https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

Comparative Evaluation of ERA5-Land and ISIMIP3 Runoff Forcing for Global River Streamflow Simulation

Julien E. S. Boulange[1], Fang Zhao[2], Simon N. Gosling[3], Yadu Pokhrel[4], Yamazaki Dai[5], Xudong Zhou[6]

[1] United graduate school of agricultural science, Tokyo University of Agriculture and Technology, Fuchu, Tokyo, Japan

[2] Key Laboratory of Geographic Information Science of the Ministry of Education, School of Geographic Sciences, East China Normal University, Shanghai 200241, China.

[3] School of Geography, University of Nottingham, UK

[4] Department of Civil and Environmental Engineering, Michigan State University, USA

[5] Global Hydrological Prediction Center, Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

[6] Institute of Hydraulic and Ocean Engineering, Ningbo University, Ningbo, China

Correspondence to: Julien E. S. Boulange (boulange@go.tuat.ac.jp) and Fang Zhao (fzhao87@foxmail.com)

Abstract. Flooding is among the most widespread natural hazards worldwide, yet many high-risk regions lack the observational data needed for effective flood planning. In these data-sparse regions, global flood models remain essential tools for estimating flood hazard, although their performance is strongly influenced by the choice of runoff forcing data. Two widely used global runoff products are the reanalysis-based ERA5-Land dataset and the ISIMIP3a multi-model hydrological ensemble. Their selection involves an inherent trade-off between high-resolution reanalysis runoff and runoff simulated by hydrological models driven by bias-corrected meteorological inputs, the latter also providing an explicit representation of uncertainty through ensemble spread. This study presents a comparative evaluation of these two products by routing both through a consistent global hydrodynamic framework (CaMa-Flood). Model performance was assessed across IPCC SREX regions against observations from 5,071 gauging stations using the Kling-Gupta Efficiency and its components, while long-term trends in low, mean, and high streamflow were evaluated from a subset of 3,135 stations with sufficient temporal coverage. Simulations forced by ERA5-Land show superior skill in reproducing observed daily streamflow, with consistently higher correlation and stronger agreement in the spatial pattern of regional streamflow trends. However, systematic biases in streamflow magnitude and a tendency to exaggerate drying trends, particularly for low streamflow, are also evident. In contrast, the ISIMIP3a ensemble shows lower skill in reproducing observed daily streamflow metrics but provides more conservative and observation-consistent estimates of long-term trends. Ensemble averaging further improves robustness, with simulated trend ranges more frequently overlapping observational uncertainty bounds, albeit at the expense of dampened variability and extremes. Differences between native and spatially aggregated ERA5-Land runoff were negligible within the present modelling framework. Overall, the results demonstrate that no single runoff product is

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

universally optimum: ERA5-Land is well suited for reproducing historical streamflow dynamics, whereas ISIMIP3a is particularly valuable for robust assessments of long-term hydrological change and uncertainty.

1 Introduction

Flooding is among the most widespread natural hazards, with far-reaching societal impacts. Between 2002 and 2022, global economic losses from flooding were estimated at $223 billion (Samadi et al., 2025). The burden of flood exposure falls disproportionately on low-income countries, with approximately 1.61 billion people directly exposed to 1-in-100 flood events (Rentschler et al., 2022). While flood-related mortality has declined since 1975, largely due to advancement in flood risk management (Jonkman et al., 2024), flood hazard is intensifying, driven by climate change and population growth (Rogers et al., 2025). Observed change in hydrological extremes are consistent with theoretical expectations regarding the influence of global warming on the hydrological cycle (Blöschl et al., 2019), and have been attributed to anthropogenic climate change (Gudmundsson et al., 2021). When considering climate change in isolation, projections suggest that global exposure to 1-in-100 floods could increase by a factor of 4 to 14 by the end of the century, depending on emission scenarios (Hirabayashi et al., 2013). More recent estimates that incorporate both climate change and demographic trends suggest that by 2100, an average of 1.9 billion people could be exposed to such flood events, with population growth accounting for the majority of uncertainty in projections (Rogers et al., 2025).

In response to the escalating climate-related risks, including flooding, most nations have ratified the Paris Climate Agreement, committing to limit global warming to well below +2 °C, and preferably to no more than +1.5 °C, above pre-industrial levels (1850–1900). However, current national climate pledges put the world on a trajectory toward approximately +2.7 °C of warming above pre-industrial level by 2100, highlighting the urgent need for ambitious and targeted measures to reduce future flood risk. As the world enters an era of unprecedented flood hazard, high-resolution flood modelling that incorporates the latest historical and future conditions is critical for supporting evidence-based flood risk management (Shamsudduha, 2025).

Global flood models (GFMs) play a key role in this effort by enabling the spatially and temporally consistent assessment of flood hazard and associated impacts (Bain et al., 2023). However, GFMs are subject to inherent limitations and approximations, including uncertainties stemming from cascading physical processes, that constrain their accuracy and may compromise their practical utility for decision-making (Trigg et al., 2016). A fundamental component of reliable flood modelling is the representation of Earth's terrain via digital elevation models (DEMs). Global DEMs suffer from systematic biases due to vegetation, built infrastructure, and sensor limitations, prompting the development of various post-processed products to mitigate these errors (Uhe et al., 2025; Yamazaki et al., 2019). In addition, the choice of runoff routing schemes has been shown to strongly modulate the magnitude and timing of hydrological extremes (Zhao et al., 2017). Among the various sources of uncertainty in GFMs, the selection of runoff input product is a dominant factor, accounting for up to 80% of total uncertainty in flood depth estimates, surpassing even methodological differences in defining flood events of specific

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

return periods (Zhou et al., 2021). This finding is corroborated by other independent studies, which report substantial discrepancies in simulating historical flood extents across different runoff forcing datasets (Mester et al., 2021).

Recognizing the critical role of runoff in flood modelling, Zhou et al. (2021) suggested that runoff datasets that are bias-corrected against streamflow may significantly improve the ability of GFMs to reproduce historical streamflow and enhance their overall reliability. However, such products remain limited and only few studies have used bias-corrected runoff data (Dang and Pokhrel, 2024; Wang et al., 2025). In practice, many studies instead rely on runoff simulated by global hydrological models driven by bias-corrected climate inputs (Boulange et al., 2021; Gu et al., 2020; Zhao et al., 2017), such as those developed by the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), where bias correction is applied to meteorological inputs rather than directly to runoff. Consequently, users are often faced with selecting between runoff products that differ substantially in their underlying assumptions and characteristics. Reanalysis-based datasets such as ERA5-Land provide comparatively high spatial resolution and near-real-time coverage, which may better capture fine-scale hydrological patterns (Muñoz-Sabater et al., 2021). In contrast, ensemble hydrological products, such as ISIMIP3, employ bias-adjusted meteorological forcing and multi-model hydrological ensembles designed to reduce systematic forcing biases while preserving large-scale climatic trends and representing structural uncertainty (Hempel et al., 2013; Lange, 2021). Despite the widespread use of both datasets, their relative strengths and limitations for streamflow simulation have not yet been systematically compared at the global scale.

This study therefore presents a comparative evaluation of river streamflow simulations derived from ERA5-Land and ISIMIP3 runoff products. We assess their ability to reproduce observed streamflow dynamics and regime characteristics across diverse hydroclimatic regions, with particular emphasis on the implications of differing dataset structures, forcing strategies, and spatial resolution for large-scale hydrological applications.

2 Materials and Methods

2.1 Flood simulations

2.1.1 Forcing data

The ERA5-Land dataset, developed by the Copernicus Climate Data and Services (https://cds.climate.copernicus.eu/), is a state-of-the-art global reanalysis product providing hourly data at 0.1° spatial resolution (approximately 9 km) from 1950 to near present (Muñoz-Sabater et al., 2021). ERA5-Land is produced by replaying the land component of the ECMWF forecasting system using atmospheric forcing from the ERA5 reanalysis, enabling a higher-resolution and temporally consistent representation of terrestrial water and energy fluxes than the parent ERA5 product. Runoff is generated by the HTESSEL land-surface model, in which surface runoff is represented through an ARNO-type variable infiltration capacity scheme and subsurface drainage is parametrized following Darcy's law (Balsamo et al., 2008). As HTESSEL was developed primarily for numerical weather prediction, ERA5-Land runoff should be viewed as a physically based by-product of land-

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

surface reanalysis rather than the explicit output of a dedicated hydrological model. Furthermore, ERA5-Land does not explicitly enforce consistency between precipitation and evaporation fluxes, which are simulated independently and are not directly constrained by in situ observations (Wang et al., 2026).

Previous studies have shown that ERA5-Land generally reproduces large-scale runoff dynamics reasonably well, although performance degrades in arid, mountainous, and high-latitude regions where precipitation biases and snow-related processes remain challenging (Chaudhari and Pokhrel, 2022; Dutta and Markonis, 2024; Liu et al., 2024). Despite these limitations, ERA5-Land is widely used as a reference runoff dataset in large-scale hydrological applications (Hamitouche et al., 2025). Additional background on previous evaluations of ERA5-Land performance is provided in Sect. S1.

ISIMIP3a constitutes the historical simulation protocol of ISIMIP Phase 3 (www.isimip.org/protocol/#isimip3a), designed to support model evaluation and attribution studies using observation-driven, bias-adjusted meteorological forcing at 0.5° × 0.5° spatial resolution (Frieler et al., 2024; Lange, 2019, 2021). Consequently, the daily forcing fields are temporally consistent with observed historical climate variability. We employ ISIMIP3a hydrological simulations forced by the first priority observational dataset GSWP3 W5E5, which merges W5E5 v2.0 for 1979–2019 (Cucchi et al., 2020; Lange, 2021) with GSWP3 v1.09 for 1901–1978 (Kim, 2017). To avoid potential inhomogeneities arising from the transition between GSWP3 and W5E5, our analysis is restricted to 1979–2019. W5E5 v2.0 is based on ERA5 reanalysis and bias adjusted using the WATCH Forcing Data methodology to align near surface meteorological variables with observation based climatology from CRU TS v4.04 and GPCP v2020 (Lange, 2021). Accordingly, ISIMIP3a simulations are driven by bias corrected precipitation, whereas ERA5-Land runoff simulations rely on uncorrected ERA5 precipitation. This distinction in forcing construction is likely to influence the propagation of precipitation-related errors into simulated runoff and represents one of several important distinctions between ISIMIP3a and ERA5-Land.

We use daily runoff from nine ISIMIP3a global hydrological models (Frieler et al., 2024): CLASSIC, CWatM, H08, HydroPy, JULES-W2, MIROC-INTEG-LAND, ORCHIDEE-MICT, WaterGAP2-2e, and WEB DHM-SG (see Table S1 for model references). Among these, WaterGAP2-2e is calibrated against observed streamflow (Müller Schmied et al., 2024), and H08 applies climate zone specific parameter optimization to enhance its ability to reproduce observed streamflow (Yoshida et al., 2022). All simulations are drawn from the "obsclim + histsoc" experiment, which represents observed climate forcing combined with time varying direct human influences such as water abstraction, dams and reservoirs, following the ISIMIP3a protocol (Frieler et al., 2024). Detailed descriptions of model structures, parameterizations, representation of human influences, and eventual calibration strategies are provided in Telteu et al. (2021) and Müller Schmied et al. (2025).

2.1.2. Global flood simulation

The Catchment-based Macro-scale Floodplain model (CaMa-Flood) is a large-scale river hydrodynamics model that solves a local inertial form of the shallow water equations to simulate river streamflow, water surface elevation, and flood inundation (Yamazaki, 2025; Yamazaki et al., 2011, 2013). Simulations are performed along a river network derived from MERIT

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

Hydro (Yamazaki et al., 2019). A defining feature is the unit catchment framework, in which each grid cell represents a catchment that aggregates channel and floodplain topography, with a uniform water level assumed within the cell. This structure allows representation of backwater effects and floodplain storage while maintaining computational efficiency at continental scale. The model has undergone continuous development, including representations of river bifurcation (Yamazaki et al., 2014), dynamic sea level boundary conditions (Eilander et al., 2020; Ikeuchi et al., 2015, 2017), dam operation (Hanazaki et al., 2022), and most recently levees (Zhao et al., 2025).

In this study, we use CaMa Flood v4.30 at 0.1° resolution to increase river network density and enable evaluation at more discharge stations compared to earlier 0.25° applications. Default parameter settings are applied, and the dam module is enabled. A standard spin up procedure is conducted by repeating the first year of runoff forcing five times to stabilize channel and floodplain storage. The model is driven by daily runoff from ERA5 Land (0.1°) and ISIMIP3a global hydrological models (0.5°) respectively, both conservatively remapped to the native 0.1°-unit catchment network.

2.2 Evaluation of simulations

The Catchment-based Macro-scale Floodplain model (CaMa-Flood) is a large-scale river hydrodynamics model that solves a local inertial form of the shallow water equations to simulate river streamflow, water surface elevation, and flood inundation (Yamazaki, 2025; Yamazaki et al., 2011, 2013). Simulations are performed along a river network derived from MERIT Hydro (Yamazaki et al., 2019). A defining feature is the unit catchment framework, in which each grid cell represents a catchment that aggregates channel and floodplain topography, with a uniform water level assumed within the cell. This structure allows representation of backwater effects and floodplain storage while maintaining computational efficiency at continental scale. The model has undergone continuous development, including representations of river bifurcation (Yamazaki et al., 2014), dynamic sea level boundary conditions (Eilander et al., 2020; Ikeuchi et al., 2015, 2017), dam operation (Hanazaki et al., 2022), and most recently levees (Zhao et al., 2025).

In this study, we use CaMa Flood v4.30 at 0.1° resolution to increase river network density and enable evaluation at more discharge stations compared to earlier 0.25° applications. Default parameter settings are applied, and the dam module is enabled. A standard spin up procedure is conducted by repeating the first year of runoff forcing five times to stabilize channel and floodplain storage. The model is driven by daily runoff from ERA5 Land (0.1°) and ISIMIP3a global hydrological models (0.5°) respectively, both conservatively remapped to the native 0.1°-unit catchment network.

2.2 Evaluation of simulations

2.2.1 Streamflow observations

Streamflow observations were obtained from the Global Runoff Data Centre (GRDC), operating under the auspices of the World Meteorological Organization. GRDC provides one of the most comprehensive collections of quality-controlled river discharge records worldwide, supporting hydrological research at the global scale. The full database, accessed on 14 March

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

2025, contains data for 10,836 gauging stations. The temporal coverage varies by station with records beginning as early as 1806 and extending through 2024.

Stations were subsequently matched to the MERIT Hydro river network (Yamazaki et al., 2019) following a multi-step filtering procedure (Zhou et al., 2025). First, we excluded stations with less than 5 years of available data (not necessary consecutive) during the period 1979–2019, as well as stations draining a catchment smaller than 360 km^2. The minimum data-length requirement ensures robust estimation of the Kling-Gupta Efficiency (*KGE*), used to evaluate the model's ability to reproduce observed daily streamflow dynamics, and the trends in annual low, mean, and high streamflow estimation. The minimum catchment threshold avoids including basins that are insufficiently resolved at the model's spatial resolution of 0.1° (approximately 11 km at the equator), thereby maintaining consistency between simulated and observed hydrological processes (Beck et al., 2017).

For the remaining eligible stations, an initial candidate grid-cell was identified from the reported coordinates, and neighboring cells were subsequently explored to determine the optimal allocation. The search radius was adapted according to the station reported drainage area, allowing larger rivers to be matched over a broader spatial window. For each candidate grid cell, an allocation score was computed by combining (i) the relative error in upstream drainage area and (ii) the great-circle distance between the reported station coordinates and the grid-cell center. The drainage-area mismatch was expressed as a relative difference and adjusted to ensure symmetric penalization of over- and under-estimation while the location mismatch was scaled as a function of catchment size to avoid excessive penalization for large basins. Stations allocated to grid cells with drainage areas substantially smaller than the reported value (<5%) were excluded from the analysis. After applying all temporal, drainage-area, and allocation-quality filters, 5,099 stations remained for analysis (Fig. S1).

2.2.2 Performance assessment of simulated streamflow

The adequacy of the simulated daily streamflow was evaluated using the Kling-Gupta Efficiency (*KGE*), which was introduced as an improvement over the Nash and Sutcliffe efficiency (*NSE*). The *NSE* has been criticized for its sensitivity to high flows and squared-error structure which can obscure systematic biases in variability and mean flow (Gupta et al., 2009; Mizukami et al., 2019). The *KGE* formulation adopted here (Eq. 1) follows Gupta et al., (2009), with the revised interpretation proposed by Kling et al., (2012):

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

where r is the Pearson correlation coefficient between observed and simulated daily streamflow, α is the variability ratio defined as the standard deviation of simulations divided by that of observations, and β is the bias ratio defined as the mean simulated streamflow divided by the mean observed streamflow.

KGE can be interpreted as the Euclidean distance in a three-dimensional space by correlation, variability, and bias components, with an optimal value of 1 corresponding to perfect agreement ($r = \alpha = \beta = 1$). By explicitly decomposing model performance into these three components, *KGE* provides a more balanced assessment of model skill across different flow regimes than *NSE*. Furthermore, the contribution of correlation, variability, and bias to a low *KGE* value can be directly

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

diagnosed. Because some allocated stations did not provide sufficient overlapping daily observations during the evaluation period, *KGE* could be computed for 5,071 stations.

2.2.3 Trend estimation

Long-term trends in simulated and observed annual streamflow metrics were estimated using the Theil-Sen slope estimator (Sen, 1968), a robust non-parametric method that is insensitive to outliers and does not assume normally distributed residuals (Gudmundsson et al., 2019). For all grid cells and allocated stations, daily streamflow was first aggregated to annual statistics, including the annual mean streamflow as well as the 10th and 90th percentiles (Q10 and Q90), representing low- and high-streamflow conditions, respectively. Following established practice, a valid year was defined as a year with at least 350 valid observations, and a minimum of 28 valid years was required for slope estimation (Gudmundsson et al., 2021), resulting in a total of 3,135 stations retained for the trend analysis.

The Theil-Sen slope was computed as the median of all pairwise slopes between annual values and expressed as percentage change per decade (Eq. 2):

$$trend\ (\%\ \mathrm{decade}^{-1}) = \frac{b \times 10}{\overline{Q}} \times 100 \quad (2)$$

where b is the Theil-Sen slope (in m^3 year^{-1}) and $\bar{Q}$is the long-term mean streamflow (m^3 year^{-1}) of the considered metric (low-, mean- and high-streamflow).

2.2.4 Regional assessment

The regional assessment is based on the 26 core SREX regions defined by the Intergovernmental Panel on Climate Change (IPCC). Station-level metrics were aggregated within each region to characterize large-scale performance and trends (Fig. S1). Regional medians were computed for ERA5-Land and GRDC observations, while multi-model means were calculated for the ISIMIP simulations. Following the literature, only regions with at least 80 stations were retained to ensure statistical robustness (Gudmundsson et al., 2021). Note that because the number of stations where *KGE* could be computed and slope was different, the availability of regions across the two assessments is affected.

Uncertainty in regional flow trend estimates was assessed using non-parametric bootstrap resampling at the station level. Within each SREX region, station-specific slope estimates were resampled with replacement, preserving the original sample size, and the regional median was recalculated for each replicate. Separate bootstrap distributions were generated for low-, mean-, and high-streamflow trends. The resulting bootstrap distributions were used to derive central estimates (median) and uncertainty bounds (2.5th and 97.5th percentiles), corresponding to a 95% uncertainty interval. A larger number of bootstrap replicates was used for the ERA5-based simulation and observations (B = 9000) compared to the ISIMIP3 ensemble (B = 1000), as the latter already incorporates variability across nine independent hydrological models, allowing robust uncertainty estimation at lower computational cost.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

3 Results

3.1 Model Performance Assessment

3.1.1 Spatial Assessment of Regional Performance

The median *KGE* across valid SREX regions for the ERA5-Land and ISIMIP3a forcings is shown in Fig. 1. ERA5-Land generally exhibits acceptable to good performance, with only two regions showing slightly negative median *KGE* values (-0.04 in CAM and -0.08 in NEB). Two additional regions (SAF and WSA) display only marginally positive scores (0.05 and 0.08, respectively). In comparison, simulations based on the ISIMIP3a multi-model ensemble display lower overall skill (Fig. 1b). Eight regions have negative median *KGE* values, while the remaining regions consistently score below ERA5-Land. Nevertheless, substantial inter-model variability is evident within ISIMIP3a (Fig. S2), with pronounced regional differences in model performance and no single model outperforming others across all regions. In SAF and WSA, none of the ISIMIP3a models yield a median *KGE* above 0, while in SAU only two models achieve positive *KGE* values.

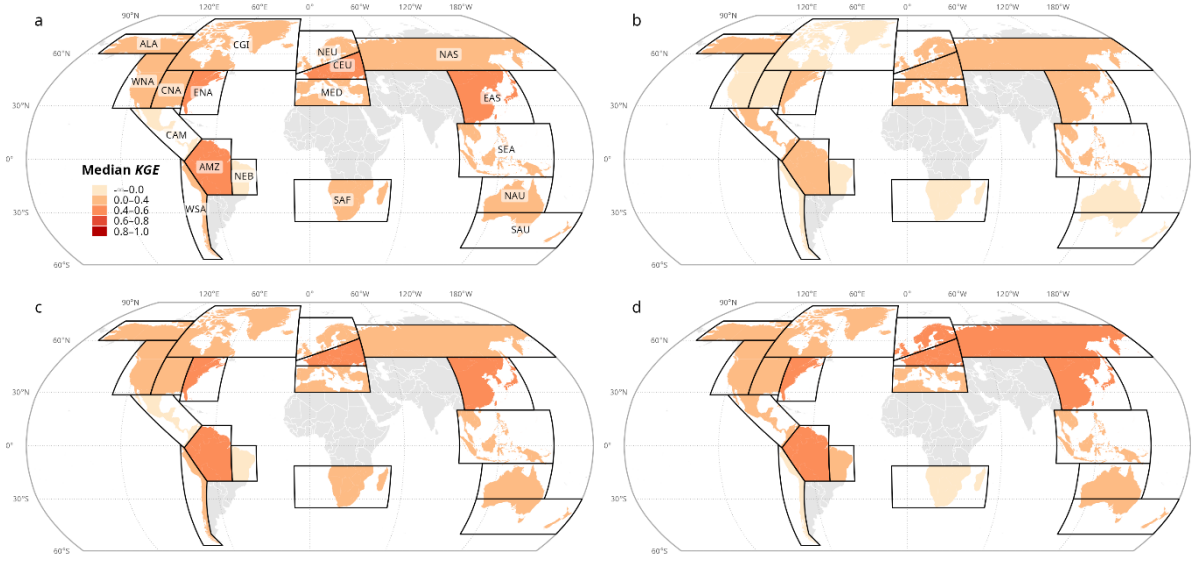


Figure 1: Median Kling–Gupta Efficiency (KGE) across IPCC SREX regions for simulations forced with (a) ERA5-Land, (b) ISIMIP3a ensemble mean, (c) ERA5-Land upscaled at 0.5° resolution and (d) the best-performing ISIMIP3a model selected independently for each region. Panel (b) represents the unweighted multi-model mean, whereas panel (d) illustrates the upper bound of attainable skill within the ISIMIP3a ensemble.

Selecting the best-performing ISIMIP3a model for each SREX region substantially improves performance. Under this optimal-model experiment, median *KGE* values become comparable to ERA5-Land in several regions, with AMZ and ENA regions exceeding 0.5, although substantial regional difference remain. An additional simulation was performed using ERA5-Land runoff conservatively remapped from its native 0.1° to a 0.5° grid prior to routing (hereinafter “upscaled ERA5-Land”). Despite the coarser forcing, simulated streamflow performance remained nearly identical to that obtained with the

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

native-resolution ERA5-Land forcing (Fig. 1c; Fig. S4), indicating limited sensitivity of simulated streamflow skill to this spatial aggregation within the present modelling framework.

3.1.2 *KGE* component analysis

Decomposition of the median *KGE* into correlation, bias, and variability components (Sect. 2.2.2) provides further insight into the contrasting behaviour of ERA5-Land and ISIMIP3a runoff forcings across SREX regions (Fig. 2). For ERA5-Land, correlation values exceeded 0.6 across all regions (Fig. 2a), indicating robust agreement in the day-to-day variability and timing of simulated streamflow. Bias patterns are more regionally dependent (Fig. 2d), with underestimation in several drier regions (e.g., MED, SAU, NAU) and overestimation in some wetter or mountainous regions (e.g., CAM, WSA, SEA). Variability ratios are close to unity in most regions, although NEB shows clear overestimation of daily streamflow variability, with weaker but similar tendencies in AMZ and WSA (Fig. 2g). For the ISIMIP3a ensemble median, correlation values are lower and spatially uniform than for ERA5-Land (Fig. 2b), indicating weaker reproduction of daily streamflow timing. Bias is the dominant source of error in several regions, with pronounced overestimation in CNA, NEB, and SAF, and more moderate positive bias in CEU, NAU, and SAU (Fig. 2e). In contrast, variability ratios of the ensemble mean remain comparatively close to unity in many regions (Fig. 2h), suggesting more balanced representation of streamflow variability despite lower temporal correspondence. The best-performing ISIMIP3a model selected independently for each SREX region substantially improves KGE components relative to the ensemble mean (Figs. 2c, f, i). In many regions, correlation approaches ERA5-Land levels while bias moves closer to unity, confirming that ensemble averaging can mask the skill of individual models and that model performance is strongly region dependent.

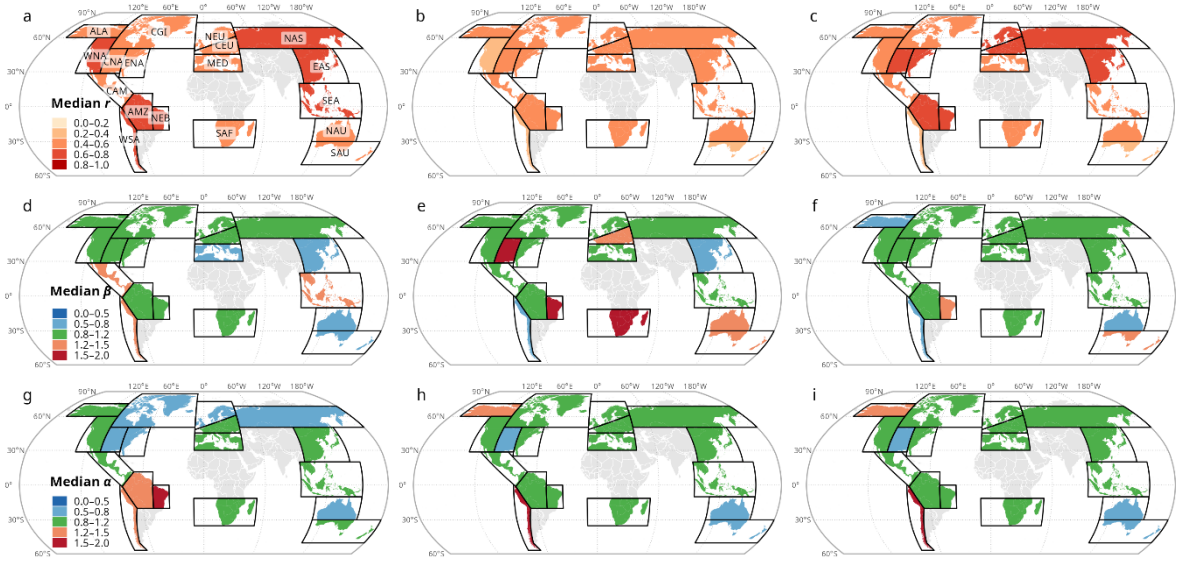


Figure 2: Decomposition of *KGE* across valid IPCC SREX regions for simulations forced with (a, d, g) ERA5-Land, (b, e, h) ISIMIP3a ensemble mean, and (c, f, i) the best-performing ISIMIP3a model selected independently for each region. The top row shows the correlation component (r), the middle row the bias component (β), and the bottom row the variability component (α) of KGE.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

3.2 Streamflow trend evaluation across datasets

3.2.1 Regional patterns of streamflow trends

The spatial patterns of observed annual trends in low, mean, and high streamflow across valid SREX regions are shown in Fig. 3a–c. Consistent with previous work (e.g., Gudmundsson et al., 2019)., the sign of change is generally consistent across the entire streamflow regime within individual regions. Wetting trends are evident in CNA, CGI, WNA, ENA, and SAF (0.62–2.36 % decade^{-1}), whereas drying trends dominate in NEB, MED, CEU, and SAU, with the strongest declines observed in NEB and SAU (-9.58 and -7.54 % decade^{-1}, respectively). Simulations forced by ERA5-Land broadly reproduce the observed spatial distribution of streamflow trends (Fig. 3d–f), with strong cross-region correlations (r=0.85–0.89), although leave-one-out sensitivity tests indicate moderate dependence on individual regions with pronounced hydroclimatic signals (Sect. S1, Table S2). However, ERA5-Land tends to amplify negative trends, particularly for low streamflow, indicating a systematic tendency toward stronger drying signals than observed (Fig. 4). Disagreements in trend direction are mainly confined to regions with weak observed trends. Upscaled ERA5-Land produces nearly identical trend patterns to the native-resolution forcing (Fig. S9), indicating limited sensitivity of simulated streamflow trends to runoff input resolution within the present modelling framework.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

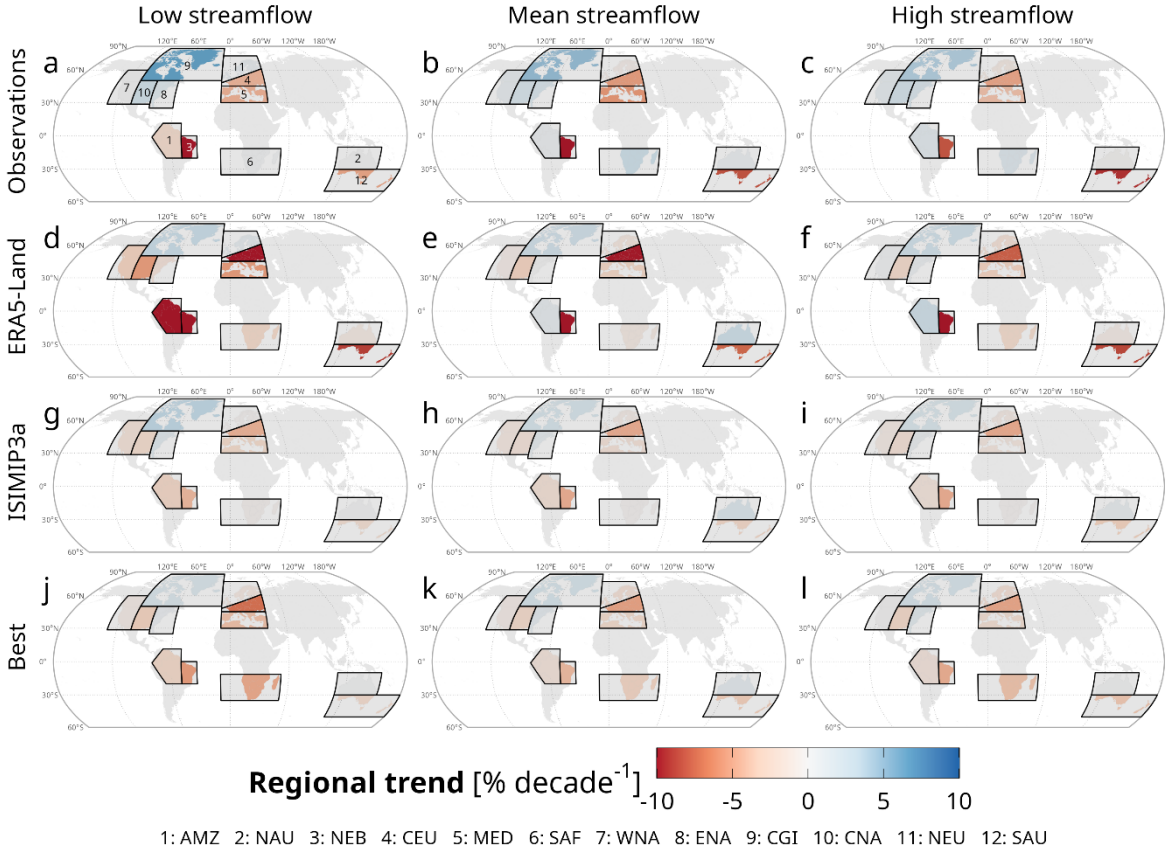
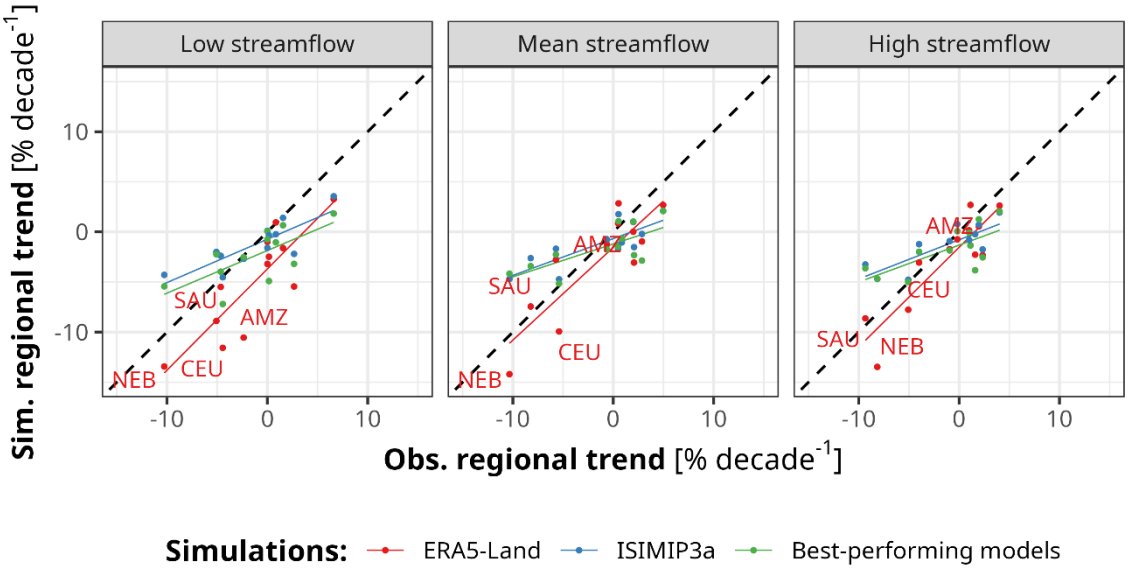


Figure 3: Regional trends in low, mean, and high streamflow expressed as percent change per decade across valid IPCC SREX region. Rows show results derived from observations, ERA5-Land, ISIMIP3a, and best-performing ISIMIP3a model ("Best") selected independently for each region (top to bottom).

285 The ISIMIP3a ensemble mean also captures the broad direction of regional change (Fig. 3g–i), with correlations of 0.80–0.82 across the streamflow spectrum. Compared to ERA5-Land, however, simulated trends are generally weaker in magnitude, indicating a more attenuated representation of long-term hydroclimatic change, while still showing sensitivity to individual regions in the leave-one-out analysis, particularly for low flows (Table S2). Selecting, the best performing ISIMIP3a model for each SREX region based on daily streamflow skill does not improve trend reproduction (Fig. 3j–l).
290 Correlations decline slightly (r=0.69–0.71), and trend magnitudes become more dispersed, suggesting that models performing well for daily dynamics do not necessarily provide superior long-term trend fidelity (Fig. 4). Overall, the results reveal contrasting behavior between the two runoff products: ERA5-Land better reproduces the spatial pattern of observed trends but tends to exaggerate drying signals, whereas ISIMIP3a provides more conservative trend estimates while preserving broad regional consistency.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026



295

Figure 4: Comparison of regional streamflow trends derived from observations and model-based datasets (ERA5-Land, ISIMIP3a, Best-performing model) for low, mean, and high flow regimes. Each point corresponds to an IPCC region. The dashed line represents perfect agreement (1:1), and solid lines indicate linear fits.

3.2.2 Uncertainty and consistency of regional streamflow trends

300 Regional median trends and correlation metrics alone do not fully capture the internal variability among stations within each SREX region. To address this, Fig. 5 presents bootstrap-derived uncertainty ranges for regional streamflow trends, enabling assessment of the robustness of simulated and observed changes. Observed regional trends are generally well constrained, with relatively narrow 95th confidence intervals that in most regions do not overlap zero, indicating coherent regional-scale changes rather than trends driven by a small subset of stations.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

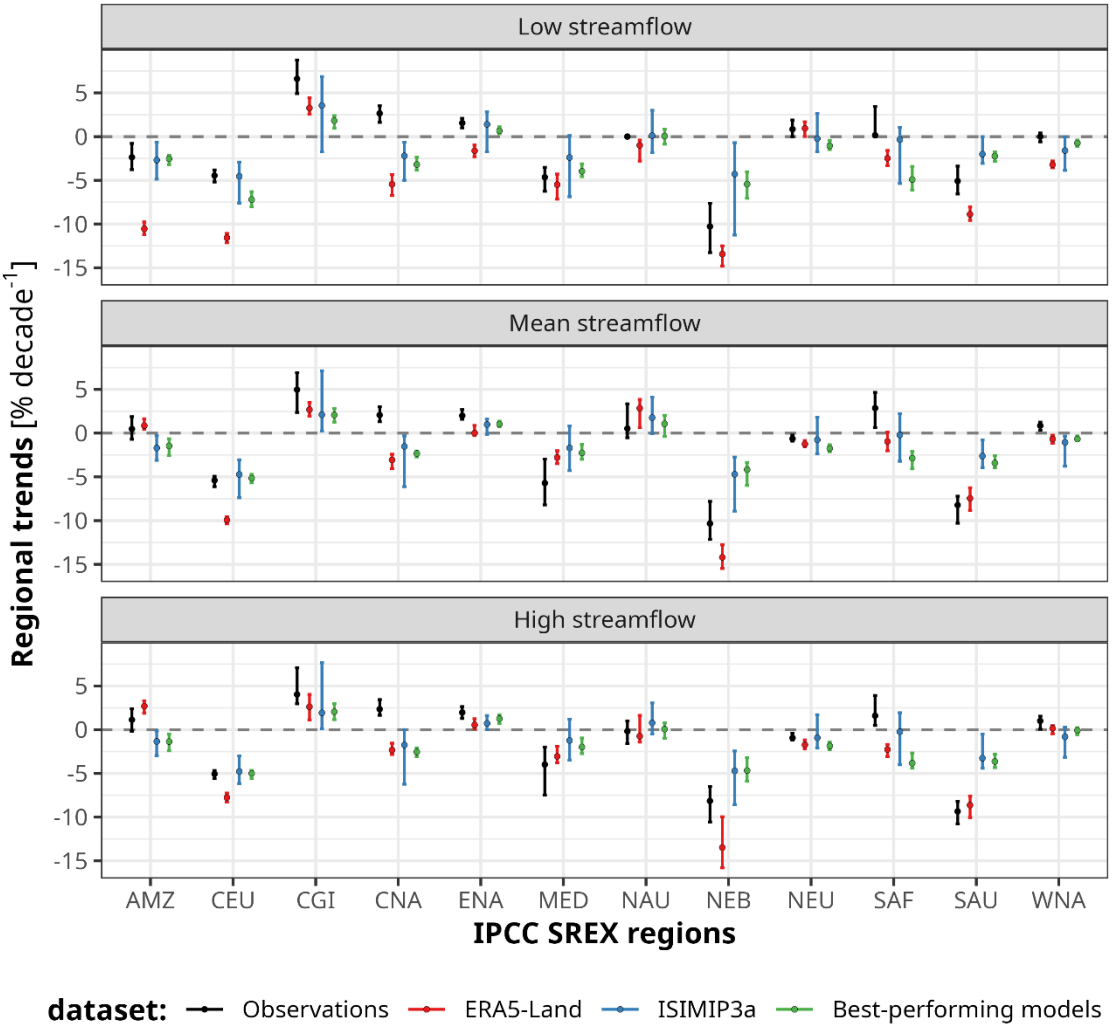


Figure 5: Uncertainty range of streamflow trends across IPCC SREX regions. Error bars indicate the lower and upper bounds of trend estimates across datasets, while points denote the central tendency. Results are shown separately for low, mean, and high streamflow metrics; the dashed grey line marks zero trend.

ERA5-Land reproduces many observed trend directions, but its uncertainty ranges are frequently shifted toward stronger negative values, consistent with the drying bias identified in Fig. 4. In several regions, simulated confidence intervals fall partly or entirely outside the observational range, indicating that systematic discrepancies that extend beyond median estimates. Upscaled ERA5-Land exhibits nearly identical behaviour. In contrast, the ISIMIP3a ensemble shows wider uncertainty ranges due to inter-model spread, yet its confidence intervals overlap the observational ranges in more regions and across all flow metrics (Table 1). Although ISIMIP3a generally produces weaker trend magnitudes, its regional estimates more frequently remain within observational uncertainty bounds. Selecting the best-performing ISIMIP3a hydrological model based on daily streamflow skill does not systematically improve trend uncertainty estimates, further

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

indicating that strong performance for daily streamflow simulation does not necessarily translate into improved reproduction on long-term trends. Overall, the two runoff products show contrasting characteristics: ERA5-Land provides narrower uncertainty ranges but systematically stronger drying trend estimates, whereas ISIMIP3a yields broader yet more observation-consistent ranges. This distinction is relevant for flood-risk applications, where both the realism and robustness of long-term hydrological trends are important.

Table 1: Percentage of IPCC SREX regions (n=12) for which simulated streamflow trends fall within the 95% confidence interval of observations. Higher values indicate greater consistency with observational uncertainty.

Dataset	Overlap (%)			
	Low	Mean	High	Overall
ERA5-Land	25	50	66.7†	47.2
Upscaled ERA5-Land	25	50	58.3†	44.4
ISIMIP3a	83.3	75	83.3	80.5
Best-performing models	41.7	41.7	50.0	44.5

4 Discussion

4.1 Contrasting streamflow behavior of ERA5-Land and ISIMIP-derived runoff

The results reveal clear and complementary differences between streamflow simulations forced by ERA5-Land and ISIMIP3a runoff. ERA5-Land generally provides superior reproduction of daily streamflow dynamics, with higher *KGE* values and stronger correlation across most SREX regions. This suggests that reanalysis-derived runoff can provide highly competitive forcing for large-scale river streamflow simulations. Although ERA5-Land runoff is generated by the HTESSEL land-surface model rather than by a dedicated hydrological model, its temporally coherent meteorological forcing and higher spatial resolution may help preserve runoff variability relevant for streamflow timing and event-scale fluctuations. In contrast, the lower overall skill of the ISIMIP3a ensemble appears strongly influenced by inter-model structural differences (Figs. S6–8). Considerable spread among individual ISIMIP models indicates that ensemble averaging combines both realistic runoff signals and model-specific errors, thereby reducing apparent skill in daily streamflow metrics. This interpretation is supported by the marked improvement obtained when the best-performing model is selected independently for each region, with several individual models achieving performance comparable to ERA5-Land. The comparison of *KGE* components further highlights distinct strengths of the two products. ERA5-Land more consistently reproduces the temporal timing of daily streamflow, whereas ISIMIP3a generally yields more balanced variability ratios and more less regionally erratic bias patterns. Nevertheless, previous studies have also reported persistent overestimation of runoff and streamflow in ISIMIP-based simulations for some regions (Hattermann et al., 2017; Veldkamp et al., 2018; Zaherpour et al., 2018), suggesting the application of precipitation bias correction alone does not eliminate runoff biases.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

The trend analysis reinforces these contrasting characteristics. ERA5-Land better reproduces the spatial pattern of observed regional wetting and drying signals, but frequently exaggerates drying trends, particularly for low flows. Pronounced drying signals exceeding -10 % decade^{-1} are simulated in CEU across all flow regimes and in AMZ for low flow. This tendency is consistent with recent concerns regarding long-term terrestrial water-cycle trends in ERA5-Land and may reflect residual imbalances in reanalysis forcing (Wang et al., 2026). The trend-bias analysis (Table 2) further shows that ERA5-Land and upscaled ERA5-Land exhibit negative biases relative to the observed trends in 77.8% of regions on average across streamflow regimes. Moreover, ERA5-Land exhibits changes in bias sign between low, mean, and high streamflow in 5 of 12 regions, suggesting that model errors are not uniform across the streamflow spectrum. In contrast, the ISIMIP3a ensemble provides weaker but more observation-consistent trend magnitudes, with confidence intervals overlapping observations in more regions. Ensemble averaging therefore appears to dampen extremes and long-term variability while improving robustness of regional trend estimates. Persistent low skill in arid and topographically complex regions such as SAF, WSA, and SAU further indicates that both runoff products continue to face challenges where streamflow is strongly controlled by intermittent rainfall, transmission losses, snowmelt timing, groundwater interactions, or human regulation which are notoriously difficult to represent in global-scale modelling systems (Heinicke et al., 2024; Zaherpour et al., 2018).

Table 2: Percentage of IPCC SREX regions (*n*=12) exhibiting negative bias across flow regimes, and percentage of regions where the sign of the bias varies across low, mean, and high streamflow. Higher values in the last column indicate lower consistency of model bias across flow regimes.

Dataset	Negative bias (%)				Inconsistent sign (%)
	Low	Mean	High	Overall	
ERA5-Land	91.7	66.7	75.0	77.8	41.7
Upscaled ERA5-Land	91.7	66.7	75.0	77.8	41.7
ISIMIP3a	66.7	58.3	50.0	58.3	16.7
Best-performing models	66.7	58.3	58.3	61.1	8.3

4.2 Implication for large-scale hydrology, limitation, and future work

The contrasting behavior of the two runoff products has practical implications for global hydrological and flood applications. Where the objective is to reproduce historical daily streamflow dynamics and flood-event responses, ERA5-Land provides strong standalone performance in the present experiments. Nevertheless, the skill achieved by several individual ISIMIP3a models and by the best-model experiment indicates that ensemble-based approaches using regional weighting or intelligent model selection may substantially narrow this gap (Pastén-Zapata et al., 2022; Zaherpour et al., 2019). Conversely, for assessing long-term regional change, quantifying structural uncertainty, or exploiting more explicit representations of runoff generation and terrestrial water balance, ISIMIP3a offers distinct advantages. The negligible differences between native and

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

upscaled ERA5-Land further indicate that runoff input resolution was of secondary importance within the present CaMa-Flood configuration. Because runoff forcing is conservatively remapped onto a 0.1° unit-catchment river network, runoff magnitude and temporal dynamics appear more influential than forcing resolution for reproducing daily streamflow and long-term streamflow trends. However, this result is likely model-dependent and should not be generalized to flood models operating directly at coarser spatial scales.

Several limitations should nevertheless be acknowledged. First, the comparison is conditioned on the CaMa-Flood routing framework, and alternative routing models may respond differently to runoff timing, magnitude, and spatial structure. Second, station density and record availability vary across SREX regions (Fig. S1), potentially affecting regional representativeness. Third, the study compares integrated runoff products whose differences arise simultaneously from forcing data, model formulation, and ensemble design, preventing strict causal attribution to any single factor. Future work could therefore evaluate these runoff products across multiple routing models, incorporating additional reanalysis and hydrological datasets, and develop hybrid ensemble approaches that retain the temporal realism of ERA5-Land while leveraging the uncertainty characterization and long-term robustness of ISIMIP-type multi-model systems.

5 Conclusion

This study provided a comparative evaluation of river streamflow simulations forced by two widely used runoff global runoff products, ERA5-Land and ISIMIP3a, within a consistent CaMa-Flood modelling framework. The results reveal contrasting but complementary strengths, with important implications for large-scale hydrological and flood applications. ERA5-Land demonstrates strong skill in reproducing daily streamflow dynamics, particularly the timing and short-term variability of streamflow, while also capturing the broad spatial pattern of observed regional trends. These characteristics make it a strong standalone runoff forcing dataset for applications focused on historical hydrograph behaviour and event-scale variability. However, ERA5-Land also exhibits systematic regional biases and a tendency to amplify drying trends, especially for low flows, indicating limitations for long-term hydroclimatic assessments.

In contrast, the ISIMIP3a ensemble mean generally shows lower skill for daily streamflow metrics, but provides more conservative and observation-consistent estimates of long-term streamflow trends. Its multi-model framework offers an explicit representation of structural uncertainty and benefits from bias-corrected climate forcing and dedicated hydrological model formulations. Moreover, the strong performance achieved by several individual ISIMIP3a models suggests that ensemble skill could be further improved through weighting or intelligent model-selection strategies. The limited differences between native and upscaled ERA5-Land indicate that, within the present CaMa-Flood configuration, runoff magnitude and temporal dynamics exert stronger controls on simulated streamflow than the difference between 0.1° and 0.5° forcing resolution. This suggests that improving runoff realism may be more beneficial than increasing forcing resolution alone, although this conclusion is likely model dependent.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

Overall, no single runoff product is universally optimal. Rather, dataset selection should be guided by the intended application: reanalysis-based products such as ERA5-Land are well suited for reproducing historical streamflow dynamics, whereas ensemble hydrological products such as ISIMIP3a are particularly valuable for robust assessments of long-term change and uncertainty. Future work should extend this comparison to water level and floodplain simulations to determine how runoff-product differences propagate to flood hazard metrics.

Code and data availability

The source code and documentation for the CaMa-Flood model are publicly available on GitHub (github.com/global-hydrodynamics/CaMa-Flood_v4), which additional benchmarking-related tools archived on Zenodo (10.5281/zenodo.14287024 and 10.5281/zenodo.10893741).

Streamflow observations from the Global Runoff Data Centre (GRDC) are available at grdc.bafg.de. ERA5-Land forcing data are openly available through the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets), while ISIMIP3 forcing data can be obtained from the ISIMIP data repository (https://data.isimip.org).

Author contributions

J.E.S.B.: Conceptualization, Formal analysis, Funding acquisition, Investigation, Visualization, Writing (original draft preparation); F.Z.: Conceptualization, Formal analysis, Investigation, Writing (original draft preparation) ; S.N.G.: Conceptualization, Writing (review and editing) ; Y.P: Conceptualization, Writing (review and editing) ; Y.D.: Writing (review and editing) ; X.Z.: Writing (review and editing).

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This project was partially supported by the Institute of Global Innovation Research (GIR) at Tokyo University of Agriculture and Technology.

References

Bain, R. L., Shaw, M. J., Geheran, M. P., Tavakoly, A. A., Wahl, M. D., and Zsoter, E.: Intercomparison of global ERA reanalysis products for streamflow simulations at the high-resolution continental scale, J. Hydrol., 616, 128624, https://doi.org/10.1016/j.jhydrol.2022.128624, 2023.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

Balsamo, G., Viterbo, P., Hurk, B. J. J. van den, Hirschi, M., Betts, A., and Scipal, K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, ECMWF Tech. Memo., 28, https://doi.org/10.21957/yzyeh0vlw, 2008.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art\hack\newline hydrological models, Hydrol. Earth Syst. Sci., 21, 2881–2903, https://doi.org/10.5194/hess-21-2881-2017, 2017.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate both increases and decreases European river floods, Nature, 573, 108–111, https://doi.org/10.1038/s41586-019-1495-6, 2019.

Boulange, J., Hanasaki, N., Yamazaki, D., and Pokhrel, Y.: Role of dams in reducing global flood exposure under climate change, Nat. Commun., 12, 417, https://doi.org/10.1038/s41467-020-20704-0, 2021.

Chaudhari, S. and Pokhrel, Y.: Alteration of River Flow and Flood Dynamics by Existing and Planned Hydropower Dams in the Amazon River Basin, Water Resour. Res., 58, e2021WR030555, https://doi.org/10.1029/2021WR030555, 2022.

Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies, Earth Syst. Sci. Data, 12, 2097–2120, https://doi.org/10.5194/essd-12-2097-2020, 2020.

Dang, H. and Pokhrel, Y.: Evolution of river regimes in the Mekong River basin over 8 decades and the role of dams in recent hydrological extremes, Hydrol. Earth Syst. Sci., 28, 3347–3365, https://doi.org/10.5194/hess-28-3347-2024, 2024.

Dutta, R. and Markonis, Y.: Does ERA5-land capture the changes in the terrestrial hydrological cycle across the globe?, Environ. Res. Lett., 19, 024054, https://doi.org/10.1088/1748-9326/ad1d3a, 2024.

Eilander, D., Couasnon, A., Ikeuchi, H., Muis, S., Yamazaki, D., Winsemius, H. C., and Ward, P. J.: The effect of surge on riverine flood hazard and impact in deltas globally, Environ. Res. Lett., 15, 104007, https://doi.org/10.1088/1748-9326/ab8ca6, 2020.

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., del Rocío Rivas López, M., Otto, C., Reyer, C. P. O., Karger, D. N., Malle, J. T., Treu, S., Menz, C., Blanchard, J. L., Harrison, C. S., Petrik, C. M., Eddy, T. D., Ortega-Cisneros, K., Novaglio, C., Rousseau, Y., Watson, R. A., Stock, C., Liu, X., Heneghan, R., Tittensor, D., Maury, O., Büchner, M., Vogt, T., Wang, T., Sun, F., Sauer, I. J., Koch, J., Vanderkelen, I., Jägermeyr, J., Müller, C., Rabin, S., Klar, J., Vega del Valle, I. D., Lasslop, G., Chadburn, S., Burke, E., Gallego-Sala, A., Smith, N., Chang, J., Hantson, S., Burton, C., Gädeke, A., Li, F., Gosling, S. N., Müller Schmied, H., Hattermann, F., Wang, J., Yao, F., Hickler, T., Marcé, R., Pierson, D., Thiery, W., Mercado-Bettín, D., Ladwig, R., Ayala-Zamora, A. I., Forrest, M., and Bechtold, M.: Scenario setup and forcing data for impact model evaluation and impact attribution within the third round of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3a), Geosci. Model Dev., 17, 1–51, https://doi.org/10.5194/gmd-17-1-2024, 2024.

Gu, X., Zhang, Q., Li, J., Chen, D., Singh, V. P., Zhang, Y., Liu, J., Shen, Z., and Yu, H.: Impacts of anthropogenic warming and uneven regional socio-economic development on global river flood risk, J. Hydrol., 590, 125262, https://doi.org/10.1016/j.jhydrol.2020.125262, 2020.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., and Seneviratne, S. I.: Observed Trends in Global Indicators of Mean and Extreme Streamflow, Geophys. Res. Lett., 46, 756–766, https://doi.org/10.1029/2018GL079725, 2019.

Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., Leonard, M., Liu, J., Müller Schmied, H., Papadimitriou, L., Pokhrel, Y., Seneviratne, S. I., Satoh, Y., Thiery, W., Westra, S., Zhang, X., and Zhao, F.: Globally observed trends in mean and extreme river flow attributed to climate change, Science, 371, 1159–1162, https://doi.org/10.1126/science.aba3996, 2021.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hamitouche, M., Fosser, G., Anav, A., He, C., and Lin, T.-S.: Impact of runoff schemes on global flow discharge: a comprehensive analysis using the Noah-MP and CaMa-Flood models, Hydrol. Earth Syst. Sci., 29, 1221–1240, https://doi.org/10.5194/hess-29-1221-2025, 2025.

Hanazaki, R., Yamazaki, D., and Yoshimura, K.: Development of a Reservoir Flood Control Scheme for Global Flood Models, J. Adv. Model. Earth Syst., 14, e2021MS002944, https://doi.org/10.1029/2021MS002944, 2022.

Hattermann, F. F., Krysanova, V., Gosling, S. N., Dankers, R., Daggupati, P., Donnelly, C., Flörke, M., Huang, S., Motovilov, Y., Buda, S., Yang, T., Müller, C., Leng, G., Tang, Q., Portmann, F. T., Hagemann, S., Gerten, D., Wada, Y., Masaki, Y., Alemayehu, T., Satoh, Y., and Samaniego, L.: Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins, Clim. Change, 141, 561–576, https://doi.org/10.1007/s10584-016-1829-4, 2017.

Heinicke, S., Volkholz, J., Schewe, J., Gosling, S. N., Müller Schmied, H., Zimmermann, S., Mengel, M., Sauer, I. J., Burek, P., Chang, J., Kou-Giesbrecht, S., Grillakis, M., Guillaumot, L., Hanasaki, N., Koutroulis, A., Otta, K., Qi, W., Satoh, Y., Stacke, T., Yokohata, T., and Frieler, K.: Global hydrological models continue to overestimate river discharge, Environ. Res. Lett., 19, 074005, https://doi.org/10.1088/1748-9326/ad52b0, 2024.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth Syst Dynam, 4, 219–236, https://doi.org/10.5194/esd-4-219-2013, 2013.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nat. Clim. Change, 3, 816–821, https://doi.org/10.1038/nclimate1911, 2013.

Ikeuchi, H., Hirabayashi, Y., Yamazaki, D., Kiguchi, M., Koirala, S., Nagano, T., Kotera, A., and Kanae, S.: Modeling complex flow dynamics of fluvial floods exacerbated by sea level rise in the Ganges–Brahmaputra–Meghna Delta, Environ. Res. Lett., 10, 124011, https://doi.org/10.1088/1748-9326/10/12/124011, 2015.

Ikeuchi, H., Hirabayashi, Y., Yamazaki, D., Muis, S., Ward, P. J., Winsemius, H. C., Verlaan, M., and Kanae, S.: Compound simulation of fluvial floods and storm surges in a global coupled river-coast flood model: Model development and its application to 2007 Cyclone Sidr in Bangladesh, J. Adv. Model. Earth Syst., 9, 1847–1862, https://doi.org/10.1002/2017MS000943, 2017.

Jonkman, S. N., Curran, A., and Bouwer, L. M.: Floods have become less deadly: an analysis of global flood fatalities 1975–2022, Nat. Hazards, 120, 6327–6342, https://doi.org/10.1007/s11069-024-06444-0, 2024.

Kim, H.: Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1), 2017.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424–425, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0), Geosci. Model Dev., 12, 3055–3070, https://doi.org/10.5194/gmd-12-3055-2019, 2019.

Lange, S.: ISIMIP3BASD, , https://doi.org/10.5281/zenodo.4686991, 2021.

Liu, L., Yi, Y., Jiang, H., Ran, Y., and Chen, D.: ERA5-Land overestimates runoff coefficient but underestimates runoff recession rate in the central Tibetan permafrost region, J. Hydrol. Reg. Stud., 53, 101792, https://doi.org/10.1016/j.ejrh.2024.101792, 2024.

Mester, B., Willner, S. N., Frieler, K., and Schewe, J.: Evaluation of river flood extent simulated with multiple global hydrological models and climate forcings, Environ. Res. Lett., 16, 094010, https://doi.org/10.1088/1748-9326/ac188d, 2021.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., and Kumar, R.: On the choice of calibration metrics for "high-flow" estimation using hydrologic models, Hydrol. Earth Syst. Sci., 23, 2601–2614, https://doi.org/10.5194/hess-23-2601-2019, 2019.

Müller Schmied, H., Trautmann, T., Ackermann, S., Cáceres, D., Flörke, M., Gerdener, H., Kynast, E., Peiris, T. A., Schiebener, L., Schumacher, M., and Döll, P.: The global water resources and use model WaterGAP v2.2e: description and evaluation of modifications and new features, Geosci. Model Dev., 17, 8817–8852, https://doi.org/10.5194/gmd-17-8817-2024, 2024.

Müller Schmied, H., Gosling, S. N., Garnsworthy, M., Müller, L., Telteu, C.-E., Ahmed, A. K., Andersen, L. S., Boulange, J., Burek, P., Chang, J., Chen, H., Gudmundsson, L., Grillakis, M., Guillaumot, L., Hanasaki, N., Koutroulis, A., Kumar, R., Leng, G., Liu, J., Liu, X., Menke, I., Mishra, V., Pokhrel, Y., Rakovec, O., Samaniego, L., Satoh, Y., Shah, H. L., Smilovic, M., Stacke, T., Sutanudjaja, E., Thiery, W., Tsilimigkras, A., Wada, Y., Wanders, N., and Yokohata, T.: Graphical representation of global water models, Geosci. Model Dev., 18, 2409–2425, https://doi.org/10.5194/gmd-18-2409-2025, 2025.

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., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Pastén-Zapata, E., Pimentel, R., Royer-Gaspard, P., Sonnenborg, T. O., Aparicio-Ibañez, J., Lemoine, A., Pérez-Palazón, M. J., Schneider, R., Photiadou, C., Thirel, G., and Refsgaard, J. C.: The effect of weighting hydrological projections based on the robustness of hydrological models under a changing climate, J. Hydrol. Reg. Stud., 41, 101113, https://doi.org/10.1016/j.ejrh.2022.101113, 2022.

Rentschler, J., Salhab, M., and Jafino, B. A.: Flood exposure and poverty in 188 countries, Nat. Commun., 13, 3527, https://doi.org/10.1038/s41467-022-30727-4, 2022.

Rogers, J. S., Maneta, M. P., Sain, S. R., Madaus, L. E., and Hacker, J. P.: The role of climate and population change in global flood exposure and vulnerability, Nat. Commun., 16, 1287, https://doi.org/10.1038/s41467-025-56654-8, 2025.

Samadi, V., Fowler, H. J., Lamond, J., Wagener, T., Brunner, M., Gourley, J., Moradkhani, H., Popescu, I., Wasko, C., Wright, D., Wu, H., Zhang, K., Arias, P. A., Duan, Q., Nazemi, A., van Oevelen, P. J., Prein, A. F., Roundy, J. K., Saberian,

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

M., and Umutoni, L.: The Needs, Challenges, and Priorities for Advancing Global Flood Research, WIREs Water, 12, e70026, https://doi.org/10.1002/wat2.70026, 2025.

Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, J. Am. Stat. Assoc., 63, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934, 1968.

Shamsudduha, M.: Redefining flood hazard and addressing emerging risks in an era of extremes, Npj Nat. Hazards, 2, 29, https://doi.org/10.1038/s44304-025-00082-7, 2025.

Telteu, C.-E., Müller Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., Boulange, J. E. S., Andersen, L. S., Grillakis, M., Gosling, S. N., Satoh, Y., Rakovec, O., Stacke, T., Chang, J., Wanders, N., Shah, H. L., Trautmann, T., Mao, G., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Samaniego, L., Wada, Y., Mishra, V., Liu, J., Döll, P., Zhao, F., Gädeke, A., Rabin, S. S., and Herz, F.: Understanding each other's models: an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication, Geosci. Model Dev., 14, 3843–3878, https://doi.org/10.5194/gmd-14-3843-2021, 2021.

Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., Yamazaki, D., Hirabayashi, Y., Pappenberger, F., Dutra, E., Ward, P. J., Winsemius, H. C., Salamon, P., Dottori, F., Rudari, R., Kappes, M. S., Simpson, A. L., Hadzilacos, G., and Fewtrell, T. J.: The credibility challenge for global fluvial flood risk analysis, Environ. Res. Lett., 11, 094014, https://doi.org/10.1088/1748-9326/11/9/094014, 2016.

Uhe, P., Lucas, C., Hawker, L., Brine, M., Wilkinson, H., Cooper, A., Saoulis, A. A., Savage, J., and Sampson, C.: FathomDEM: an improved global terrain map using a hybrid vision transformer model, Environ. Res. Lett., 20, 034002, https://doi.org/10.1088/1748-9326/ada972, 2025.

Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, Environ. Res. Lett., 13, 055008, https://doi.org/10.1088/1748-9326/aab96f, 2018.

Wang, K., Liu, X., Cui, P., Zhang, Y., Xie, J., Liu, C., and Gosling, S. N.: China's nationwide streamflow decline driven by landscape changes and human interventions, Sci. Adv., 11, eadu8032, https://doi.org/10.1126/sciadv.adu8032, 2025.

Wang, K., Mou, W., Cai, H., Yin, C., Li, Y., Allan, R. P., Beck, H., Dai, A., Miralles, D. G., Ryu, D., and Pappenberger, F.: ERA5 overestimates land drying trend from 1980 to 2023 by more than 100%, Sci. Bull., https://doi.org/10.1016/j.scib.2026.01.045, 2026.

Yamazaki, D.: Advancing global river hydrodynamics simulations by catchment-based macro-scale floodplain modeling approach, Geosci. Lett., 12, 72, https://doi.org/10.1186/s40562-025-00452-z, 2025.

Yamazaki, D., Kanae, S., Kim, H., and Oki, T.: A physically based description of floodplain inundation dynamics in a global river routing model, Water Resour. Res., 47, https://doi.org/10.1029/2010WR009726, 2011.

Yamazaki, D., de Almeida, G. A. M., and Bates, P. D.: Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map, Water Resour. Res., 49, 7221–7235, https://doi.org/10.1002/wrcr.20552, 2013.

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., and Bates, P. D.: Regional flood dynamics in a bifurcating mega delta simulated in a global river model, Geophys. Res. Lett., 41, 3127–3135, https://doi.org/10.1002/2014GL059744, 2014.

https://doi.org/10.5194/egusphere-2026-2739
Preprint. Discussion started: 5 June 2026

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resour. Res., 55, 5053–5073, https://doi.org/10.1029/2019WR024873, 2019.

Yoshida, T., Hanasaki, N., Nishina, K., Boulange, J., Okada, M., and Troch, P. A.: Inference of Parameters for a Global Hydrological Model: Identifiability and Predictive Uncertainties of Climate-Based Parameters, Water Resour. Res., 58, e2021WR030660, https://doi.org/10.1029/2021WR030660, 2022.

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., and Wada, Y.: Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts, Environ. Res. Lett., 13, 065015, https://doi.org/10.1088/1748-9326/aac547, 2018.

Zaherpour, J., Mount, N., Gosling, S. N., Dankers, R., Eisner, S., Gerten, D., Liu, X., Masaki, Y., Müller Schmied, H., Tang, Q., and Wada, Y.: Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models, Environ. Model. Softw., 114, 112–128, https://doi.org/10.1016/j.envsoft.2019.01.003, 2019.

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., Stacke, T., Ciais, P., Chang, J., Ducharne, A., Guimberteau, M., Wada, Y., Kim, H., and Yamazaki, D.: The critical role of the routing scheme in simulating peak river discharge in global hydrological models, Environ. Res. Lett., 12, 075003, https://doi.org/10.1088/1748-9326/aa7250, 2017.

Zhao, G., Yamazaki, D., Tanaka, Y., Zhou, X., Li, S., Hu, Y., Hirabayashi, Y., Neal, J., and Bates, P.: Developing a Levee Module for Global Flood Modeling With a Reach-Level Parameterization Approach, Water Resour. Res., 61, e2024WR039790, https://doi.org/10.1029/2024WR039790, 2025.

Zhou, X., Ma, W., Echizenya, W., and Yamazaki, D.: The uncertainty of flood frequency analyses in hydrodynamic model simulations, Nat Hazards Earth Syst Sci, 21, 1071–1085, https://doi.org/10.5194/nhess-21-1071-2021, 2021.

Zhou, X., Yamazaki, D., Revel, M., Zhao, G., and Modi, P.: Benchmark Framework for Global River Models, J. Adv. Model. Earth Syst., 17, e2024MS004379, https://doi.org/10.1029/2024MS004379, 2025.