Benefits of Calibrating a Global Hydrological Model for Regional Analyses of Flood and Drought Projections: A Case Study of the Yangtze River Basin

Fang Zhao[1,2,3,4] [iD], **Ning Nie**[5], **Yang Liu**[1,2], **Congrui Yi**[1,2], **Luca Guillaumot**[4,6] [iD], **Yoshihide Wada**[7,8] [iD], **Peter Burek**[4] [iD], **Mikhail Smilovic**[4], **Katja Frieler**[3] [iD], **Matthias Buechner**[3] [iD], **Jacob Schewe**[3] [iD], and **Simon N. Gosling**[9] [iD]

[1]Key Laboratory of Geographic Information Science of the Ministry of Education, School of Geographic Sciences, East China Normal University, Shanghai, China, [2]Key Laboratory of Spatial-temporal Big Data Analysis and Application of Natural Resources in Megacities, Ministry of Natural Resources, Shanghai, China, [3]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany, [4]Water Security Research Group, Biodiversity and Natural Resources Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, [5]College of Atmospheric Sciences, Lanzhou University, Lanzhou, China, [6]BRGM, Orléans, France, [7]Biodiversity and Natural Resources Program, International Institute for Applied Systems Analysis, Laxenburg, Austria, [8]Climate and Livability, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [9]School of Geography, University of Nottingham, Nottingham, UK

Abstract Uncalibrated global hydrological models are primarily used to inform projections of flood and drought changes under global warming and their impacts, but it remains unclear how model calibration might benefit these projections. Using the Yangtze River Basin as a case study, we compare projected changes in flood and drought frequencies and their impacts—area, population, and gross domestic product affected—at various warming levels, from uncalibrated and calibrated simulations with the Community Water Model. These projections are driven by 10 General Circulation Models (GCMs) from Coupled Model Intercomparison Project Phase 6, within the Inter-Sectoral Impact Model Intercomparison Project framework. Calibration significantly improves simulated discharge, yet the impact of calibration under climate change on projected increases in flood frequency and their associated impacts is minor, in contrast to its notable role in drought projections. We further quantify the relative contribution of GCMs, emission scenarios, and calibration approaches to the projected impacts, finding that GCMs primarily drive projected flood changes, while emission scenarios and calibration contribute more significantly to the variance in drought projections after 2050. The differing sensitivities to calibration are attributed to the dominance of extreme precipitation in flood generation and the influence of long-term evapotranspiration trends on drought occurrence. The findings imply that future projections of relative changes in flood frequency and risks based on uncalibrated hydrological models are likely still quite reliable for warm and humid regions. However, careful calibration and model improvement is crucial for enhancing the reliability of future drought impact assessments.

1. Introduction

Global warming is expected to significantly alter the distribution, variability, and intensity of precipitation, consequently changing flood and drought risk characteristics (Donat et al., 2016; Hirabayashi et al., 2013; Kundzewicz, Su, Wang, Xia, et al., 2019; C. Li et al., 2019, 2021; Pendergrass & Hartmann, 2014; Pfahl et al., 2017; Prein et al., 2017; Prudhomme et al., 2014; Tanoue et al., 2021; Thompson et al., 2021; J. Zhang et al., 2021). Assessing global flood and drought risks at global warming levels (GWLs) is vital for aiding the implementation of the ambitious mitigation and adaptation goals established by the Paris Agreement (UNFCCC, 2015) and the Sustainable Development Goals (United Nations Office for Disaster Risk Reduction, 2015). Projections indicate that, even with 2°C global warming, the number of people affected by floods and droughts will multiply by the end of this century (Dottori et al., 2018; Lange et al., 2020; Willner et al., 2018). Given these projections, providing stakeholders with reliable and locally relevant information on future high-risk areas for droughts and floods is critically important to facilitate effective climate change adaptation and mitigation strategies.

Due to limited available observation data and regional model simulations, projections of flood and drought changes and their impacts in many regions primarily rely on global hydrological models (GHMs). While these

models often use coarse spatial resolutions (50×50 km grids or larger) and lack regionalized parameters leading to suboptimal performance across various regions (Beck et al., 2016; Zaherpour et al., 2018), they remain essential tools for large-scale hydrological projections. Consequently, many studies employ a modeling chain that includes GHMs to project flood impacts, such as the number of people affected or economic damages (Arnell & Gosling, 2016; Dottori et al., 2018; Hirabayashi et al., 2013). Despite multiple sources of uncertainty in the modeling chain, GHMs generally reproduce large-scale observed impact patterns with reasonable accuracy (Sauer et al., 2021). However, decision-makers often hesitate to trust these models due to their suboptimal performance in replicating historical hydrological conditions, which can limit their effectiveness in informing future climate change adaptation, disaster prevention, and reduction measures.

Calibration of GHMs is often limited due to data availability, substantial computational demands and the significant time required to calibrate models globally. In contrast, regional hydrological models are frequently calibrated to better represent key hydrological processes (Krysanova et al., 2018). However, applying regional models to long-term scenarios and extensive areas, especially where observational data is lacking, is both costly and challenging. Furthermore, areas with significant human influence often lack reliable ensemble regional high-resolution simulation results. For example, the regional water sector of the Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP) often focuses on sparsely populated upper reaches of river basins rather than the entire basin (Frieler et al., 2017). This limits the practical application of drought and flood impact assessments in the more populous middle and lower reaches.

Several studies have assessed the reliability of GHM projections by comparing them with regional hydrological models across various large river basins (Gosling et al., 2011, 2017; Hattermann et al., 2017; Kumar et al., 2022). Despite a considerable bias in long-term average monthly discharge, mostly uncalibrated global hydrological models often show similar climate variability sensitivity to regional counterparts, especially when evaluating percentage changes of the future period relative to the historical period (Hattermann et al., 2017). However, Krysanova et al. (2018) reported that in regions significantly influenced by snowmelt, the disparities in projected changes between the models become more pronounced. These comparisons often overlook managed basins and the hydrological impacts of extreme events. Additionally, varied model structures and forcing data make it challenging to isolate the impact of model calibration on the differences observed between global and regional hydrological model projections (Müller Schmied et al., 2024; Telteu et al., 2021).

Numerous studies have directly investigated how various aspects of model calibration affect regional hydrological model simulations, including the selection of calibrated hydrological variables (Adhikary et al., 2019; Bai et al., 2018; S. Huang et al., 2020; Koch et al., 2020; Müller Schmied et al., 2024; Nkiaka et al., 2018; Tarasova et al., 2016), the time period (Ekmekcioglu et al., 2022; Shin & Jung, 2022), the timescale of calibration data (Setti et al., 2022), model parameters (Demirel et al., 2013; Qi et al., 2019; Q. Wu et al., 2017), and calibration metrics (Mizukami et al., 2019; Munoz-Castro et al., 2023; Seiller et al., 2017). For instance, Seiller et al. (2017) found that the choice of calibration metrics significantly influences the diagnosis and understanding of climate impacts on water resources in Canada. S. Huang et al. (2020) discovered that different calibration methods can lead to contradictory directions of change for the model ensemble simulated high flow (Q10), median flow, and low flow (Q90) in the Mississippi basin, while the simulated directions of changes are mostly consistent for the Yellow and Rhine basins. However, few studies have examined the implications of calibration for model projections of extreme events with high return periods, such as 100-year return period floods and droughts (Kumar et al., 2022). These events are often the focus of impact assessments due to their high potential damage to society compared to less extreme events. Analyzing how calibration affects the impact assessment of floods and droughts at different GWLs, similar to assessments in the IPCC AR6 and SR1.5 reports (IPCC, 2018, 2023a), can provide highly relevant information for decision-makers. Additionally, discussions on the underlying mechanisms of how calibration affects projected changes in hydrological extremes are limited.

The Yangtze River Basin (YRB), a region of profound human, economic, and environmental significance, sustains over 400 million people and nearly 40% of China's gross domestic product (GDP), playing a pivotal role in global economic and social development (Y. Li et al., 2021; J. Lu et al., 2019). Historically plagued by frequent drought and flood events, particularly in its densely populated middle and lower reaches, hydrological extremes occurred over the YRB have led to significant socio-economic impacts, including the devastating 1998 flood that caused direct economic losses of approximately 23 billion United States dollars (Chen et al., 2018; Qi et al., 2022; F. Yu et al., 2009; Zong & Chen, 2000). Recent projections suggest an intensification of these extremes due to

global warming, with increases in both drought severity and flood frequency, posing an even greater threat to the region (Kundzewicz, Su, Wang, Wang, et al., 2019; Sun et al., 2019; Xiong et al., 2022; Z. Yu et al., 2018; J. Zhao et al., 2022). Despite the urgent need, spatially explicit projections that integrate the various impacts of both extremes on the YRB are still lacking. This deficiency underscores the necessity of assessing the benefits of calibration in hydrological modeling to enhance the reliability of such projections. As a region emblematic of the challenges posed by both floods and droughts, the YRB offers a great opportunity to evaluate the differences between calibrated and uncalibrated models in projecting changes in hydrological extremes. Insights from this comparative analysis are invaluable for improving public safety measures and formulating robust disaster prevention and reduction strategies, not only for the YRB but potentially for other regions worldwide.

The main objectives of this study are to: (a) evaluate the Community Water Model (CWatM)'s performance under various experimental settings, including uncalibrated simulation and simulations calibrated with different metrics; (b) examine the effect of calibration on projected flood and drought frequency at different GWLs; (c) assess the influences of calibration on projections of the total basin area, population, and GDP affected by droughts and floods at different GWLs; and (d) analyze the contribution of methodological choices to the spread among projected impacts.

Under the third phase of the ISIMIP framework (ISIMIP3), various global models are employed for hydrological projections using the latest bias-adjusted Coupled Model Intercomparison Project Phase 6 (CMIP6) climate scenarios. Among these models, CWatM operates globally at a spatial resolution of 30′ and regionally at a higher resolution of approximately 10 km (Burek, Satoh, et al., 2020). Like other regional hydrological models, the accuracy of regional CWatM simulations can be significantly enhanced through model calibration, offering an opportunity to assess the impact of calibration on a model actively participating in global hydrological projection comparisons.

Using the YRB as the study region, we first identify the change signals in the frequency of 100-year return period floods and droughts and their regionally aggregated impacts, projected by uncalibrated CWatM. Then we compare them with calibrated CWatM simulations to evaluate to what extent calibration would affect quantification of the pure effect of climate change impacts at different levels of global warming at the regional scale. Specifically, we investigate how flood and drought frequency and impact change with respect to different levels of global warming above pre-industrial levels, focusing on three GWLs: 1.5, 2, and 3°C warming. The first two levels are mitigation targets set out in the Paris Agreement, while 3°C warming is likely to occur by the end of the twenty-first century if sufficient mitigation strategies are not implemented (UNFCCC, 2016).

2. Materials and Methods

2.1. Study Region

The YRB, covering a total area of 1.8 million km², accounts for 18.8% of China's land area (Figure 1). As one of the world's most densely populated river basins, it has undergone rapid socio-economic development over the past 50 years. The basin's terrain is characterized by a descending gradient from over 7,000 m in the source area to below 100 m in the lower reaches. Influenced by the southeast monsoon, the basin is predominantly warm and humid, featuring an average annual precipitation of 1,100 mm, mostly concentrated during the summer season. Stream discharge peaks in summer (Figure S1 in Supporting Information S1), when the contribution of snowmelt to runoff is minimal (Figure S2 in Supporting Information S1). Precipitation distribution varies significantly across the basin, with mean annual precipitation less than 400 mm in the source region and over 1,600 mm in the downstream area (H. Yang et al., 2018). Since the devastating 1998 flood, the Yangtze River has experienced a surge in water-related infrastructure development. Notably, the Three Gorges Dam (TGD), situated upstream of Yichang, boasts a reservoir storage capacity of 39.3 billion m³ (reached in 2009) and has served multiple purposes including flood control, power generation, irrigation, and navigation since 2003. The reservoir markedly influences seasonal streamflow variations, terrestrial water storage, and extensive rice-paddy agriculture in the basin (Nie, Zhang, Chen, et al., 2021; Nie, Zhang, Liu, et al., 2021; Wang et al., 2013). Cuntan, Yichang, Hankou, and Datong, as depicted in Figure 1, are the major hydrological stations along the main stream of the Yangtze River. Specifically, Yichang serves as the outlet of the upper reaches, while discharge changes at Datong station typically reflect runoff changes of the entire basin (Fremme & Sodemann, 2019; Y. Huang et al., 2015; Nie, Zhang, Liu, et al., 2021). The mean annual discharge at Datong station is 8.9×10^{11} m³, equivalent to an annual basin average of 495 mm, representing approximately 33% of China's total surface runoff (D. Zhang et al., 2016).
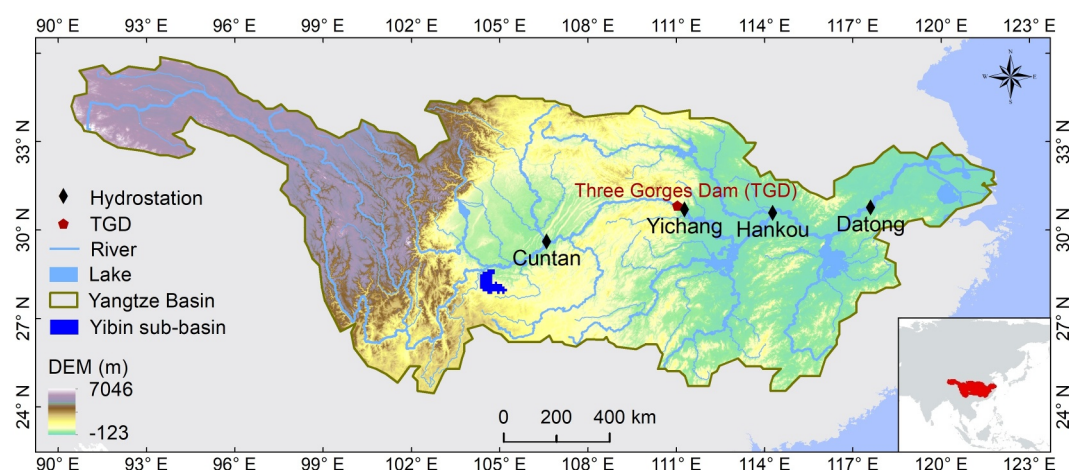
Figure 1. Geographic location of the study region, the Yangtze River Basin (YRB). The Digital Elevation Model (DEM) data illustrate the topographical variation across the basin.

As of 2015, the basin's total population was around 457 million (Klein Goldewijk et al., 2017), with a higher concentration in the middle and lower reaches (Figure S3 in Supporting Information S1).

2.2. Methodology and Data

2.2.1. Meteorological Forcing

ISIMIP used the W5E5 v2.0 data set (Lange et al., 2021), which combines WFDE5 v2.0 (Cucchi et al., 2020) with ERA5 ocean data (Hersbach et al., 2020), to bias-adjust 10 General Circulation Models (GCMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6) using an updated trend-preserving quantile mapping method (ISIMIP3BASD v2.5.0 (Lange, 2019, 2021)). These 10 GCMs are listed in Table S1 in Supporting Information S1.

In ISIMIP Phase 3b, bias-adjusted pre-industrial control (piControl), historical, and future scenarios from 10 GCMs were provided to support cross-sectoral impact simulations (Frieler et al., 2024). The piControl simulation in CMIP6, with a minimum of 500 years for each GCM, was driven by non-evolving pre-industrial forcing, reflecting the unforced variability of the climate system (Eyring et al., 2016). ISIMIP adjusted 500 years of CMIP6 piControl simulation data from each of the 10 GCMs and renamed the simulation years from 0001 to 0500 to 1601–2100 (representing pre-industrial climate conditions with 1850 as the base year), facilitating easier handling of the calendar in the netCDF files. ISIMIP selected three shared socioeconomic pathway (SSP)-based future scenarios from the Tier 1 CMIP6 scenarios, representing a range of forcing pathways: high end (SSP5-8.5), medium to high end (SSP3-7.5), and low end (SSP1-2.6) (O'Neill et al., 2016). For this study, we used 200 years of the 500-year piControl (1851–2050), in addition to historical (1971–2014) and future (2015–2100) data from ISIMIP to drive our simulations (further details are in Section 2.2.3).

The meteorological variables required as input for CWatM included precipitation, relative humidity, long- and short-wave downward surface radiation fluxes, maximum, minimum, and average 2 m temperature, 10 m wind speed, and surface pressure, all available at 30′ spatial and daily temporal resolution. Additionally, the same W5E5 v2.0 data set used by ISIMIP for bias-adjustment was employed for model calibration and validation.

2.2.2. Hydrological Modeling

The recently developed CWatM v1.1 (Burek, Satoh, et al., 2020), participating in ISIMIP3, is an integrated hydrological and water resources model. The model simulates key natural hydrological processes, including snow accumulation and melting, interception, and evapotranspiration (using the Penman-Monteith method as the default for potential evapotranspiration) across six land cover classes: forest, grassland, paddy-irrigated, nonpaddy-irrigated, urban sealed, and water. It models soil infiltration (using the Xinanjiang approach), pref-erential bypass flow, three-layer soil moisture redistribution and capillary rise, linear reservoir groundwater

storage, lakes, runoff concentration within grid cells, and kinematic wave river routing. CWatM also considers human activities such as water use from irrigation, domestic, industrial, and livestock sectors, reservoir regulation (including the Three Gorges Reservoir), and land cover change. Necessary input data, including elevation, flow direction, and soil pedo-transfer data, are prepared at 30′ (∼50 km) and 5′ (∼10 km) spatial resolution to facilitate simulations globally or for any region on a daily time step. For the 5′ version, by default, the CWatM model internally regridded the coarser meteorological forcing data to 5′ using the delta change method (Moreno & Hasenauer, 2016; Mosier et al., 2018) based on high-resolution monthly data from WorldClim version 2 (Fick & Hijmans, 2017).

For the ISIMIP3 simulations, the DEAP evolutionary computation framework (Fortin et al., 2012) was employed to calibrate a set of globally uniform parameters to optimize the average modified Kling-Gupta efficiency (KGE, Kling et al., 2012) performance across more than 1,000 river basins worldwide (F. Zhao et al., 2017). To be consistent with other studies, we used the same model version and parameter set from the ISIMIP3 simulations for our uncalibrated simulations (Table S1 in Supporting Information S1). While the globally calibrated parameters optimized the average performance, for some basins, the KGE is lower compared to the simulated discharge using the default parameter set. We found this to be the case for the YRB, which is not one of the basins used for the global calibration; therefore, we refer to our experiment using globally calibrated parameters as the uncalibrated experiment for the YRB. We used the default global data sets provided with CWatM at 5′ resolution as the basis for framing conditions (Table 1), with some modifications (for a complete list of model input data, refer to Burek, Satoh, et al. (2020)). The flow direction near the Yangtze Delta was manually adjusted to include Shanghai and its vicinity in the study region (Figure 1). For land cover data at 5′ resolution, we harmonized the 30′ annual land cover data from 1901 to 2018 with the model's default 5′ land cover for 2000, preserving both the ISIMIP historical land cover changes at 30′ and the finer spatial features at 5′. In line with the ISIMIP3b design (https://protocol.isimip.org/#ISIMIP3b/water_global), while the socio-economic data (including land cover, water use, and reservoirs) were time-varying for historical simulations, they were held constant at year-2015 levels for both pre-industrial control and future simulations. Similar to Lange et al. (2020), keeping socioeconomics data the same for both pre-industrial control and future simulations allows us to quantify the pure effect of climate change.

2.2.3. Experiment Design

This study examines the impact of calibration on simulated flood and drought changes and assesses the applicability of uncalibrated model simulations for regional-scale flood and drought impact assessments. We conducted four sets of experiments at 5′ spatial resolution, adhering to the ISIMIP3a (historical reanalysis) and ISIMIP3b (including piControl, historical, and 3-SSP future simulations) protocols (www.isimip.org/protocol).

The first of the four experiment sets, unCal, consists of uncalibrated simulations using the same model parameters as in ISIMIP (Table S1 in Supporting Information S1). The other three sets, Cal_L, Cal_S, and Cal_H, differ from the first in that they involve 12 model parameters calibrated using different objective functions (Table 2). The choice of objective function is critical; our aim is not to identify the most optimal calibration metrics but rather to assess whether projected flood and drought changes differ when using various metrics that could reasonably be adopted in regional simulation studies. In the experiment set Cal_S, we optimized the KGE (Kling et al., 2012), which has become popular in recent studies (Dahri et al., 2021; Hirpa et al., 2018; Liu, 2020; Mizukami et al., 2019). The KGE is expressed as:

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\alpha-1)^2}, \beta = \frac{\mu_s}{\mu_o}, \alpha = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{1}$$

Here, r is Pearson's correlation coefficient, β represents the bias computed as the ratio of the simulated to the observed mean (μ), and α is the variability ratio, given by the ratio of the simulated to observed coefficients of variation (σ/μ), where σ is the standard deviation. The subscripts "s" and "o" correspond to the simulated values and observed values, respectively.

In the other two calibrated experiments (Cal_L and Cal_H), we aimed to optimize low flow and high flow performance, respectively, to more accurately represent droughts and floods, while ensuring robust overall performance of simulated streamflow. Mizukami et al. (2019) demonstrated that optimizing based on a single application-specific metric could lead to poor performance in even closely related metrics. To circumvent this

Table 1
Data Sets Used in This Study

Variables	Data sources	Periods	Original spatial resolution	Temporal resolution	Source
DEM	MERIT Hydro (Yamazaki et al., 2019)	/	3"	/	http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/
Land cover	Forest Land Cover (Hansen et al., 2013)	1901–2018	1"	Yearly	http://earthenginepartners.appspot.com/science-2013-global-forest
	Impervious area (Elvidge et al., 2007)	2000	30"	/	ftp://ftp.ngdc.noaa.gov/DMSP/
	Irrigated areas (Döll & Siebert, 2002; Siebert et al., 2005, 2010)	2000	5'	/	https://www.fao.org/aquastat/en/geospatial-information/global-maps-irrigated-areas
	Cropland/pasture/urban: ISIMIP3 (Volkholz & Ostberg, 2022)	1901–2018	5'	Yearly	https://data.isimip.org/10.48364/ISIMIP.571261.2
Soil	Hydraulic Parameters, Data for SoilGrids, 0–0.05 m, 0.15–0.3 m, 0.6–1.0 m (Dai, Wei, et al., 2019; Dai, Xin, et al., 2019)	/	30"	/	http://globalchange.bnu.edu.cn/research/soil5.jsp#cite
Flow direction	DRT (H. Wu et al., 2011)	/	5'	/	ftp://ftp.ntsg.umt.edu/pub/data/DRT/
Meteorological forcing	Historical: W5E5 (Cucchi et al., 2020; Lange et al., 2021)	1901–2019	30'	Daily	https://data.isimip.org/10.48364/ISIMIP.342217
	Downscaling to 5' WorldClim version2 (Fick & Hijmans, 2017)	1970–2000 Mean	30"	Monthly Climatology	https://www.worldclim.org/
	Historical: 10 GCMs (Lange & Büchner, 2021; Lange et al., 2024)	1901–2014	30'[a]	Daily	https://data.isimip.org/10.48364/ISIMIP.842396.1; https://doi.org/10.48364/ISIMIP.581124.4
	Future: 10 GCMs, 3 SSPs (Lange & Büchner, 2021; Lange et al., 2024)	2015–2100	30'[a]	Daily	https://data.isimip.org/10.48364/ISIMIP.842396.1; https://doi.org/10.48364/ISIMIP.581124.4
	pre-industrial Control (piControl): 10 GCMs (Lange & Büchner, 2021; Lange et al., 2024)	1851–2050	30'[a]	Daily	https://data.isimip.org/10.48364/ISIMIP.842396.1; https://doi.org/10.48364/ISIMIP.581124.4
Discharge	National Hydrological Yearbook of the Yangtze River Basin	2003–2018	/	Daily	Changjiang Water Resources Commission of the Ministry of Water Resources of China (see Open Research)
Livestock water demand	Gridded livestock densities (Steinfeld et al., 2006; Wint & Robinson, 2007); Livestock per country (FAO, 2012)	1960–2015	5'	Monthly	https://www.fao.org/faostat/en/#data

Table 1
Continued

Variables	Data sources	Periods	Original spatial resolution	Temporal resolution	Source
Industry water demand	Gridded industrial water data (Shiklomanov, 1997; Sutanudjaja et al., 2017; Wada et al., 2016)	1960–2015	5'	Monthly	https://zenodo.org/records/1045339#.XWUr7E2P5aR
Domestic water demand	Domestic water withdrawal per capita (FAO, 2012; Gleick et al., 2009)	1960–2015	5'	Monthly	https://www.fao.org/faostat/en/#data
Reservoir	HydroLakes database (Lehner et al., 2011; Messager et al., 2016)	Until 2011	shapefile	Yearly	http://www.hydrosheds.org/page/hydrolakes
Population	HYDE3.2 (Klein Goldewijk et al., 2017)	2015	5'	/	ftp://ftp.pbl.nl/hyde/hyde3.2/
GDP	Gridded global data sets (Kummu et al., 2018, 2020)	2015	5'	/	https://doi.org/10.5061/dryad.dk1j0

^aDownscaled and bias-corrected by the ISIMIP team to 30' from CMIP6 GCMs' original resolution.

Table 2
Optimization Goals and Objective Functions for Four Experiment Sets

Experiment set	unCal	Cal_L	Cal_S	Cal_H
Optimization Goal	N/A	Low flow; adequate overall performance	Overall performance	High flow; adequate overall performance
Objective Function	N/A	NSElog; KGE[a]	KGE	NSE1%; KGE[a]

[a]KGE is considered a secondary objective to ensure adequate overall performance.

issue, we employed a multi-objective calibration framework where the primary objective was to optimize the model's performance for low flow (or high flow), with the overall performance metric (KGE) considered as a secondary objective to ensure acceptable general performance.

For the low flow optimization experiment (Cal_L), we employed the logarithmic transformation of Nash-Sutcliffe efficiency (NSE) as the primary calibration metric (Farmer & Vogel, 2016; Oudin et al., 2006), expressed as:

$$NSElog = 1 - \frac{\sum_{i=1}^{N}(\log(Q_{s,i}) - \log(Q_{o,i}))^2}{\sum_{i=1}^{N}(\log(Q_{s,i}) - \overline{\log(Q_o)})^2} \tag{2}$$

Here, $Q_{s,i}$ and $Q_{o,i}$ represent the simulated and observed streamflows at time step i, $\overline{Q_o}$ is the mean observed flow over the simulation period (time steps 1 to N), and log represents the natural logarithmic transform. We opted for NSE over KGE for the transformed streamflow due to reported numerical instabilities and unit-dependence in the logarithmic transformation of KGE (Santos et al., 2018). For the high flow optimization experiment (Cal_H), we used NSE for the highest 1% daily discharge (NSE1%) as the primary calibration metric, along with KGE as the secondary objective to ensure overall performance. This approach effectively captures the magnitude of observed flood peaks and maintains overall good performance.

The multi-objective calibration identified solutions from the entire space, and those identified solutions together form the Pareto front, where improving one objective function necessitates a reduction in another. In multi-objective calibration, it's common to select the solution closest to the Utopian point to balance all objectives. However, since our focus was to optimize model performance specifically for low flow (Cal_L) or high flow (Cal_H) to improve the representation of droughts or floods, we prioritized the primary objective metric (NSElog for Cal_L or NSE1% for Cal_H). Therefore, for each set of non-dominated solutions, we selected one calibrated parameter set with the best performance in the primary objective metric. Our approach effectively resembles a single-objective calibration within a multi-objective framework. This means we concentrated on one primary metric while still considering overall performance through the secondary objective (KGE). It should be noted that due to computational limitations, the evolutionary algorithm might identify local optima rather than the global optimum, meaning that performance in the primary objective might not always be fully optimized. An example of Pareto fronts is shown in Figure S4 in Supporting Information S1.

A cascade calibration was performed at the four stations along the Yangtze River: Cuntan, Yichang, Hankou, and Datong. The calibration started at the upstream station, with the outflow from each station being used as the inflow for the next downstream station. Due to significant alterations in discharge after 2003 resulting from the construction of the Three Gorges Dam, the calibration period was set from 2008 to 2016 (calendar years), including the extensive 2016 summer flooding (W. Zhang et al., 2016). We performed split-sample validation, where the periods from 2003 to 2007 and 2017–2018 were used together for validation.

Comparisons between calibrated and uncalibrated simulations directly quantify the effect of calibration on YRB hydrology. The calibration process starts with an initial population size of 256, followed by subsequent generations, each having a population size of 32. The number of generations is set to 30, which has been found to be sufficient to achieve convergence for stations (Burek, Satoh, et al., 2020). However, studies aiming to produce more robust future projections should consider significantly increasing the number of generations to enhance the likelihood of finding the global optimum. While our calibration approach may not have identified the global optimum at all stations due to computational limitations, potentially affecting the absolute values of our projections, we believe that this does not significantly impact the relative differences observed between the calibration experiments, which are the primary focus of our study.

Depending on the subbasin size, a single simulation of 14 years (5 years as spin-up time and 9 years for comparing to observed data) takes around 30–60 min. After an initial 256 simulations for the general population, another 960 simulations are run (30 generations times 32 pool sizes). Altogether, these 1,216 simulations are run on 32 CPU cores in parallel sessions in around 20–40 hr. Out of approximately 100 parameters in the CWatM model, we calibrated 12 key hydrological parameters suggested in the CWatM manual (Burek, Smilovic, et al., 2020), which significantly influence streamflow simulations. Calibrating additional parameters would significantly increase computation time. The final calibrated parameters for each subbasin are detailed in Table S1 in Supporting Information S1. Note that in each calibrated experiment, parameters should be considered collectively rather than as individual values, due to their interdependence, where changes in one parameter may be compensated by others (Bárdossy, 2007).

In total, not counting calibration and validation simulations, we performed 200 simulations (4 experiment sets × 10 GCMs, each with piControl, historical, and 3 future scenarios). The 40 piControl simulations served as a baseline for determining the magnitude of 100-year floods and droughts at each grid cell. We calculated the change signal for the corresponding historical and future scenario simulations separately. The pure effect of climate change on the frequency and impacts of floods and droughts is derived from differences between scenario and baseline simulations. We examined the change signals as opposed to absolute values themselves because they are often the focus of impact studies on extremes (IPCC, 2023c), and it controls for systematic differences in the model setups. To balance between computational cost and a reasonable estimate of 100-year events' extreme value statistics under a pre-industrial climate, we chose 200 years (1851–2050) of the 500-year piControl forcing provided by ISIMIP to drive the four experiment sets. Similarly, we conducted historical (1971–2014) and future (2015–2100) simulations using bias-adjusted CMIP6 forcing data from ISIMIP.

2.2.4. Computing Flood and Drought Frequency Change at Different Warming Levels

In this study, we used the 100-year return period of annual maximum discharges and 7-day minimum discharges to represent flood and drought conditions, respectively. The 100-year return period of annual maximum discharge is widely recognized as a benchmark for flood representation in both academic literature (Hirabayashi et al., 2013) and public services (e.g., Federal Emergency Management Agency flood maps at https://www.fema.gov/flood-maps). During such large floods, peak flows can exceed flood defense capacities in many areas worldwide (Scussolini et al., 2016). For consistency, we applied a 100-year return period threshold for droughts as well, but based on annual 7-day minimum discharge (minimum value of the mean discharge of any seven consecutive days each year), following an approach similar to Su et al. (2020). At each grid cell for every simulation, we fitted the generalized extreme value (GEV) distribution (Jenkinson, 1955) to the 200-year pre-industrial time series of annual maximum daily and minimum 7-day discharges using L-moment estimators (Hosking & Wallis, 1997) to determine the magnitudes of 100-year floods and droughts under pre-industrial climate conditions (PI-100 floods and droughts). Unlike methods that fit historical records, our approach provides a more robust estimate of 100-year extremes and avoids potential frequency biases that can arise when the historical period is used both for fitting extreme functions and as a reference for future changes (F. Zhao et al., 2024). We selected the GEV distribution for both flood and drought analyses based on its satisfactory overall goodness-of-fit, as indicated by the probability plot correlation coefficient (Vogel, 1986) for each grid cell's fit (Figures S5 and S6 in Supporting Information S1).

We assessed flood and drought changes in the YRB at GWLs to link global mean temperature (GMT) changes with regional-scale societal consequences, thereby better communicating the urgency of climate change mitigation and adaptation. For each GCM, the 1.5, 2, and 3°C GWLs were defined as the first instances when the 31-year running means of projected global average annual mean temperature reached these levels above those of the pre-industrial period. The center years of these 31-year time slices were computed by ISIMIP to facilitate consistent impact study comparisons (https://www.isimip.org/protocol/isimip3b-temperature-thresholds-and-time-slices/, Table S3 in Supporting Information S1). We extracted simulated discharge for the 31-year periods centered on these years, and fitted their annual maximum daily and minimum 7-day discharges with the GEV distribution using the same procedure as the piControl simulations. For each piControl simulation (10 GCM × 4 experiment sets), we computed the return period of the PI-100 floods and droughts at all grid cells for each of the future simulations representing different GWLs. For each experiment set, we pooled all the computed results to quantify changes in flood and drought frequency at each GWL. The pooling was done under the assumption that the relationship between GMT change and flood (drought) change is scenario-independent.

Additionally, we identified simulated historical and future annual maximum (minimum 7-day) discharges exceeding PI-100 flood (drought) thresholds at each grid cell as occurrences of the corresponding extreme events. We then combined the GCM- and experiment-specific occurrences of extreme events with socio-economic data. For each year, we calculated the YRB's area, population, and GDP (population and GDP were held at 2015 levels) affected by PI-100 floods and droughts at least once. Then, we used all historical and future simulations to quantify changes in flood and drought impacts at different levels of global warming.

Previous research indicates that numerous temperature and precipitation extremes scale with GMT regardless of emission scenarios (Seneviratne et al., 2016). In this study, we investigated whether flood and drought exposure in the YRB exhibit similar scaling relationships with GMT (changes in flood or drought exposure per GMT increase). For each GCM and experiment set, we computed averages of relative impact changes (area, population, and GDP affected by PI-100 floods or droughts relative to 1971–2000) and GMT anomalies (relative to pre-industrial levels) over 10-year moving windows. The moving average values were then assigned to the last year of each window period (e.g., the value for 2000 represents the average during 1991–2000).

2.2.5. Decomposing Relative Contribution to the Variance of Projected Impacts

To quantify the relative importance of GCMs, SSPs, and experiment sets (EXP) to the projected impacts (total area/population/GDP affected in the YRB) of future floods and droughts, we analyzed the total variance of the full projection ensemble (varTotal) as a measure of uncertainty. We assumed the total variance comprises the sum of the variance across all GCMs (varGCM) after averaging all GCM-driven simulations, the variance across all SSPs (varSSP) after averaging all simulations for each SSP, and the variance across all experiments (varEXP) after averaging all simulations for each experiment, along with a cross term (cross.terms.) describing the covariance between GCM, SSP, and EXP responses. This cross term is assumed to represent the difference from unity, as detailed in Equation 3.

$$1 = \frac{\text{varGCM}}{\text{varTotal}} + \frac{\text{varSSP}}{\text{varTotal}} + \frac{\text{varEXP}}{\text{varTotal}} + \text{cross.terms.} \tag{3}$$

Given that varGCM may exhibit a larger proportion due to unbalanced samples (10 GCMs, 3 SSPs, and 4 EXPs), we randomly selected 4 GCMs out of the 10 (yielding a total of 210 combinations) to compute the variance terms in Equation 3 and present the median values from all samples.

3. Results

3.1. Model Performance for the Four Experiment Sets

Figure 2 illustrates the model's performance at four stations along the Yangtze River for the four experiment sets (unCal, Cal_L, Cal_S, and Cal_H) in the calibration period (2008–2016). The uncalibrated simulation exhibits unsatisfactory KGE scores, ranging between 0.41 and 0.54 at the four stations. The annual mean discharge is underestimated by approximately 21%–31%, and peak discharges at upstream stations generally lag behind observed peaks by about 3 months. Model performance in simulating low flow is also poor, as evidenced by the low NSElog scores.

Calibration significantly improved model performance, with all three calibrated experiments achieving KGE scores ranging between 0.84 and 0.91 for the calibration period and 0.81 to 0.90 for the validation period. Biases were within ±10% for the calibration period and ±15% for the validation period at the four stations (results for the validation period are shown in Figure S7 in Supporting Information S1). After calibration, the simulated mean seasonal cycles aligned more closely with observations. According to the quantile-quantile (Q-Q) plot, the Cal_H experiment captures the magnitude of high flows well. However, the other two calibrated experiments significantly overestimate the magnitude of the largest peak flow events at Hankou and Datong, while they tend to underestimate these events at the upstream stations. The average KGE score among the stations is highest in the Cal_S experiment. Similarly, the Cal_L experiment shows the best representation of low flow, as indicated by the NSElog scores.
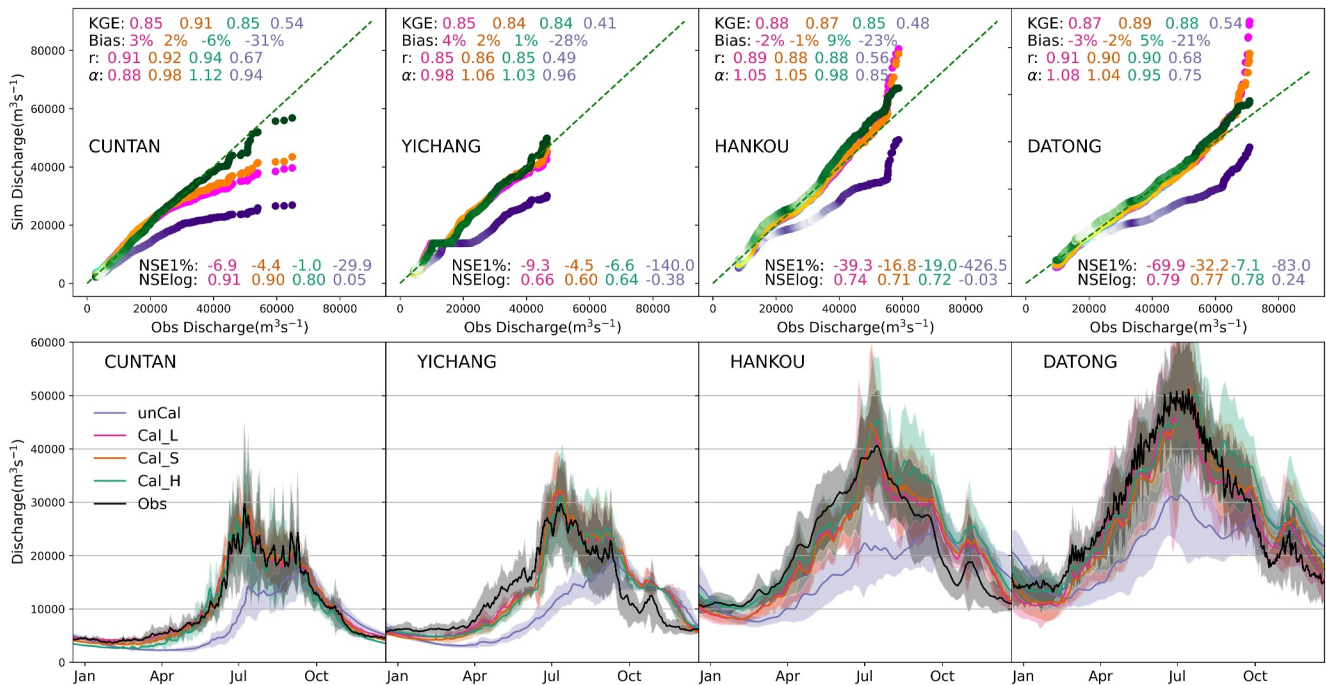
**Figure 2.** Model performance at four stations for the four experiment sets (unCal, Cal_L, Cal_S, and Cal_H). The top panels display Quantile-Quantile (Q–Q) density plots comparing daily observed and simulated discharge distributions for the calibration period (2008–2016). Brighter colors indicate higher data density. The lower panels illustrate the mean seasonal cycles of observed (black) and simulated daily discharge for the same period (solid lines represent the median values, and shadings denote the one standard deviation range).

3.2. Effect of Calibration on Projected Flood and Drought Frequency

Using projected results from the uncalibrated and calibrated experiments, we examined how frequently floods (droughts) with magnitudes greater than the PI-100 floods (droughts) would occur at 1.5, 2, and 3°C warming above pre-industrial levels at the four stations and the likelihood of projected changes following IPCC terminology. The IPCC quantifies the agreement among model ensemble members, which commonly serves as the basis for assessing the likelihood of projected changes. A projected change is assessed as "*likely*" if 66%–100% of the model ensemble members agree on the direction of the change, "*very likely*" with 90%–100% model agreement, and "*virtually certain*" with 99%–100% model agreement (IPCC, 2023b).

According to the ensemble median results from the unCal experiment, the PI-100 floods are projected to occur slightly more frequently at the downstream stations (Hankou, Datong) and less frequently upstream (Cuntan). However, agreement on future flood change directions among the GCM-SSP combinations, as indicated by the numbers in the box plots (Figure 3b, dark gray bars), is lower than 66% in most cases (except for Cuntan at 1.5°C warming). For all stations and GWLs, there is a slight tendency for more frequent floods at higher warming levels, as indicated by shorter return periods of the PI-100 floods (Figure 3a, dark gray bars). In contrast, the PI-100 droughts are projected to occur about once every 30 years at most stations and GWLs based on the ensemble median, except for Yichang at 1.5 and 2°C warming. The projected drought frequency increases are *likely* in most cases and *very likely* for Hankou and Datong at 3°C warming. There is no clear tendency regarding whether droughts will become more frequent after passing 1.5°C warming.

Based on the ensemble median results, calibration leads to an increase in projected flood frequency at some stations and GWLs but a decrease in other cases. The three calibrated experiments consistently increase flood frequency at Cuntan, decrease it at Yichang, and generally reduce flood frequency at Hankou and Datong at 1.5 and 2°C warming. In 4 out of the 12 cases (4 stations × 3 GWLs), the agreement level regarding flood frequency change is consistently altered across all three calibrated experiments. Based on any of the three calibrated simulations, it is *likely* that there will be more frequent PI-100 floods at Cuntan at 3°C warming and less frequent PI-100 floods at Yichang at 1.5 and 2°C warming, both of which have low agreement levels under the unCal experiment. The PI-100 floods are no longer *likely* to occur at Cuntan at 1.5°C warming after calibration. The
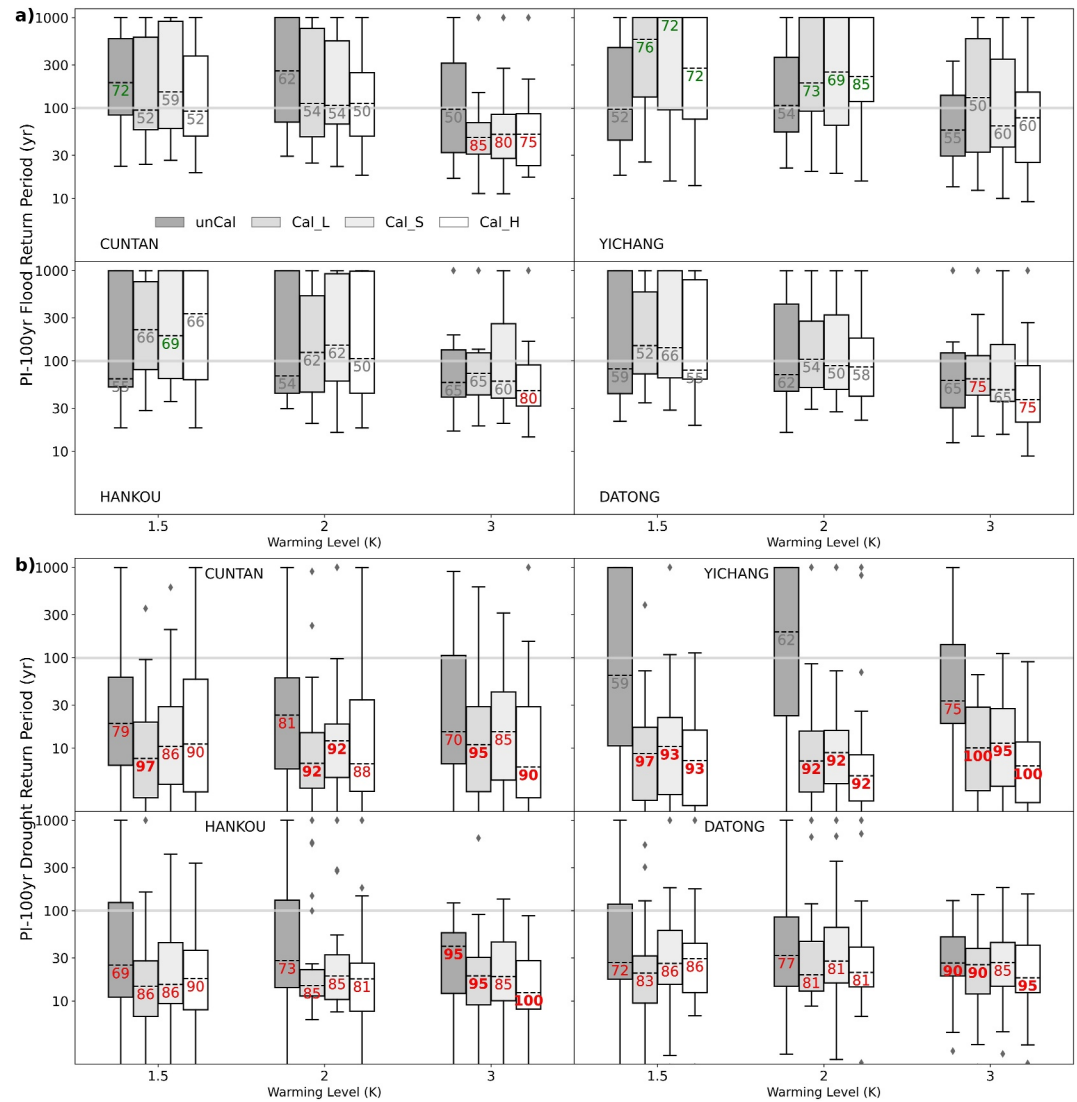
Figure 3. Projected return periods of the pre-industrial 100-year (PI-100) flood (a) and drought (b) at four stations at various warming levels and for experiment sets. The height of the boxes represents the interquartile range (75th–25th percentile), with the dashed line within each box denoting the median value. Solid lines depict the maximum and minimum return periods across all GCM-SSP combinations. A total of 29, 26, 20 GCM-SSP combinations contribute to the boxes for 1.5, 2, and 3°C warming, respectively (Table S3 in Supporting Information S1). Model consistencies are represented as percentage numbers within the boxes, indicating the proportion of models agreeing on the direction of change. Directions of *likely* changes, with model agreement ranging from 66% to 100%, are indicated by color: green for longer return periods or less frequent events, and red for shorter return periods or more frequent events. Bold numbers signify *very likely* changes, with model agreement ranging from 90% to 100% (IPCC, 2023b). Note that numbers are rounded to the nearest integer, meaning a value shown as 66 may actually be just below the threshold and thus depicted in gray.

Cal_H experiment, optimized for peak flow performance, detects two additional *likely* changes to a higher frequency at Hankou and Datong at 3°C warming (Figure 3a). Results for each individual GCM can be found in Figure S8 in Supporting Information S1.

In most cases, the PI-100 droughts are projected to become even more frequent after calibration, occurring approximately once every 10 years at Cuntan, Yichang, and Hankou based on the ensemble median results. Calibration generally reduces model spread across all cases, as evidenced by the narrower interquartile ranges compared to the uncalibrated simulation. The likelihood of projected drought frequency increases is also greatly enhanced at Cuntan and Yichang. At the Yichang station, it is *very likely* that droughts will occur more frequently according to all three calibrated experiments, and the change is *virtually certain* at 3°C warming based on the
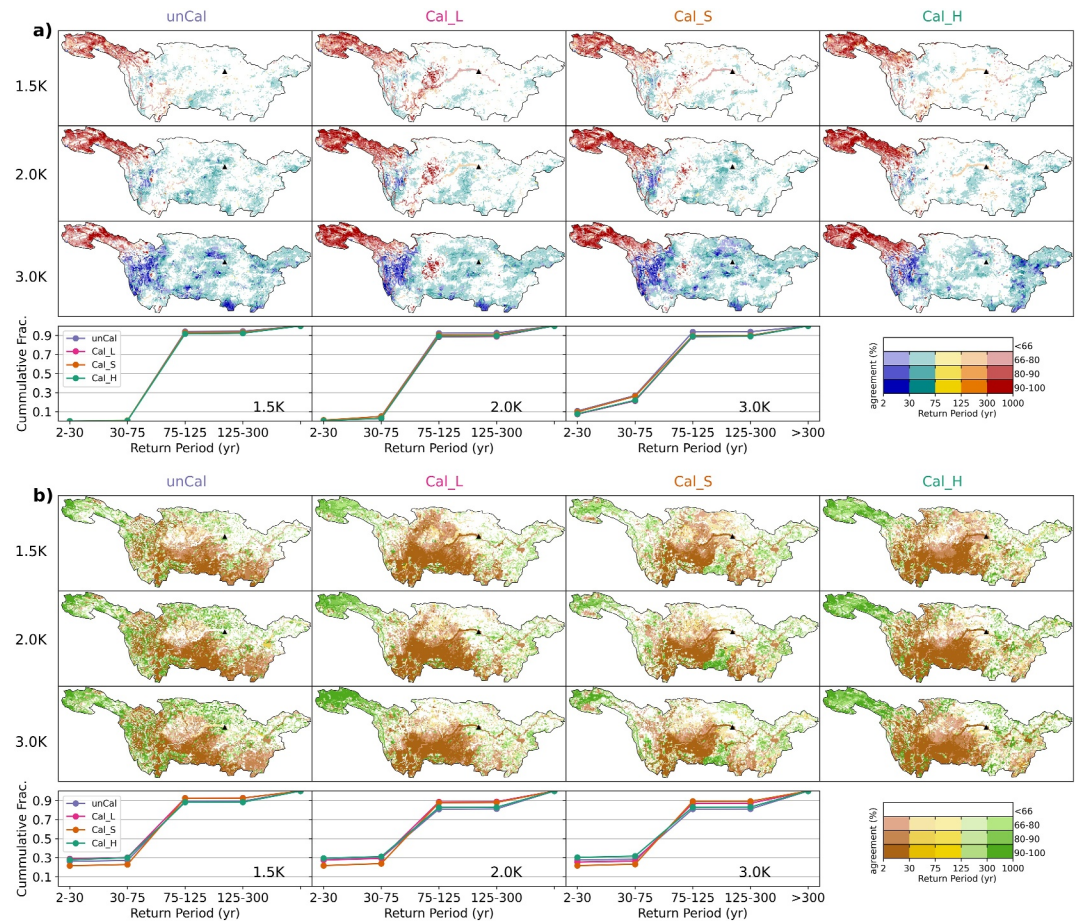
Figure 4. Projected changes in (a) flood and (b) drought frequency across the entire basin at 1.5, 2, and 3°C warming compared to the piControl simulation, under the unCal, Cal_L, Cal_S, and Cal_H experiments. *Upper three rows*: Color hues indicate the multimodel median change, while saturation levels represent the agreement on the direction of change across all GCM-SSP combinations (percentage of model runs agreeing on the direction). The location of Three Gorges Dam is marked with a triangle. *Bottom row*: Cumulative basin area fraction represented under each return period category.

Cal_L and Cal_H experiments. At the Cuntan station, the frequency of droughts becomes *very likely* to increase in the majority of cases based on the three calibrated experiments. After calibration, there is still no clear tendency regarding the sensitivity to drought frequency change at the three GWLs (Figure 3b).

Discharge changes at the stations indicate future floods and droughts around the main river channel. To further assess changes at the local level, we analyze future changes in flood and drought frequency and the consistency of the GCMs at each grid across the entire basin (Figure 4). Blue (brown) colors indicate that the ensemble generally agrees that the frequency of PI-100 floods (droughts) will increase, while red (green) colors indicate agreement on a decrease in the frequency of PI-100 floods (droughts). Lighter colors indicate lower agreement among the ensemble members, and white color represents areas where fewer than 66% of the ensemble members agree on the direction of change.

The multi-model median across all GCMs for the four experiments (Figure 4a) indicates that, generally, the frequency of PI-100 floods is projected to increase in the middle and lower parts of the basin, and decrease in the northern part of the upper Yangtze basin. At higher GWLs, there are stronger and more consistent changes in flood frequency. Specifically, a robust increase in flood frequency is observed in areas around the TGD, in larger parts of the downstream area, and in some of the southwestern region. According to the unCal experiment, the PI-100 floods are projected to occur every 75 years or more frequently in approximately 0.5%, 3%, and 22% of the entire basin at 1.5, 2, and 3°C warming, respectively (Figure 4a, last row).
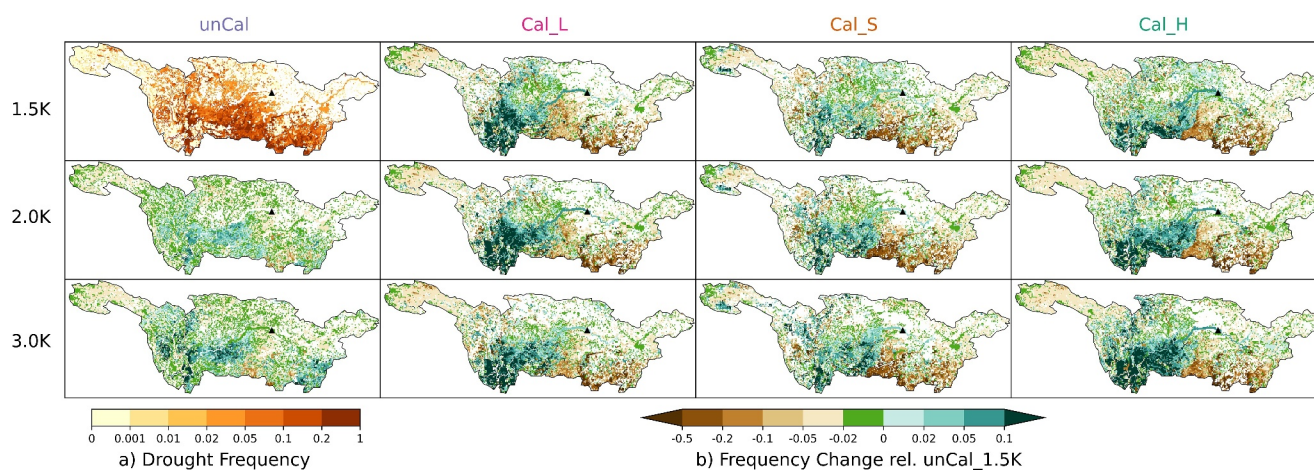
Figure 5. Differences in projected drought frequency changes across the entire basin at 1.5, 2, and 3°C warming levels for all experiments, relative to the projected drought frequency from the unCal simulation at 1.5°C warming (first panel). The multi-model median values are depicted. Colorbar (a) corresponds to the first panel (top left), while colorbar (b) applies to all subsequent panels. The location of Three Gorges Dam is marked with triangle.

Drought frequency is projected to decrease in the northwestern region of the basin and increase in the southern part. The spatial patterns of drought frequency changes appear less sensitive to GWLs. According to the unCal experiment, the PI-100 droughts are projected to occur more than once every 30 years in about 25% of the basin area at all GWLs (Figure 4b, last row).

For some experiments and GWLs, calibration leads to differences, including a decrease in flood frequency along the main river channel in the center of the basin, and in the central-west area of the basin. However, the large-scale patterns of flood frequency change are similar across the four experiment sets. In all experiments, there is a clear expansion in the area of increasing flood frequency as GWL increases. Conversely, the overall increase in drought frequency in the region depends on the choice of calibration method. After calibration, the previously high model agreement in the southeastern region diminishes; the central-north region shifts from a decreasing to no consistent change or increasing drought frequency; and the middle section of the Yangtze River main stream is projected to experience more frequent droughts with high model consistency. Furthermore, calibration may influence the pattern of drought frequency changes associated with increasing GMT. While all three calibrated simulations increase drought frequency in the southwestern part of the basin and decrease it in the southeastern part, notably more frequent droughts at higher GMT levels are observed only in the Cal_H and the uncalibrated simulation (Figure 5).

3.3. Effect of Calibration on the Projected Area Affected by Flood and Drought

Next, we evaluate how the basin total area affected by PI-100 floods (droughts) changes with different warming levels, computed by aggregating for each year the total area of the grids with magnitudes greater than the PI-100 floods (droughts) occurring at least once. Results indicate that the area affected by floods scales nearly linearly with global warming from 0.5°C up to over 3°C, after which the rate of increase in affected area accelerates (Figure 6). To quantitatively assess the scaling relationship, we apply the Akaike Information Criterion (AIC) (Akaike, 1998) across the four experiment sets and three SSPs to determine whether a linear, exponential, or polynomial model best fits data in this temperature range. Results show that the linear model is preferred in most cases (Table S4 in Supporting Information S1).

Based on the unCal experiment under the SSP5-8.5 scenario, the ensemble median projected area affected by PI-100 floods relative to the 1971–2000 mean increases by an additional 0.4%, 0.5%, 1.4%, and 3.6% of the basin's total area at 1.5, 2, 3, and 4°C warming. The model spread increases significantly at higher warming levels, partly due to the decreasing number of GCM-SSP combinations. Projections under different SSPs generally indicate a roughly similar trend of increasing area affected as GMT increases.

Conversely, according to the unCal experiment under the SSP5-8.5 scenario, the ensemble median projected additional area affected by PI-100 droughts relative to the 1971–2000 mean increases with global warming up
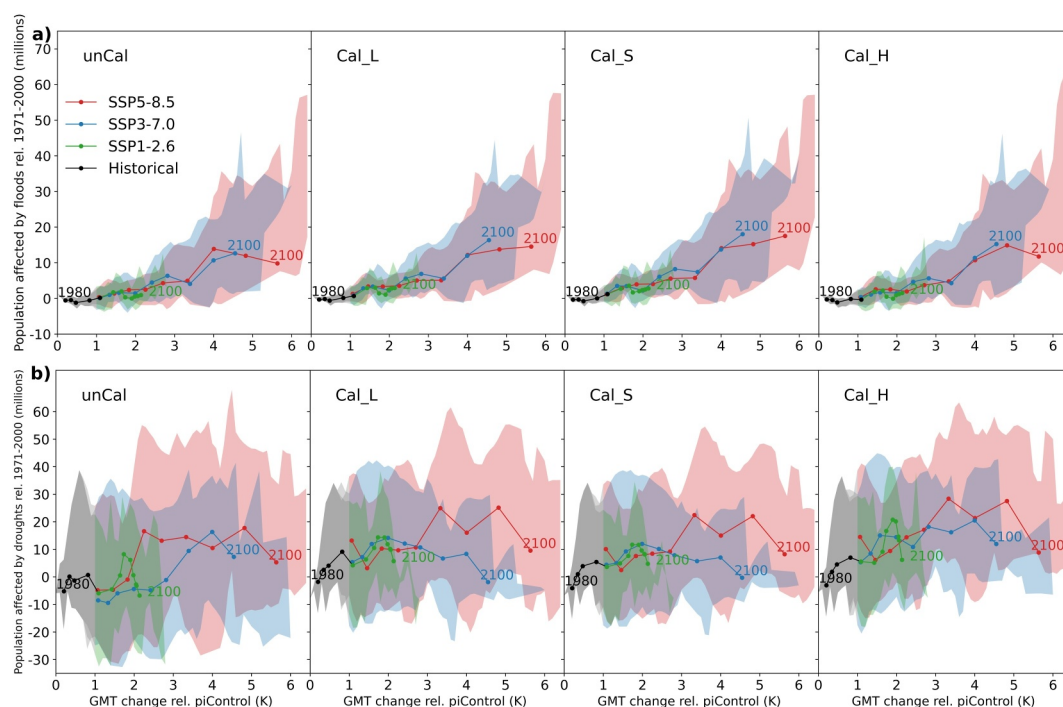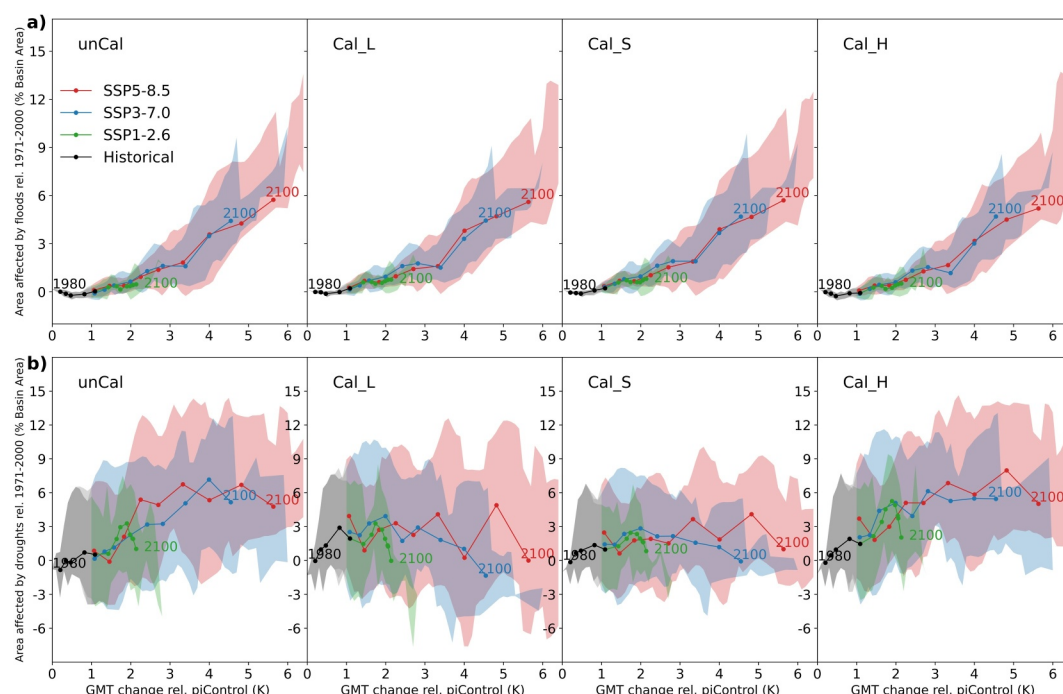
Figure 6. (a) Scaling between relative changes in the projected area affected by PI-100 flood and (b) drought and changes in global mean temperature (GMT) under the unCal, Cal_L, Cal_S, and Cal_H experiments. GMT changes are relative to the piControl global mean, and projected area changes are relative to the mean of 1971–2000. The solid black line denotes the ensemble median for historical runs until 2020 (combined with SSP5-8.5 for 2015–2020), while the solid red, blue, and green lines represent the ensemble medians of future projections under the SSP5-8.5, SSP3-7.0, and SSP1-2.6 scenarios, respectively. The colored shaded areas indicate the ±1 standard deviation (SD) model spread (capped at the minimum value) for all simulations under their respective scenarios.

until 3.5°C warming, then the increase plateaus. At 1.5, 2, and 3°C warming, the area affected by PI-100 droughts is projected to expand by an additional 0%, 3.7% and 7.6%, respectively. The model spread is notably larger than that of floods at all warming levels. There is also more variance among the SSPs than that of floods.

Calibration has only a minor effect on the total basin area exposed to floods, projecting an additional 0.7% (0.8%/0.4%), 0.7% (0.9%/0.5%), 1.5% (1.6%/1.4%), and 3.8% (3.9%/3.1%) of the basin's total area affected by PI-100 floods at 1.5, 2, 3, and 4°C warming according to the Cal_L (Cal_S/Cal_H) experiment. The model spread is similar across different experiments. The scaling relationships for each individual GCM-SSP combination also demonstrate a relatively minor overall effect of calibration, with more noticeable variability at higher degrees of warming (Figure S9 in Supporting Information S1).

In contrast to floods, the different experiment sets significantly impact the total area of the basin exposed to drought (Figure 6b). For the Cal_L and Cal_S experiments, drought exposure remains unchanged or even slightly reduces as GMT increases. Conversely, for the unCal and Cal_H experiments, drought exposure increases approximately linearly until around 3.5°C warming. The scaling relationships for each individual GCM-SSP combination demonstrate a much larger overall effect of calibration, with increased model spread and potential changes in direction at higher degrees of warming (Figure S10 in Supporting Information S1). Unlike floods, where the affected area significantly expands at higher degrees of warming, the increase in the area affected by drought with warming based on the unCal or Cal_H experiment is primarily driven by more frequent drought occurrences in the same regions (Figure 5).

Although the GWL approach has been widely adopted in recent IPCC assessments, the hydrological response for droughts at the same GWL can vary depending on whether that warming level is reached earlier or later due to cumulative hydrological changes. To assess how this might affect our results, we also evaluated how the basin-wide total area affected by PI-100 floods and droughts changes over time. Our key findings remain robust: calibration significantly affects projected changes in drought-affected areas more than flood-affected areas, and

Figure 7. (a) Scaling between relative changes in the projected population affected by PI-100 flood and (b) drought and changes in global mean temperature under the unCal, Cal_L, Cal_S, and Cal_H experiments. Symbols and color schemes follow Figure 6.

flood-affected areas consistently increase over time in all calibrated experiments (Figure S11 in Supporting Information S1). Using this time-based approach, different calibration experiments show more agreement in the general direction of projected ensemble median changes for droughts. Compared to the unCal and Cal_H experiments, the Cal_L and Cal_S experiments project approximately half the increase in additional area affected by droughts toward the end of the century.

3.4. Effect of Calibration on the Projected Population and GDP Affected by Flood and Drought

The implications of projected changes in floods and droughts for human society can be considered by assessing the present and future population or GDP at risk. For floods, we calculated the sum of the population and GDP on 5′ grids where the annual maximum/minimum discharge exceeds the PI-100-year flood or drought threshold. As only low elevation areas in these 5′ grids will be potentially inundated when floods occur, this calculation also includes populations and GDP indirectly affected by disruptions such as traffic or activity interruptions. For droughts, while an earlier study assumed only the rural population was exposed (Lange et al., 2020), we consider all individuals on those grids to be affected, including urban populations impacted by increased agricultural product prices and water level declines in rivers and reservoirs. Additionally, we used constant 2015 population and GDP data for the computation (in line with the ISIMIP3 protocol) and maintained consistency in the calculations for both floods and droughts, as our primary interest lies in determining whether the effect of calibration differs between these two extremes.

Similar to the affected area (Figure 6a), the population affected by PI-100 floods generally increases with global warming (Figure 7a). According to the ensemble median projections under the unCal experiment, an additional 1.6, 2.1, 4.4, and 13.8 million people and an additional 18, 24, 59, and 159 billion dollars (2011 US$PPP) GDP in the YRB will be affected by PI-100 floods at 1.5, 2, 3, and 4°C warming. Changes in population and GDP affected by floods scale approximately linearly with global warming from 0.5°C up to over 3°C (Table S4 in Supporting Information S1), after which the model variability becomes more pronounced compared to the affected area (Figure 6). Different SSPs affect the scaling between increasing population or GDP affected and GMT increases more than that of the affected area, especially when global warming is 4°C above the pre-industrial level.

For droughts, the ensemble median projected population and GDP affected also tend to increase with global warming based on the unCal experiment. However, impacts of different SSPs on population and GDP affected by droughts are more pronounced (Figure 7b and Figure S12b in Supporting Information S1). We observe similar patterns between population and GDP affected by droughts in the YRB, as densely populated regions are also the most affluent (Figure S3 in Supporting Information S1).

Calibration does not alter the projection that the population (Figure 7a) or GDP (Figure S12a in Supporting Information S1) affected by floods generally increases under global warming, projecting an additional 3.5 (3.3/2.7), 3.2 (4.1/1.5), 5.2 (6.2/3.9), and 12.1 (14.1/10.7) million people affected by PI-100 floods according to the Cal_L (Cal_S/Cal_H) experiment at 1.5, 2, 3, and 4°C warming under the SSP5-8.5 scenario, respectively. Similarly, the additional multi-model median GDP affected by floods is 37 (34/32), 37 (45/21), 73 (85/51), and 131 (155/119) billion (2011 US$ PPP) based on the Cal_L (Cal_S/Cal_H) experiment under the SSP5-8.5 scenario at 1.5, 2, 3, and 4°C warming, respectively. The AIC scores support a near-linear scaling relationship from 0.5°C up to over 3°C, which does not change in most cases (Table S4 in Supporting Information S1).

Unlike the affected area, the impact of calibration on population and GDP affected by PI-100 droughts depends on the SSP scenarios. Under the SSP5-8.5 scenario, calibration does not change the direction of the projected increase in population or GDP affected by PI-100 droughts. However, under the SSP3-7.0 scenario, population or GDP affected by PI-100 droughts does not increase with global warming for the Cal_L or Cal_S experiments. Similar to the affected area, the scaling relationships for each individual GCM-SSP combination demonstrate a significantly larger overall effect of calibration on drought impacts (Figures S13–S16 in Supporting Information S1).

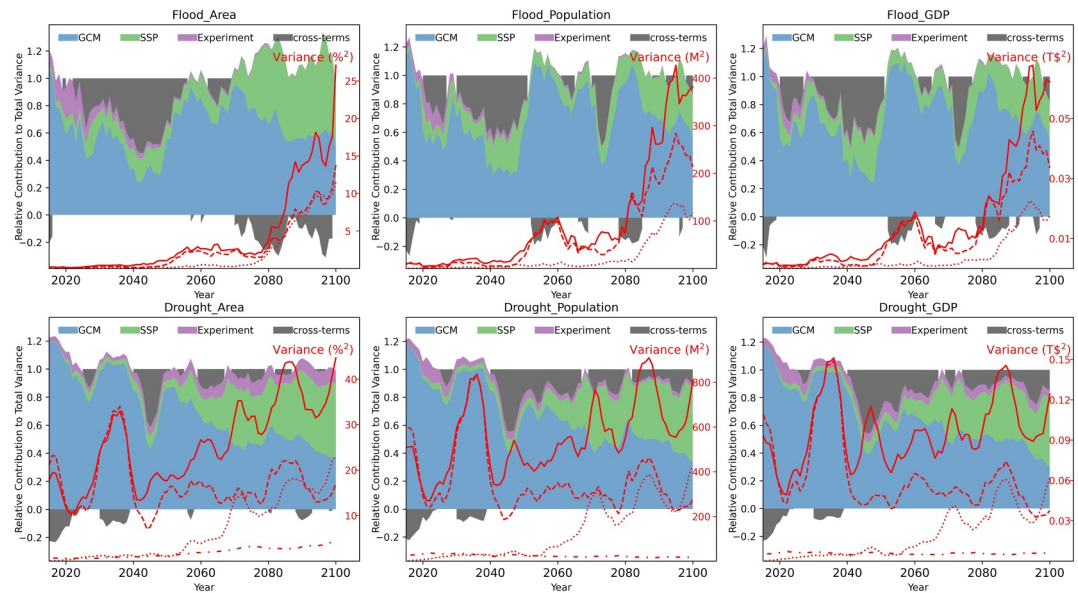### 3.5. Relative Contribution to the Variance of Projected Impacts

To further quantify the relative importance of GCMs, SSPs, and experiment sets on projected impacts (total area, population, and GDP affected in the YRB) of floods and droughts, we examined the total variance of the full projection ensemble as a measure of uncertainty. With the exception of GDP affected by drought, the overall variance of ensemble projections generally increases over time, though this increase levels off or even reverses around mid-century (Figure 8, solid red lines). For flood impacts, the most substantial increase in variance is observed in the last two decades of the 21st century. A relatively linear increase in variance over time is observed for drought impacts after 2040, accompanied by significant interannual variations.

Breaking down the overall variance in projections into GCM, SSP, and experiment components, we find that for the area affected by floods, the GCM component predominates until the latter part of the 21st century. The contribution from the SSP component increases rapidly after 2070, with its importance becoming comparable to that of GCM in the final decades of the 21st century. The share of variance induced by the experiment component is small or negligible compared to other components after 2050, suggesting that calibration plays a minor role in the uncertainty of projected flood impacts at higher warming levels. The cross-terms contribute a relatively small fraction of the overall variance, except for the period between 2020 and 2050. For population or GDP affected by floods, the share of variance attributed to SSP also increases after 2080, yet its importance remains notably smaller than that attributed to GCM.

For droughts, the GCM component dominates the overall variance until the last few decades of the 21st century, when the SSP component overtakes it for the affected area, population, and GDP. The fluctuation in total variance is primarily attributed to the GCM component in the first half of the century and increasingly to the SSP component in the latter half, with larger interannual variations for population and GDP affected. Although still relatively small (contributing on average about 9%, 3%, and 6% of the total variance for the area, population, and GDP affected over the whole period, respectively), the share of variance induced by the experiment component is notably larger for droughts compared to floods after 2050. The cross-terms contribute a relatively small portion (under 20%) of the overall variance most of the time, except around 2050. Generally, we observe a larger role of different experiments on drought impacts compared to floods, aligning with the more considerable changes in drought frequency observed after model calibration (Figure 5).

A significant portion of the variance in the GCM and SSP components originates from the different times each model reaches global warming levels under the three SSPs (Table S3 in Supporting Information S1). This issue becomes more pronounced toward the end of the 21st century, as the GMT is projected to increase by a wide range

Figure 8. Relative contributions of GCMs, SSPs, and experiment sets to the ensemble projections of basin total area, population, and gross domestic product affected by floods or droughts. Red lines represent the absolute variance of the whole ensemble (solid), the GCM share (dashed), the SSP share (dotted), and the experiment share (dash-dot, only shown for drought). Absolute variance scores are adjusted for each panel and are therefore not directly comparable. Median variance shares from the 210 combinations of 4 GCMs (randomly selecting 4 GCMs from the 10 in each combination) × 3 SSPs × 4 experiments are depicted. The variance shares of GCMs, SSPs, and experiments typically do not sum to the total variance, as these three sources of uncertainty are not entirely independent. The difference from the total variance is depicted in gray and referred to as "cross-terms" (refer to Equation 3).

from 1.4 K (GFDL-ESM4, SSP1-2.6) to 7.0 K (UKESM1-0-LL, SSP5-8.5) compared to pre-industrial climate conditions. Therefore, while calibration exhibits a significant effect on the area, population, and GDP affected at different warming levels (Figures S10, S14, and S16 in Supporting Information S1), its variance remains relatively small compared to the variance from GCMs and SSPs in future decades.

For a more comprehensive assessment, we also conducted the same analyses for two additional model combinations at the 25th and 75th percentiles, ranked by average varGCM over the entire period (Figures S17 and S18 in Supporting Information S1). We observe three robust findings: a notably larger role of calibration on the affected area, population, and GDP for droughts compared to floods; an increase in the relative importance of SSP toward the end of the 21st century; and a general increase in variance terms over time. The overall and GCM variance exhibit a decreasing trend for drought-affected GDP in the 25th percentile varGCM case (lower contribution of varGCM), and remain steady or decline for drought-affected area, population, and GDP in the 75th percentile varGCM case (higher contribution of varGCM). This is consistent with the earlier finding that the model spread of drought impacts does not increase at higher degrees of warming (Figures 6 and 7; Figure S12 in Supporting Information S1).

4. Discussion

4.1. Why Calibration's Impact on Flood Changes in the YRB Is Relatively Minor

Our findings indicate a projected increase in the area affected by floods in the YRB as global warming intensifies, consistent across all experiment sets. Conversely, the expansion of areas affected by droughts is observed only in some experiments (Figure 5). This suggests that changes in floods and droughts have different sensitivities to calibration, which we refer to as the asymmetric effect of calibration. The observed asymmetric effect of calibration is primarily due to the significant role of extreme precipitation in flood generation, while long-term trends in evapotranspiration influence drought occurrence. To illustrate this, we conducted further analyses on Yibin (Figure 1), a representative upstream subbasin in the YRB projected to experience more frequent floods and
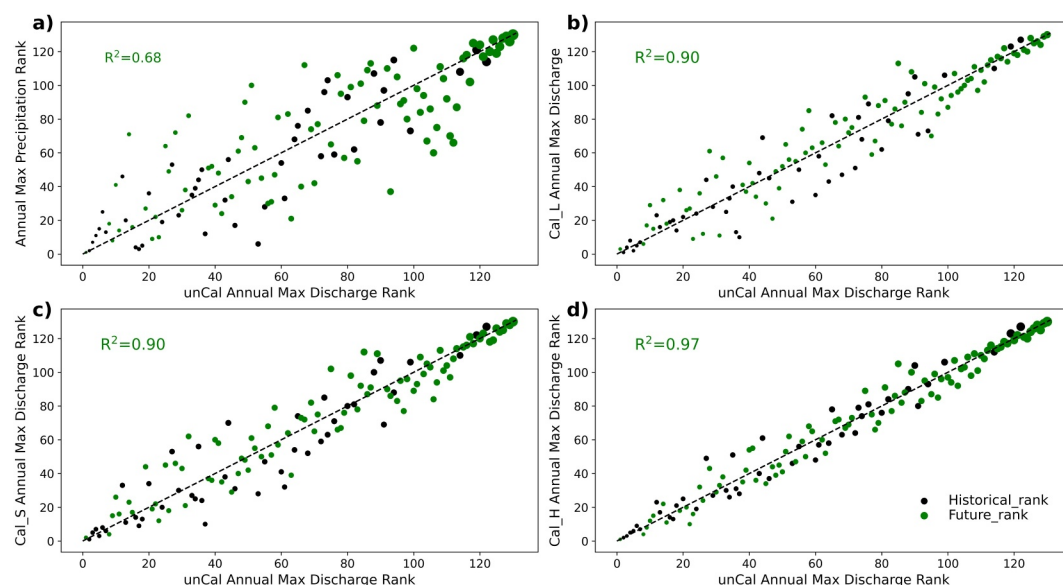
Figure 9. (a) Rank correlation between uncalibrated annual maximum daily discharge and annual maximum daily precipitation, (b) calibrated annual maximum daily discharge from the Cal_L, (c) Cal_S, and (d) Cal_H experiments from 1971 to 2100 for the Yibin subbasin in the upstream of the YRB (Figure 1), based on one GCM-SSP combination (UK-ESM1-0-LL, SSP5-8.5). Black dots represent the ranks for historical years (1971–2014), while green dots represent future years (2015–2100). Higher ranks indicate larger events, and the size of the dots corresponds to the magnitude of events.

droughts, using experiment results under a specific GCM-SSP combination (UK-ESM1-0-LL, SSP5-8.5). The findings from these analyses will be examined in detail here and in Section 4.2.

Specifically, we ranked annual maximum precipitation and discharge (higher numbers indicate larger values) and depicted their relationship in the unCal experiment (Figure 9a). A relatively strong correlation ($R^2 = 0.68$) confirms extreme rainfall's dominant effect on floods, evidenced by years with the largest maximum discharge consistently ranking high in annual maximum precipitation. Subsequently, we investigated how calibration affects the ranking of maximum discharge. The results show a high correlation ($R^2 \geq 0.9$) between calibrated and uncalibrated maximum discharge rankings, with the highest correlation ($R^2 = 0.97$) in the Cal_H experiment, which optimizes high-flow model performance. This suggests that calibration has a minor effect on the relative rank of flood events. This is especially evident in the largest floods, where calibration barely changes the rank: the largest dots in the upper right corner of panels b–d (Figure 9) lie very close to the 1:1 line.

Notably, consistent with the Clausius–Clapeyron equation (Pall et al., 2007), the annual maximum precipitation in the YRB has exhibited an approximate 7% increase per degree of global warming, based on median sensitivity across the stations (M. Lü et al., 2018). As the occurrence of the largest floods is closely linked to extreme precipitation, their frequency is expected to increase with the anticipated rise in extreme rainfall intensity. Since precipitation serves as a component of climate forcing, changes in extreme flood frequency and impacts driven by precipitation remain unaffected by calibration. Although calibration can notably alter flood magnitude by modifying factors such as infiltration, runoff concentration, soil, and reservoir storage (see Text S1, Figures S19, S21, S22, and Table S5 in Supporting Information S1 for a more detailed discussion), it simultaneously affects both the historical reference and future periods. Our analyses show that this does not significantly impact the relative ranks of the largest flood events. Consequently, our examination of changes in the 100-year flood frequency and its impacts reveals a relatively minor effect of calibration. However, the results may differ for less extreme floods (e.g., 2-year return period), where antecedent conditions play a more significant role. Additionally, in dry or snow-dominated basins, where floods are more influenced by evapotranspiration and snowmelt trends, calibration may have a considerably larger impact.
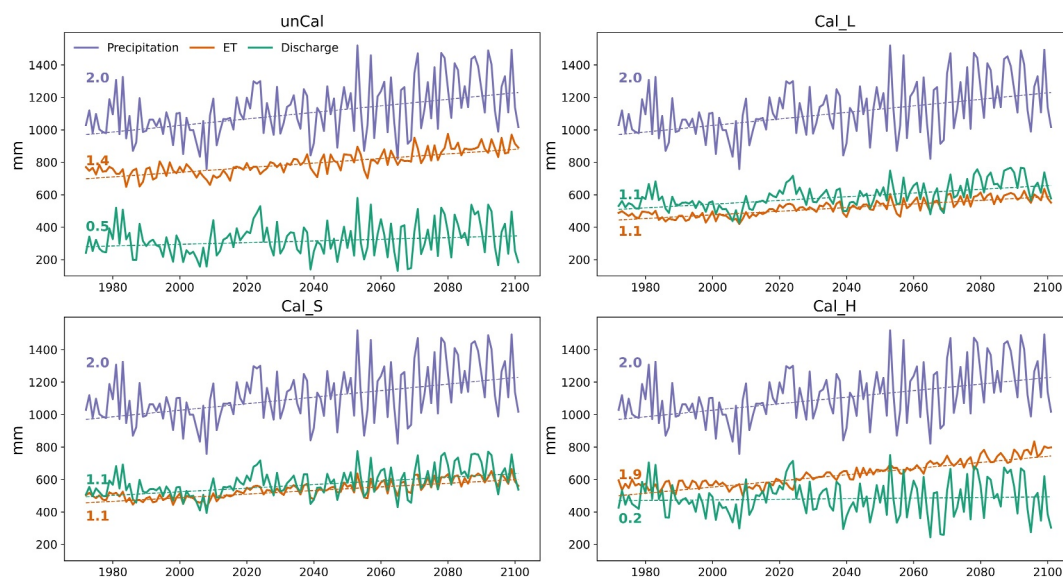
Figure 10. Annual mean precipitation, total evapotranspiration (ET), and discharge from 1971 to 2100 for the Yibin subbasin in the upstream of the YRB (Figure 1), based on one GCM-SSP combination (UK-ESM1-0-LL, SSP5-8.5), for the unCal and three calibrated experiments. Sen's slope trendlines (dashed) and their magnitudes are depicted.

4.2. Why Calibration has a Larger Impact on Drought Changes in the YRB

Hydrological droughts often develop from dry antecedent conditions, closely related to the basin's annual hydrological balance. The increasing frequency and impacts of droughts found in some experiments are *likely* related to enhanced evapotranspiration under a warmer climate. Precipitation and evaporation directly influence runoff, and their trends significantly impact changes in floods and droughts. Using the same data described in Section 4.1, we demonstrate that while annual precipitation notably increases by 2 mm/year, the annual mean discharge in the uncalibrated run increases only slightly by 0.5 mm/year, due to a 1.4 mm/year increase in annual total evapotranspiration (Figure 10). Calibration does not affect the trend of precipitation (2 mm/year), but the trend of evapotranspiration is either suppressed (to 1.1 mm/year in Cal_L and Cal_S) or elevated (to 1.9 mm/year in Cal_H). As a result, trends in annual mean discharge, which approximate the difference between trends in precipitation and evapotranspiration, change accordingly. In summary, calibration can alter the trend in evapotranspiration, which in turn affects the trend in discharge and consequently influences the relative change in drought frequency.

Unlike the case with floods, the calibrated experiments show significant differences in the direction of changes in drought magnitude (see Text S2, Figure S20, and Table S6 in Supporting Information S1 for more detailed discussion). Previous studies (Munoz-Castro et al., 2023; Najafi et al., 2011; Seiller et al., 2017; Vansteenkiste et al., 2014) reported similar findings, highlighting that calibration has a more pronounced impact on low flows or in drier basins. We anticipate that the asymmetric effect of calibration on floods and droughts identified in this study might also be present in other warm and wet basins, like those in Southeastern Asia and tropical Africa, where accurate model calibration is frequently challenging due to limited data availability.

4.3. Limitations and Future Implications

We have addressed the projection uncertainty arising from the selection of meteorological forcing data (primarily different GCMs), SSP scenarios, and model parameters (Figure 8). However, we only quantified uncertainty in the pure effect of climate change impacts, yet potentially more significant uncertainty may arise from future human activities and socioeconomic changes. Given our primary focus on the asymmetric effect of calibration, we did not account for the impact of future changes in population, GDP, and water consumption, which could vary significantly under different socioeconomic scenarios. Previous studies (Chen et al., 2014; Nie, Zhang, Liu, et al., 2021; S. Yang et al., 2015) indicate that water consumption in socioeconomic activities has a minimal net impact on streamflow in the YRB, while the regulation of large reservoirs can effectively mitigate extreme floods

and droughts. Nonetheless, the significant and consistent increases in areas exposed to floods and droughts under global warming, as projected in this study, underscore the necessity for robust climate mitigation and adaptation strategies to address the challenges of future floods and droughts in the YRB. Regarding the asymmetric effect of calibration on projected flood and drought changes, considering that the calibrated model demonstrates robust performance and the experiments use consistent forcing data in the simulations, we consider that the main conclusions of this study are reliable.

We demonstrate that calibration can significantly improve the simulation of discharge in the YRB. Calibration also enhances the robustness of projected increases in drought frequency along the main river of the Yangtze and alters the detection of flood frequency changes in several instances. While calibration clearly benefits the reliability of projected changes in floods and droughts, it is important to note that over-tuning a model through calibration may compromise its performance under changed climate conditions (outside of its calibration zone) and the credibility of projections (Viney et al., 2009). Even a well-calibrated model may not be reliable under significant climate change (Merz et al., 2011), as future climate conditions may extend beyond the range for which the parameters were calibrated. For example, this includes scenarios where hydrological processes are influenced by management practices that were not included in the calibration. Therefore, improved process representation, including management practices, and enhanced calibration methods are both necessary for a more robust projection of the climate change impact on regional hydrological extremes. Nevertheless, fully optimizing the model is unlikely to alter the key findings of our study.

Using the YRB as an example, we find that projected drought changes are highly sensitive to model calibration parameters. Calibration has been identified to have a significant impact in more arid areas and on low flows (Chegwidden et al., 2019). We demonstrate that even for a large river basin with a wet climate, trends in region-aggregated drought impacts can have different directions depending on the choice of objective functions used during calibration. Future studies are necessary to better evaluate how and why hydrological models differ in their sensitivities to changes in low flows, and to better understand how climate change will impact future droughts. Improved process understanding and careful model calibration are both necessary to provide robust projections of future drought changes and to guide mitigation and adaptation efforts. The Surface Water and Ocean Topography (SWOT) satellite, launched in 2022, will vastly expand measurements of global rivers, providing crucial data for model calibration and improving our understanding of hydrological extremes (Biancamaria et al., 2016).

Conversely, flood changes in this warm and humid basin are predominantly influenced by precipitation. Similar to Najafi et al. (2011), we find that uncertainty in hydrological model parameters is significantly lower than the uncertainty in the climate projections from GCMs used as hydrological model forcing data. Concurrently, for this basin and based on our model ensemble, the influence of model parameters on projected flood changes under climate change is relatively small. Similarly, Niraula et al. (2015) concluded that calibration was less important for analyzing the relative change in streamflow due to climate change compared to that due to land use change. Since precipitation is not altered by calibration, it is reasonable to deduce that calibration likely also only has a minor effect on changes in the 100-year flood frequency and its impacts in other warm and humid regions where extreme precipitation play a significant role in flood generation (see Section 4.1 for detailed discussion). This suggests that projected large-scale increases in flood frequency and impacts in those regions, such as parts of East Asia, Southeast Asia, and tropical Africa, from uncalibrated hydrological models, could still be quite reliable. These regions have also been identified as consistently high flood-risk areas in previous studies (Lange et al., 2020). The calibration workflow in our study can be readily applied in future research across similar regions to assess reliability of projections from uncalibrated models. Decision-makers and practitioners should closely monitor and evaluate the increasing flood risk in these regions due to global warming and promptly implement climate mitigation and adaptation strategies. This includes prioritizing resources for high-resolution regional calibrated simulations in high flood-risk regions to provide high-quality and locally relevant projections to guide mitigation and adaptation measures. However, hydro-climatic conditions may vary seasonally in other regions, and factors beyond precipitation could be critical for flood generation. Particularly in cold climate regions, different calibration methods can significantly impact hydrological projections, as evidenced in previous studies for the Upper Mississippi (S. Huang et al., 2020) and Canada (Seiller et al., 2017). Insights from regional calibrated models in these areas could help to improve process understanding and better regionalization of GHMs.

5. Conclusions

Using four sets of experiments with varied model parameters, we investigated the effect of model calibration on projected flood and drought frequency, as well as on the affected area, population, and GDP in the YRB using the hydrological model CWatM. The results demonstrate that calibration significantly enhances model performance in discharge simulation, aligning simulated seasonal cycles more closely with observations. Calibration also increases projection consistency at stations along the main stream of the Yangtze River. We observe that calibration has a more significant effect on projected changes in drought frequency and impacts compared to projected flood changes in the YRB. Furthermore, we quantified the relative contributions of GCMs, SSPs, and calibration experiments to the projected impacts of floods and droughts in the 21st century by analyzing the total variance of the projection ensemble. The findings indicate that the GCM component predominates in projected flood impact changes. In terms of projected drought impact changes, the experiment component plays a more significant role compared to its influence on projected flood impact changes.

Flood frequency in the YRB is projected to consistently increase with global warming. Notably, areas upstream of the Three Gorges Dam, along with some regions in the southwest and downstream of the basin, are *likely* to experience more frequent floods. It is imperative to implement effective measures in these regions to mitigate the increasing risk to population and GDP from floods. Conversely, drought frequency is expected to rise in the southwestern part of the basin. However, this result warrants cautious interpretation due to the substantial influence of calibration on projected drought frequency and impacts. We advocate for increased research efforts aimed at enhancing model process representation and calibration methods to ensure more reliable projections of drought impacts.

The asymmetric effect of calibration on projected flood and drought changes is primarily influenced by the relative contribution of evapotranspiration trends to the regional water cycle. In the YRB, a warm and wet region, floods are predominantly driven by extreme rainfall, resulting in a minor effect of calibration on flood frequency. Our findings suggest that in warm and humid regions, large-scale increases in flood risks projected by uncalibrated hydrological models could still be quite reliable, providing valuable insights for prompt mitigation and adaptation efforts. However, to effectively implement adaptation strategies at the local level, it's essential to prioritize resources for calibrating high-resolution regional models in these high-risk areas. Such calibration enhances the accuracy of projections at finer scales, aiding in the development of targeted and effective adaptation measures.

This work highlights the crucial link between model calibration and accurate projection of drought and flood impacts, offering valuable insights for related simulation studies. Given the complicated impact of calibration parameters and different model structures, we hope this study can inspire other model teams to verify the applicability of our findings by conducting similar analyses in more basins.

Data Availability Statement

Different data sets used in this study were obtained from various sources described in Table 1, in Section 2.2.1 and Section 2.2.2. The in situ 2003–2018 daily streamflow observations at Cuntan, Yichang, Hankou and Datong used in this study are obtained from Chinese Government agency (Changjiang Water Resources Commission of the Ministry of Water Resources of China). Due to restrictions of the data license, this data is not available from the authors. However, the authors can provide the readers with feasible ways to obtain the data. The CWatM model is publicly accessible via https://cwatm.iiasa.ac.at/. Figures were made with Matplotlib version 3.3.3 (Caswell et al., 2020; Hunter, 2007), available under the Matplotlib license at https://matplotlib.org/.

References

Adhikary, P. P., Sena, D. R., Dash, C. J., Mandal, U., Nanda, S., Madhu, M., et al. (2019). Effect of calibration and validation decisions on streamflow modeling for a heterogeneous and low runoff-producing River Basin in India. *Journal of Hydrologic Engineering*, *24*(7), 05019015. https://doi.org/10.1061/(asce)he.1943-5584.0001792

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer.

Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, *134*(3), 387–401. https://doi.org/10.1007/s10584-014-1084-5

Bai, P., Liu, X., & Liu, C. (2018). Improving hydrological simulations by incorporating GRACE data for model calibration. *Journal of Hydrology*, *557*, 291–304. https://doi.org/10.1016/j.jhydrol.2017.12.025

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, *11*(2), 703–710. https://doi.org/10.5194/hess-11-703-2007

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, *52*(5), 3599–3622. https://doi.org/10.1002/2015WR018247

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, *37*(2), 307–337. https://doi.org/10.1007/s10712-015-9346-y

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., et al. (2020). Development of the Community Water Model (CWatM v1.04)—A high-resolution hydrological model for global and regional assessment of integrated water resources management. *Geoscientific Model Development*, *13*(7), 3267–3298. https://doi.org/10.5194/gmd-13-3267-2020

Burek, P., Smilovic, M., Guillaumot, L., de Bruijn, J., Greve, P., Satoh, Y., et al. (2020). *Community Water Model CWatM Manual. IIASA Report*. IIASA.

Caswell, T. A., Droettboom, M., Lee, A., Hunter, J., Andrade, E. S. d., Firing, E., et al. (2020). matplotlib/matplotlib: REL: v3.3.3 [Software]. *Zenodo*. https://zenodo.org/records/4268928

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., et al. (2019). How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates. *Earth's Future*, *7*(6), 623–637. https://doi.org/10.1029/2018EF001047

Chen, J., Liu, Y. J., Pan, T., Liu, Y. H., Sun, F. B., & Ge, Q. S. (2018). Population exposure to droughts in China under the 1.5°C global warming target. *Earth System Dynamics*, *9*(3), 1097–1106. https://doi.org/10.5194/esd-9-1097-2018

Chen, J., Wu, X., Finlayson, B. L., Webber, M., Wei, T., Li, M., & Chen, Z. (2014). Variability and trend in the hydrology of the Yangtze River, China: Annual precipitation and runoff. *Journal of Hydrology*, *513*, 403–412. https://doi.org/10.1016/j.jhydrol.2014.03.044

Cucchi, M., Weedon, G., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., et al. (2020). WFDE5: Bias adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*, *12*(3), 2097–2120. https://doi.org/10.5194/essd-12-2097-2020

Dahri, Z. H., Ludwig, F., Moors, E., Ahmad, S., Ahmad, B., Ahmad, S., et al. (2021). Climate change and hydrological regime of the high-altitude Indus basin under extreme climate scenarios. *Science of the Total Environment*, *768*, 144467. https://doi.org/10.1016/j.scitotenv.2020.144467

Dai, Y., Wei, N., Yuan, H., Zhang, S., Shangguan, W., Liu, S., et al. (2019). Evaluation of soil thermal conductivity schemes for use in land surface modeling. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3454–3473. https://doi.org/10.1029/2019MS001723

Dai, Y., Xin, Q., Wei, N., Zhang, Y., Shangguan, W., Yuan, H., et al. (2019). A global high-resolution data set of soil hydraulic and thermal properties for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *11*(9), 2996–3023. https://doi.org/10.1029/2019MS001784

Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2013). Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research*, *49*(7), 4035–4053. https://doi.org/10.1002/wrcr.20294

Döll, P., & Siebert, S. (2002). Global modeling of irrigation water requirements. *Water Resources Research*, *38*(4), WR000355. https://doi.org/10.1029/2001wr000355

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. *Nature Climate Change*, *6*(5), 508–513. https://doi.org/10.1038/nclimate2941

Dottori, F., Szewczyk, W., Ciscar, J. C., Zhao, F., Alfieri, L., Hirabayashi, Y., et al. (2018). Increased human and economic losses from river flooding with anthropogenic warming. *Nature Climate Change*, *8*(9), 781–786. https://doi.org/10.1038/s41558-018-0257-z

Ekmekcioglu, O., Demirel, M. C., & Booij, M. J. (2022). Effect of data length, spin-up period and spatial model resolution on fully distributed hydrological model calibration in the Moselle basin. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *67*(5), 759–772. https://doi.org/10.1080/02626667.2022.2046754

Elvidge, C. D., Tuttle, B. T., Sutton, P. S., Baugh, K. E., Howard, A. T., Milesi, C., et al. (2007). Global distribution and density of constructed impervious surfaces. *Sensors*, *7*(9), 1962–1979. https://doi.org/10.3390/s7091962

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

FAO. (2012). FAOSTAT online database [Dataset]. *FAO*. Retrieved from https://www.fao.org/faostat/en/#data

Farmer, W. H., & Vogel, R. M. (2016). On the deterministic and stochastic use of hydrologic models. *Water Resources Research*, *52*(7), 5619–5633. https://doi.org/10.1002/2016WR019129

Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fortin, F. A., De Rainville, F. M., Gardner, M. A., Parizeau, M., & Gagne, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, *13*(1), 2171–2175. https://doi.org/10.1109/MIS.2010.134

Fremme, A., & Sodemann, H. (2019). The role of land and ocean evaporation on the variability of precipitation in the Yangtze River valley. *Hydrology and Earth System Sciences*, *23*(6), 2525–2540. https://doi.org/10.5194/hess-23-2525-2019

Frieler, K., Lange, S., Piontek, F., Reyer, C. P. O., Schewe, J., Warszawski, L., et al. (2017). Assessing the impacts of 1.5°C global warming – Simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geoscientific Model Development*, *10*(12), 4321–4345. https://doi.org/10.5194/gmd-10-4321-2017

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., del Rocío Rivas López, M., et al. (2024). 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). *Geoscientific Model Development*, *17*(1), 1–51. https://doi.org/10.5194/gmd-17-1-2024

Gleick, P. H., Cooley, H., Cohen, M. J., Morikawa, M., Morrison, J., & Palaniappan, M. (2009). *The worlds water 2008–2009, The biennial report on freshwater resources*. Island Press.

Gosling, S. N., Taylor, R. G., Arnell, N. W., & Todd, M. C. (2011). A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models. *Hydrology and Earth System Sciences*, *15*(1), 279–294. https://doi.org/10.5194/hess-15-279-2011

Gosling, S. N., Zaherpour, J., Mount, N. J., Hattermann, F. F., Dankers, R., Arheimer, B., et al. (2017). A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1°C, 2°C and 3°C. *Climatic Change*, *141*(3), 577–595. https://doi.org/10.1007/s10584-016-1773-3

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hattermann, F. F., Krysanova, V., Gosling, S. N., Dankers, R., Daggupati, P., Donnelly, C., et al. (2017). Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. *Climatic Change*, *141*(3), 561–576. https://doi.org/10.1007/s10584-016-1829-4

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., et al. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/nclimate1911

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. https://doi.org/10.1016/j.jhydrol.2018.09.052

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press.

Huang, S., Shah, H., Naz, B. S., Shrestha, N., Mishra, V., Daggupati, P., et al. (2020). Impacts of hydrological model calibration on projected hydrological changes under climate change-a multi-model assessment in three large river basins. *Climatic Change*, *163*(3), 1143–1164. https://doi.org/10.1007/s10584-020-02872-6

Huang, Y., Salama, M. S., Krol, M. S., Su, Z., Hoekstra, A. Y., Zeng, Y., & Zhou, Y. (2015). Estimation of human-induced changes in terrestrial water storage through integration of GRACE satellite detection and hydrological modeling: A case study of the Yangtze River basin. *Water Resources Research*, *51*(10), 8494–8516. https://doi.org/10.1002/2015wr016923

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

IPCC. (2018). Summary for Policymakers. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, et al. (Eds.), *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (pp. 1–32). World Meteorological Organization.

IPCC. (2023a). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

IPCC. (2023b). Summary for Policymakers. In Core Writing Team, H. Lee, & J. Romero (Eds.), *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–34). IPCC.

IPCC. (2023c). Weather and climate extreme events in a changing climate. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1513–1766). Cambridge University Press.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158–171. https://doi.org/10.1002/qj.49708134804

Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth System Science Data*, *9*(2), 927–953. https://doi.org/10.5194/essd-9-927-2017

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

Koch, H., Silva, A. L. C., Liersch, S., de Azevedo, J. R. G., & Hattermann, F. F. (2020). Effects of model calibration on hydrological and water resources management simulations under climate change in a semi-arid watershed. *Climatic Change*, *163*(3), 1247–1266. https://doi.org/10.1007/s10584-020-02917-w

Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., & Kundzewicz, Z. W. (2018). How the performance of hydrological models relates to credibility of projections under climate change. *Hydrological Sciences Journal*, *63*(5), 696–720. https://doi.org/10.1080/02626667.2018.1446214

Kumar, A., Gosling, S. N., Johnson, M. F., Jones, M. D., Zaherpour, J., Kumar, R., et al. (2022). Multi-model evaluation of catchment- and global-scale hydrological model simulations of drought characteristics across eight large river catchments. *Advances in Water Resources*, *165*, 104212. https://doi.org/10.1016/j.advwatres.2022.104212

Kummu, M., Taka, M., & Guillaume, J. H. A. (2018). Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Scientific Data*, *5*(1), 180004. https://doi.org/10.1038/sdata.2018.4

Kummu, M., Taka, M., & Guillaume, J. H. A. (2020). Data from: Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015 [Dataset]. *Dryad*. https://doi.org/10.5061/dryad.dk1j0

Kundzewicz, Z. W., Su, B. D., Wang, Y. J., Wang, G. J., Wang, G. F., Huang, J. L., & Jiang, T. (2019). Flood risk in a range of spatial perspectives - From global to local scales. *Natural Hazards and Earth System Sciences*, *19*(7), 1319–1328. https://doi.org/10.5194/nhess-19-1319-2019

Kundzewicz, Z. W., Su, B. D., Wang, Y. J., Xia, J., Huang, J. L., & Jiang, T. (2019). Flood risk and its reduction in China. *Advances in Water Resources*, *130*, 37–45. https://doi.org/10.1016/j.advwatres.2019.05.020

Lange, S. (2019). Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0). *Geoscientific Model Development*, *12*(7), 3055–3070. https://doi.org/10.5194/gmd-12-3055-2019

Lange, S. (2021). ISIMIP3BASD (2.5.0) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.4686991

Lange, S., & Büchner, M. (2021). ISIMIP3b bias-adjusted atmospheric climate input data (v1.1) [Dataset]. *ISIMIP Repository*. https://doi.org/10.48364/ISIMIP.842396.1

Lange, S., Menz, C., Gleixner, S., Cucchi, M., Weedon, G. P., Amici, A., et al. (2021). WFDE5 over land merged with ERA5 over the ocean (W5E5 v2.0) [Dataset]. *ISIMIP Repository*. https://doi.org/10.48364/ISIMIP.342217

Lange, S., Quesada-Chacón, D., & Büchner, M. (2024). Secondary ISIMIP3b bias-adjusted atmospheric climate input data (v1.4) [Dataset]. *ISIMIP Repository*. https://doi.org/10.48364/ISIMIP.581124.4

Lange, S., Volkholz, J., Geiger, T., Zhao, F., Vega, I., Veldkamp, T., et al. (2020). Projecting exposure to extreme climate impact events across six event categories and three spatial scales. *Earth's Future*, *8*(12), e2020EF001616. https://doi.org/10.1029/2020ef001616

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502. https://doi.org/10.1890/100125

Li, C., Zwiers, F., Zhang, X., Chen, G., Lu, J., Li, G., et al. (2019). Larger increases in more extreme local precipitation events as climate warms. *Geophysical Research Letters*, *46*(12), 6885–6891. https://doi.org/10.1029/2019gl082908

Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., & Wehner, M. (2021). Changes in annual extremes of daily temperature and precipitation in CMIP6 models. *Journal of Climate*, *34*(9), 3441–3460. https://doi.org/10.1175/jcli-d-19-1013.1

Li, Y., Yan, D., Peng, H., & Xiao, S. (2021). Evaluation of precipitation in CMIP6 over the Yangtze River Basin. *Atmospheric Research*, *253*, 105406. https://doi.org/10.1016/j.atmosres.2020.105406

Liu, D. (2020). A rational performance criterion for hydrological model. *Journal of Hydrology*, *590*, 125488. https://doi.org/10.1016/j.jhydrol.2020.125488

Lu, J., Wang, G., Gong, T., Hagan, D. F. T., Wang, Y., Jiang, T., & Su, B. (2019). Changes of actual evapotranspiration and its components in the Yangtze River valley during 1980-2014 from satellite assimilation product. *Theoretical and Applied Climatology*, *138*(3–4), 1493–1510. https://doi.org/10.1007/s00704-019-02913-w

Lü, M., Wu, S. J., Chen, J., Chen, C., Wen, Z., & Huang, Y. (2018). Changes in extreme precipitation in the Yangtze River basin and its association with global mean temperature and ENSO. *International Journal of Climatology*, *38*(4), 1989–2005. https://doi.org/10.1002/joc.5311

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, *47*(2), W02531. https://doi.org/10.1029/2010wr009505

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geostatistical approach. *Nature Communications*, *7*(1), 13603. https://doi.org/10.1038/ncomms13603

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614. https://doi.org/10.5194/hess-23-2601-2019

Moreno, A., & Hasenauer, H. (2016). Spatial downscaling of European climate data. *International Journal of Climatology*, *36*(3), 1444–1458. https://doi.org/10.1002/joc.4436

Mosier, T. M., Hill, D. F., & Sharp, K. V. (2018). Update to the Global Climate Data package: Analysis of empirical bias correction methods in the context of producing very high resolution climate projections. *International Journal of Climatology*, *38*(2), 825–840. https://doi.org/10.1002/joc.5213

Müller Schmied, H., Gosling, S. N., Garnsworthy, M., Müller, L., Telteu, C. E., Ahmed, A. K., et al. (2024). Graphical representation of global water models. *EGUsphere*, *2024*, 1–51. https://doi.org/10.5194/egusphere-2024-1303

Munoz-Castro, E., Mendoza, P. A., Vasquez, N., & Vargas, X. (2023). Exploring parameter (dis)agreement due to calibration metric selection in conceptual rainfall-runoff models. *Hydrological Sciences Journal*, *68*(12), 1754–1768. https://doi.org/10.1080/02626667.2023.2231434

Najafi, M. R., Moradkhani, H., & Jung, I. W. (2011). Assessing the uncertainties of hydrologic model selection in climate change impact studies. *Hydrological Processes*, *25*(18), 2814–2826. https://doi.org/10.1002/hyp.8043

Nie, N., Zhang, W., Chen, H., Zhao, D., & Liu, M. (2021). Temporal variation characteristics and attribution analysis of terrestrial water storage change in the Yangtze River basin. *Advances in Water Science*, *32*(3), 87–98. https://doi.org/10.14042/j.cnki.32.1309.2021.03.008

Nie, N., Zhang, W., Liu, M., Chen, H., & Zhao, D. (2021). Separating the impacts of climate variability, land-use change and large reservoir operations on streamflow in the Yangtze River basin, China, using a hydrological modeling approach. *International Journal of Digital Earth*, *14*(2), 231–249. https://doi.org/10.1080/17538947.2020.1812740

Niraula, R., Meixner, T., & Norman, L. M. (2015). Determining the importance of model calibration for forecasting absolute/relative changes in streamflow from LULC and climate changes. *Journal of Hydrology*, *522*, 439–451. https://doi.org/10.1016/j.jhydrol.2015.01.007

Nkiaka, E., Nawaz, N. R., & Lovett, J. C. (2018). Effect of single and multi-site calibration techniques on hydrological model performance, parameter estimation and predictive uncertainty: A case study in the Logone catchment, Lake Chad basin. *Stochastic Environmental Research and Risk Assessment*, *32*(6), 1665–1682. https://doi.org/10.1007/s00477-017-1466-0

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., & Michel, C. (2006). Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, *42*(7), W07410. https://doi.org/10.1029/2005WR004636

Pall, P., Allen, M. R., & Stone, D. A. (2007). Testing the Clausius–Clapeyron constraint on changes in extreme precipitation under CO_2 warming. *Climate Dynamics*, *28*(4), 351–363. https://doi.org/10.1007/s00382-006-0180-2

Pendergrass, A. G., & Hartmann, D. L. (2014). Changes in the distribution of rain frequency and intensity in response to global warming. *Journal of Climate*, *27*(22), 8372–8383. https://doi.org/10.1175/JCLI-D-14-00183.1

Pfahl, S., O'Gorman, P. A., & Fischer, E. M. (2017). Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, *7*(6), 423–427. https://doi.org/10.1038/nclimate3287

Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C. H., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. *Nature Climate Change*, *7*(1), 48–53. https://doi.org/10.1038/nclimate3168

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., et al. (2014). Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(9), 3262–3267. https://doi.org/10.1073/pnas.1222473110

Qi, W., Feng, L., Yang, H., Liu, J., Zheng, Y., Shi, H., et al. (2022). Economic growth dominates rising potential flood risk in the Yangtze River and benefits of raising dikes from 1991 to 2015. *Environmental Research Letters*, *17*(3), 034046. https://doi.org/10.1088/1748-9326/ac5561

Qi, W., Zhang, C., Fu, G., Sweetapple, C., & Liu, Y. (2019). Impact of robustness of hydrological model parameters on flood prediction uncertainty. *Journal of Flood Risk Management*, *12*(S1), e12488. https://doi.org/10.1111/jfr3.12488

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018

Sauer, I. J., Reese, R., Otto, C., Geiger, T., Willner, S. N., Guillod, B. P., et al. (2021). Climate signals in river flood damages emerge under sound regional disaggregation. *Nature Communications*, *12*(1), 2128. https://doi.org/10.1038/s41467-021-22153-9

Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., & Ward, P. J. (2016). FLOPROS: An evolving global database of flood protection standards. *Natural Hazards and Earth System Sciences*, *16*(5), 1049–1061. https://doi.org/10.5194/nhess-16-1049-2016

Seiller, G., Roy, R., & Anctil, F. (2017). Influence of three common calibration metrics on the diagnosis of climate change impacts on water resources. *Journal of Hydrology*, *547*, 280–295. https://doi.org/10.1016/j.jhydrol.2017.02.004

Seneviratne, S. I., Donat, M. G., Pitman, A. J., Knutti, R., & Wilby, R. L. (2016). Allowable CO_2 emissions based on regional and impact-related climate targets. *Nature*, *529*(7587), 477–483. https://doi.org/10.1038/nature16542

Setti, S., Barik, K. K., Merz, B., Agarwal, A., & Rathinasamy, M. (2022). Investigating the impact of calibration timescales on streamflow simulation, parameter sensitivity and model performance for Indian catchments. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *67*(5), 661–675. https://doi.org/10.1080/02626667.2022.2036340

Shiklomanov, I. A. (1997). *Assessment of water resources and water availability in the world*. WMO.

Shin, M.-J., & Jung, Y. (2022). Using a global sensitivity analysis to estimate the appropriate length of calibration period in the presence of high hydrological model uncertainty. *Journal of Hydrology*, *607*, 127546. https://doi.org/10.1016/j.jhydrol.2022.127546

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., & Portmann, F. T. (2010). Groundwater use for irrigation – A global inventory. *Hydrology and Earth System Sciences*, *14*(10), 1863–1880. https://doi.org/10.5194/hess-14-1863-2010

Siebert, S., Döll, P., Hoogeveen, J., Faures, J. M., Frenken, K., & Feick, S. (2005). Development and validation of the global map of irrigation areas. *Hydrology and Earth System Sciences*, *9*(5), 535–547. https://doi.org/10.5194/hess-9-535-2005

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., & de Haan, C. (2006). *Livestock's long shadow: Environmental issues and options* (Vol. 24, No. 4, pp. 15–17). Food & Agriculture Organization.

Su, Z., Sun, X., Devineni, N., Lall, U., Hao, Z., & Chen, X. (2020). The effects of pre-season high flows, climate, and the Three Gorges Dam on low flow at the Three Gorges Region, China. *Hydrological Processes*, *34*(9), 2088–2100. https://doi.org/10.1002/hyp.13714

Sun, F., Mejia, A., Zeng, P., & Che, Y. (2019). Projecting meteorological, hydrological and agricultural droughts for the Yangtze River basin. *Science of the Total Environment*, *696*, 134076. https://doi.org/10.1016/j.scitotenv.2019.134076

Sutanudjaja, E., van Beek, R., Wanders, N., Wada, Y., Bosmans, J., Drost, N., et al. (2017). PCR-GLOBWB 2 input files version 2017_11_beta_1 [Dataset]. *Zenodo*. https://zenodo.org/records/1045339

Tanoue, M., Taguchi, R., Alifu, H., & Hirabayashi, Y. (2021). Residual flood damage under intensive adaptation. *Nature Climate Change*, *11*(10), 823–826. https://doi.org/10.1038/s41558-021-01158-8

Tarasova, L., Knoche, M., Dietrich, J., & Merz, R. (2016). Effects of input discretization, model complexity, and calibration strategy on model performance in a data-scarce glacierized catchment in Central Asia. *Water Resources Research*, *52*(6), 4674–4699. https://doi.org/10.1002/2015wr018551

Telteu, C. E., Müller Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., et al. (2021). Understanding each other's models: An introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication. *Geoscientific Model Development*, *14*(6), 3843–3878. https://doi.org/10.5194/gmd-14-3843-2021

Thompson, J. R., Gosling, S. N., Zaherpour, J., & Laizé, C. L. R. (2021). Increasing risk of ecological change to major rivers of the world with global warming. *Earth's Future*, *9*(11), e2021EF002048. https://doi.org/10.1029/2021EF002048

UNFCCC. (2015). Adoption of the Paris Agreement: Proposal by the President. Retrieved from https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

UNFCCC. (2016). Aggregate effect of the intended nationally determined contributions: An update. Retrieved from https://unfccc.int/resource/docs/2016/cop22/eng/02.pdf

United Nations Office for Disaster Risk Reduction. (2015). Sendai Framework for Disaster Risk Reduction 2015–2030. Retrieved from https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

Vansteenkiste, T., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F., & Willems, P. (2014). Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. *Journal of Hydrology*, *519*, 743–755. https://doi.org/10.1016/j.jhydrol.2014.07.062

Viney, N. R., Bormann, H., Breuer, L., Bronstert, A., Croke, B. F. W., Frede, H., et al. (2009). Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions. *Advances in Water Resources*, *32*(2), 147–158. https://doi.org/10.1016/j.advwatres.2008.05.006

Vogel, R. M. (1986). The probability plot correlation coefficient test for the normal, lognormal, and gumbel distributional hypotheses. *Water Resources Research*, *17*(1), 111–117. https://doi.org/10.1029/WR022i004p00587

Volkholz, J., & Ostberg, S. (2022). ISIMIP3a landuse input data (v1.2) [Dataset]. *ISIMIP Repository*. https://doi.org/10.48364/ISIMIP.571261.2

Wada, Y., de Graaf, I. E. M., & van Beek, L. P. H. (2016). High-resolution modeling of human and climate impacts on global water resources. *Journal of Advances in Modeling Earth Systems*, *8*(2), 735–763. https://doi.org/10.1002/2015MS000618

Wang, J., Sheng, Y., Gleason, C. J., & Wada, Y. (2013). Downstream Yangtze River levels impacted by Three Gorges Dam. *Environmental Research Letters*, *8*(4), 044012. https://doi.org/10.1088/1748-9326/8/4/044012

Willner, S., Levermann, A., Zhao, F., & Frieler, K. (2018). Adaptation required to preserve future high-end river flood risk at present levels. *Science Advances*, *4*(1), eaao1914. https://doi.org/10.1126/sciadv.aao1914

Wint, G. R. W., & Robinson, T. P. (2007). *Gridded Livestock of the World 2007*. FAO.

Wu, H., Kimball, J. S., Mantua, N., & Stanford, J. (2011). Automated upscaling of river networks for macroscale hydrological modeling. *Water Resources Research*, *47*(3), W03517. https://doi.org/10.1029/2009WR008871

Wu, Q., Liu, S., Cai, Y., Li, X., & Jiang, Y. (2017). Improvement of hydrological model calibration by selecting multiple parameter ranges. *Hydrology and Earth System Sciences*, *21*(1), 393–407. https://doi.org/10.5194/hess-21-393-2017

Xiong, J., Guo, S., Yin, J., Ning, Z., Zeng, Z., & Wang, R. (2022). Projected changes in terrestrial water storage and associated flood potential across the Yangtze River basin. *Science of the Total Environment*, *817*, 152998. https://doi.org/10.1016/j.scitotenv.2022.152998

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*(6), 5053–5073. https://doi.org/10.1029/2019WR024873

Yang, H., Yang, S., Xu, K., Milliman, J., Wang, H., Yang, Z., et al. (2018). Human impacts on sediment in the Yangtze River: A review and new perspectives. *Global and Planetary Change*, *162*, 8–17. https://doi.org/10.1016/j.gloplacha.2018.01.001

Yang, S., Xu, K., Milliman, J., Yang, H., & Wu, C. (2015). Decline of Yangtze River water and sediment discharge: Impact from natural and anthropogenic changes. *Scientific Reports*, *5*(1), 12581. https://doi.org/10.1038/srep12581

Yu, F., Chen, Z., Ren, X., & Yang, G. (2009). Analysis of historical floods on the Yangtze River, China: Characteristics and explanations. *Geomorphology*, *113*(3–4), 210–216. https://doi.org/10.1016/j.geomorph.2009.03.008

Yu, Z., Gu, H., Wang, J., Xia, J., & Lu, B. (2018). Effect of projected climate change on the hydrological regime of the Yangtze River Basin, China. *Stochastic Environmental Research and Risk Assessment*, *32*(1), 1–16. https://doi.org/10.1007/s00477-017-1391-2

Zaherpour, J., Gosling, S. N., Mount, N., Schmied, H. M., Veldkamp, T. I. E., Dankers, R., et al. (2018). Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environmental Research Letters*, *13*(6), 065015. https://doi.org/10.1088/1748-9326/aac547

Zhang, D., Zhang, Q., Werner, A. D., & Liu, X. (2016). GRACE-based hydrological drought evaluation of the Yangtze River Basin, China. *Journal of Hydrometeorology*, *17*(3), 811–828. https://doi.org/10.1175/JHM-D-15-0084.1

Zhang, J., Xu, W., Liao, X., Zong, S., & Liu, B. (2021). Global mortality risk assessment from river flooding under climate change. *Environmental Research Letters*, *16*(6), 064036. https://doi.org/10.1088/1748-9326/abff87

Zhang, W., Jin, F.-F., Stuecker, M. F., Wittenberg, A. T., Timmermann, A., Ren, H.-L., et al. (2016). Unraveling El Niño's impact on the East Asian Monsoon and Yangtze River summer flooding. *Geophysical Research Letters*, *43*(21), 11375–11382. https://doi.org/10.1002/2016GL071190

Zhao, F., Lange, S., Goswami, B., & Frieler, K. (2024). Frequency bias causes overestimation of climate change impacts on global flood occurrence. *Geophysical Research Letters*, *51*(16), e2024GL108855. https://doi.org/10.1029/2024GL108855

Zhao, F., Veldkamp, T. I. E., Frieler, K., Schewe, J., Ostberg, S., Willner, S., et al. (2017). The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters*, *12*(7), 075003. https://doi.org/10.1088/1748-9326/aa7250

Zhao, J., He, S., & Wang, H. (2022). Historical and future runoff changes in the Yangtze River Basin from CMIP6 models constrained by a weighting strategy. *Environmental Research Letters*, *17*(2), 024015. https://doi.org/10.1088/1748-9326/ac3f61

Zong, Y., & Chen, X. (2000). The 1998 flood on the Yangtze, China. *Natural Hazards*, *22*(2), 165–184. https://doi.org/10.1023/a:1008119805106