

## Highlights

### **Intercomparison of runoff and river discharge reanalysis datasets: Insights from the Upper Jinsha River, an alpine river at the eastern edge of the Tibetan Plateau**

Shuanglong Chen, Heng Yang, Hui Zheng

- The calibration of land surface models exerts a significant influence on the accuracy of runoff and river discharge estimations
- The runoff characteristics-based calibration method is highly effective for runoff estimation
- High-resolution simulations can yield distinct runoff characteristics compared to their low-resolution counterparts
- High-resolution river network delineation is essential for the concurrent enhancement of runoff and discharge estimations
- Intercomparison can identify deficiencies within model configurations and inform subsequent refinements

# Intercomparison of runoff and river discharge reanalysis datasets: Insights from the Upper Jinsha River, an alpine river at the eastern edge of the Tibetan Plateau

Shuanglong Chen<sup>a,1</sup>, Heng Yang<sup>b,1</sup>, Hui Zheng<sup>c,\*</sup>

<sup>a</sup>*Baihetan Hydropower Plant, China Three Gorges Corporation, 615421, Sichuan, China*

<sup>b</sup>*Science and Technology Research Institute, China Three Gorges Corporation, 100038, Beijing, China*

<sup>c</sup>*Institute of Atmospheric Physics, Chinese Academy of Sciences, 100029, Beijing, China*

---

## Abstract

The availability of runoff and river discharge reanalysis datasets is rapidly increasing. These datasets use different meteorological forcing, land surface models, river routing methods, river networks, calibration data, and calibration methods. This diversity calls for intercomparison. This study intercompares ERA5-Land, GloFAS, GRFR, and CNRD in the Upper Jinsha River, an alpine river on the eastern edge of the Tibetan Plateau. We include a high-resolution runoff simulation using the NoahMP land surface model and ERA5-Land meteorological data in the intercomparison to assess the influence of meteorological inputs, spatial resolution, and calibration on runoff estimation. The runoff from these datasets is rerouted on a high-resolution river network delineated from the 3-arcsecond MERIT-Hydro dataset, enabling us to evaluate the effects of routing models and river network delineations on discharge estimates. Our analysis is based on observations at three gauging

---

\*Corresponding author

Email addresses: chen\_shuanglong@ctg.com.cn (Shuanglong Chen), yang\_heng2@ctg.com.cn (Heng Yang), hzheng\_iap@outlook.com (Hui Zheng)

<sup>1</sup>These authors contributed equally to this work and should be considered co-first authors.

stations—Zhimenda, Gangtuo, and Benzilan—at monthly, daily, and hourly scales. Results indicate that the calibration of land surface or hydrological models significantly impacts runoff and river discharge estimates, potentially more than meteorological data. Calibrated datasets, GloFAS and GRFR, outperform the others, despite being driven by different forcings. The runoff characteristics-based calibration method, as seen in GRFR, yields superior performance, with rerouted hourly river discharge using GRFR runoff achieving a high Kling-Gupta Efficiency of approximately 0.9. Accurate river network delineation with high-resolution hydrography data is crucial. The high performance of GloFAS discharge, contrasting with its runoff, is likely attributed to the coarse routing grid resolution. These insights can guide further improvements in the reanalyses.

*Keywords:* river discharge, runoff, model intercomparison, the upper Jinsha River

---

## 1. Introduction

The proliferation of global- and continental-scale runoff and discharge reanalysis datasets has been a significant development in hydrological research. Datasets such as the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA5) (Hersbach et al., 2020), ERA5-Land (Muñoz-Sabater et al., 2021), the Global Flood Awareness System (GloFAS) (Alfieri et al., 2020; Harrigan et al., 2020), and the Global Reach-Level Flood Reanalysis (GRFR) (Yang et al., 2021; Lin et al., 2019) offer continuous and spatially comprehensive estimates of runoff and river discharge on a global scale. Additionally, regional datasets like the National Water Model (NWM) of the United States (Cosgrove et al., 2024), the European Flood Early Warning Systems (Najafi et al., 2024), and the China Natural Runoff Dataset (CNRD) (Gou et al., 2021; Miao et al., 2022) provide runoff and discharge estimates tailored to specific regions. These datasets have significantly advanced our scientific understanding of river systems (Feng et al., 2021; Collins et al., 2024)

and their complex interactions with other components of the Earth system (Liu et al., 2024). Furthermore, they have been instrumental in supporting decision-making processes, particularly in the area of water-resources utilization and flood protection (Cosgrove et al., 2024; Coughlan de Perez et al., 2016).

These reanalysis datasets of runoff and river discharge are generated through a complex chain of methodologies that include meteorological forcing data, land surface models or hydrological models, river routing models, river network delineations, and calibrations. The diversity within these components can significantly influence the estimation of runoff and river discharge. Studies have demonstrated that meteorological forcing data can differ remarkably (Sun et al., 2018; Henn et al., 2018). The use of different meteorological forcing data (Renard et al., 2010; Bai et al., 2024), in combination with varying model structure (Butts et al., 2004), spatial resolutions (Haddeland et al., 2002; Barnhart et al., 2024; Li et al., 2022), parameterizations (Zheng et al., 2019, 2023), and parameters (Li et al., 2020) of land surface models or hydrological models, can lead to disparate estimates of runoff (Saxe et al., 2021; Beck et al., 2017). Furthermore, these differences in runoff estimation can be further amplified by the methods employed in river routing (Fry et al., 2014; Mai et al., 2022; Ahmed et al., 2023).

The diversity presented in the estimation methodologies call for careful intercomparison before the use of reanalysis datasets. In the past decades, a variety of intercomparison studies have been conducted with different objectives. Land surface model comparisons such as the Project for Intercomparison of Land-surface Parameterization Schemes (Henderson-Sellers et al., 1993) and the Land Surface, Snow and Soil Moisture Model Intercomparison Project (van den Hurk et al., 2016) conducted with the same meteorological forcing data greatly help to reveal the weakness of the participant models and the interactions of land surface processes. The Distributed Model Intercomparison Project (Reed et al., 2004; Smith et al., 2004) compared distributed models with lumped models. This project revealed the

superiority of distributed routing models in reproducing flood peaks and reported the importance of calibration on distributed rainfall-runoff models. Beck et al. (2017) compares ten hydrological models at the global scale. Their results further confirm the importance of calibration of model parameters in accurate runoff estimation. The Continental Hydrological Intercomparison Project (Tijerina et al., 2021) compares two high-resolution, physics-based models at the continental scale. The intercomparison promotes high-resolution modeling to enhance topographic processing and to improve topographic representation over complex topography. There are also intercomparisons with regional focuses. The Great Lakes Runoff Intercomparison Project (Mai et al., 2022) and the Nelson Model Intercomparison Project under the Integrated Modelling Program of Canada initiative (Ahmed et al., 2023) are two examples. These intercomparison studies aim to select the best-performing estimation methodologies for the region and guide further improvements of operations.

The Upper Jinsha River, serving as the source of the Changjiang River, is notable for its pronounced vertical drop and substantial hydroelectric potential. The river is located on the eastern edge of the Tibetan Plateau. To date, thirteen hydroelectric facilities have been constructed or are in the planning stages along the main stem of the Upper Jinsha River, collectively possessing an installed capacity of approximately 14.6 gigawatts. The river originates from the Tibetan Plateau. The source area of the river consists of a major part of the Sanjiangyuan National Nature Reserve, the highest and largest nature reserve in China yet. The operations of the hydroelectric facilities and the natural reserves demand accurate estimation of water resources. However, hydrometeorological observations are scarce in the region. The complex topography, the high-alpine environment, and the strong demand for accurate runoff and river discharge estimates make the Upper Jinsha River an ideal testbed for runoff and river discharge reanalysis datasets.

This study is designed to intercompare runoff and river discharge reanalysis data obtained from ERA5-Land, GloFAS, GRFR, and CNRD. As a pilot effort of such reanalysis

intercomparisons, this study focuses specifically on the Upper Jinsha River. The primary objective is to assess the effectiveness and identify any shortcomings of the methodologies employed in these reanalyses. The findings from this intercomparison are expected to provide valuable insights that will guide future enhancements in the estimation of runoff and river discharge within the region.

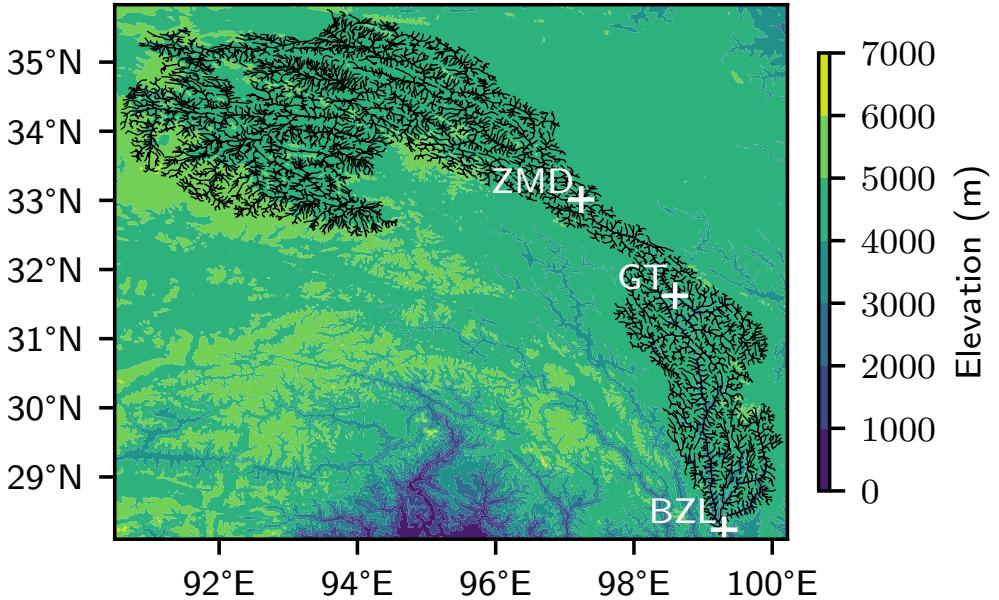
## 2. Data and methods

### 2.1. Study area and river network delineation

Figure 1 illustrates the terrain and river network of the upper Jinsha River. The study area encompasses the region upstream of the Benzilan gauging station. The elevation within the study area ranges from approximately 6600 m at its highest point to approximately 2000 m at its lowest. The river network was extracted from the Multi-Error-Removed Improved-Terrain Hydrography (MERIT-Hydro) dataset (Yamazaki et al., 2019, 2017). MERIT-Hydro is a globally recognized dataset providing hydrologically conditioned elevation, flow direction, and flow accumulation area data, with a spatial resolution of 3 arc-seconds. The extraction of the river network was based on the flow accumulation area, using a threshold of 25 km<sup>2</sup>, which resulted in the identification of 5,787 river reaches within the study area.

### 2.2. Reanalysis Datasets

Table 1 presents an overview of the datasets employed in this study. ERA5-Land (Muñoz-Sabater et al., 2021) is an enhanced global dataset, an improvement over ERA5 (Hersbach et al., 2020). This dataset features a substantial enhancement in spatial resolution, with an increase from ERA5's 31 km to 0.1° globally. A variety of land surface variables, including runoff, are simulated using the Carbon Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land (CHTESSEL) model (Boussetta et al., 2021; Balsamo et al., 2009), operating at an hourly temporal resolution.



**Fig. 1.** The study domain and river network of the source region of the upper Jinsha River. The white crosses indicate the location of the gauging stations: ZMD for Zhimenda, GT for Gangtuo, BT for Batang, and BZL for Benzilan.

The GloFAS historical reanalysis version 4.0 (Alfieri et al., 2020; Harrigan et al., 2020) is another dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), with a focus on hydrological variables. GloFAS shares the same spatial resolution of  $0.1^\circ$  as ERA5-Land. It is driven by ERA5 meteorological data, which is highly similar to ERA5-Land. The primary distinctions between GloFAS and ERA5-Land lie in the hydrological model employed, the temporal resolution, and the output variables. GloFAS utilizes the LISFLOOD model (Knijff et al., 2010) at a daily time step. Unlike ERA5-Land, the parameters of LISFLOOD are calibrated against daily river discharge observations using an evolutionary algorithm (Hirpa et al., 2018). LISFLOOD simulates both runoff and river discharge, whereas ERA5-Land provides only runoff estimates.

The GRFR dataset (Yang et al., 2021) is a comprehensive global dataset providing runoff and river discharge data at a 3-hourly temporal resolution. The runoff is simulated using the Variable Infiltration Capacity (VIC) model (Liang et al., 1994) at a spatial resolution of  $0.05^\circ$ , which is driven by precipitation data from the Multi-Source Weighted-Ensemble

Precipitation (MSWEP) dataset (Beck et al., 2019) and other meteorological variables from ERA5. The VIC model's parameters are calibrated (Yang et al., 2019) at each  $0.05^{\circ}$  grid cell against a machine-learning-derived runoff characteristics dataset produced by Beck et al. (2015). The bias in the VIC-simulated runoff is further corrected using the same runoff characteristics dataset, following the methodology of Lin et al. (2019). The calibrated and bias-corrected runoff data are then utilized to drive the Routing Application for Parallel computation of Discharge (RAPID) river routing model (David et al., 2011) for simulating river discharge. The parameters of RAPID are estimated using a statistical relationship that correlates discharge with river width, water depth, channel slope, and celerity (Lin et al., 2019). The river network employed in the routing process is also derived from the MERIT Hydro dataset (Lin et al., 2021).

The China Natural Runoff Dataset (CNRD) (Gou et al., 2021; Miao et al., 2022) comprises runoff data at a  $0.25^{\circ}$  spatial resolution. This dataset is simulated using the VIC model, driven by the China Meteorological Forcing Dataset (He et al., 2020). The parameters of the VIC model were calibrated against river discharge observations from approximately 200 catchments across China and then extrapolated to ungauged river basins using the multiscale regionalization method (Samaniego et al., 2010).

The quality of runoff and river discharge reanalyses is influenced by various factors, including meteorological forcing, land surface model/hydrological model, river network routing model, model resolution, and parameter settings. In this study, we produced a new dataset for comparison. As detailed in Table 1, our dataset boasts the highest spatial and temporal resolution, employs an advanced land surface model, and utilizes the most sophisticated river routing method among the datasets examined. The dataset is generated using the Noah with Multi-Parameterization (NoahMP) land surface model (Niu et al., 2011), which is driven by ERA5-Land meteorological data. The static parameters for NoahMP are derived from the Weather Research and Forecasting geographic static data (<https://www2.mmm.ucar.edu/>

`wrf/users/download/get_sources_wps_geog.html`). NoahMP operates at a spatial resolution of  $0.01^\circ$  from 2008 to 2016. The initial condition for 2008 is sourced from ERA5-Land. The first year is allocated for model spin-up, with the simulated runoff from 2009 to 2016 used for analysis. NoahMP offers a variety of parameterization schemes for land surface processes, and we adopted the same schemes as those used in the NWM of the United States (Cosgrove et al., 2024). The runoff is routed using the variable-parameter Muskingum–Cunge (MC) method on the river network depicted in Figure 1 to estimate river discharge. The routing algorithm mirrors that of the NWM, assuming a trapezoidal channel cross-section. Channel cross-section geometry and roughness parameters are sourced from the default parameter lookup table of the NWM (Read et al., 2023) and the Strahler stream order for each river reach within the network. In theory, the MC routing method would outperform the kinematic wave routing method used in GloFAS by also additionally considering diffusive wave dynamics, while the Muskingum routing method used in GRFR would be the least performing among the three. It is important to note that this comparison is theoretical, and actual performance can vary significantly with model implementation and parameter settings

### *2.3. River discharge observations*

We utilized river discharge observations from three gauging stations along the mainstream of the upper Jinsha River: Zhimenda (ZMD), Gangtuo (GT), and Benzilan (BZL), covering the period from 2009 to 2016. The selection of this timeframe is justified by the minimal presence of hydroelectric facilities, which ensures that the river discharge data are largely unaffected by human interventions.

The dataset from Zhimenda is compiled at a daily resolution. The dataset comprises 1,477 records, with 49% of the days within the 2009 to 2016 period missing. In contrast, Gangtuo and Benzilan provide data at an hourly resolution, accumulating 52,246 hours and 69,011 hours, respectively. The respective missing data rates for these stations are 25%

for Gangtuo and 1.6% for Benzilan. Given the nature of our analysis, we have chosen not to impute the missing values, opting instead to focus solely on the complete data records. This decision was made to preserve the integrity of the dataset and to mitigate the potential biases that could be introduced by gap-filling methods.

To our knowledge, the daily river discharge observations at Zhimenda have been used to calibrate GloFAS and GRFR, while the observations at Gangtuo and Benzilan have not been incorporated into the datasets analyzed in this study.

**Table 1.** The datasets used in this study.

Dataset	Time period	Temporal Resolution	Runoff Resolution	River Routing	References
<i>Publically available datasets</i>					
ERA5-Land	1950–Present	Hourly	0.10°	N/A	Muñoz-Sabater et al. (2021)
GloFAS v4.0	1979–Present	Daily	0.10°	LISFLOOD	Harrigan et al. (2020)
GRFR v1.0	1980–2019	3-hourly	0.05°	RAPID	Yang et al. (2021)
CNRD v1.0	1961–2018	Monthly	0.25°	N/A	Miao et al. (2022)
<i>Datasets produced in this study</i>					
ERA5-Land/ NoahMP	2009–2016	Hourly	0.01°	MC	This study

#### 2.4. Evaluation methods

As shown in Table 1, the datasets vary in spatial and temporal resolution. River discharge estimates are not available in every dataset examined. To address these inconsistencies, we remapped the runoff onto a common 0.01° grid and refined both daily and monthly values to hourly values, as depicted in Figure 2. The spatial remapping was conducted using a first-order conservative method. For temporal refinement, we used the fraction of hourly ERA5-Land runoff to the accumulated runoff within a day or a month to redistribute the daily or monthly values. Subsequently, the runoff from all datasets was routed using the same Muskingum–Cunge (MC) method, as previously described, to estimate river discharge.

For GloFAS and GRFR, which provide river discharge estimates, we also refined the 3-hourly or daily values to hourly values using the same refinement method as the runoff.

The river discharge estimates derived from the Muskingum–Cunge method, as well as the temporally refined river discharge data from GRFR and GloFAS, are then compared with the observations at the gauging stations. This evaluation is conducted across three temporal scales—monthly, daily, and hourly—to comprehensively assess the accuracy and consistency of the estimates with the actual observations.

The structure and parameters of the river routing model significantly influence the river discharge estimates. To isolate the impact of river routing on these estimates, we not only compared the model-derived river discharge with observations at gauging stations but also aggregated the runoff above these stations for a more direct comparison. This aggregation helps to isolate the effects of routing by focusing on the input runoff rather than the routing process itself. The aggregated runoff is calculated by summing the runoff within the upstream area of each gauging station. The delineation of the upstream area was based on the flow direction data from the MERIT-Hydro dataset.

## *2.5. Evaluation metrics*

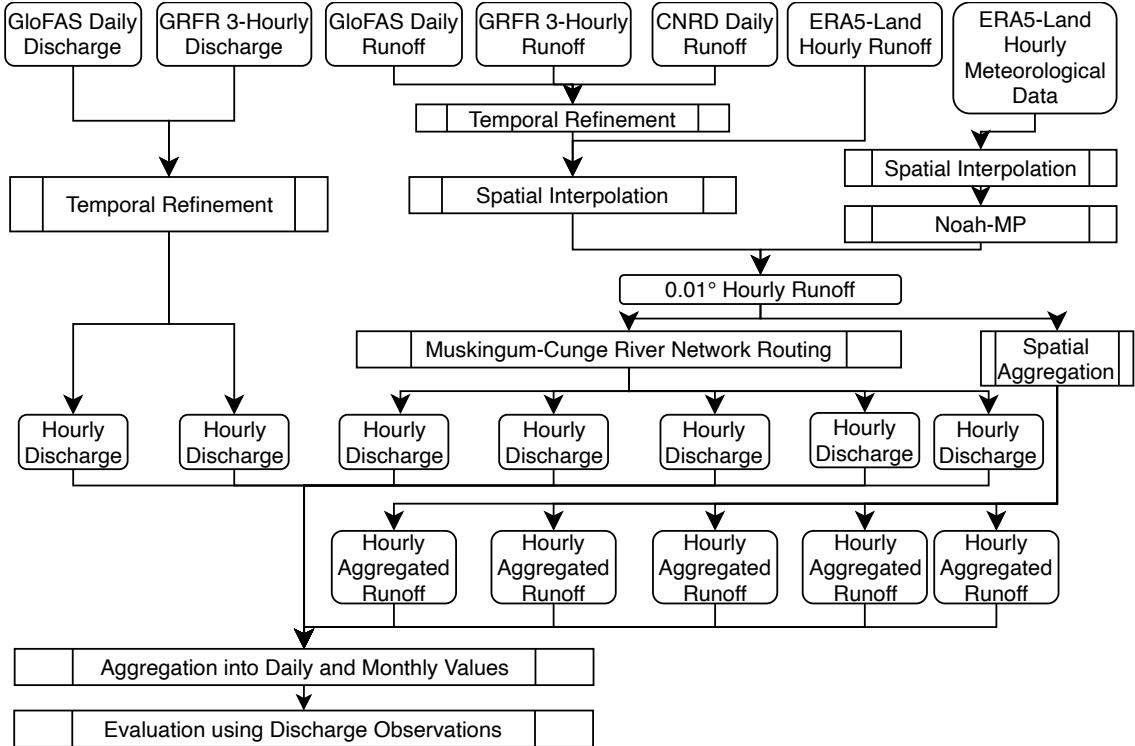
We employed the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) to assess the performance of the reanalysis datasets. The KGE is a comprehensive metric that encapsulates the correlation coefficient, the ratio of the standard deviations, and the ratio of the mean values of both observed and simulated data. The KGE is formulated as follows:

$$\text{KGE} = 1 - \sqrt{G_1^2 + G_2^2 + G_3^2}, \quad (1)$$

$$G_1 = r - 1, \quad (2)$$

$$G_2 = \frac{\sigma_s}{\sigma_o} - 1, \quad (3)$$

$$G_3 = \frac{\mu_s}{\mu_o} - 1, \quad (4)$$



**Fig. 2.** Workflow of this study.

where  $r$  is the correlation coefficient,  $\sigma_s$  and  $\sigma_o$  are the standard deviations of the simulated and observed data, respectively, and  $\mu_s$  and  $\mu_o$  are the mean values of the simulated and observed data, respectively. The KGE ranges from  $-\infty$  to 1, with a value of 1 indicating a perfect match between the simulated and observed data. A KGE value significantly less than 1 indicates a poor match.

The relative contribution of the three terms in the KGE can be calculated using the following formula (Gupta et al., 2009):

$$g_i = \frac{G_i^2}{G_1^2 + G_2^2 + G_3^2}, \quad \text{for } i = 1, 2, 3, \quad (5)$$

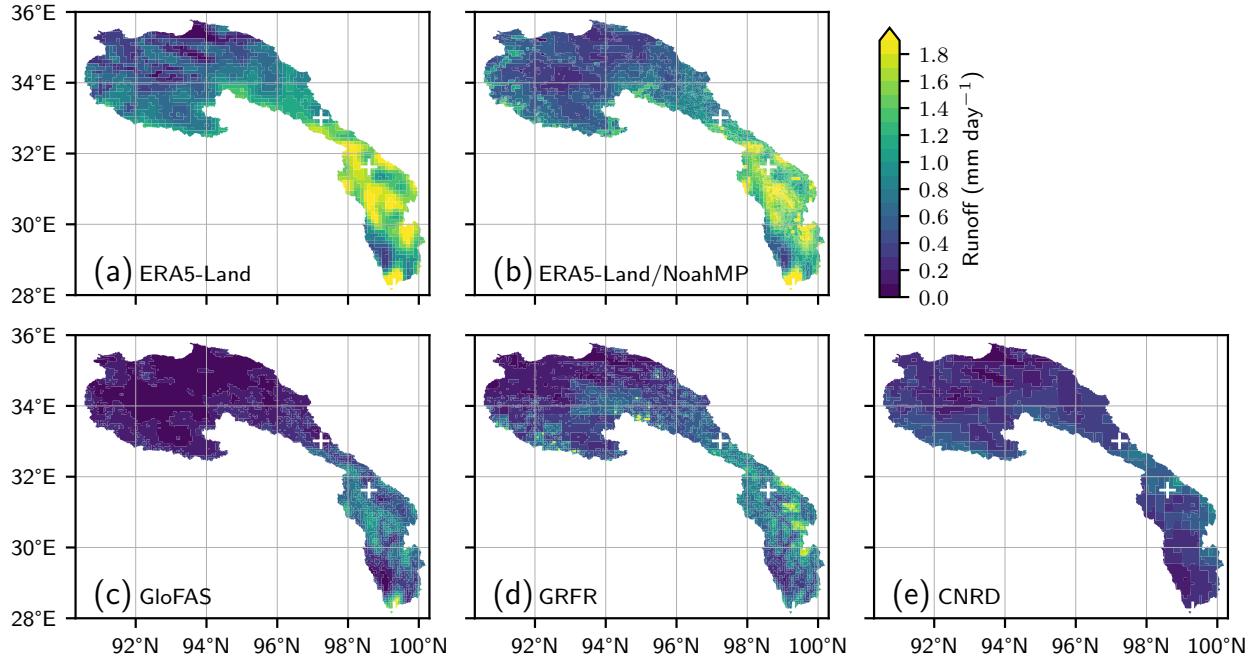
where  $g_1$ ,  $g_2$ , and  $g_3$  are the relative contributions of correlation coefficient, variability, and bias, respectively. A higher value indicates a greater contribution of the corresponding term to the deviation of the KGE from 1.

### 3. Results and discussion

#### 3.1. Runoff

Figure 3 geographically compares the runoff estimates within the Benzilan catchment. The five examined datasets fall into three categories: the first includes ERA5-Land and ERA5-Land/NoahMP, the second includes GloFAS and GRFR, and the third includes CNRD. The first group utilizes uncalibrated land surface models, the second group employs extensively calibrated hydrological models, and the third group features a hydrological model with parameters that are regionalized rather than individually calibrated. It is intriguing to note that ERA5-Land, ERA5-Land/NoahMP, and GloFAS exhibit significant differences, despite their meteorological forcing data being similar. On the other hand, GloFAS and GRFR, which are driven by different meteorological forcing data, show close spatial patterns. This comparison indicates that the calibration of land surface models or hydrological models has a profound impact on runoff estimation, more than the meteorological data.

Figure 4 intercompares the monthly variations and annual cycle of the aggregated runoff above Zhimenda, Gangtuo, and Benzilan. Since the river flow travel time is typically smaller than a month (Allen et al., 2018), it is generally safe to neglect the impact of river routing at the monthly time scale. Both ERA5 and CNRD overestimate the runoff at all examined stations, but the reasons for this overestimation differ. The overestimation of ERA5 is consistent with previous studies, which link it to overestimated precipitation in high altitude regions (Bian et al., 2019, 2020; Sun et al., 2021; Wang et al., 2024). In contrast, the overestimation of CNRD is primarily observed in the upstream area of Zhimenda. Given that CMFD, the meteorological data that drive CNRD, has shown closeness to in-situ precipita-



**Fig. 3.** Intercomparison of the spatial distribution of multiyear averaged runoff from 2009 to 2016.

tion observations in this region (Wang et al., 2024), the overestimation of the CNRD runoff is likely due to less constrained hydrological model parameters. CNRD uses parameters calibrated at gauged basins and extrapolates these parameters to ungauged basins using multiscale regionalization functions (Gou et al., 2021; Miao et al., 2022). The scarcity of river discharge gauges on the Tibetan Plateau means that this extrapolation can lead to bias.

The runoff from ERA5-Land/NoahMP exhibits two peaks, particularly in the high-altitude regions above Zhimenda. In contrast, ERA5-Land, which utilizes the same meteorological data, shows an insignificant rise in runoff during Spring. The discrepancy between ERA5-Land/NoahMP and ERA5-Land would be attributed to the spatial resolution of the land surface model. The higher resolution of ERA5-Land/NoahMP allows for a better capture of the vertical gradient of land surface temperature. The combination of overestimated precipitation in the ERA5-Land meteorological data and increased snowmelt in low-altitude model grids (Barnhart et al., 2024) results in a factitious Spring runoff peak that is not

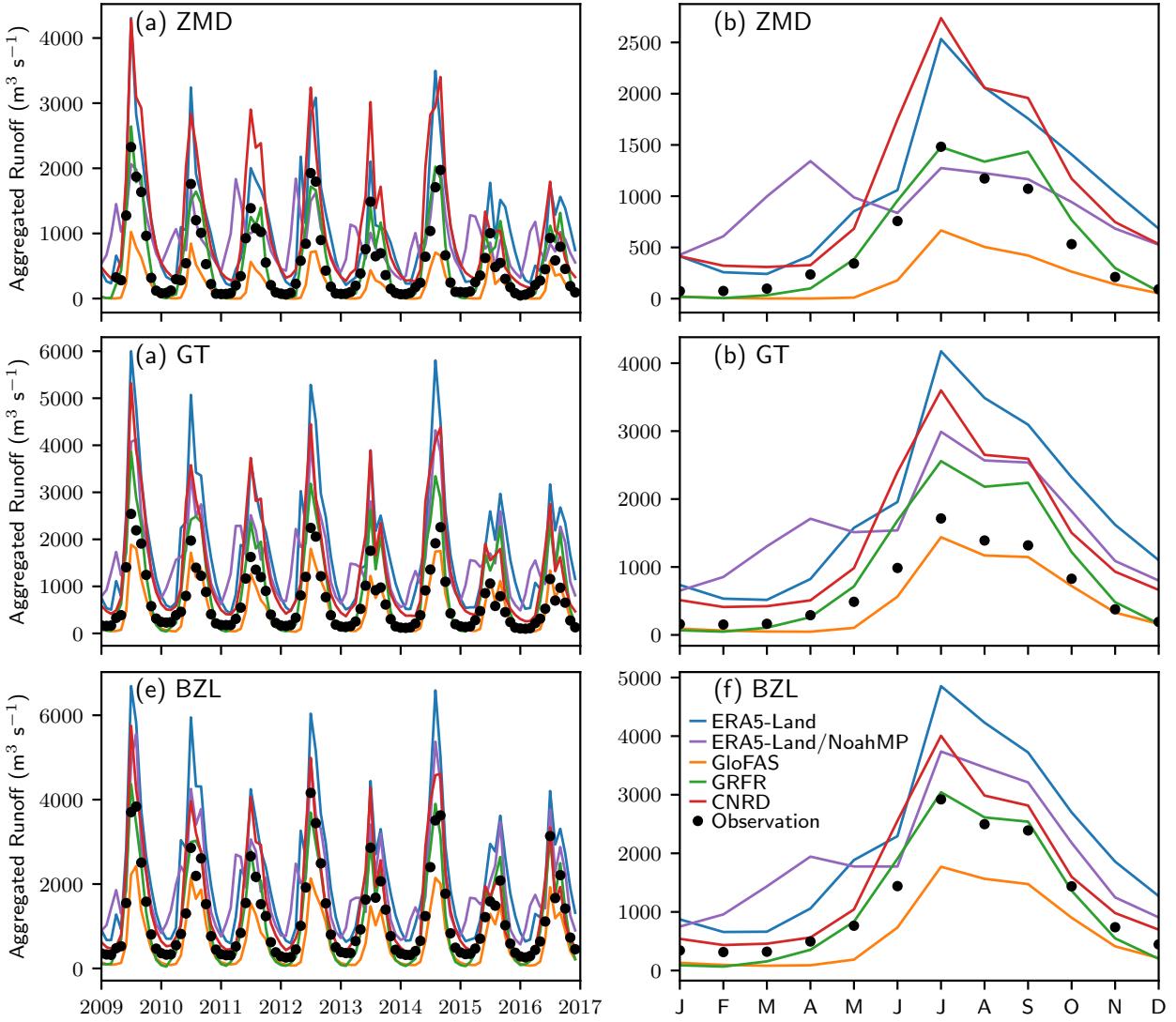
observed in actual data.

The two calibrated datasets, GloFAS and GRFR, exhibit distinct behaviors. This difference is even evident at Zhimenda, where observations are publicly available through the Global Runoff Data Centre and have been utilized in both datasets. The distinct calibration approaches of the two datasets likely account for this difference. GloFAS is calibrated against river discharge observations (Hirpa et al., 2018), while its runoff is not directly calibrated. Since river discharge can be calculated by multiplying the catchment area by runoff, different estimates of the catchment area will lead to different runoff values after the calibration of river discharge. In contrast, GRFR’s runoff is directly calibrated against a machine-learning-derived runoff characteristics dataset (Yang et al., 2019). The results suggest that the calibration method employed by GRFR is highly effective, yielding the best-performing runoff among the examined datasets.

Figure 5 presents the performance of the aggregated runoff in comparison with discharge observations at the monthly time scale. ERA5-Land consistently underperforms across all stations and time scales, corresponding to the overestimation shown in Figure 4. CNRD also performs poorly in high-altitude areas above Gangtuo due to overestimation, but its performance improves at Benzilan. The performance of CNRD slightly outperforms the high-resolution ERA5-Land/NoahMP. Among the examined datasets, the two calibrated datasets, GloFAS and GRFR, perform the best at almost every station examined, highlighting the importance of calibration for land surface models or hydrological models. Notably, GRFR’s performance is exceptionally good, with a KGE above 0.8 at Zhimenda and Benzilan. The high performance of GRFR suggests the effectiveness of the characteristics-based calibration method.

### 3.2. Streamflow

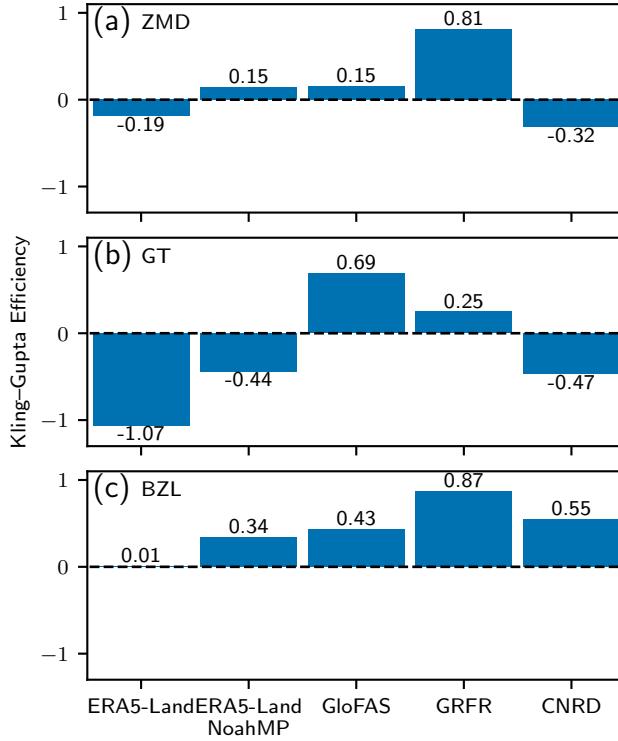
Figure 6 compares the monthly discharge estimations with observations. The patterns generally follow the runoff pattern shown in Figure 4. The overestimation observed in



**Fig. 4.** Intercomparison of the aggregated runoff at the gauging stations. (a) monthly aggregated runoff at Zhimenda; (b) annual cycle of the aggregated runoff at Zhimenda; (c) monthly aggregated runoff at Gangtuo; (d) annual cycle of the aggregated runoff at Gangtuo; (e) monthly aggregated runoff at Benzilan; and (f) annual cycle of the aggregated runoff at Benzilan.

ERA5-Land and CNRD closely mirrors their runoff overestimation. The artificial Spring peak, which is also present in the ERA5-Land/NoahMP discharge, aligns with the runoff pattern.

It is interesting to note that the GloFAS and GRFR discharge behave differently from the runoff. Our rerouted GloFAS runoff (GloFAS-MC in Figure 6) can reproduce the underestimated patterns observed for the runoff (Figure 4), confirming that GloFAS underestimates

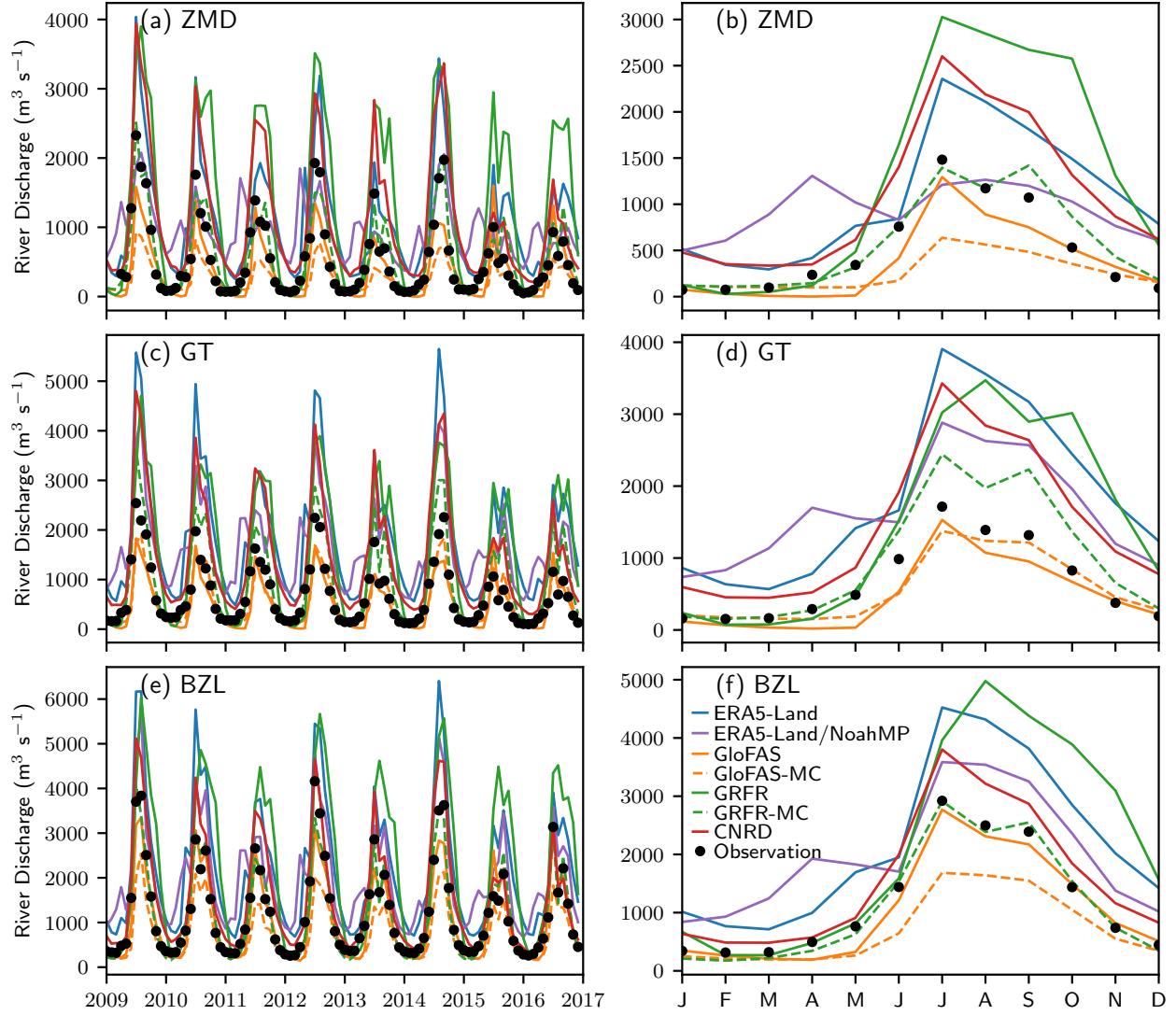


**Fig. 5.** Kling–Gupta Efficiency of the monthly aggregated runoff. (a) the measures at Zhimenda, (b) the measures at Gangtuo, and (c) the measures at Benzilan.

runoff. However, the underestimation of runoff is no longer presented in the river discharge. GloFAS closely reproduced the observed river discharge (Figure 6). The contradiction between the runoff underestimation and the unbiased discharge estimation suggests that the spatial resolution of GloFAS’s routing grid plays a role. GloFAS is routed on a  $\frac{1}{10}^\circ$  grid cell basis, whereas the river basins delineated in this study (as shown in Figure 1) are based on  $\frac{1}{1200}^\circ$  grids. The resolution of GloFAS may be too coarse to accurately delineate the river basins. However, this error is mitigated in the discharge due to the calibration, and the impact is transferred to errors in the runoff.

Despite the high skill of GRFR’s runoff demonstrated in Figure 4, the skill of GRFR’s discharge is unexpectedly unsatisfactory. GRFR remarkably overestimates discharge at all gauging stations. Meanwhile, our estimated discharge (GRFR-MC) using the GRFR runoff and the Muskingum–Cunge routing method can closely reproduce the observed discharge

patterns. Since neither the dynamics nor the parameters of the river routing model should have such significant impacts on monthly discharge, the unexpected overestimation in GRFR is likely attributable to some inadequate configurations of the river routing model.



**Fig. 6.** Same as Figure 4, but for river discharge.

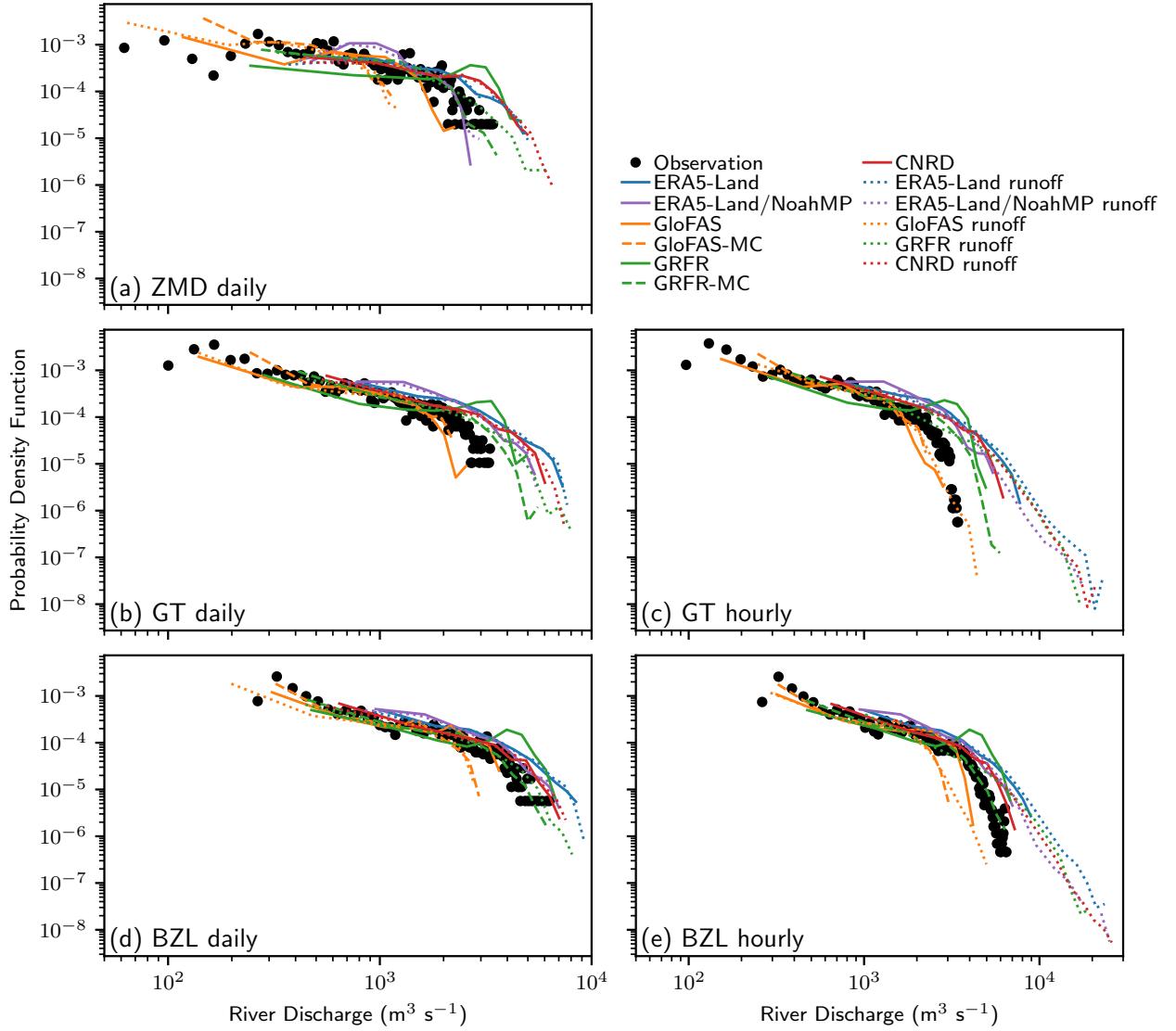
Figure 7 presents the river discharge estimates compared with both daily and hourly observations. The aggregated runoff is also displayed for comparison. The difference between the aggregated runoff and the river discharge reveals the impact of river routing. The comparison shows that river routing has a marginal impact in low flow conditions and at

the monthly time scale. The impact increases with the flow rate and on shorter time scales. Significant impacts of the river routing are found on the hourly high flow estimations at all gauging stations. This finding suggests the indispensable role of river routing in flood forecasting.

The differences among the examined river discharge estimates are mainly exhibited in high flow conditions. However, these differences do not vary significantly with the time scale of the estimates (i.e., daily or hourly). The GRFR river discharge estimation suspiciously indicates a high probability of discharge at approximately  $4000 \text{ m}^3 \text{s}^{-1}$  at all the examined stations, whereas such a high probability is not found in the other estimates including GRFR-MC. This suspicious pattern is another strong evident of inadequate configurations within GRFR's routing process. The GloFAS estimate can well reproduce the observed discharge probability patterns. Our rerouted river discharge using the GloFAS runoff and the Muskingum–Cunge method deviates from the observed pattern of discharge probability. The departure is transferred by the calibration of river discharge from the imprecision in the estimation of river basin areas.

Figure 8 illustrates the skill scores for the estimation of river discharge across monthly, daily, and hourly temporal scales. At the monthly scale, the river discharge estimation's skill closely mirrors that of the runoff, as depicted in Figure 5. The coincide hints that the primary determinant of river discharge variability within the study area is the runoff estimation, rather than the river routing process. An exception to this pattern is observed with GloFAS, where the divergence in discharge skill scores from those of runoff is likely due to the coarse and imprecise river routing grid employed, as previously discussed.

Among the examined dataset, GloFAS, GloFAS-MC, and GRFR-MC perform the best at all gauging stations and across all time scales. Proper calibration of the land surface model and the hydrological model is essential for accurate river discharge estimation. As revealed in Figure 9, the primary reasons for the poor performance of the examined datasets,



**Fig. 7.** Intercomparison of the probability distribution of river discharge and aggregated runoff at the river gauging stations. (a) daily estimates at Zhimenda; (b) hourly estimates at Zhimenda; (c) daily estimates at Gangtuo; (d) hourly estimates at Gangtuo; (e) daily estimates at Benzilan; (f) hourly estimates at Benzilan. Dotted lines denote the aggregated runoff. Solid lines denote the river discharge estimates provided by the examined datasets. Dashed lines denote the river discharge estimates using the runoff and the Muskingum–Cunge routing method. Zhimenda is not shown in the hourly estimates due to the lack of hourly observations.

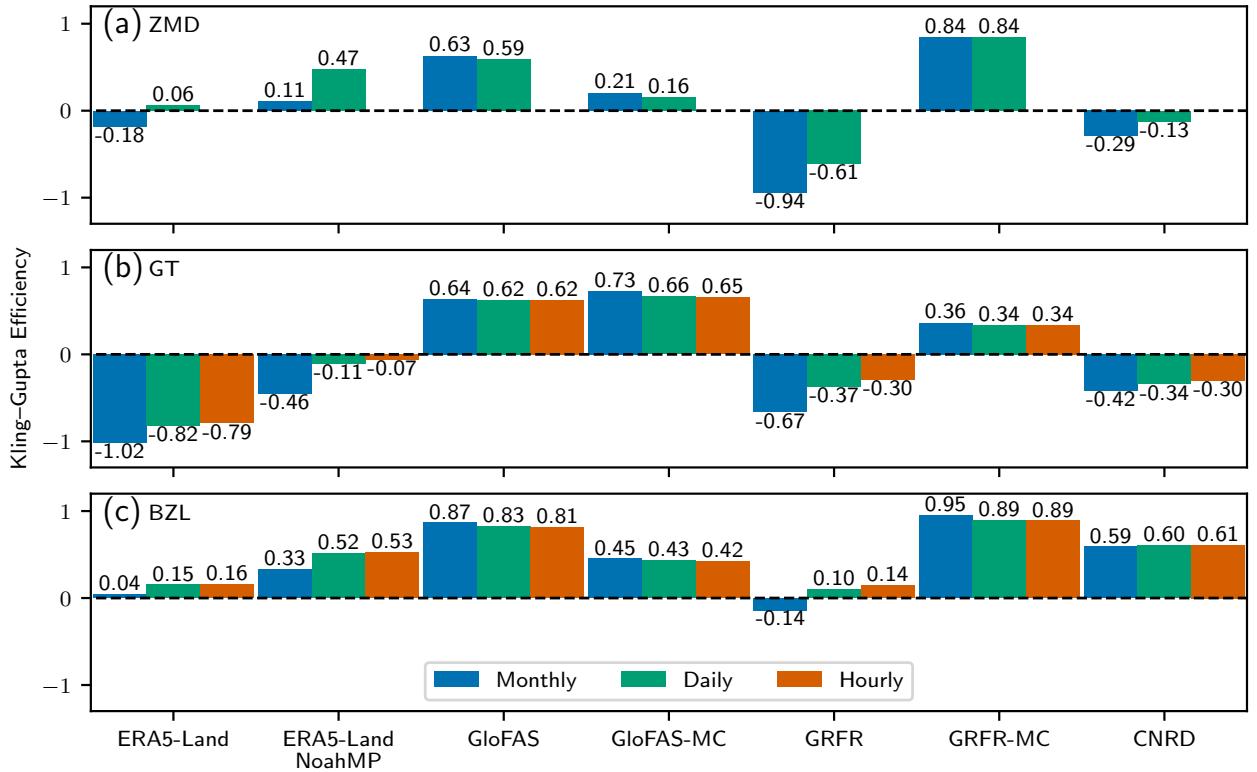
except for GRFR-MC, are errors in reproducing the observed mean and variability of river discharge. GRFR-MC can effectively reproduce the observed mean and variability of river discharge. The KGE of GRFR-MC's hourly discharge can reach a high score of approximately 0.9 at Benzinan, which again confirms the effectiveness of the characteristics-based

calibration method. The deviation of the KGE from 1 for GRFR-MC is mainly attributed to the correlation coefficient, which can be further improved through the calibration of the river routing parameters. The exceptionally high skill of GRFR-MC again confirms the effectiveness of the characteristics-based calibration method.

It is worth to note that the KGE (Figure 8) and the relative contribution within the KGE (Figure 9) are consistent for both hourly and daily estimates. This consistency suggests that calibration at the daily time scale could potentially be applied to the hourly time scale. The KGE of river discharge is also consistent between daily and monthly estimates for most datasets, with the exception of the high-resolution ERA5-Land/NoahMP estimation. The increase in KGE from the monthly to daily time scale indicates that a more accurate representation of the spatial distribution of runoff and river routing could improve the temporal variations in river discharge estimates, aligning with the concept of the time of concentration (Beven, 2020). However, these findings are based on a limited number of gauging stations and a limited time period. Further studies are needed to confirm these preliminary observations.

#### 4. Conclusions

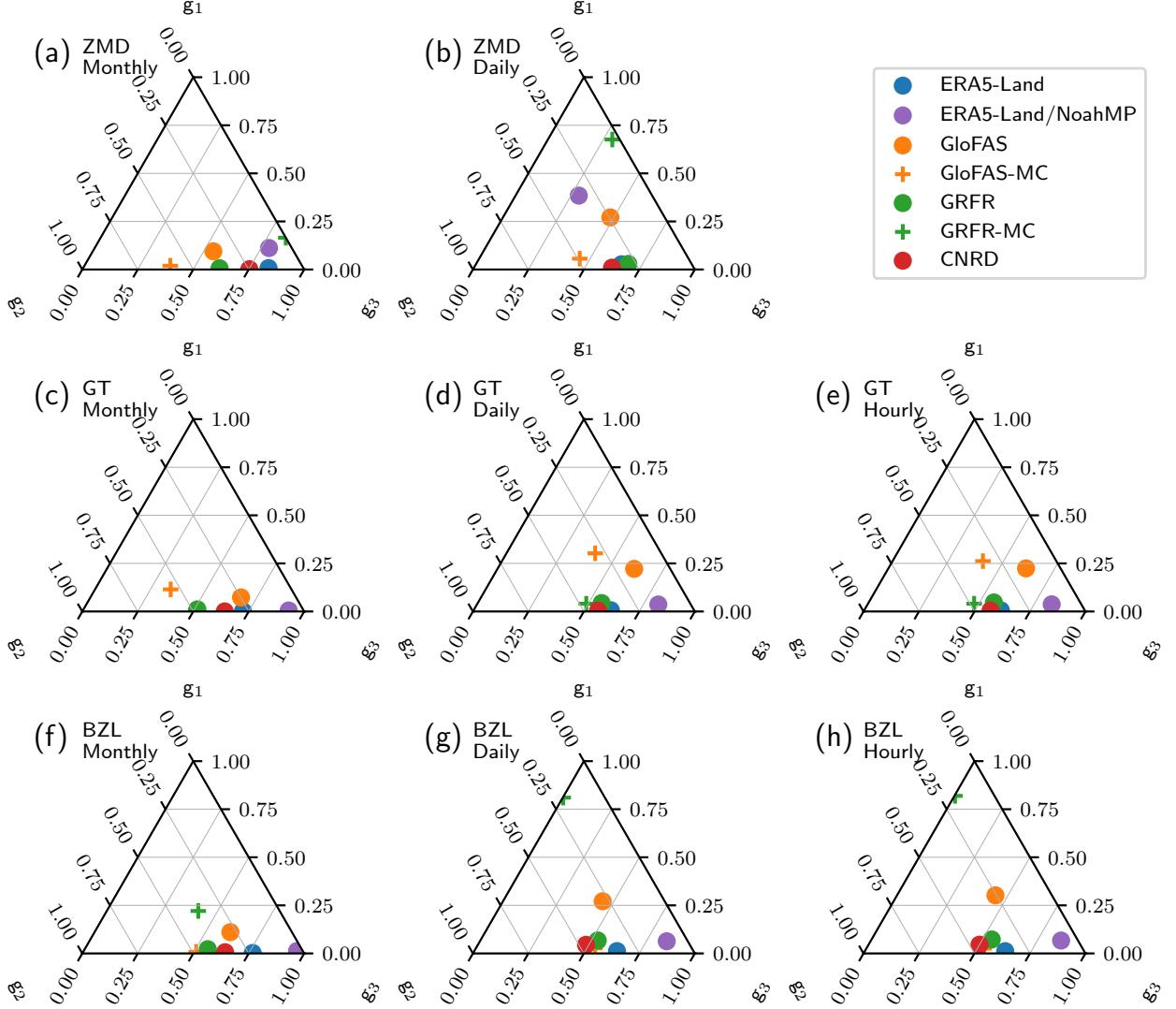
This study presents a pilot intercomparison of several publicly available large-domain reanalysis datasets of runoff and river discharge in the Upper Jinsha River basin. The datasets under scrutiny include ERA5-Land, GloFAS, GRFR, and CNRD. To enhance the understanding of the effects of calibration, spatial resolution, and meteorological forcing on runoff estimation, a high-resolution runoff simulation has been incorporated into the comparison. The runoff from the examined datasets is rerouted on a high-resolution river network delineated from the 3-arcsecond MERIT-Hydro dataset. This rerouting aids in identifying the impacts of different river routing models and river network delineations on river discharge estimation. The estimated runoff and river discharge are compared with observations at



**Fig. 8.** Kling-Gupta Efficiency of the river discharge estimation. (a) the measures at the Zhimenda station, (b) the measures at the Gangtuo station, and (c) the measures at the Benzilan station.

three gauging stations—Zhimenda, Gangtuo, and Benzilan—spanning monthly, daily, and hourly time scales. The principal findings of this study are as follows:

- The examined runoff and river discharge reanalysis datasets display significant discrepancies. When compared with observations, the Kling-Gupta Efficiency of river discharge closely aligns with that of runoff. The primary determinant of river discharge variability within the study area is runoff estimation, rather than the river routing process. The data and models used in runoff estimation need further scrutinization.
- Variations in runoff estimates are predominantly attributed to the calibration processes of land surface and hydrological models. Among the datasets, GloFAS and GRFR, which have been calibrated, demonstrate superior performance in runoff estimation.



**Fig. 9.** Ternary plot showing the relative contribution of correlation coefficient ( $g_1$ ), variability ( $g_2$ ), and bias ( $g_3$ ) to the deviation of the Kling–Gupta Efficiency of simulated river discharge from one. The three rows of the plot represent the estimation at Zhimenda, Gangtuo, and Benzilan, respectively. The three columns of the plot represent the results of the monthly, daily, and hourly estimates, respectively.

GloFAS and GRFR show similar runoff patterns, despite being driven by different meteorological forcing data. In contrast, the runoff estimates from GloFAS markedly diverge from those of ERA5-Land, even though the meteorological forcing data are similar.

- The runoff characteristics-based calibration method shows to be highly effective. GRFR,

which employs this method, exhibits the best performance among the examined datasets in terms of runoff estimation. An uncalibrated re-routing using the Muskingum-Cunge method with GRFR runoff can achieve the best performance in estimated river discharge among the examined datasets. However, the river discharge of GRFR exhibits unexpected deterioration, likely due to inadequate configurations in the river routing model.

- The river routing process shows its significance during high flow conditions and at time scales shorter than a month. This finding underscores the importance of incorporating a river routing model in flood forecasting. Conversely, the impact of river routing is minimal in low flow conditions and at the monthly time scale. The similarity in monthly aggregated runoff and river discharge estimates allows for an examination of the adequacy of the river routing model and the river network delineation.
- A high-resolution river network for flow routing is crucial for concurrent runoff and river discharge estimation. The intercomparison shows a contradiction between GloFAS's high skill in river discharge estimation and a relatively low skill in runoff estimation. The contradiction is likely due to its coarse river routing grid. The coarse grid introduces errors in catchment area estimation, which are then transferred to the runoff estimates during the calibration of river discharge.
- High-resolution runoff simulations exhibit distinct runoff characteristics compared to their low-resolution counterparts. While the calibration of high-resolution runoff models can ameliorate these discrepancies, inherent differences in parameter settings between the two resolutions may still result in persistent divergences within other facets of the hydrological cycle.

In summary, this study underscores the critical role of model calibration within hydrological modeling and advocates for the integration of high-resolution hydrography data. The

results provide significant insights into the enhancement of runoff and streamflow estimation accuracy, which are essential for effective water resource management and the early warning of flood risks in the Upper Jinsha River, and likely in other regions with analogous hydrological attributes.

### **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.

### **Data availability**

The MERIT-Hydro dataset was obtained from [https://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT\\_Hydro/](https://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/) (accessed on 2024-11-10). The ERA5-Land dataset was obtained from <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land>. The GloFAS v4.0 dataset was obtained from <https://ewds.climate.copernicus.eu/datasets/cems-glofas-historical> (accessed on 2024-11-10). The GRFR v1.0 dataset was obtained from <https://www.reachhydro.org/home/records/grfr> (accessed on 2024-11-10). The COMIDs used to extract the river discharge from GRFR are 43037713 for Zhimenda, 43046824 for Gangtuo, and 43066085 for Benzilan, respectively. The CNRD v1.0 dataset was obtained from [https://doi.org/10.11888/Atmos\\_tpdc.272864](https://doi.org/10.11888/Atmos_tpdc.272864) (accessed on 2024-11-10). The river discharge observations were obtained from the China Three Gorges Corporation; however, the authors do not have permission to share the data.

### **Acknowledgements**

This study is supported by China Three Gorges Corporation (contract Z532302035), the National Key Research and Development Program of China (2023YFF0805501), and the Natural Science Foundation of China (grants 42075165 and 42275178). We thank for the

technical support of the National Large Scientific Infrastructure “Earth System Numerical Simulation Facility” (<https://cstr.cn/31134.02.EL>).

## References

- Ahmed, M.I., Stadnyk, T., Pietroniro, A., Awoye, H., Bajracharya, A., Mai, J., Tolson, B.A., Shen, H., Craig, J.R., Gervais, M., Sagan, K., Wruth, S., Koenig, K., Lilhare, R., Déry, S.J., Pokorny, S., Venema, H., Muhammad, A., Taheri, M., 2023. Learning from hydrological models' challenges: A case study from the Nelson basin model intercomparison project. *Journal of Hydrology* 623, 129820. doi:10.1016/j.jhydrol.2023.129820.
- Alfieri, L., Lorini, V., Hirpa, F.A., Harrigan, S., Zsoter, E., Prudhomme, C., Salamon, P., 2020. A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X* 6, 100049. doi:10.1016/j.hydroa.2019.100049.
- Allen, G.H., David, C.H., Andreadis, K.M., Hossain, F., Famiglietti, J.S., 2018. Global estimates of river flow wave travel times and implications for low-latency satellite data. *Geophysical Research Letters* 45, 7551–7560. doi:10.1029/2018GL077914.
- Bai, F., Wei, Z., Wei, N., Lu, X., Yuan, H., Zhang, S., Liu, S., Zhang, Y., Li, X., Dai, Y., 2024. Global assessment of atmospheric forcing uncertainties in the Common Land Model 2024 simulations. *Journal of Geophysical Research: Atmospheres* 129, e2024JD041520. doi:10.1029/2024JD041520.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., Betts, A.K., 2009. A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *Journal of Hydrometeorology* 10, 623–643. doi:10.1175/2008JHM1068.1.
- Barnhart, T.B., Putman, A.L., Heldmyer, A.J., Rey, D.M., Hammond, J.C., Driscoll, J.M., Sexstone, G.A., 2024. Evaluating distributed snow model resolution and meteorology parameterizations against streamflow observations: finer is not always better. *Water Resources Research* 60, e2023WR035982. doi:10.1029/2023WR035982.
- Beck, H.E., van Dijk, A.I.J.M., de Roo, A., Dutra, E., Fink, G., Orth, R., Schellekens, J., 2017. Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrology and Earth System Sciences* 21, 2881–2903. doi:10.5194/hess-21-2881-2017.
- Beck, H.E., de Roo, A., van Dijk, A.I.J.M., 2015. Global maps of streamflow characteristics based on observations from several thousand catchments. *Journal of Hydrometeorology* 16, 1478–1501. doi:10.1175/JHM-D-14-0155.1.

- Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., van Dijk, A.I.J.M., McVicar, T.R., Adler, R.F., 2019. MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society* 100, 473–500. doi:10.1175/BAMS-D-17-0138.1.
- Beven, K.J., 2020. A history of the concept of time of concentration. *Hydrology and Earth System Sciences* 24, 2655–2670. doi:10.5194/hess-24-2655-2020.
- Bian, Q., Xu, Z., Zhao, L., Zhang, Y.F., Zheng, H., Shi, C., Zhang, S., Xie, C., Yang, Z.L., 2019. Evaluation and intercomparison of multiple snow water equivalent products over the Tibetan Plateau. *Journal of Hydrometeorology* 20, 2043–2055. doi:10.1175/JHM-D-19-0011.1.
- Bian, Q., Xu, Z., Zheng, H., Li, K., Liang, J., Fei, W., Shi, C., Zhang, S., Yang, Z.L., 2020. Multiscale changes in snow over the Tibetan Plateau during 1980–2018 represented by reanalysis data sets and satellite observations. *Journal of Geophysical Research: Atmospheres* 125, e2019JD031914. doi:10.1029/2019JD031914.
- Boussetta, S., Balsamo, G., Arduini, G., Dutra, E., McNorton, J., Choulga, M., Agustí-Panareda, A., Beljaars, A., Wedi, N., Muñoz-Sabater, J., de Rosnay, P., Sandu, I., Hadade, I., Carver, G., Mazzetti, C., Prudhomme, C., Yamazaki, D., Zsoter, E., 2021. ECLand: The ECMWF land surface modelling system. *Atmosphere* 12, 723. doi:10.3390/atmos12060723.
- Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* 298, 242–266. doi:10.1016/j.jhydrol.2004.03.042.
- Collins, E.L., David, C.H., Riggs, R., Allen, G.H., Pavelsky, T.M., Lin, P., Pan, M., Yamazaki, D., Meentemeyer, R.K., Sanchez, G.M., 2024. Global patterns in river water storage dependent on residence time. *Nature Geoscience* 17, 433–439. doi:10.1038/s41561-024-01421-5.
- Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., Clark, E., Cabell, R., Casiday, N., Cui, Z., Eicher, K., Fall, G., Feng, X., Fitzgerald, K., Frazier, N., George, C., Gibbs, R., Hernandez, L., Johnson, D., Jones, R., Karsten, L., Kefelegn, H., Kitzmiller, D., Lee, H., Liu, Y., Mashriqui, H., Mattern, D., McCluskey, A., McCreight, J.L., McDaniel, R., Midekisa, A., Newman, A., Pan, L., Pham, C., RafieeiNasab, A., Rasmussen, R., Read, L., Rezaeianzadeh, M., Salas, F., Sang, D., Sampson, K., Schneider, T., Shi, Q., Sood, G., Wood, A., Wu, W., Yates, D., Yu, W., Zhang, Y., 2024. NOAA’s National Water Model: Advancing operational hydrology through continental-scale modeling. *Journal of the American Water Resources Association* 60, 247–272. doi:10.1111/1752-1688.13184.
- David, C.H., Maidment, D.R., Niu, G.Y., Yang, Z.L., Habets, F., Eijkhout, V., 2011. River network routing

- on the NHDPlus dataset. *Journal of Hydrometeorology* 12, 913–934. doi:10.1175/2011JHM1345.1.
- Feng, D., Gleason, C.J., Lin, P., Yang, X., Pan, M., Ishitsuka, Y., 2021. Recent changes to Arctic river discharge. *Nature Communications* 12, 6917. doi:10.1038/s41467-021-27228-1.
- Fry, L.M., Gronewold, A.D., Fortin, V., Buan, S., Clites, A.H., Luukkonen, C., Holtschlag, D., Diamond, L., Hunter, T., Seglenieks, F., Durnford, D., Dimitrijevic, M., Subich, C., Klyszejko, E., Kea, K., Restrepo, P., 2014. The Great Lakes Runoff Intercomparison Project Phase 1: Lake Michigan (GRIP-M). *Journal of Hydrology* 519, 3448–3465. doi:10.1016/j.jhydrol.2014.07.021.
- Gou, J., Miao, C., Samaniego, L., Xiao, M., Wu, J., Guo, X., 2021. CNRD v1.0: a high-quality natural runoff dataset for hydrological and climate studies in China. *Bulletin of the American Meteorological Society* 102, E929–E947. doi:10.1175/BAMS-D-20-0094.1.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* 377, 80–91. doi:10.1016/j.jhydrol.2009.08.003.
- Haddeland, I., Matheussen, B.V., Lettenmaier, D.P., 2002. Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model. *Water Resources Research* 38, 29–1–29–10. doi:10.1029/2001WR000854.
- Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., Pappenberger, F., 2020. GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth System Science Data* 12, 2043–2060. doi:10.5194/essd-12-2043-2020.
- He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X., 2020. The first high-resolution meteorological forcing dataset for land process studies over China. *Scientific Data* 7, 25. doi:10.1038/s41597-020-0369-y.
- Henderson-Sellers, A., Yang, Z.L., Dickinson, R.E., 1993. The Project for Intercomparison of Land-surface Parameterization Schemes. *Bulletin of the American Meteorological Society* 74, 1335–1349. doi:10.1175/1520-0477(1993)074<1335:TPFIOL>2.0.CO;2.
- Henn, B., Newman, A.J., Livneh, B., Daly, C., Lundquist, J.D., 2018. An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology* 556, 1205–1219. doi:10.1016/j.jhydrol.2017.03.008.
- Hersbach, H., Bell, B., Berriesford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellán, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G.D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani,

- R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P.d., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146, 1999–2049. doi:10.1002/qj.3803.
- Hirpa, F.A., Salamon, P., Beck, H.E., Lorini, V., Alfieri, L., Zsoter, E., Dadson, S.J., 2018. Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology* 566, 595–606. doi:10.1016/j.jhydrol.2018.09.052.
- van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S.I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M.J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D.M., Weedon, G.P., Ellis, R., Hagemann, S., Mao, J., Flanner, M.G., Zampieri, M., Materia, S., Law, R.M., Sheffield, J., 2016. LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome. *Geoscientific Model Development* 9, 2809–2832. doi:10.5194/gmd-9-2809-2016.
- Knijff, J.M.V.D., Younis, J., Roo, A.P.J.D., 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science* 24, 189–212. doi:10.1080/13658810802549154.
- Li, J., Chen, F., Lu, X., Gong, W., Zhang, G., Gan, Y., 2020. Quantifying contributions of uncertainties in physical parameterization schemes and model parameters to overall errors in noah-mp dynamic vegetation modeling. *Journal of Advances in Modeling Earth Systems* 12, e2019MS001914. doi:10.1029/2019MS001914.
- Li, L., Bisht, G., Leung, L.R., 2022. Spatial heterogeneity effects on land surface modeling of water and energy partitioning. *Geoscientific Model Development* 15, 5489–5510. doi:10.5194/gmd-15-5489-2022.
- Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres* 99, 14415–14428. doi:10.1029/94JD00483.
- Lin, P., Pan, M., Beck, H.E., Yang, Y., Yamazaki, D., Frasson, R., David, C.H., Durand, M., Pavelsky, T.M., Allen, G.H., Gleason, C.J., Wood, E.F., 2019. Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research* 55, 6499–6516. doi:10.1029/2019WR025287.
- Lin, P., Pan, M., Wood, E.F., Yamazaki, D., Allen, G.H., 2021. A new vector-based global river network dataset accounting for variable drainage density. *Scientific Data* 8, 28. doi:10.1038/s41597-021-00819-9.

- Liu, M., Raymond, P.A., Lauerwald, R., Zhang, Q., Trapp-Müller, G., Davis, K.L., Moosdorf, N., Xiao, C., Middelburg, J.J., Bouwman, A.F., Beusen, A.H.W., Peng, C., Lacroix, F., Tian, H., Wang, J., Li, M., Zhu, Q., Cohen, S., van Hoek, W.J., Li, Y., Li, Y., Yao, Y., Regnier, P., 2024. Global riverine land-to-ocean carbon export constrained by observations and multi-model assessment. *Nature Geoscience* 17, 896–904. doi:10.1038/s41561-024-01524-z.
- Mai, J., Shen, H., Tolson, B.A., Gaborit, E., Arsenault, R., Craig, J.R., Fortin, V., Fry, L.M., Gauch, M., Klotz, D., Kratzert, F., O'Brien, N., Princz, D.G., Rasiya Koya, S., Roy, T., Seglenieks, F., Shrestha, N.K., Temgoua, A.G.T., Vionnet, V., Waddell, J.W., 2022. The Great Lakes Runoff Intercomparison Project Phase 4: the Great Lakes (GRIP-GL). *Hydrology and Earth System Sciences* 26, 3537–3572. doi:10.5194/hess-26-3537-2022.
- Miao, C., Gou, J., Fu, B., Tang, Q., Duan, Q., Chen, Z., Lei, H., Chen, J., Guo, J., Borthwick, A.G.L., Ding, W., Duan, X., Li, Y., Kong, D., Guo, X., Wu, J., 2022. High-quality reconstruction of China's natural streamflow. *Science Bulletin* 67, 547–556. doi:10.1016/j.scib.2021.09.022.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data* 13, 4349–4383. doi:10.5194/essd-13-4349-2021.
- Najafi, H., Shrestha, P.K., Rakovec, O., Apel, H., Vorogushyn, S., Kumar, R., Thober, S., Merz, B., Samaniego, L., 2024. High-resolution impact-based early warning system for riverine flooding. *Nature Communications* 15, 3726. doi:10.1038/s41467-024-48065-y.
- Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres* 116, D12109. doi:10.1029/2010JD015139.
- Coughlan de Perez, E., van den Hurk, B., van Aalst, M.K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., Zsoter, E., 2016. Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences* 20, 3549–3560. doi:10.5194/hess-20-3549-2016.
- Read, L.K., Yates, D.N., McCreight, J.M., Rafieeinab, A., Sampson, K., Gochis, D.J., 2023. Development and evaluation of the channel routing model and parameters within the National Water Model. *JAWRA*

- Journal of the American Water Resources Association 59, 1051–1066. doi:10.1111/1752-1688.13134.
- Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.J., DMIP Participants, 2004. Overall Distributed Model Intercomparison Project results. *Journal of Hydrology* 298, 27–60. doi:10.1016/j.jhydrol.2004.03.031.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research* 46, W05521. doi:10.1029/2009WR008328.
- Samaniego, L., Bárdossy, A., Kumar, R., 2010. Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. *Water Resources Research* 46, W02506. doi:10.1029/2008WR007695.
- Saxe, S., Farmer, W., Driscoll, J., Hogue, T.S., 2021. Implications of model selection: a comparison of publicly available, conterminous US-extent hydrologic component estimates. *Hydrology and Earth System Sciences* 25, 1529–1568. doi:10.5194/hess-25-1529-2021.
- Smith, M.B., Seo, D.J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., Cong, S., 2004. The Distributed Model Intercomparison Project (DMIP): motivation and experiment design. *Journal of Hydrology* 298, 4–26. doi:10.1016/j.jhydrol.2004.03.040.
- Sun, H., Su, F., Yao, T., He, Z., Tang, G., Huang, J., Zheng, B., Meng, F., Ou, T., Chen, D., 2021. General overestimation of ERA5 precipitation in flow simulations for High Mountain Asia basins. *Environmental Research Communications* 3, 121003. doi:10.1088/2515-7620/ac40f0.
- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.L., 2018. A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Reviews of Geophysics* 56, 79–107. doi:10.1002/2017RG000574.
- Tijerina, D., Condon, L., FitzGerald, K., Dugger, A., O'Neill, M.M., Sampson, K., Gochis, D., Maxwell, R., 2021. Continental hydrologic intercomparison project, phase 1: A large-scale hydrologic model comparison over the continental united states. *Water Resources Research* 57, e2020WR028931. doi:10.1029/2020WR028931.
- Wang, X., Zhou, J., Ma, J., Luo, P., Fu, X., Feng, X., Zhang, X., Jia, Z., Wang, X., Huang, X., 2024. Evaluation and comparison of reanalysis data for runoff simulation in the data-scarce watersheds of alpine regions. *Remote Sensing* 16, 751. doi:10.3390/rs16050751.
- Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P.D., Allen, G.H., Pavelsky, T.M., 2019. MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research* 55, 5053–5073. doi:10.1029/2019wr024873.

- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S., Bates, P.D., 2017. A high-accuracy map of global terrain elevations. *Geophysical Research Letters* 44, 5844–5853. doi:10.1002/2017GL072874.
- Yang, Y., Pan, M., Beck, H.E., Fisher, C.K., Beighley, R.E., Kao, S.C., Hong, Y., Wood, E.F., 2019. In quest of calibration density and consistency in hydrologic modeling: distributed parameter calibration against streamflow characteristics. *Water Resources Research* 55, 7784–7803. doi:10.1029/2018wr024178.
- Yang, Y., Pan, M., Lin, P., Beck, H.E., Zeng, Z., Yamazaki, D., David, C.H., Lu, H., Yang, K., Hong, Y., Wood, E.F., 2021. Global reach-level 3-hourly river flood reanalysis (1980–2019). *Bulletin of the American Meteorological Society* 102, E2086–E2105. doi:10.1175/BAMS-D-20-0057.1.
- Zheng, H., Fei, W., Yang, Z.L., Wei, J., Zhao, L., Li, L., Wang, S., 2023. An ensemble of 48 physically perturbed model estimates of the 1/8° terrestrial water budget over the conterminous United States, 1980–2015. *Earth System Science Data* 15, 2755–2780. doi:10.5194/essd-15-2755-2023.
- Zheng, H., Yang, Z.L., Lin, P., Wei, J., Wu, W.Y., Li, L., Zhao, L., Wang, S., 2019. On the sensitivity of the precipitation partitioning into evapotranspiration and runoff in land surface parameterizations. *Water Resources Research* 55, 95–111. doi:10.1029/2017WR022236.