate discharge for the whole UYR. The results demonstrate that though all HMs perform well at Cuntan, substantial differences in performance exist for the upstream stations compared with Cuntan. Except for Cuntan, NSE, KGE, and R² are lower and RSR is higher at all other three stations, indicating a worse HM performance there.

Therefore, we can conclude that it is possible to reach satisfactory results in the upstream subcatchments and also to improve performance in the downstream part by using the multi-

Table 2 Criteria of fitting for three hydrological models in the calibration and validation periods based on the Cuntan single-station calibration procedure conducted using the simple approach (upper part), and model performances at other three stations (lower part) in the same periods (without calibration); good criteria values are displayed in italics

Performance at Cuntan gauge										
Station	Criteria	Calibration (1979–1996)			Validation (1997–2004)			Validation (2005–2013)		
		HBV-D	VIC	SWAT	HBV-D	VIC	SWAT	HBV-D	VIC	SWAT
Cuntan	NSE	<i>0.93</i>	<i>0.91</i>	<i>0.97</i>	<i>0.94</i>	<i>0.89</i>	<i>0.97</i>	<i>0.90</i>	<i>0.88</i>	<i>0.95</i>
	R2	<i>0.96</i>	<i>0.99</i>	<i>0.99</i>	<i>0.98</i>	<i>0.98</i>	<i>0.98</i>	<i>0.96</i>	<i>0.97</i>	<i>0.98</i>
	KGE	<i>0.95</i>	0.68	<i>0.94</i>	<i>0.90</i>	0.67	<i>0.95</i>	<i>0.92</i>	0.66	<i>0.91</i>
	RSR	<i>0.26</i>	<i>0.3</i>	<i>0.17</i>	<i>0.24</i>	<i>0.34</i>	<i>0.18</i>	<i>0.31</i>	<i>0.35</i>	<i>0.23</i>
Performance at other gauges										
Station	Criteria	1979–1996			1997–2004			2005–2013		
		HBV-D	VIC	SWAT	HBV-D	VIC	SWAT	HBV-D	VIC	SWAT
Pingshan	NSE	<i>0.79</i>	<i>0.86</i>	<i>0.84</i>	<i>0.88</i>	<i>0.87</i>	<i>0.90</i>	0.67	<i>0.81</i>	<i>0.81</i>
	R2	<i>0.92</i>	<i>0.97</i>	<i>0.96</i>	<i>0.94</i>	<i>0.97</i>	<i>0.96</i>	<i>0.91</i>	<i>0.96</i>	<i>0.95</i>
	KGE	<i>0.83</i>	0.65	<i>0.84</i>	0.89	0.63	0.87	0.77	0.6	0.81
	RSR	<i>0.46</i>	<i>0.38</i>	<i>0.40</i>	<i>0.35</i>	<i>0.36</i>	<i>0.31</i>	<i>0.57</i>	<i>0.43</i>	<i>0.44</i>
Shigu	NSE	0.59	0.62	0.68	0.66	0.62	<i>0.76</i>	0.33	0.63	<i>0.81</i>
	R2	<i>0.90</i>	<i>0.94</i>	<i>0.95</i>	<i>0.92</i>	<i>0.96</i>	<i>0.94</i>	<i>0.90</i>	<i>0.96</i>	<i>0.94</i>
	KGE	0.64	0.66	0.48	0.64	0.61	0.55	0.51	0.66	0.66
	RSR	0.64	0.61	<i>0.47</i>	<i>0.58</i>	0.62	<i>0.49</i>	0.82	0.61	<i>0.43</i>
Zhimenda	NSE	-0.60	-1.43	0.57	-2.51	-3.96	0.14	-1.36	-1.5	0.17
	R2	<i>0.89</i>	<i>0.91</i>	<i>0.93</i>	<i>0.89</i>	<i>0.92</i>	<i>0.93</i>	<i>0.89</i>	<i>0.94</i>	<i>0.94</i>
	KGE	-0.22	0.07	0.54	-0.59	-0.20	0.52	-0.31	0.06	0.46
	RSR	1.26	1.56	0.58	1.87	2.23	0.93	1.54	1.58	0.91

gauge comprehensive calibration procedure. This demonstrates the benefit of using all available discharge data for the hydrological analysis. Parameterizations of the models derived from the comprehensive approach make parameters for the whole basin more representative compared to the simple method.

The long-term monthly mean discharges for the calibration and validation periods at four stations after the comprehensive evaluation are shown in Fig. 2a, b. The average annual precipitation in this area is abundant, and due to differences in drainage areas, the range of annual discharge is spanning from 429 m³/s at Zhimenda to 10,670 m³/s at Cuntan. Generally, the long-term average seasonal dynamics of discharge is characterized by a single-peak (i.e., unimodal) curve. Some underestimation in discharge during the flood season (July–September) is evident in the HBV-D and SWAT simulations, and a slight overestimation by VIC. In the dry season, underestimation in discharge at the beginning of the year is obvious for VIC at Cuntan and Pingshan, and for SWAT at Shigu. Performance of all models in the validation period is slightly weaker than in the calibration period.

In addition, the annual simulated values of Q10 and Q90 in the calibration and validation periods were compared to those estimated from the observed time series (Fig. 2c, d). The results show that all HMs have a good performance for high-flow conditions (Q10) in all four gauges. Correlation coefficients between the observed and modeled Q10 range between 0.7 and 0.9 at all stations, and the biases become larger when stations move further upstream along the flow path. The highest bias is over 20% at Zhimenda for the simulation runs with all HMs. The biases at Pingshan and Shigu range from -11.5 to 3.4% and from -12.5 to 11.2%,

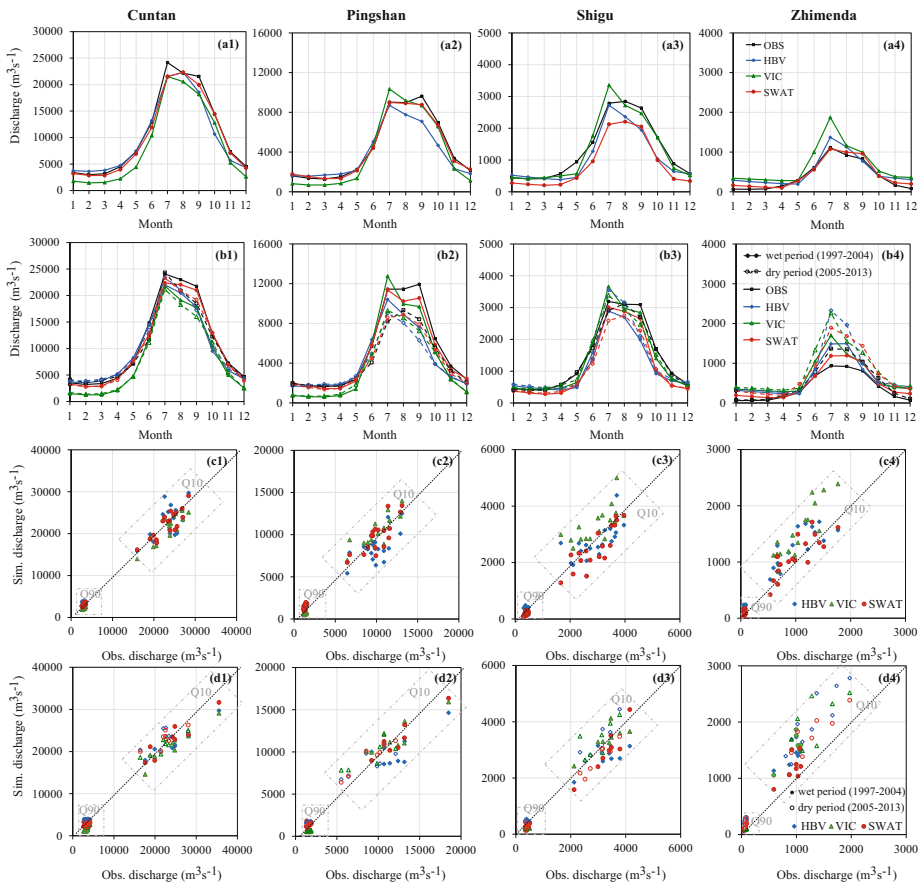


Fig. 2 Calibration (a, c) and validation (b, d) for the average monthly discharge (a, b), annual Q10 (high flow) and Q90 (low flow) simulated by three hydrological models HBV-D, VIC, and SWAT driven by the EWEMBI forcing data at four hydrological stations in the Upper Yangtze River Basin: Cuntan; Pingshan; Shigu; Zhimenda, based on multi-gauge comprehensive calibration procedure

respectively. The bias at Cuntan is lower, between -7.5 and 2.9% . On the contrary, for the low-flow conditions (Q90), a slightly weaker performance can be seen. The biases at most gauges range from -30 to 8% with respect to SWAT and HBV-D, while for VIC, they are generally over 40% . The correlation coefficients for VIC are about $0.2-0.3$, for SWAT $0.4-0.6$, and the values for HBV-D are in between. Based on the comparison results, we can conclude that the VIC performance is a little weaker than that of HBV-D and SWAT, mostly showing an overestimation in discharge compared with observed results. The lowest correlation and the largest bias are found at Zhimenda, and other criteria also show similar findings.

In general, the annual average river discharges simulated by all three models have captured the inter-annual and decadal characteristics of the observations quite well (Fig. S4 in the supplementary material), with the correlation coefficients ranging from 0.8 to 0.86 for HBV-D and SWAT, and a little lower but still satisfactory value of about 0.65 for VIC during the evaluation period.

Furthermore, Sen's slopes were calculated to reveal the increasing or decreasing trends, and Mann-Kendall test was used for testing their significance in order to evaluate the HM performance in terms of trends. Both the observed and simulated mean annual discharges at the Cuntan station declined in the period 1979–2013 with a slope of 0.88–1.29 mm/year (Table S3). There were no statistically significant trends found in the observed and SWAT simulated annual discharges (agreement), whereas HBV-D and VIC showed significant negative trends at the significance level of 0.05. The observed Q10 at the Cuntan station significantly decreased by 5.32 mm/year, and the simulated Q10 decreased ranging from 2.47 mm/year in SWAT to 5.42 in HBV-D, both being statistically significant (agreement). The observed Q90 slightly increased by about 1.01 mm/year, whereas the modeling results showed an increase by 0.2 mm/year or as a negligible decrease. Though the differences in magnitudes exist, the directions of trends in the observed and modeled discharges are generally consistent. As for other stations, no trends could be found in all observed series, and this was reproduced by most of HMs.

The evapotranspiration is another important hydrological variable, which was selected to check the performance of HMs. The comparison results between the satellite-based (GLEAM) and model-simulated evapotranspiration are shown in Table S4. An apparent upward trend was detected in the annual average GLEAM evapotranspiration time series for the 1980–2013 period with a slope of 0.36 mm/year at 0.01 significant level, and the increasing trends were also detected in all simulations, albeit with different rates. The positive trend was overestimated by VIC (0.64 mm/year), and underestimated by SWAT (0.19 mm/year) and HBV-D (0.14 mm/year). The spatial correlation coefficients between the average evapotranspiration values from GLEAM and each model for the 1980–2013 period ranged from 0.84 to 0.90, indicating similar spatial patterns. The same four metrics as for discharge were used for evaluation of monthly evapotranspiration time series. The values of R^2 were over 0.96 for all HMs; the highest NSE and the lowest RSR were 0.79 and 0.46 for the VIC model, respectively, indicating its better performance compared to the other two models.

Summarizing the HM evaluation at multiple sites, we can conclude that the results are acceptable in terms of NSE, KGE, R^2 , and RSR for river discharge, high flow, and evapotranspiration. All three HMs could be used in MMC in order to get more credible discharge results from climate change impact studies. However, the results for Q90 are weaker. For all stations, the average Q10 is nearly 10 times higher than Q90. Therefore, low flows get small weights in the calibration process, leading to bias in simulations.

In general, SWAT outperforms HBV-D and VIC in terms of variations in discharge, whereas VIC outperforms two other models for variations in evapotranspiration. Based on the observed multi-year averaged annual flows at the four hydrological stations during 1979–2013, the proportions of discharge at each station to the outlet discharge were computed. Considering this together with the evaluation results at each station, different HMs were weighted by using the MLR method, and the following weighting coefficients for discharge simulated by SWAT, HBV-D, and VIC were obtained: 0.7, 0.2, and 0.1, respectively. Consequently, a weighted multi-model ensemble discharge can be produced for the reference and future periods based on that.

3.3 Climate impact assessment

The long-term dynamics of discharge in the historical and future periods under four RCP scenarios based on the comprehensive calibration/validation method are shown in Fig. 3a. The annual mean discharge shows slightly decreasing trend with a slope of -0.5 mm/year ($p < 0.05$) in the historical period, and it is projected to significantly increase under all future climate change scenarios with different degrees, ranging from 0.2 to 0.8 mm/year ($p < 0.01$). The average rates of increase are 0.8 and 0.6 mm/year under RCP4.5 and RCP8.5, respectively, which is more than twice as high as that of RCP2.6 and RCP6.0. No significant transition years have been detected under RCP4.5 and RCP8.5 over the twenty-first century, whereas around 2055, the long-term trend is obviously changing under the other two RCPs, 2.6 and 6.0. The varying features of discharge are highly consistent with the detected results for the annual average precipitation simulations (more detailed information on the GCMs simulations of temperature and precipitation are given in Supplementary, Fig. S5). Relative to the reference period of 1970–1999, the mean discharge is projected to increase by 4.1–10.5% under the RCP scenarios (5.7%, 10.5%, 4.1%, and 8.5% under RCP2.6, RCP4.5, RCP6.0, and RCP8.5, respectively) at the end-century (2070–2099) over the UYR.

The river flow at the Cuntan station was also projected based on the simple model evaluation method, aiming at answering the question on how impact model evaluation influences the results of climate change impact assessment, and dynamics of the annual mean discharge are presented in Fig. 3b. Compared to the comprehensive approach based results (Fig. 3a), the general tendency of the projected discharge is the same for both methods. Both show the increasing discharge; however, the increase is stronger based on the simple model evaluation method. Relative to the historical period of 1970–1999, changes in the annual mean discharge are projected to be 12.9%, 15.7%, 6.3%, and 15.2% under RCP2.6, RCP4.5, RCP6.0, and RCP8.5, respectively, based on the simple method, in the 2070–2099 period. Notably, they are roughly doubled compared to projections based on the comprehensive evaluation simulated results. The greatest increases in the annual mean discharge are projected to occur under RCP2.6 and RCP8.5 for the both methods, and the wider range of variations based on the simple method among RCPs at the end of the century (Fig. 3b) is an indication of a larger spread in model simulations in this case.

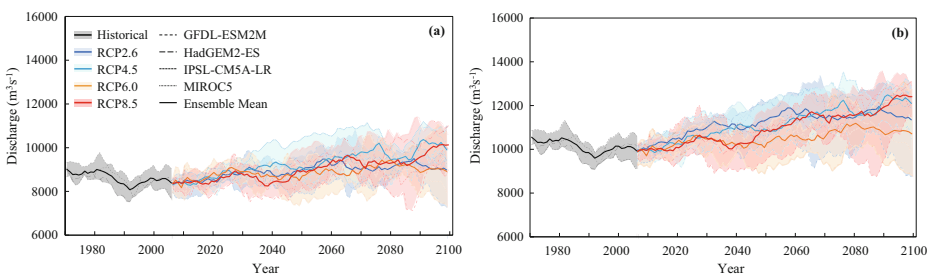


Fig. 3 Annual mean discharge in the historical and future periods under four RCP scenarios (RCP2.6, RCP4.5, RCP6.0, RCP8.5) simulated by three hydrological models (HBV-D, VIC, SWAT) based on both comprehensive (a) and simple (b) evaluation approaches at the Cuntan station for 1970–2099

The projected changes in mean annual discharge, Q10, and Q90 for the middle and end of century relative to the reference period have been calculated based on both the comprehensive and simple evaluation methods (Fig. 4). No matter which method is used, the relative differences in mean average discharge and Q10 clearly show significant increases, which is an indication of facing wetter conditions and more frequent flood events in the UYR. The Q90 is projected to decrease under all RCPs based on the comprehensive method, and it is consistent with the simple method based analysis for RCPs 6.0 and 8.5. After the comprehensive evaluation, the Q10 is projected to increase by 3.9–10.7% under the RCP scenarios in the period of 2070–2099, and the Q90 is projected to decrease by 3.1–11.8% for the same period. The absolute values are projected to be much higher at the end of the twenty-first century compared to the mid-century. Thus, the extreme flows are projected to increase in terms of severity. These results imply that both flood and drought disasters might become more severe with time by the end of the century compared to the past.

The greatest differences in the changes for the annual mean discharge between simulations based on two methods are projected under RCP2.6, as well as the end-century period for RCP8.5 (all around 7%) (Fig. 4c). The differences for Q10 range from 1.9 to 7.4%, reaching maximum under RCP8.5 by the end of this century (Fig. 4f). For Q90, the differences are higher, reaching 16.2% under RCP2.6 in 2070–2099 (Fig. 4i).

In general, if the simple model evaluation method is applied, the magnitudes of mean average and Q10 discharges are projected to increase, whereas notable reduction in Q90 could be expected (Fig. 4). According to the comprehensive model calibration/validation based analysis, all projected changes are expected to be softer: mean discharge would increase by 4–10%, both floods and droughts would become more severe than now, and the related spreads of projections notably smaller.

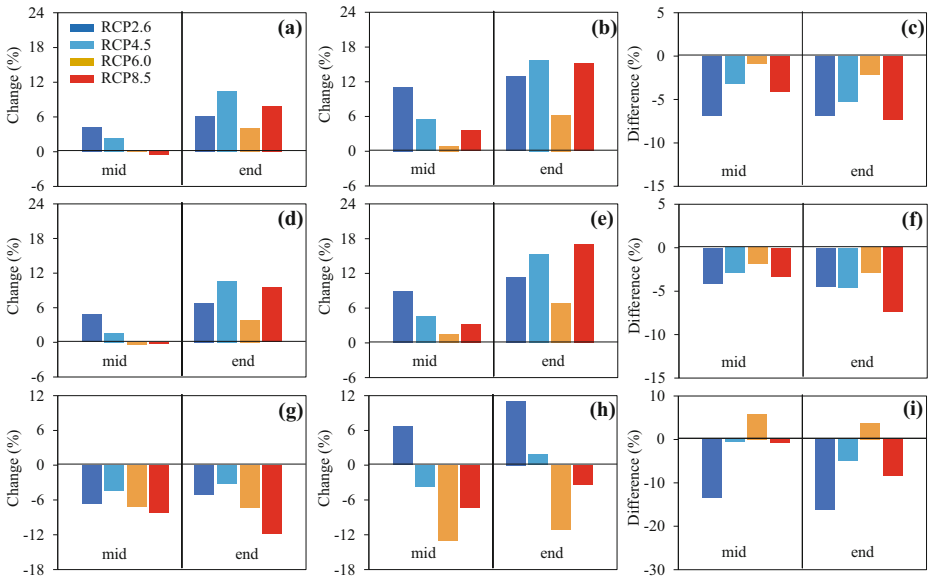


Fig. 4 Projected changes (%) in the mean annual discharge (a, b), Q10 (d, e), and Q90 (g, h) simulated based on the comprehensive (a, d, g) and simple (b, e, h) evaluation approaches, and differences between them (c, f, i) under four RCPs at the mid and end of century (mid: 2030–2059; end: 2070–2099) relative to the reference period (1970–1999) at the Cuntan station

4 Discussion, conclusions, and outlook

This paper focuses on investigating whether and how the impact model evaluation influences results of climate impact assessment via application of three semi-distributed HMs (HBV-D, VIC, and SWAT) evaluated using two methods driven by climate projections from four downscaled and bias-corrected GCMs applied within the ISIMIP2b simulation round for the Upper Yangtze River Basin (UYR). Two HM evaluation methods, the comprehensive evaluation approach (multi-site and multi-variable with contrasting climate periods and trend analysis) proposed by Krysanova et al. (2018) and the traditional or simple HM calibration/validation approach (only for runoff at one site), have been used for simulating discharge, and a comparison of differences in the projected impacts on annual average discharge, high/low flows, and mean seasonal dynamics of river flow between two methods has been done.

The results show that in the UYR, according to the combined evaluation using multiple goodness-of-fit criteria (NSE, KGE, R^2 , RSR) for two variables (runoff and evapotranspiration) at multiple sites located in the mainstream (Zhimenda, Shigu, Pingshan, and Cuntan), all HMs perform generally better based on the comprehensive evaluation method in comparison to the simple method. Parameterizations of the models derived from the comprehensive approach make parameters for the whole basin more representative in comparison to the simple one.

All three models performed well at the three gauges after the comprehensive evaluation, but weaker at the most upstream station Zhimenda. The drainage area of the Zhimenda sub-basin roughly accounts for 16% of the drainage area of the whole UYR basin, and provides only about 4% of the total river flow in the basin on average, according to the observations. The Zhimenda sub-basin is located in the high altitude area, with elevations mostly above 4000 m. It includes approximately 1276 km² of glaciers (Shi et al. 2005) accounting for only 0.9% of its area (0.15% of the whole UYR basin area), and permafrost is widespread in this sub-basin, occupying about 75% of its area (Wang et al. 2017). The density of rain gauges is low in this mountainous area, and wind speed can also strongly influence the observed precipitation, in particular, the solid precipitation (snowfall) at higher altitudes, because of complex terrain (Zhao et al. 2015; Sun et al. 2018). Though glacial meltwater and ice meltwater from thawing permafrost soils likely present important water sources locally, precipitation is the main runoff source for Zhimenda mountainous sub-basin (Liu et al. 2009; Chen et al. 2014; Fang et al., 2017; Qi et al. 2019). Small errors in precipitation may translate into major changes in surface runoff estimates, further influencing the HM performance at the Zhimenda station. The model performance at the Shigu station is also slightly weaker compared to the two stations downstream, and this might be explained partly by similar conditions as at the upstream area of the Zhimenda.

Under the conditions of increasing temperature and precipitation, streamflow in UYR is projected to increase in the twenty-first century, which is consistent with previous studies conducted by Su et al. (2017) and Gao et al. (2020), which performed HMs parameter estimation by optimizing the objective function at a single monitoring station only. In this study, for further understanding how HM evaluation approaches influence the results of climate impact assessment, river flow was projected by using two model evaluation methods. The general tendencies of projected impacts based on two calibration/validation methods are the same. They show an increase in mean and high flows and decrease in low flow, but there are notable differences in the magnitudes of changes. The greatest differences in the projected changes in the annual mean discharge are under RCP2.6 and RCP8.5. The difference in the

projected changes for annual mean discharge is around 7%, and it is about 7.4% for Q10 under RCP8.5, while for Q90, difference is the highest among these three indicators, reaching up to 16.2% under RCP2.6. Therefore, as models after the comprehensive evaluation are more robust, have more representative parameters for the catchment, and climate change impacts based on them differ in magnitude from the results based on the simple method, we can conclude that the results based on the comprehensive evaluation are more trustworthy. Besides, we have found that the projection spreads are smaller at the end-century based on the comprehensive method.

The extreme flows are projected to become more severe with time based on both methods. Mean discharges are projected to be higher under all RCPs, especially for the summer season (Fig. S6, S7). According to the comprehensive method based analysis, both flood and drought disasters are projected to become more severe but softer, and the related spreads smaller, in comparison with projections based on the simple method.

The multi-model ensemble of three models was applied with the model weighting scheme, by weighting each HM based on its performance. SWAT was given a higher weight compared to the other two models in this study, after considering all statistics such as the discrepancy between observed and simulated discharge, the ability of reproducing of observed trends, simulating extreme flows, evapotranspiration, etc. However, it is worth noting that SWAT was not an optimal model in all cases, it outperformed HBV-D and VIC in terms of variations in discharge, whereas VIC outperformed the other two models regarding variations in evaporation. Anyway, we can conclude that using multiple models with weighting coefficients is a more reliable method in comparison with an ensemble mean approach for obtaining more credible climate change impact assessment results.

Despite the fact that the EWEMBI reanalysis data captures the spatial patterns of climatic variables very well (with a spatial correlation coefficient greater than 0.94), underestimation of precipitation and minimum temperature, and overestimation of mean temperature may indicate drier conditions in the UYR compared to the actual state (Fig. S1–S3 in supplementary material). This may influence simulation of hydrological regimes, and the HM performance could be further affected. However, this might also induce a better performance in relatively dry period compared to the wet one, though the values of goodness-of-fit estimators during the calibration period are a little worse than for the validation period, particularly in the dry period (shown in Tables 1 and 2 for some HMs).

According to the observations, the correlation coefficient between annual precipitation and streamflow at the Cuntan station is over 0.8, and the runoff coefficient is over 0.5 in the UYR, meaning that the streamflow is strongly related to precipitation, and this is consistent with many previous studies (e.g., Chen et al. 2014). The increase in temperature and substantial changes in precipitation can alter regional water balances and hydrological regimes. As the UYR is extremely rich in hydropower, more and more hydraulic structures (reservoirs, dams, and dikes) are under construction or at the preplanning stage for the upper reaches of the Yangtze River (Chu et al. 2019). Therefore, human activities could play an increasingly important role for streamflow in the basin in the future, and should be considered in the modeling.

A drawback that was ignored in this study is the assumption of constant land use/land cover. However, since land-use and/or land-cover changes are strongly influenced by human activities, the accurate simulation of future land use patterns is needed in future impact studies;

otherwise, the magnitude and uncertainty of climate change impact results on water resources could be affected.

By using the models evaluated using the comprehensive method, the spreads of model projections were narrowed. Besides, under the guidance of the comprehensive evaluation method (checking for climate contrasting periods, intermediate gauges, additional variables, etc.), the models parameterized after this evaluation became more robust, and more representative for describing hydrological processes in the basin, compared to the models calibrated with the simple method. Generally, the impact results based on the weighted HMs reflect the model performance in the historical period, and the enhanced model calibration and validation have been expected to improve the credibility of projections under climate change (Krysanova et al. 2018; Zaherpour et al. 2018), and it was concluded in our study, but the degree of improvement remains uncertain. This is the question that cannot be answered adequately unless a better and more reliable approach is developed for determining the model weighting strategies. Further efforts are necessary to explore the model weighting approaches and strategies. Besides, as parameters obtained from HM calibration could be affected by several factors such as correlations among parameters, sensitivity of parameters, these may lead to the so-called equifinality (Beven, 1996). However, it makes sense to use more sets of suitable HM parameters for the uncertainty analysis, and this will be implemented in our future work.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10584-020-02929-6>.

Acknowledgments The authors would like to thank the ISIMIP modeling group for providing the climate data and methodological guidance.

Funding This study was jointly supported by the National Key Research and Development Program of China (2017YFA0603701), the National Natural Science Foundation of China (42005126), China Postdoctoral Science Foundation funded project (2020M671536), Jiangsu Planned Projects for Postdoctoral Research Funds (2019K096), and High-level Talent Recruitment Program of the Nanjing University of Information Science and Technology (NUIST).

References

- Abbaspour KC et al (2004) Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone J* 3:1340–1352
- Arnold JG et al (1998) Large area hydrologic modeling and assessment part I: model development. *JAWRA J Am Water Resour Assoc* 34:73–89
- Arsenault R et al (2015) A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *J Hydrol* 529:754–767
- Bartholomé E, Belward AS (2005) GLC2000: a new approach to global land cover mapping from Earth observation data. *Int J Remote Sens* 26:1959–1977
- Beaulieu E et al (2012) High sensitivity of the continental-weathering carbon dioxide sink to future climate change. *Nat Clim Chang* 2:346
- Bastola S et al (2011) The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. *Adv Water Resour* 34:562–576
- Bergström S, Forsman A (1973) Development of a conceptual deterministic rainfall-runoff mode. *Nord Hydrol* 4: 240–253
- Beven K (1996) The limits of splitting: hydrology. *Sci Total Environ* 183:89–97
- Bosshard T et al (2013) Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Resour Res* 49:1523–1536
- Braud I et al (2010) The use of distributed hydrological models for the Gard 2002 flash flood event: analysis of associated hydrological processes. *J Hydrol* 394:162–181

- Chai Y et al (2019) Influence of climate variability and reservoir operation on streamflow in the Yangtze River. *Sci Rep* 9:1–10
- Chen J et al (2014) Variability and trend in the hydrology of the Yangtze River, China: annual precipitation and runoff. *J Hydrol* 513:403–412
- Chiang LC et al (2014) Assessing SWAT's performance in the Kaskaskia River watershed as influenced by the number of calibration stations used. *Hydrol Process* 28:676–687
- Chu PH et al (2019) Prospects of hydropower industry in the Yangtze River Basin: China's green energy choice. *Renew Energy* 131:1168–1185
- Cloke HL, Pappenberger F (2009) Ensemble flood forecasting: a review. *J Hydrol* 375:613–626
- Eghdamirad S et al (2019) The influence of dependence in characterizing multi-variable uncertainty for climate change impact assessments. *Hydrol Sci J* 64:731–738
- Ejder T et al (2016) Effects of climate change on annual streamflow of Kocabaş Stream (Çanakkale, Turkey). *J Scient Res Rep* 11:1–11
- Fang YH et al (2017) Study of the spatiotemporal characteristics of meltwater contribution to the total runoff in the upper Changjiang River basin. *Water* 9:165
- FAO/IIASA/ISRIC/ISS-CAS/JRC (2009) Harmonized world soil database (version 1.1. FAO: IIASA, Rome: Laxenburg
- Franco ACL, Bonumá NB (2017) Multi-variable SWAT model calibration with remotely sensed evapotranspiration and observed flow. *RBRH*, 22
- Frieler K et al (2017) Assessing the impacts of 1.5°C global warming—simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geosci Model Dev* 10:4321–4345
- Gao C et al (2020) A 439-year simulated daily discharge dataset (1861–2299) for the upper Yangtze River, China. *Earth Syst Sci Data* 12:387–402
- Gosling SN et al (2017) A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1°C, 2°C and 3°C. *Clim Chang* 141:577–595
- Guan Y et al (2015) Trends and variability of daily temperature extremes during 1960–2012 in the Yangtze River Basin, China. *Glob Planet Chang* 124:79–94
- Hattermann FF et al (2017) Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. *Clim Chang* 141:561–576
- Hempel S et al (2013) A trend-preserving bias correction—the ISI-MIP approach. *Earth Syst Dynam* 4:219–236
- Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems, vol 45. Elsevier, Amsterdam
- hoc AG-Boden (2006) *Bodenkundliche Kartieranleitung, 5. Verbesserte und erweiterte Auflage*, Bundesanstalt für Geowissenschaften und Rohstoffe, Stuttgart
- Huang SC et al (2017) Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. *Clim Chang* 141:381–397
- IPCC (2013) *Climate change 2013: the physical basis*. In: Contribution of working group to the fifth assessment report of the IPCC. Cambridge University Press, New York
- Jarvis A et al (2008) Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available at: <http://srtm.csi.cgiar.org>.
- Jones PW (1999) First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon Weather Rev* 127:2204–2210
- Kendall MG (1975) *Rank correlation methods*. Griffin, London
- Knutti R, Sedláček J (2013) Robustness and uncertainties in the new CMIP5 climate model projections. *Nat Clim Chang* 3:369–373
- Krysanova V et al (1999) Modelling river discharge for large drainage basins: from lumped to distributed approach. *Hydrol Sci J* 44:313–331
- Krysanova V et al (2018) How the performance of hydrological models relates to credibility of projections under climate change. *Hydrol Sci J* 63:696–672
- Krysanova V, Hattermann FF (2017) Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results. *Clim Chang* 141:363–379
- Krysanova V et al (2016) Assessment of climate change impacts on water resources. In: *Handbook of Applied Hydrology*, p 1440
- Kundzewicz ZW et al (2018) Uncertainty in climate change impacts on water resources. *Environ Sci Pol* 79:1–8
- Lange S (2018) Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. *Earth Syst Dynam* 9:627–645
- Liang X et al (1994) A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J Geophys Res - Atmos* 99:14415–14428
- Liu LL, Du JJ (2017) Documented changes in annual runoff and attribution since the 1950s within selected rivers in China. *Adv Clim Chang Res* 8:37–47

- Liu SY et al (2009) Estimation of glacier runoff and future trends in the Yangtze River source region, China. *J Glaciol* 55:353–362
- Martens B et al (2016) Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture. *Int J Appl Earth Obs* 48:6–162
- Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13:245–259
- Miralles DG et al (2011) Global land-surface evaporation estimated from satellite-based observations. *Hydrol Earth Syst Sci* 15:53–469
- Miralles DG et al (2014) Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat Geosci* 7:45–349
- Moriassi DN et al (2015) Hydrologic and water quality models: performance measures and evaluation criteria. *Trans ASABE* 58:163–1785
- Motavita DF et al (2019) The comprehensive differential split-sample test: a stress-test for hydrological model robustness under climate variability. *J Hydrol* 573:501–515
- National Bureau of Statistics (2011) China statistical yearbook 2011. National Bureau of Statistics, China Statistics Press, Beijing (in Chinese)
- Puertes C et al (2019) Explaining the hydrological behaviour of facultative phreatophytes using a multi-variable and multi-objective modelling approach. *J Hydrol* 575:395–407
- Qi J et al (2019) Coupled snow and frozen ground physics improves cold region hydrological simulations: an evaluation at the Upper Yangtze River Basin (Tibetan Plateau). *J Geophys Res-Atmos* 124:12985–13004
- Ren ZH et al (2010) Quality control procedures for hourly precipitation data from automatic weather stations in China. *Meteorol Month* 36:123–132 (in Chinese)
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 63:1379–1389
- Shamseldin AY et al (2007) A comparative study of three neural network forecast combination methods for simulated river flows of different rainfall—runoff models. *Hydrol Sci J* 52:896–916
- Shi Y et al (2005) A concise China glacier inventory. Shanghai Science Popularization Press, Shanghai (in Chinese)
- Su BD et al (2017) Impacts of climate change on streamflow in the upper Yangtze River basin. *Clim Chang* 141: 533–546
- Sun J et al (2013) Hydrological impacts of climate change in the upper reaches of the Yangtze River Basin. *Quat Int* 304:62–74
- Sun QH et al (2018) A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev Geophys* 56:79–107
- Taylor KE et al (2012) An overview of CMIP5 and the experiment design. *B AM Meteorol Soc* 93:485–498
- Teng J et al (2012) Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff. *J Hydrometeorol* 13:122–139
- Tomer MD, Schilling KE (2009) A simple approach to distinguish land-use and climate-change effects on watershed hydrology. *J Hydrol* 376:24–33
- Uyanik GK, Güler N (2013) A study on multiple linear regression analysis. *Procedia Soc Behav Sci* 106:234–240
- Vetter T et al (2015) Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. *Earth Syst Dynam* 6:17–43
- Vetter T et al (2017) Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Clim Chang* 141:419–433
- Wang W et al (2019) Attribution of decreasing annual and autumn inflows to the Three Gorges Reservoir, Yangtze River: climate variability, water consumption or upstream reservoir operation? *J Hydrodyn* 579: 124180
- Wang X et al (2017) Effects of permafrost degradation on the hydrological regime in the source regions of the Yangtze and Yellow Rivers, China. *Water* 9:897
- Warszawski L et al (2014) The inter-sectoral impact model intercomparison project (ISI-MIP): project framework. *Proc Natl Acad Sci U S A* 111:3228–3232
- Wu J, Gao X (2013) A gridded daily observation dataset over China region and comparison with the other datasets. *Chin J Geophys* 56:1102–1111
- Zaherpour J et al (2018) Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environ Res Lett* 13:065015
- Zhao C et al (2015) Validation of the accuracy of different precipitation datasets over Tianshan mountainous area. *Adv Meteorol* 2015

Affiliations

Shanshan Wen¹ · Buda Su^{1,2,3} · Yanjun Wang¹ · Jianqing Zhai² · Hemin Sun⁴ · Ziyang Chen¹ · Jinlong Huang¹ · Anqian Wang³ · Tong Jiang^{1,3}

✉ Yanjun Wang
yjwang78@163.com

✉ Tong Jiang
jiangtong@cma.gov.cn

Shanshan Wen
wensans@foxmail.com

Buda Su
subd@cma.gov.cn

¹ Institute for Disaster Risk Management / School of Geographical Science, Nanjing University of Information Science & Technology, Nanjing 210044, China

² National Climate Center, China Meteorological Administration, Beijing 100081, China

³ State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

⁴ Beijing Meteorological Bureau, Beijing 100089, China