Science of the Total Environment 765 (2021) 144494

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Intercomparison of ten ISI-MIP models in simulating discharges along the Lancang-Mekong River basin

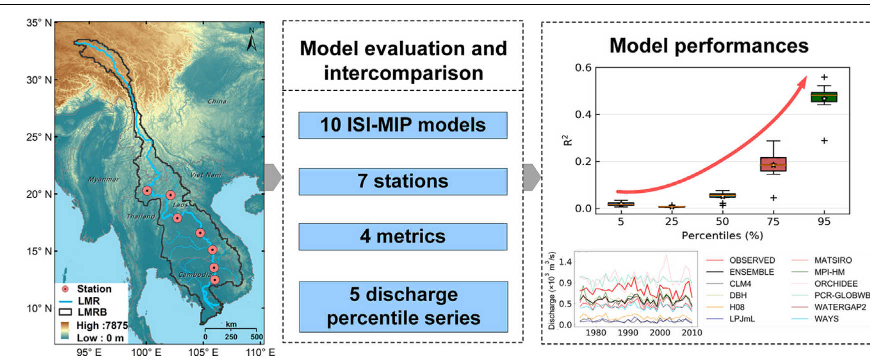
He Chen [a,b], Junguo Liu [b,*], Ganquan Mao [b], Zifeng Wang [b], Zhenzhong Zeng [b], Aifang Chen [b], Kai Wang [b], Deliang Chen [c]

[a] *School of Environment, Harbin Institute of Technology, Harbin 150090, China*
[b] *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*
[c] *Regional Climate Group, Department of Earth Sciences, University of Gothenburg, Gothenburg 40530, Sweden*

HIGHLIGHTS

- Ten models were evaluated in term of discharge simulation at different percentiles.
- Poor performances and larger uncertainties at lower discharge percentiles
- Simulation performances improved as discharge percentiles increased.
- Simulation performances improved from upstream to downstream.
- Poor performances in discharge simulation for the river sections close to estuary

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 20 October 2020
Received in revised form 3 December 2020
Accepted 9 December 2020
Available online 24 December 2020

Editor: Kuishuang Feng

Keywords:
Hydrological models
Land surface models
ISI-MIP
Model evaluation

ABSTRACT

Water resources are of strategic importance for socioeconomic development. Many hydrological models (HMs) and land surface models (LSMs) have been developed for water resources assessment. However, systematic evaluation of discharge simulation from multiple models is still lacking in the Lancang-Mekong River basin. Here, we evaluated the performances of ten HMs and LSMs by evaluating their simulated discharge against observations at the basin scale. The selected models were within the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP2a) framework driven by Global Soil Wetness Project 3 (GSWP3) climate forcing data. Five discharge percentile series were used to evaluate the model performances for low, mean, and high flows. The intercomparison according to four statistical criteria revealed considerable differences exist in model performances for different discharge percentiles, indicating a large uncertainty caused by the choice of models with different degree of physical complexity and sensitivity to the quality of the input data. The models generally performed better for high flow than for low flow. Furthermore, the models generally performed better in downstream than in upstream, with the exception of close to the estuary, where complex processes involving interactions between freshwater and saline water are present. It is not surprising that the two calibrated model (WaterGAP2 and WAYS) are superior over the other models. This systematic intercomparison provides insights into the model behaviours and accuracies in discharges predicting with varying intensities, which can aid in quantifying uncertainties in water resources simulation at the basin scale.

* Corresponding author.
E-mail addresses: liujg@sustech.edu.cn, water21water@yahoo.com, junguo.liu@gmail.com (J. Liu).

https://doi.org/10.1016/j.scitotenv.2020.144494
0048-9697/

1. Introduction

Hydrological models (HMs) and land surface models (LSMs) are valuable tools in water resources simulation and management. HMs describe water storage and fluxes by defining a set of equations (Van Huijgevoort et al., 2013; Ward et al., 2014). LSMs simulate hydrological processes based on land surface energy fluxes and water balances. Parts of LSMs define water fluxes using the same concepts as HMs (Haddeland et al., 2011), while HMs focus more closely than LSMs on the description of the water cycle processes (Gudmundsson et al., 2012a). HMs and LSMs play important roles in assessment of hydrological responses to climate change, e.g., water availability changes and projections (Beran et al., 2016; Gosling and Arnell, 2016), flood hazard and risk assessments (Ward et al., 2014; Gosling et al., 2017), and drought assessments (Van Huijgevoort et al., 2013; Prudhomme et al., 2014). Hydrological metrics simulated by HMs and LSMs can guide policy decisions on water resources management and climate change (Arnell et al., 2016; He et al., 2017); thus, evaluations of the current HMs and LSMs carefully are essential. However, reports on HMs and LSMs performances for the Lancang-Mekong River basin are rare. The Lancang-Mekong River is one of the longest rivers in the world. Model evaluations for this river are critical to identifying the limitations and strengths of the models and improving the models to ensure their wide applications.

Discharge is one of the most intuitive metrics to reflect the conditions of a watershed. Evaluating the observed and simulated discharge series is a commonly used method in hydrological models assessment (Balsamo et al., 2009; Magome et al., 2015; Wang et al., 2016). Previous studies of hydrological model assessments usually utilized different statistical measures with complete data series (e.g., discharge, runoff, and evapotranspiration) and ranked the models (Faiz et al., 2018; Zaherpour et al., 2018). For a more detailed understanding of model strengths and limitations, further model evaluations were carried out from different perspectives. Runoff trend assessment using annual seven-day maxima (high flow) and annual seven-day minima (low flow) with multi-model ensemble of eight models in Europe concluded that runoff displayed significantly different spatial trends (Stahl et al., 2012). The findings also revealed different trends for different runoff flows: high flow demonstrated an increase trend in wet season, while low flow exhibited a decrease trend for dry season. To analyze the effects of climate change on hydrological processes, nine large-scale hydrological models were evaluated using the observed runoff data from 426 nearly natural catchments in Europe. The simulated runoff were evaluated with five runoff percentile series that could adequately characterize the overall flow range, and concluded that the model performance systematically decreased from high to low runoff percentiles (Gudmundsson et al., 2012a). A novel evaluation method was used to evaluate the model performances of six hydrological models in 40 river basins, models performances for different runoff percentile series were also evaluated. The study demonstrated that the models generally overestimated mean annual runoff and extreme runoff (Zaherpour et al., 2018). The methods and datasets used for hydrological model evaluation have become increasingly diverse (e.g., percentile series, high flow and low flow; remote sensing data). The highly varied evaluation methods means that there is still room in model evaluation.

The Lancang-Mekong River is the most critical trans-boundary river in Southeast Asia, and it plays an important role in regional agriculture, hydropower production, and fisheries (MRC, 2019). The discharge of the Lancang-Mekong River is essential for Southeast Asia's socio-economic development where has undergone substantial natural and socio-economic changes in the past few decades. A number of models have been used to simulate the discharge of the Lancang-Mekong River (Kingston et al., 2011; Johnston and Kummu, 2012). However, an accurate and systematic evaluation of discharge simulations for commonly used HMs and LSMs in the Lancang-Mekong River is still lacking. A systematic evaluation of currently used models with a common statistical framework is helpful to determine reliabilities and to quantify uncertainties of models in hydrological simulations. To fill the gap of hydrological model evaluation in this area, we evaluated the performances of ten models using different discharge percentile series against the observed data. The definition of percentile in this study follows a statistical concept, which is the cumulative or non-exceed frequencies in the corresponding data series. The high flows represent the direct response of catchments to hydrological events, whereas low flows were associated with retention and a slow release of water which occurred during dry periods (Smakhtin, 2001; Gudmundsson et al., 2011).

The overall aim of this study is to provide a systematically assessment of the selected HMs and LSMs with regard to their discharge simulations in the Lancang-Mekong River basin with a focus on extreme hydrology events. First, the strengths of the ten selected HMs and LSMs in extreme hydrological events simulation were assessed. We evaluated the model performances with the annual simulated discharge series at different percentiles. Then, the selected models were ranked based on overall discharge series according to the four statistical metrics at the basin scale.

2. Study area and data

2.1. Study area

The Lancang-Mekong River is one of the largest river systems in the world, spanning a length of approximately 4880 km and draining an area of 795,000 km^2 (http://www.lmcchina.org/eng/). The Lancang-Mekong River originates in the Qinghai Province on the Qinghai-Tibetan Plateau. It flows through five other countries in order of Myanmar, Thailand, Lao P.D.R, Cambodia, and Vietnam, and ends in the South China Sea. The upstream of the Lancang-Mekong River basin flows through the Tibetan Plateau and has an average elevation above 4000 m (Yao et al., 2019), and it is considered to be in a near natural state due to the low impacts from human activities. The Lancang-Mekong River basin supports more than 230 million people and contributes substantially to the socioeconomic development of the countries throughout which it flows (http://www.lmcchina.org/eng/).

The annual mean discharge of the Lancang-Mekong River is approximately 446 km^3, making it the world's eighth largest flow (MRC, 2019). The region has a wet season dominated by the southwest monsoon from May to November, contributing approximately 70% of the annual precipitation (Chen et al., 2018). The wet season precipitation provides major source of the Mekong River discharge and it is crucial for regional agriculture (Hoang et al., 2016). In addition, the snow melt from the Qinghai-Tibetan Plateau can affect the downstream discharge especially in dry season (Johnston and Kummu, 2012).

2.2. Data

2.2.1. Observed data

The observed daily river discharge series covering the time period 1975–2010 for seven hydrological stations were used as the verification data. Table 1 displays the longitude and latitude of the stations. The red dots in Fig. 1 display the locations of seven hydrological stations which

Table 1
General information of the seven hydrological stations.

Station	Longitude (°E)	Latitude (°N)	Elevation (m)	Mean annual discharge ($\times 10^3$ m^3/s)
Chiang Saen	100.083	20.273	698	2.594
Luang Prabang	102.134	19.893	690	3.760
Nong Khai	102.732	17.881	172	4.403
Mukdahan	104.733	16.583	166	7.850
Pakse	105.813	15.100	117	9.856
Stung Treng	105.950	13.533	66	12.839
Kratie	106.018	12.481	14	12.055

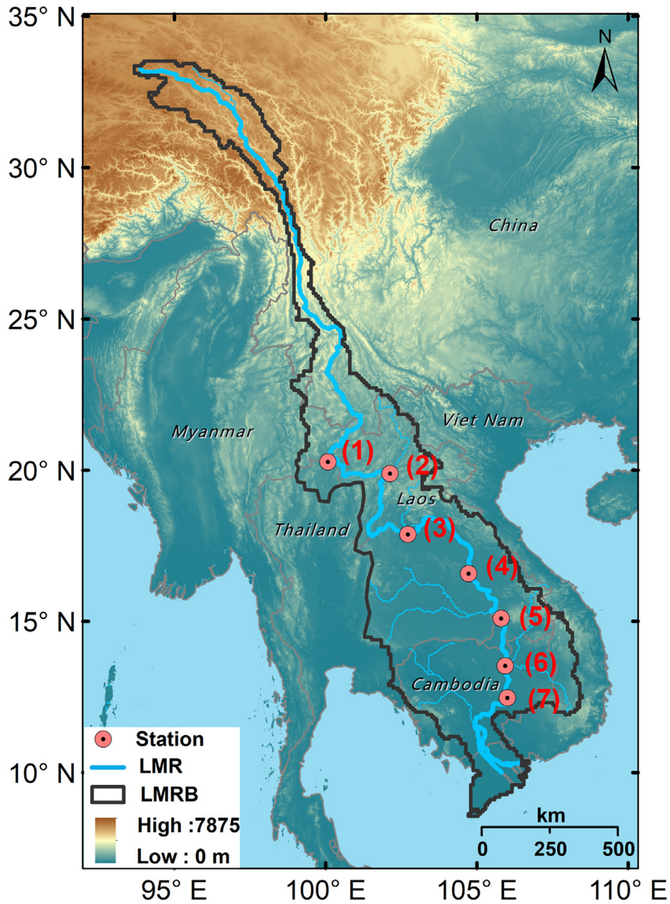


Fig. 1. Location and terrain of the Lancang-Mekong River basin (LMRB) and the Lancang-Mekong River (LMR). Red dots represent the locations of the seven hydrological stations: (1) Chiang Saen, (2) Luang Prabang, (3) Nong Khai, (4) Mukdahan, (5) Pakse, (6) Stung Treng, and (7) Kratie.

were used in this study. Fig. 2 shows the monthly mean discharge of the seven hydrological stations. It can be found that the discharge gradually increased from upstream to downstream.

2.2.2. ISI-MIP data

The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) is a climate-impact modelling initiative dominated by community. ISI-MIP contains different sections and modelling scales under a consistent framework (https://www.ISI-MIP.org/) (Schewe et al., 2014). The aim of ISI-MIP is to improve multi-scale risk management by integrating climate impacts in different sectors (Warszawski et al., 2014). ISI-MIP has several simulation rounds including ISI-MIP Fast Track, ISI-MIP2a, and ISI-MIP2b. ISI-MIP Fast Track is the first simulation round, it aimed at providing cross-sectoral projections of the impacts of different levels of global warming. ISI-MIP2a focus on model validation, with special attention to the representation of the effects of extreme weather and climate variability. ISI-MPI2b focuses on the impacts of different sectors at multiple scale and provides assessments of the impacts of 1.5 °C global warming.

In the present study, ISI-MIP2a was used for model validation and determining the effects of extreme events since it focus on model comparison and validation. The experimental data of the global water sector forced by GSWP3 data in ISI-MIP2a were selected to validate the discharge simulations of different models in this study. Global Soil Wetness Project 3 (GSWP3) provide the climate forcing data for the historical period from 1971 to 2010 (http://hydro.iis.u-tokyo.ac.jp/GSWP3/) (Kim, 2017). The GSWP3 data provides daily climate data at a 0.5° spatial resolution, and the climate metrics used to drive the model include precipitation, minimum temperature, maximum temperature, relative humidity, surface down welling long-wave radiation, surface down welling short-wave radiation and wind speed at 10 m.

Ten HMs and LSMs were selected to assess the discharge simulations for different discharge percentiles in the Lancang-Mekong River basin, including the Community Land Model version 4.0 (CLM4.0) (Leng et al., 2015), the Distributed Biosphere Hydrological model (DBH) (Tang et al., 2007), H08 (Hanasaki et al., 2017), Lund-Potsdam-Jena managed Land (LPJmL) (Sitch et al., 2003), Minimal Advanced Treatments of Surface Interaction and Runoff (MATSIRO) (Takata et al., 2003), Max Planck Institute–Hydrology Model (MPI-HM) (Stacke and Hagemann, 2012), Organizing Carbon and Hydrology in Dynamic EcosystEms (ORCHIDEE) (Guimberteau et al., 2014), PCRaster Global Water Balance (PCR-GLOBWB) (Wada et al., 2014), Water Global Assessment and Prognosis version 2 (WaterGAP2) (Alcamo et al., 2003; Muller Schmied et al., 2016), Water And ecosYstem Simulator (WAYS) (Mao and Liu, 2019). All models were driven by GSWP3 meteorological data, and the resolution of the output discharge data was 0.5° at a daily time step. The ensemble of all models also participated in the assessments. Table 2 displays the brief summaries of the ten HMs and LSMs used in this study.

3. Methods

The simulated daily discharge series for the 1975–2010 period were evaluated for different discharge percentiles against the daily observed data. Five percentile levels were selected to evaluate the model performances of discharge simulations, including the 5th percentile (Q5), 25th percentile (Q25), 50th percentile (Q50), 75th percentile (Q75)

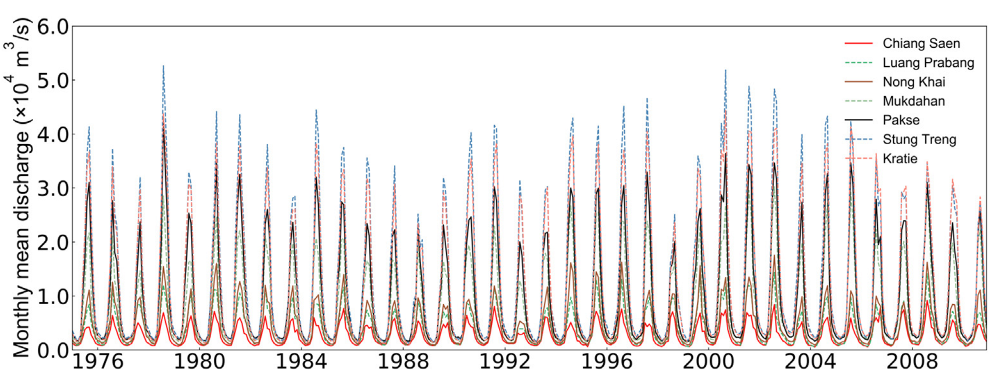


Fig. 2. Monthly mean discharge from 1975 to 2010 for Chiang Saen, Luang Prabang, Nong Khai, Mukdahan, Pakse, Stung Treng, and Kratie.

Table 2
General information of the ten evaluated HMs and LSMs.

Model	Class of model	Spatial resolution (°)	Evapotranspiration	Dam/Reservoir	Surface Runoff	Routing
CLM4	LSM	0.5	the mass transfer equation	No	TOPMODEL-based (Beven and Kirkby, 1979)	Linear reservoir, constant flow velocity
DBH	GHM	0.5	Energy balance	No	SiB2 model based (Sellers et al., 1986)	Linear reservoir model
H08	GHM	0.5	Bulk approach	Yes	An improved bucket model (Hanasaki et al., 2008)	Based on DDM30
LPJmL	GHM	0.5	Priestley–Taylor	Yes	a semi-empirical method (Haxeltine and Prentice, 1996)	Continuity equation derived from linear reservoir model, routing data according to DDM30
MATSIRO	LSM	0.5	Penman-Monteith	Yes	SiB2 model based (Sellers et al., 1986)	TRIP model based on DDM30
MPI-HM	GHM	0.5	Bulk approach	No	HD-Model based (Hagemann and Dümenil, 1998)	Linear reservoir cascade based on DDM30
ORCHIDEE	LSM	0.5	Bulk approach	No	a multi-layer soil hydrology scheme (de Rosnay et al., 2002)	STN-30p river network
WaterGAP2	GHM	0.5	Priestley–Taylor	Yes	HBV model based (Alcamo et al., 2003)	Linear reservoir, flow velocity based on Manning-Strickler based on DDM30
WAYS	GHM	0.5	Penman-Monteith	No	Xinanjiang model based (Zhao, 1992)	CaMa-Flood

and 95th percentile (Q95). In this study, the definition of percentile follows a statistical concept representing the cumulative or non-exceedance frequencies in the data series. Q5 represents extremely low discharge, Q50 represents the median of the discharge series, Q95 represents the high flow (very close to the complete discharge series). Q25 and Q75 are two supplementary discharge series that provide more discharge information. Both the observed and simulated daily discharge series were converted to annual discharge series for the five percentiles.

Different models have specific routing schemes, but the routing module in HMs and LSMs cannot accurately simulate the location of the river network. Therefore, the hydrological stations may deviate from the simulated river network. In additional, the river networks produced by the routing module in each of the hydrological model were different. The corresponding pixel of hydrological stations may not overlap with the simulated river network. To resolve this issue, we first located the initial station pixel according to the longitude and latitude of the station; then the sum of the Nash-Sutcliffe efficiency coefficient (NSE) and the squared Pearson correlation coefficient (R^2) were calculated for the located pixel and its eight surrounding pixels. The pixel with the largest sum was identified as the target pixel where the hydrological station was located.

Four statistical metrics were used to assess the models: R^2, NSE, $\Delta\mu$, and $\Delta\sigma$. R^2 is an evaluation index for the effect of fitting regression. In this study, R^2 represents the ability of models to capture temporal discharge patterns of standard deviation for different percentiles. R^2 can be calculated as follows:

$$R^2 = \left(\frac{\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)^2}} \right)^2 \quad (1)$$

NSE represents the relative magnitude of the residual variance compared with the measured data variance (Nash and Sutcliffe, 1970). NSE ranges from -∞ to 1, and the optimal value for NSE is 1. The simulated result is more reliable when the NSE value is greater than 0. Conversely, when the NSE is less than 0, the simulation is considered unreliable.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)^2} \quad (2)$$

$\Delta\mu$ represents deviation from the mean between simulated and observed discharges; it measures the models' ability to capture the average discharge magnitude. $\Delta\mu$ can be calculated as follows:

$$\Delta\mu = \frac{\frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} y_i}{n}}{\frac{\sum_{i=1}^{n} y_i}{n}} \quad (3)$$

$\Delta\sigma$ represents deviation from the standard deviation, which measures the ability to capture the amplitude of the standard deviation. It can be calculated as follows:

$$\Delta\sigma = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2} - \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)^2}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)^2}} \quad (4)$$

x_i denotes the simulated discharge values on day i, and y_i denotes the observed discharge values on day i.

These four statistical metrics were calculated for each model at all stations. Two steps were conducted to rank the model performances in discharge simulations based on the observed data of seven hydrological stations. First, we determined the optimal value for different metrics: 1 for R^2 and NSE, 0 for $\Delta\mu$ and $\Delta\sigma$. We ranked the ten models based on the distances to the optimal value for four metrics at each station, and four ranking sequences were obtained at each station. The ranking represents the score of the item, which means that the closer the value to the optimal value, the lower score of the model. 28 score series were calculated for four metrics at seven stations. We summed the 28 scores for each model, and the sum represents the performance of the models. The lower the score, the better the model performance. Similarly, we extracted the discharge series at different percentiles for all simulated and observed data and ranked the model performances for the five discharge percentiles utilizing the same method.

4. Results

4.1. Model performance at different discharge percentiles

The distribution of the simulated discharges became more concentrated as the percentile increased. Fig. 3 to Fig. 7 display the annual observed and simulated discharge series for the five discharge percentiles at the seven stations. For the low percentiles (Q5, Q25), large deviations occurred between the simulated and observed discharge series, and their annual discharge curves simulated by different models were

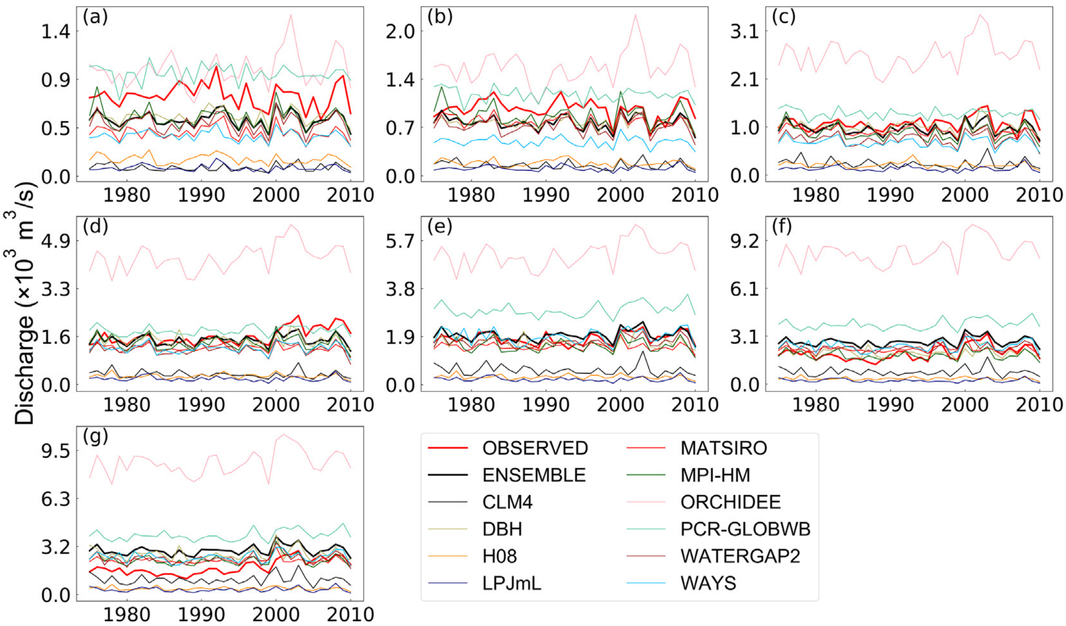


Fig. 3. Comparison of the annual discharge series of observed and simulated discharges at the 5th percentile. (a) Chiang Saen, (b) Luang Prabang, (c) Nong Khai, (d) Mukdahan, (e) Pakse, (f) Stung Treng, and (g) Kratie.

more divergent than that of high flow. The model ensemble discharge and the observed discharge displayed high consistency at most stations for all percentiles. Model ensemble series was closer to the observed series than most of the single model series. The interannual fluctuation in the discharge increased as the discharge percentile increased. A systematic error occurred for the low percentiles (Q5, Q25), where the three models CLM4, H08 and LPJmL always simulated a much smaller discharge than the observed and other models. ORCHIDEE simulated a much higher discharge than the others for Q5 and Q25. This trend became more obvious when the station was closer to the estuary. For the simulated discharge performance for Q5, ORCHIDEE was approximately twice that of the other models including the observed and ensemble models at Chiang Saen, Luang Prabang, and Nong Khai stations. Discharge of ORCHIDEE for Q5 was approximately four times of that at Stung Treng and Kratie stations. The simulated discharge of ORCHIDEE for Q25 was similar to that of Q5. For the high discharge percentile (Q95), ORCHIDEE simulated a smaller discharge than other models at all stations except for Kratie. The same trend occurred with the H08 model, which underestimated at low percentiles and overestimated at high percentiles, although not as obviously as ORCHIDEE displayed. The simulated discharges of WaterGAP2, WAYS, PCR-GLOBWB and MATSIRO were closer to the observed discharge.

Dispersion of the simulated discharge series reflect the uncertainties in discharge simulations among different models. The deviations in discharge between different models for lower percentiles (e.g., Figs. 3 and 4) were much greater than those in higher percentiles (e.g., Figs. 6 and 7). The distribution at the median (Q50) was between the extreme low and extreme high flow (Fig. 5). The large deviations between the

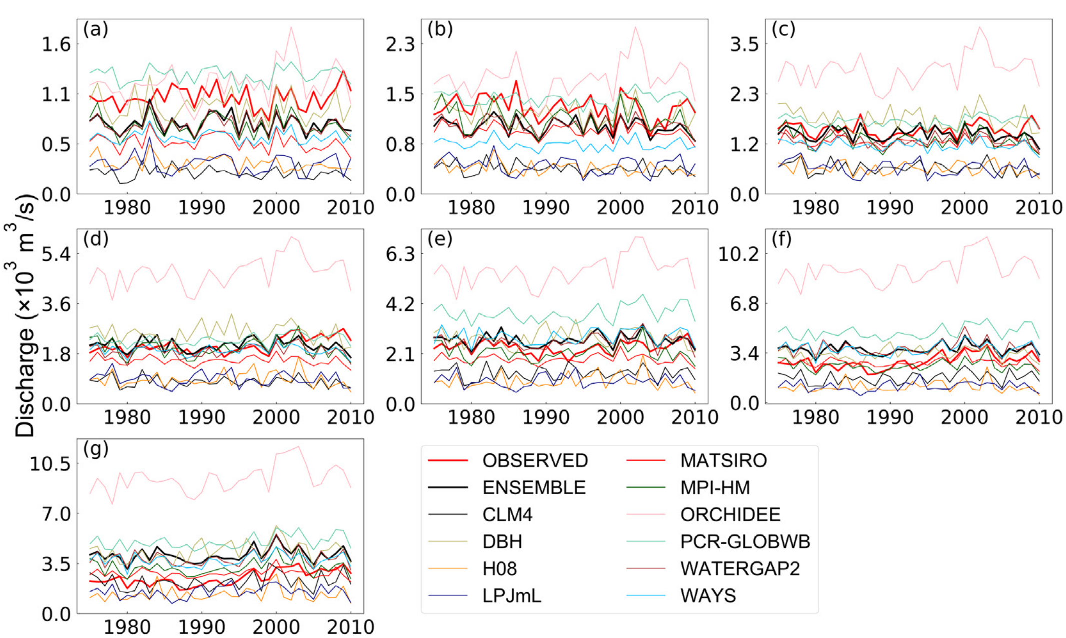


Fig. 4. Comparison of the annual discharge series of observed and simulated discharges at the 25th percentile. (a) Chiang Saen, (b) Luang Prabang, (c) Nong Khai, (d) Mukdahan, (e) Pakse, (f) Stung Treng, and (g) Kratie.

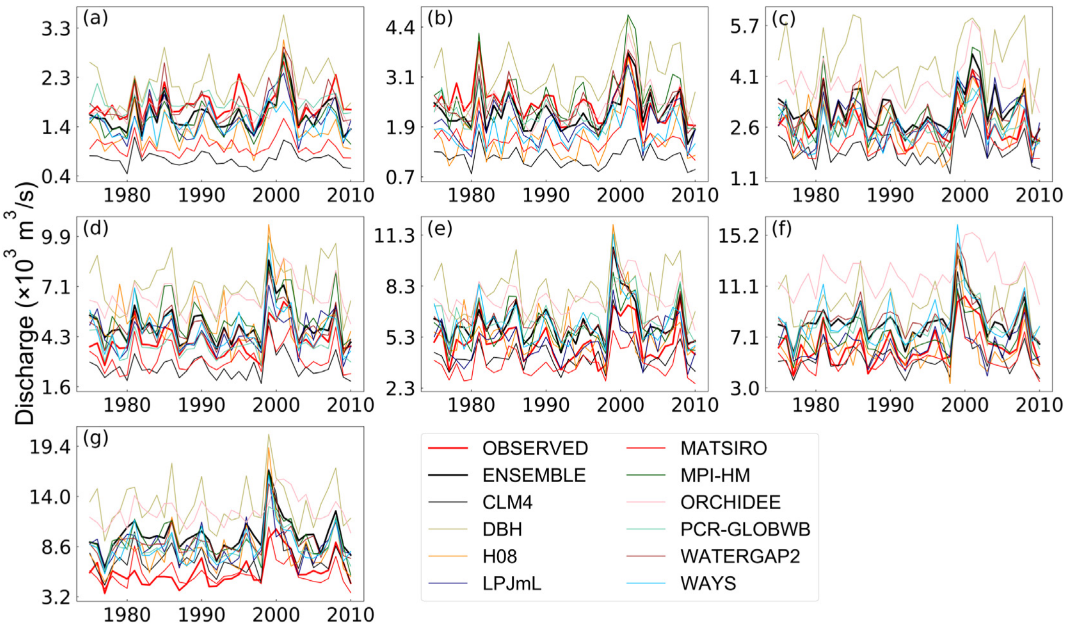


Fig. 5. Comparison of the annual discharge series of observed and simulated discharges at the 50th percentile. (a) Chiang Saen, (b) Luang Prabang, (c) Nong Khai, (d) Mukdahan, (e) Pakse, (f) Stung Treng, and (g) Kratie.

selected models indicated that, uncertainties in discharge simulation for lower percentiles were much greater than that for higher percentiles.

The annual mean series displayed similar patterns at all the stations for all the models. Fig. 8 displays the annual mean discharge series for five discharge percentiles. At all the seven stations, the simulated discharges for all the models increased as the percentiles increased, although the patterns were different especially for ORCHIDEE. In addition, the discharge increased dramatically from upstream to the downstream. The discharge increased approximately three fold from Chiang Saen station to Kratie station, indicating that a substantial amount of local contributions to discharge occurred from downstream of the Lancang-Mekong River basin. The phenomenon of underestimation at low percentiles and overestimation at high percentiles for ORCHIDEE is displayed in Fig. 3 to Fig. 7. Fig. 8 also revealed the phenomenon. The discharges simulated by ORCHIDEE for Q5 and Q25 were larger than all other simulated discharges including the observed and ensemble at all the stations except for Chiang Saen. For Q95, the discharge of ORCHIDEE was smaller than those of the other models at all stations. The discharge series simulated by CLM4 and MATSIRO were smaller than other models at most stations. H08, MPI-HM and DBH overestimated the discharge for high discharge percentiles at most stations.

Four statistical metrics indicated increasingly improvement as the percentile of the simulated discharge series increased. Fig. 9

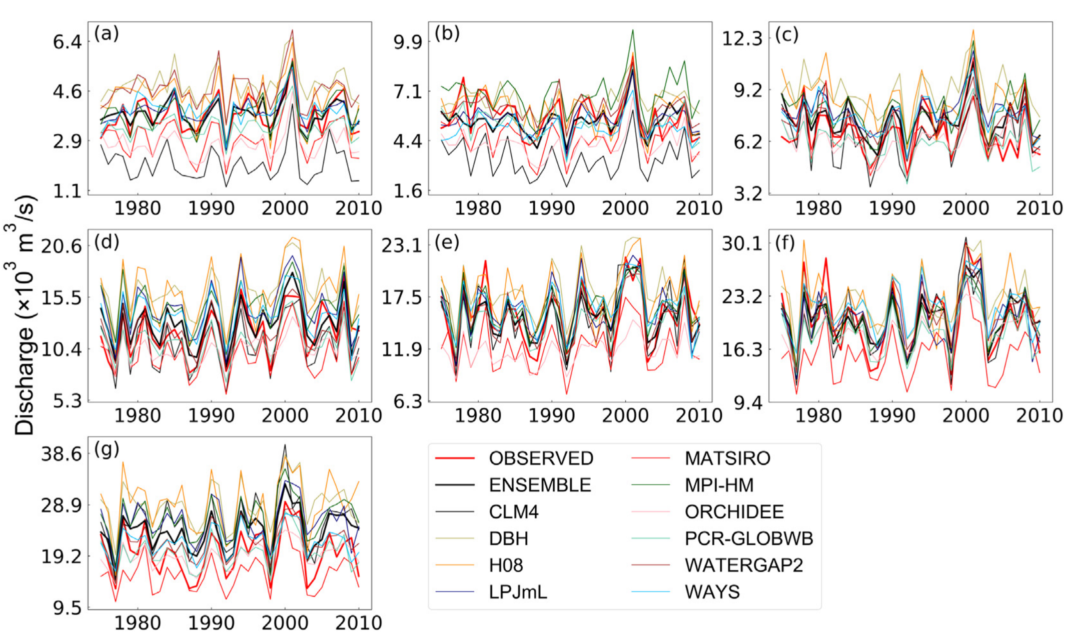


Fig. 6. Comparison of the annual discharge series of observed and simulated discharges at the 75th percentile. (a) Chiang Saen, (b) Luang Prabang, (c) Nong Khai, (d) Mukdahan, (e) Pakse, (f) Stung Treng, and (g) Kratie.

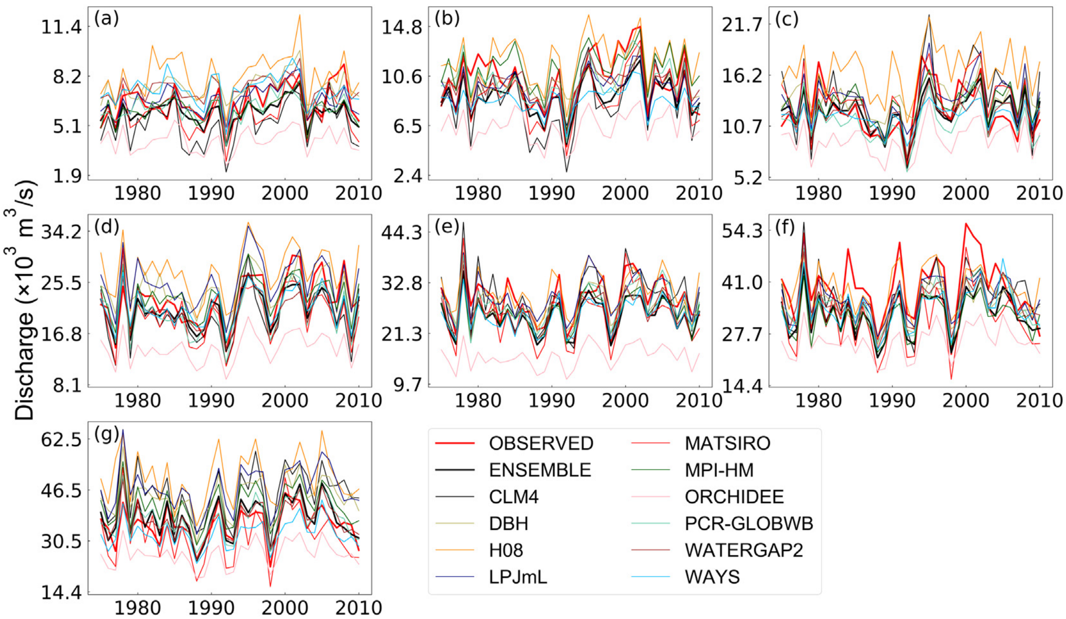


Fig. 7. Comparison of the annual discharge series of observed and simulated discharges at the 95th percentile. (a) Chiang Saen, (b) Luang Prabang, (c) Nong Khai, (d) Mukdahan, (e) Pakse, (f) Stung Treng, and (g) Kratie.

demonstrates station mean performances for different discharge percentiles using the average discharge from all the models. Fig. 9 (a) displays the R^2 for the five discharge percentiles. R^2 for Q5, Q25 and Q50 were less than 0.10, although it increased rapidly for Q75 and Q95. R^2 for Q95 was close to 0.70. The results show that the models performed poorly for discharges at low percentiles, and the fitting effect improved as the discharge percentile increased; the simulated discharges had a poor performance for Q95. NSE indicated an obvious improvement in discharge simulations along with increasing percentile (Fig. 9(b)). NSE had a very low value (less than −60) for the extreme low percentile (Q5), indicating poor predictive power for the evaluated models for low discharge percentile. The simulated discharges displayed underestimation for Q5, Q25, Q50 and Q75; slightly overestimation for Q95. The simulated discharge series had the most centralized distribution for Q95 (Fig. 9(c)). The ensemble simulated discharge unusually underestimated the discharges for most percentiles. The ability of the models to capture the amplitude of standard deviation also improved as the percentile increased, and $\Delta\sigma$ was very close to 0 for Q95 (Fig. 9(d)). In general, the performances of discharge simulations were poor for low percentiles but improved as percentile increased. Fig. 9 also indicated that the uncertainties in discharge for low percentiles were greater than for high percentiles.

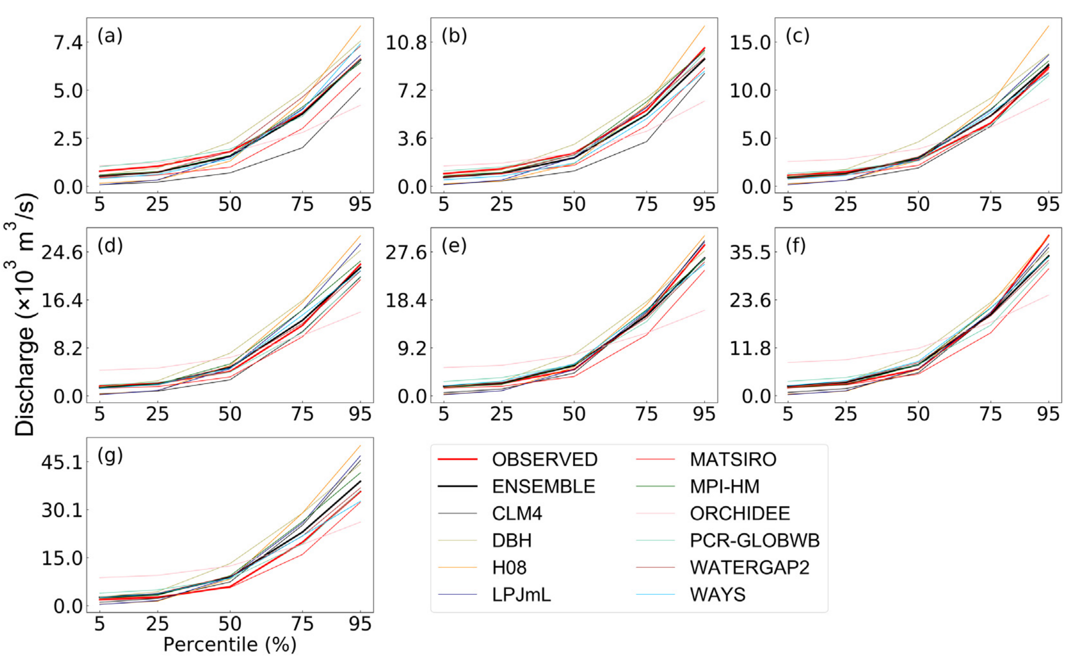


Fig. 8. Annual mean of observed and simulated discharges at different percentiles. (a) Chiang Saen, (b) Luang Prabang, (c) Nong Khai, (d) Mukdahan, (e) Pakse, (f) Stung Treng, and (g) Kratie.

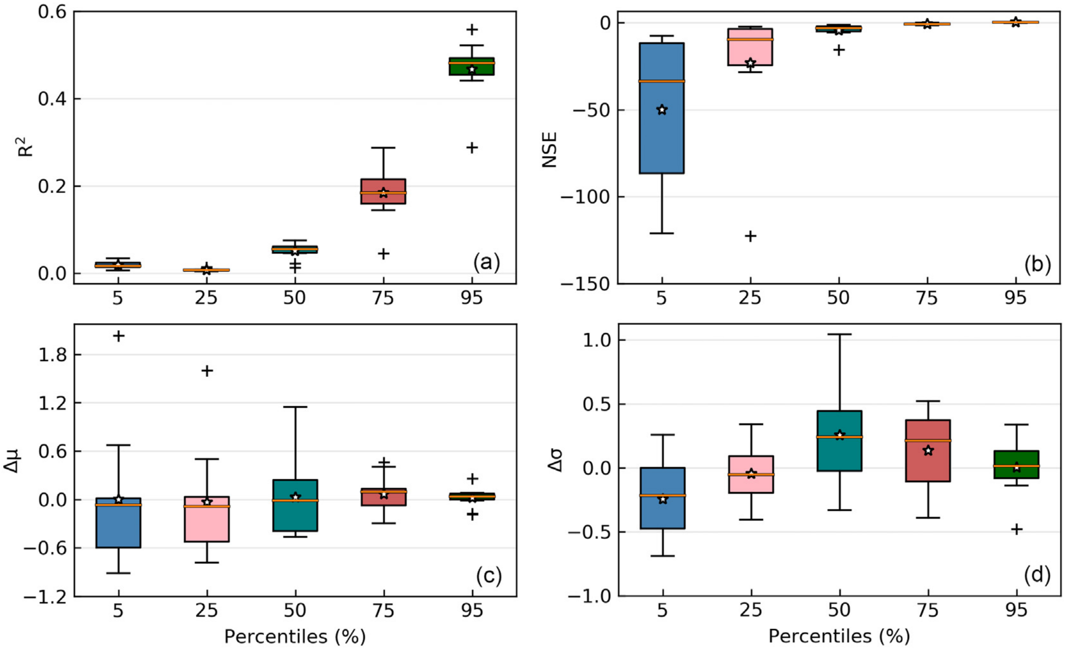


Fig. 9. Comparison of model performances with different metrics at different percentiles. (a) R^2, (b) NSE, (c) Δμ, and (d) Δσ.

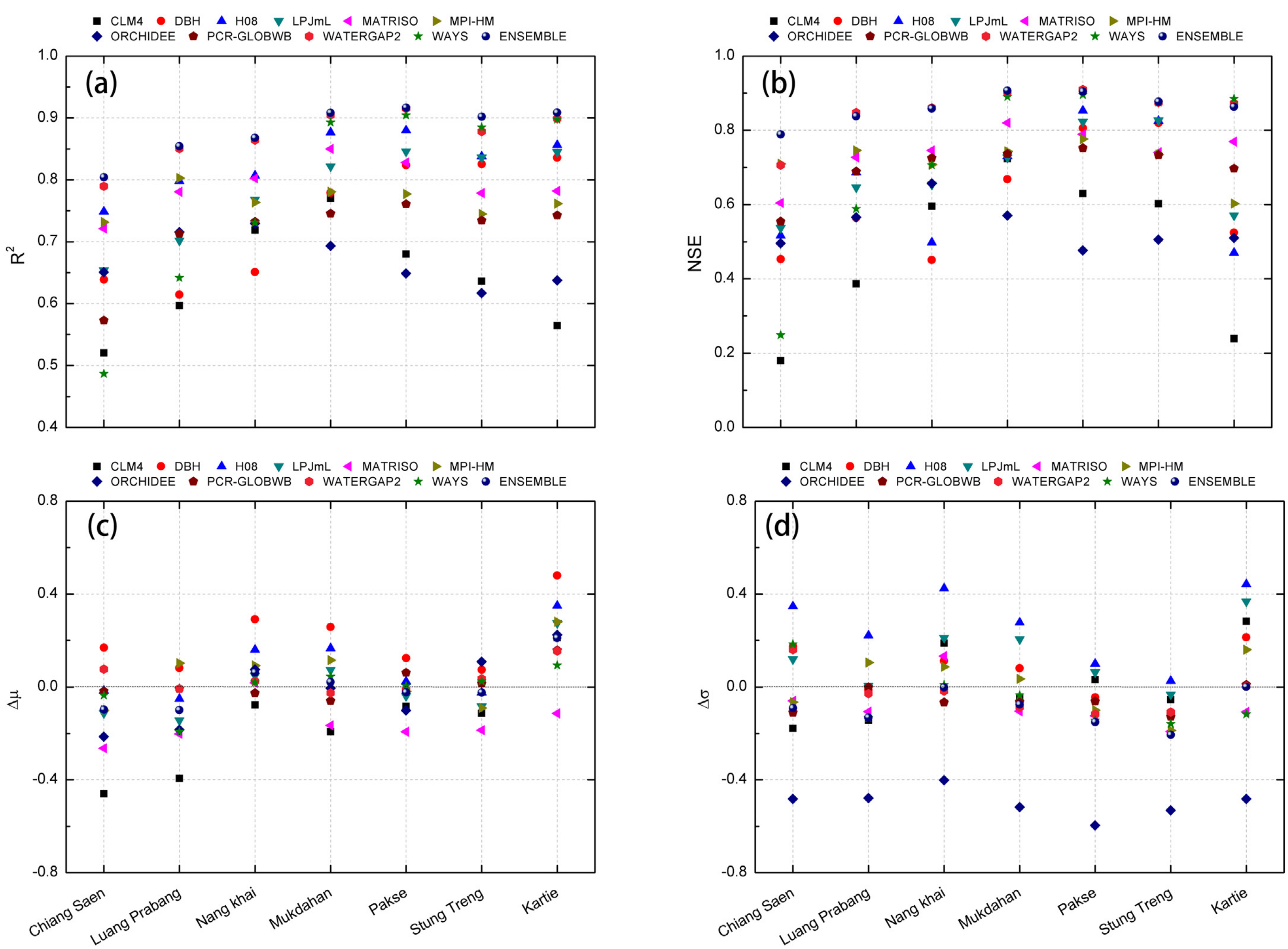


Fig. 10. Comparison of model performance with different metrics at seven hydrological stations. (a) R^2, (b) NSE, (c) Δμ, (d) Δσ. The optimal value is 1 for R^2 and NSE, 0 for Δμ and Δσ.

4.2. Model ranking based on discharge simulation performances

As for the entire discharge series, the models displayed different patterns for different stations. Fig. 10 demonstrates the performances of discharge simulations using four statistical metrics. All the models displayed significant correlations with the observed discharge series while R^2 of most models were greater than 0.60 for all stations. R^2 of several models were larger than 0.80 for stations downstream of the river including WaterGAP2, MPI-HM, H08, MATRISO and WAYS. These results mean that the simulated discharge series of all the models reproduce the observed series satisfactorily. The model ensemble series had the best performance at all the stations in contrast with the single model series. CLM4 and ORCHIDEE had the lowest R^2 for most stations (greater than 0.70 at all stations, and greater than 0.80 for some stations). R^2 at Chiang Saen was relatively low for most models. The overall trend of R^2 increased as the distance from the estuary decreased, although for stations that were closest to the estuary (Stung Treng, Kratie), the distribution of R^2 was not concentrated at the stations in the middle of the basin, such as Mukdahan (Fig. 10(a)).

Fig. 10(b) shows that all NSE values were greater than 0 (the vast majority of NSE values were greater than 0.40), which means that the simulations could be trusted. Similar to R^2, NSE of the model ensemble were greater than signal model at most stations. Model ensemble of NSE were greater than 0.80 at all stations except for Chiang Saen station. WaterGAP was the best model at all stations based on NSE, and even performed better than the model ensemble at Luang Prabang, Pakse and Kratie. NSE values for CLM4 at Chiang Saen, Luang Prabang and Kratie stations were obviously less than those of the other models. WAYS also had a lower NSE value at Chiang Saen. The NSE values of DBH, H08 and ORCHIDEE were lower than the average values.

Fig. 10(c) shows that the models had different performances for different stations as for $\Delta\mu$, which indicates the deviation from mean value between the simulated discharge series and the observed data. $\Delta\mu$ displayed negative deviations at Chiang Saen, Luang Prabang stations and positive deviations at Nang Khai and Kratie stations for most of the models. From the perspective of model performance, CLM4 and MATRISO had much higher negative deviations than the other models for upstream and downstream reaches, respectively. DBH had higher positive deviation than the other models at most stations. $\Delta\sigma$ indicates the deviation from standard deviation between the simulated discharge series and the observed data. The obvious point was that $\Delta\sigma$ of H08 and ORCHIDEE were significantly different from those of other models. $\Delta\sigma$ of H08 were larger than that the others at all stations, whereas ORCHIDEE had an opposite performance. DBH model had a large positive deviation in $\Delta\mu$ but a good performance in $\Delta\sigma$. WaterGAP2 displayed a good performance for both $\Delta\mu$ and $\Delta\sigma$.

Table 3 presents the detailed values of the four indicators for all the models at the seven stations. The performances of the ensemble series were better than that of each of single model series. As the performance of specific model. WaterGAP2 got the highest score based on the scoring system, while WAYS, PCR-GLOWBW, MPI-HM and MATRISO ranked 2, 3, 4, and 5, respectively. ORCHIDEE obtained the lowest ranking mainly due to its poor performance for $\Delta\mu$. CLM4 model did not performed well in term of NSE (0.18 for Chiang Saen and 0.24 for Kratie), and it had negative deviations for $\Delta\mu$ at Chiang Saen (−0.46) and Luang Prabang (−0.39). The simulated discharge series of CLM4 differed considerably from those of the other models. NSE and R^2 for most of the models were closer to 1 as the station location approached the estuary, but there was a drop at Kratie station. The values of $\Delta\mu$ at Chiang Saen, Luang Prabang and Pakse were negative for most models, which means that most models underestimated the magnitudes of the discharge series at these stations. However, as the station locations approached the estuary, models with negative $\Delta\mu$ values were significantly reduced, and only two models displayed negative values for $\Delta\mu$ (MATRISO and ORCHIDEE) at Kratie station. This phenomenon indicates that the model performances improved as the location of stations moved closer to the estuary.

5. Discussion

This study systematically evaluated ten HMs and LSMs using different discharge percentile series for the Lancang-Mekong River basin from 1975 to 2010. Four metrics (R^2, NSE, $\Delta\mu$, and $\Delta\sigma$) were used to assess the model performances. Model ranking based on the entire

Table 3
Model performances at seven hydrological stations based on NSE, R^2, $\Delta\mu$, and $\Delta\sigma$.

Station	Indicator	CLM4	DBH	H08	LPJmL	MATRISO	MPI-HM	ORCHIDEE	PCR-GLOWBW	WaterGAP2	WAYS	ENSEMBLE
Chiang Saen	NSE	0.18	0.45	0.52	0.54	0.60	0.71	0.50	0.55	0.71	0.25	0.79
	R^2	0.52	0.64	0.75	0.65	0.72	0.73	0.65	0.57	0.79	0.49	0.80
	$\Delta\mu$	−0.46	0.17	−0.02	−0.11	−0.26	−0.10	−0.21	−0.02	0.08	−0.04	−0.10
	$\Delta\sigma$	−0.18	0.17	0.35	0.12	−0.06	−0.07	−0.48	−0.11	0.16	0.18	−0.09
Luang Prabang	NSE	0.39	0.56	0.69	0.65	0.73	0.75	0.57	0.69	0.85	0.59	0.84
	R^2	0.60	0.61	0.80	0.70	0.78	0.80	0.72	0.71	0.85	0.64	0.85
	$\Delta\mu$	−0.39	0.08	−0.05	−0.14	−0.20	0.10	−0.18	−0.01	−0.01	−0.19	−0.10
	$\Delta\sigma$	−0.14	−0.01	0.22	0.01	−0.11	0.11	−0.48	0.00	−0.03	−0.13	−0.13
Nang Khai	NSE	0.60	0.45	0.50	0.65	0.75	0.71	0.66	0.73	0.86	0.71	0.86
	R^2	0.72	0.65	0.81	0.77	0.80	0.76	0.73	0.73	0.86	0.73	0.87
	$\Delta\mu$	−0.08	0.29	0.16	0.05	0.02	0.09	0.08	−0.03	0.02	0.02	0.06
	$\Delta\sigma$	0.19	0.11	0.42	0.21	0.13	0.09	−0.40	−0.07	−0.02	0.01	0.00
Mukdahan	NSE	0.72	0.67	0.73	0.73	0.82	0.74	0.57	0.74	0.90	0.89	0.91
	R^2	0.77	0.78	0.88	0.82	0.85	0.78	0.69	0.75	0.90	0.89	0.91
	$\Delta\mu$	−0.19	0.26	0.17	0.07	−0.17	0.12	0.00	−0.06	−0.03	0.04	0.02
	$\Delta\sigma$	−0.04	0.08	0.28	0.20	−0.10	0.04	−0.52	−0.06	−0.09	−0.04	−0.08
Pakse	NSE	0.63	0.81	0.85	0.82	0.79	0.78	0.48	0.75	0.91	0.90	0.90
	R^2	0.68	0.82	0.88	0.85	0.83	0.78	0.65	0.76	0.92	0.90	0.92
	$\Delta\mu$	−0.08	0.12	0.02	−0.04	−0.19	−0.02	−0.10	0.06	−0.01	0.00	−0.02
	$\Delta\sigma$	0.03	−0.04	0.10	0.06	−0.12	−0.10	−0.60	−0.06	−0.12	−0.14	−0.15
Stung Treng	NSE	0.60	0.82	0.82	0.83	0.74	0.73	0.51	0.73	0.87	0.87	0.88
	R^2	0.64	0.83	0.84	0.84	0.78	0.74	0.62	0.73	0.88	0.88	0.90
	$\Delta\mu$	−0.11	0.07	−0.02	−0.08	−0.19	−0.09	0.11	0.02	0.04	0.02	−0.02
	$\Delta\sigma$	−0.06	−0.11	0.03	−0.03	−0.19	−0.19	−0.53	−0.13	−0.11	−0.16	−0.21
Kratie	NSE	0.24	0.52	0.47	0.57	0.77	0.60	0.51	0.70	0.87	0.89	0.86
	R^2	0.56	0.84	0.86	0.84	0.78	0.76	0.64	0.74	0.90	0.90	0.91
	$\Delta\mu$	0.21	0.48	0.35	0.27	−0.11	0.28	0.22	0.16	0.15	0.09	0.21
	$\Delta\sigma$	0.28	0.21	0.44	0.37	−0.11	0.16	−0.48	0.01	0.00	−0.12	0.00
	RANK	9	8	7	6	5	4	10	3	1	2	

discharge series was acquired that could represent the relative abilities of the models in discharge simulations.

An interesting finding of this study was the model performances improved as percentiles increased (Fig. 9). The trends of model performances with discharge percentile were consistent with those of the previous studies (Smakhtin, 2001; Gudmundsson et al., 2011). The models displayed relatively poor performances and huge uncertainties for low percentiles. The dispersed distribution of simulated discharge in lower percentiles can be attributed to several reasons. First, the different strategies in the routing scheme, especially the differences in extreme hydrological events simulations (e.g., drought events) could lead to different simulation results under same driven data. A previous study compared the simulated discharges from nine HMs and CaMa-Flood, the simulated discharges from CaMa-Flood were driven by the HMs' simulated runoff. They concluded that the routing scheme had a considerable influence on discharge simulations (Zhao et al., 2017). Second, the low discharge percentiles were more sensitive to anthropogenic impacts. Human activities could lead to greater impacts on stream flow at low percentiles. For example, irrigation and water withdrawal in dry season for domestic and industrial usage have larger effects on stream flow at local scales (Wada et al., 2014).

It can be demonstrated that the overall performances of the models were generally good (Fig. 10), although most of the models were not calibrated. WaterGAP2 and WAYS were the only two models that had underwent calibration, which explains why these two models were the best among the ten models. The results of this study also displayed that the simulated discharge series of these two models were the best among the ten models (Table 3). This finding indicated that calibration could improve the model and generate better simulations, which was consistent with the findings of the previous studies (Mendoza et al., 2015; Bai et al., 2018). Moreover, the simulation performances of the HMs were generally better than those of the LSMs. Three LSMs used in this study including CLM4, MATRISO, and ORCHIDEE ranked 9, 5, and 10 respectively. It can be attribute to that HMs were more focus on hydrology process simulations, and generally featured more detailed descriptions of hydrology processes than LSMs.

In addition, the model performances of the discharge simulations improved as the stations approach to the estuary, but deteriorated at the station which was closest to the estuary (Kratie). As demonstrated in Table 3, most models performed better at stations located farther downstream than upstream, e.g., the values of R^2 and NSE of WAYS were closer to 1 for stations from upstream to downstream but dropped at Kratie. The model ensemble series displayed a similar trend. This pattern linked to the fact that the discharge in the downstream is greater than that in the upstream. The simulations of the routing process were more accurate in the downstream with greater discharge. Moreover, the effects of forcing data cannot be ignored. An accurate forcing data will substantially improve the simulations. The uncertainty in precipitation data can led to a significant impact on discharge simulations (Biemans et al., 2009). For the Lancang-Mekong River basin, the natural conditions vary greatly from upstream to downstream. The upstream of the Lancang-Mekong River flows through the Hengduan Mountains, and the topography of the upstream is more complex than that of the downstream area. This difference between upstream and downstream could have huge impacts on discharge simulations. The reason for the worse performance at Kratie station (which is the closest to the estuary) may be attributed to the fact that the current HMs and LSMs did not take the effects of interaction between sea water and river into consideration. Such interactions between river and sea, e.g., seawater intrusion and salt interaction can significantly change the hydrological process in the river near the estuary (Pokhrel et al., 2018). The current HMs and LSMs generally did not consider these interactions, leading to unsatisfactory simulations at Kratie.

Several models displayed systematic underestimations at low discharge percentiles for all stations in the Lancang-Mekong River basin (CLM4, LPJmL and H08). This kind of systematic error also appeared in the previous studies (Gudmundsson et al., 2012b). The study compared large-scale hydrological models with observed runoff percentiles in 426 small catchments, significantly underestimated also appeared for the above models (LPJmL and MPI-HM at Q5, MPI-HM at Q25 and Q50, and H08 at Q75). The deviation between the station and selected pixel led to extreme underestimations in the discharge simulations because incorrect simulations of flow networks resulted in incorrect routing. Since we optimized the pixel selection with R^2 and NSE to determine where the stations were located, there was no deviation between the selected pixel and station location. Beside, all the models were forced by the unified GSWP3 climate data. Thus, the systematic underestimation at low percentiles can most likely be attributed to the routing scheme.

Overall, the selected HMs and LSMs performed generally well regarding the entire discharge series for the Lancang-Mekong River basin, while the calibrated WaterGAP2 and WAYS had the best discharge simulation performances. The models had poor performances at low discharge percentiles, although the performances improved as discharge percentiles increased. This pattern reflects the inadequacy of existing models to simulate extremely dry hydrological events. In addition, the model performances in discharge simulations generally improved with the distance towards the estuary. The differences of routing schemes and different natural conditions from upstream to the downstream appear to be responsible for this phenomenon.

6. Conclusion

This study evaluated the performances of the ten ISI-MIP models in discharge simulation in the Lancang-Mekong River using five discharge percentiles. The major findings of this paper are summarized as follows:

(1) The selected models performed poorly in discharge simulation for low discharge percentiles but improved as percentiles increased. The uncertainties for discharge simulations for lower percentiles were much larger than that for higher percentiles.
(2) The model performances generally improved with the distance to the estuary for all discharge percentiles. However, the models have difficulties in simulating discharge for the river sections close to estuary.
(3) The models performed generally well in overall discharge simulations for the Lancang-Mekong River basin, and while the calibrated WaterGAP2 and WAYS had the best performances.

This study inspire us current models are still not satisfactory in extreme hydrological events modelling and seawater intrusion modelling. Hence, further development of models remain necessary for extreme hydrological event forecasts. The interactions between freshwater and saline water near to estuary need to be further considered for more accurate hydrological simulation.

CRediT authorship contribution statement

Conceptualization, Junguo Liu and He Chen; methodology, He Chen and Junguo Liu; validation, He Chen; formal analysis, He Chen; writing - original draft preparation, He Chen; writing - review and editing, He Chen, Junguo Liu, Ganquan Mao, Zifeng Wang, Zhenzhong Zeng, Aifang Chen, Kai Wang, and Deliang Chen; visualization, He Chen; supervision, Junguo Liu; funding acquisition, Junguo Liu. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20060402), the National Key Research and Development Program of China (No. 2017YFA0603704), the Pengcheng Scholar Program of Shenzhen, the National High-level Talents Special Support Plan ("Ten Thousand Talents Plan"), the Leading Innovative Talent Program for young and middle-aged scholars by the Ministry of Science and Technology of China, and the High-level Special Funding of the Southern University of Science and Technology (Grant Nos. G02296302, G02296402). The simulation discharge datasets were downloaded from ISI-MIP2a output data sets (https://www.ISI-MIP.org/outputdata/). We also thank the Mekong River Commission (MRC) to provide the observed discharge data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.144494.

References

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., Siebert, S., 2003. Development and testing of the WaterGAP 2 global model of water use and availability. Hydrol. Sci. J. 48 (3), 317–337. https://doi.org/10.1623/hysj.48.3.317.45290.

Arnell, N.W., Brown, S., Gosling, S.N., Hinkel, J., Huntingford, C., Lloyd-Hughes, B., Lowe, J.A., Osborn, T., Nicholls, R.J., Zelazowski, P., 2016. Global-scale climate impact functions: the relationship between climate forcing and impact. Clim. Chang. 134 (3), 475–487. https://doi.org/10.1007/s10584-013-1034-7.

Bai, P., Liu, X., Liu, C., 2018. Improving hydrological simulations by incorporating GRACE data for model calibration. J. Hydrol. 557, 291–304. https://doi.org/10.1016/j.jhydrol.2017.12.025.

Balsamo, G., Viterbo, P., Beijaars, A., van den Hurk, B., Hirschi, M., Betts, A.K., Scipal, K., 2009. A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the integrated forecast system. J. Hydrometeorol. 10 (3), 623–643. https://doi.org/10.1175/2008JHM1068.1.

Beran, A., Hanel, M., Nesládková, M., Vizina, A., 2016. Increasing water resources availability under climate change. Procedia Engineering 162, 448–454. https://doi.org/10.1016/j.proeng.2016.11.087.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24 (1), 43–69. https://doi.org/10.1080/02626667909491834.

Biemans, H., Hutjes, R.W.A., Kabat, P., Strengers, B.J., Gerten, D., Rost, S., 2009. Effects of precipitation uncertainty on discharge calculations for Main River basins. J. Hydrometeorol. 10 (4), 1011–1025. https://doi.org/10.1175/2008JHM1067.1.

Chen, A., Chen, D., Azorin-Molina, C., 2018. Assessing reliability of precipitation data over the Mekong River basin: a comparison of ground-based, satellite, and reanalysis datasets. Int. J. Climatol. 38 (11), 4314–4334. https://doi.org/10.1002/joc.5670.

Faiz, M.A., Liu, D., Fu, Q., Li, M., Baig, F., Tahir, A.A., Khan, M.I., Li, T., Cui, S., 2018. Performance evaluation of hydrological models using ensemble of general circulation models in the northeastern China. J. Hydrol. 565, 599–613. https://doi.org/10.1016/j.jhydrol.2018.08.057.

Gosling, S.N., Arnell, N.W., 2016. A global assessment of the impact of climate change on water scarcity. Clim. Chang. 134 (3), 371–385. https://doi.org/10.1007/s10584-013-0853-x.

Gosling, S.N., Zaherpour, J., Mount, N.J., Hattermann, F.F., Dankers, R., Arheimer, B., Breuer, L., Ding, J., Haddeland, I., Kumar, R., Kundu, D., Liu, J., van Griensven, A., Veldkamp, T.I.E., Vetter, T., Wang, X., Zhang, X., 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 (3), 577–595. https://doi.org/10.1007/s10584-016-1773-3.

Gudmundsson, L., Tallaksen, L.M., Stahl, K., 2011. Spatial cross-correlation patterns of European low, mean and high flows. Hydrol. Process. 25 (7), 1034–1045. https://doi.org/10.1002/hyp.7807.

Gudmundsson, L., Wagener, T., Tallaksen, L.M., Engeland, K., 2012a. Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. Water Resour. Res. 48 (11), 1–20. https://doi.org/10.1029/2011WR010911.

Gudmundsson, L., Tallaksen, L.M., Stahl, K., Clark, D.B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., Koirala, A.S., 2012b. Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. J. Hydrometeorol. 13 (2), 604–620. https://doi.org/10.1175/JHM-D-11-083.1.

Guimberteau, M., Ducharne, A., Ciais, P., Boisier, J.P., Peng, S., De Weirdt, M., Verbeeck, H., 2014. Testing conceptual and physically based soil hydrology schemes against observations for the Amazon Basin. Geosci. Model Dev. 7 (3), 1115–1136. https://doi.org/10.5194/gmd-7-1115-2014.

Haddeland, I., Clark, D.B., Franssen, W., Ludwig, F., Voß, F., Arnell, N.W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S.N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G.P., Yeh, P., 2011. Multimodel estimate of the global terrestrial water balance: setup and first results. J. Hydrometeorol. 12 (5), 869–884. https://doi.org/10.1175/2011JHM1324.1.

Hagemann, S., Dümenil, L., 1998. A parametrization of the lateral waterflow for the global scale. Clim. Dyn. 14 (1), 17–31. https://doi.org/10.1007/s003820050205.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., Tanaka, K., 2008. An integrated model for the assessment of global water resources - part 1: model description and input meteorological forcing. Hydrol. Earth Syst. Sci. 12 (4), 1007–1025. https://doi.org/10.5194/hess-12-1007-2008.

Hanasaki, N., Yoshikawa, S., Pokhrel, Y., Kanae, S., 2017. A global hydrological simulation to specify the sources of water used by humans. Hydrol. Earth Syst. Sci. Discuss. 08, 1–53. https://doi.org/10.5194/hess-2017-280.

Haxeltine, A., Prentice, I.C., 1996. BIOME3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. Global Biogeochemical Cycle 10 (4), 693–709. https://doi.org/10.1029/96GB02344.

He, X., Wada, Y., Wanders, N., Sheffield, J., 2017. Intensification of hydrological drought in California by human water management. Geophys. Res. Lett. 44 (4), 1777–1785. https://doi.org/10.1002/2016GL071665.

Hoang, L.P., Lauri, H., Kummu, M., Koponen, J., Vliet, M.T.H.V., Supit, I., Leemans, R., Kabat, P., Ludwig, F., 2016. Mekong River flow and hydrological extremes under climate change. Hydrol. Earth Syst. Sci. 20 (7), 3027–3041. https://doi.org/10.5194/hess-20-3027-2016.

Johnston, R., Kummu, M., 2012. Water resource models in the mekong basin: a review. Water Resources Management 26 (2), 429–455. https://doi.org/10.1007/s11269-011-9925-8.

Kim, H., 2017. Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1) [Data set], Data Integration and Analysis System (DIAS). https://doi.org/10.20783/DIAS.501 2017.

Kingston, D.G., Thompson, J.R., Kite, G., 2011. Uncertainty in climate change projections of discharge for the Mekong River Basin. Hydrology and Earth System Sciences 15, 1459–1471. https://doi.org/10.5194/hess-15-1459-2011.

Leng, G., Huang, M., Tang, Q., Leung, L.R., 2015. A modeling of irrgation effects on global surface water and groundwater resources under a changing climate. Journal of Advances in Modeling Earth Systems 7 (3), 1285–1304. https://doi.org/10.1002/2015MS000437.

Magome, J., Gusyev, M.A., Hasegawa, A., Takeuchi, K., 2015. River discharge simulation of a distributed hydrological model on global scale for the hazard quantification. MODSIM 2015, 21st International Congress on Modelling and Simulation, December, 1593–1599 https://doi.org/10.36334/modsim.2015.g5.magome.

Mao, G., Liu, J., 2019. WAYS v1: a hydrological model for root zone water storage simulation on a global scale. Geoscientific Model Development 12 (12), 5267–5289. https://doi.org/10.5194/gmd-12-5267-2019.

Mekong River Commission, 2019. State of the Basin Report 2018.

Mendoza, P.A., Clark, M.P., Mizukami, N., Newman, A.J., Barlage, M., Gutmann, E.D., Rasmussen, R.M., Rajagopalan, B., Brekke, L.D., Arnold, J.R., 2015. Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. Journal of Hydrometeorology 16 (2), 762–780. https://doi.org/10.1175/JHM-D-14-0104.1.

Muller Schmied, H., Adam, L., Eisner, S., Fink, G., Florke, M., Kim, H., Oki, T., Theodor Portmann, F., Reinecke, R., Riedel, C., Song, Q., Zhang, J., Doll, P., 2016. Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. Hydrology and Earth System Sciences 20 (7), 2877–2898. https://doi.org/10.5194/hess-20-2877-2016.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - a discussion of principles. Journal of Hydrology 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Pokhrel, Y., Shin, S., Lin, Z., Yamazaki, D., Qi, J., 2018. Potential disruption of flood dynamics in the lower mekong river basin due to upstream flow regulation. Scientific Reports 8 (1), 1–13. https://doi.org/10.1038/s41598-018-35823-4.

Prudhomme, C., Giuntoli, I., Robinson, E.L., Clark, D.B., Arnell, N.W., Dankers, R., Fekete, B.M., Franssen, W., Gerten, D., Gosling, S.N., Hagemann, S., Hannah, D.M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., Wisser, D., 2014. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proceedings of the National Academy of Sciences of the United States of America 111 (9), 3262–3267. https://doi.org/10.1073/pnas.1222473110.

de Rosnay, P., Polcher, J., Bruen, M., Laval, K., 2002. Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes. Journal of Geophysical Research: Atmospheres 107 (11). https://doi.org/10.1029/2001jd000634.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N.W., Clark, D.B., Dankers, R., Eisner, S., Fekete, B.M., Colón-González, F.J., Gosling, S.N., Kim, H., Liu, X., Masaki, Y., Portmann, F.T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., Kabat, P., 2014. Multimodel assessment of water scarcity under climate change. Proceedings of the National Academy of Sciences of the United States of America 111 (9), 3245–3250. https://doi.org/10.1073/pnas.1222460110.

Sellers, P.J., Mintz, Y., Sud, Y.D., Dalcher, A., 1986. A simple biosphere model (SiB) for use within general circulation models. Journal of the Atmospheric Science 43 (6), 505–531. https://doi.org/10.1175/1520-0469(1986)043.

Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T., Thonicke, K., Venevsky, S., 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology 9 (2), 161–185. https://doi.org/10.1046/j.1365-2486.2003.00569.x.

Smakhtin, V.U., 2001. Smakhtin 2010- low flow hydrology.pdf. Journal of HydrologyJournal of Hydrology 240, 147–186. https://doi.org/10.1016/S0022-1694(00)00340-1.
Stacke, T., Hagemann, S., 2012. Development and evaluation of a global dynamical wetlands extent scheme. Hydrology and Earth System Sciences 16 (8), 2915–2933. https://doi.org/10.5194/hess-16-2915-2012.
Stahl, K., Tallaksen, L.M., Hannaford, J., Van Lanen, H.A.J., 2012. Filling the white space on maps of European runoff trends: Estimates from a multi-model ensemble. Hydrology and Earth System Sciences 16 (7), 2035–2047. https://doi.org/10.5194/hess-16-2035-2012.
Takata, K., Emori, S., Watanabe, T., 2003. Development of the minimal advanced treatments of surface interaction and runoff. Global and Planetary Change 38 (1–2), 209–222. https://doi.org/10.1016/S0921-8181(03)00030-4.
Tang, Q., Oki, T., Kanae, S., Hu, H., 2007. The influence of precipitation variability and partial irrigation within grid cells on a hydrological simulation. Journal of Hydrometeorology 8 (3), 499–512. https://doi.org/10.1175/JHM589.1.
Van Huijgevoort, M.H.J., Hazenberg, P., Van Lanen, H.A.J., Teuling, A.J., Clark, D.B., Folwell, S., Gosling, S.N., Hanasaki, N., Heinke, J., Koirala, S., Stacke, T., Voss, F., Sheffield, J., Uijlenhoet, R., 2013. Global multimodel analysis of drought in runofffor the second half of the twentieth century. Journal of Hydrometeorology 14 (5), 1535–1552. https://doi.org/10.1175/JHM-D-12-0186.1.
Wada, Y., Wisser, D., Bierkens, M.F.P., 2014. Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth System Dynamics 5 (1), 15–40. https://doi.org/10.5194/esd-5-15-2014.
Wang, W., Lu, H., Yang, D., Sothea, K., Jiao, Y., Gao, B., Peng, X., Pang, Z., 2016. Modelling hydrologic processes in the Lancang-Mekong River basin using a distributed model driven by satellite precipitation and rain gauge observations. PLoS ONE 11 (3), 1–19. https://doi.org/10.1371/journal.pone.0152229.
Ward, P.J., Jongman, B., Weiland, F.S., Bouwman, A., Van Beek, R., Bierkens, M.F.P., Ligtvoet, W., Winsemius, H.C., 2014. Assessing flood risk at the global scale: Model setup, results, and sensitivity. Environmental Research Letters 8 (4). https://doi.org/10.1088/1748-9326/8/4/044019.
Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., Schewe, J., 2014. The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework. Proceedings of the National Academy of ences of the United States of America, 2014 111 (9), 3228–3232. https://doi.org/10.1073/pnas.1312330110.
Yao, T., Wu, G., Xu, B., Wang, W., Gao, J., An, B., 2019. Asian water tower change and its impacts. Bulletin of Chinese Academy of Sciences 11, 1203–1209. https://doi.org/10.16418/j.issn.1000-3045.2019.11.003.
Zaherpour, J., Gosling, S.N., Mount, N., Schmied, H.M., Veldkamp, T.I.E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J., Wada, Y., 2018. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environmental Research Letters 13 (6), 1–23. https://doi.org/10.1088/1748-9326/aac547.
Zhao, R., 1992. The Xinanjiang model applied in China. Journal of Hydrology 135 (1–4), 371–381. https://doi.org/10.1016/0022-1694(92)90096-E.
Zhao, F., Veldkamp, T.I.E., Frieler, K., Schewe, J., Ostberg, S., Willner, S., Schauberger, B., Gosling, S.N., Schmied, H.M., Portmann, F.T., Leng, G., Huang, M., Liu, X., Tang, Q., Hanasaki, N., Biemans, H., Gerten, D., Satoh, Y., Pokhrel, Y., Stack, Y., Ciais, P., Chang, J., Ducharne, A., Guimberteau, M., Wada, Y., Kim, H., Yamazaki, D., 2017. The critical role of the routing scheme in simulating peak river discharge in global hydrological models. Environmental Research Letters 12 (7). https://doi.org/10.1088/1748-9326/aa7250.