



## Enhancing hydrological hazard early warning: A 60-day streamflow forecasting framework integrating deep learning and process-based modeling

Zhijie Liu<sup>1,2</sup>, Hanbo Yang<sup>1,2,\*</sup>, Dawen Yang<sup>1,2</sup>

5 <sup>1</sup>Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China

<sup>2</sup>State Key Laboratory of Hydrosience and Engineering, Tsinghua University, Beijing 100084, China

\*Correspondence: Hanbo Yang (yanghanbo@tsinghua.edu.cn)

10 **Abstract:** Reliable medium- and long-term streamflow forecasting is a cornerstone of hydrological hazard early warning and water resources management, yet achieving accurate predictions with sufficient lead time remains a formidable challenge. This study proposes a 60-day streamflow forecasting framework to strengthen early warning capabilities by systematically integrating a convolutional neural network (CNN) for bias correction of precipitation forecasts from the UK Met  
15 Office (UKMO) numerical weather prediction model, the Geomorphology-Based Eco-Hydrological Model (GBEHM) for streamflow simulation, and an autoregressive with exogenous input (ARX) model for statistical post-processing. Applying the proposed framework to the Upper Yangtze River Basin, results indicate that the CNN model reduces the areal-averaged precipitation root mean square error (RMSE) by around 35% and elevates the temporal correlation coefficient (TCC) from  
20 0.62 to 0.74 against raw UKMO forecasts across the 60-day horizon, with performance gains amplifying at longer lead times. Subsequently, when driving the GBEHM with corrected precipitation and applying ARX post-processing, the streamflow forecasts exhibit substantial enhancements with a reduction in RMSE of 36%, a decrease in relative error (RE) from 48.2% to 17.4%, and an increase in Nash–Sutcliffe efficiency (NSE) from 0.33 to 0.72 compared to those  
25 driven by raw forecasts in terms of 60-day mean performance. Error decomposition identifies precipitation forecast errors which intensify with lead time as the dominant source of uncertainty for medium- and long-term streamflow forecasting, while confirming that hydrological model uncertainty remains a significant component, highlighting that the selection of a robust hydrological model is crucial for enhancing the reliability and predictive skill of the streamflow forecasts. By  
30 systematically leveraging the CNN to mitigate drifting meteorological biases, the GBEHM to capture physical catchment dynamics, and the ARX to minimize residual errors, the proposed framework extends the effective early warning horizon to 60 days with high volumetric accuracy and temporal consistency, providing vital decision support for flood and drought risk management and regional water security.

35

### 1 Introduction

The escalating frequency and intensity of extreme hydrological events, such as devastating floods and prolonged droughts, present significant challenges to global water security and socioeconomic stability within the context of climate change (Swain et al., 2025; Sutanto et al., 2025; Kreibich et al., 2022; Tabari, 2020). Reliable medium- and long-term streamflow forecasting, typically referring  
40 to predictions with lead times ranging from 3 days to 1 year (Shao et al., 2024), is recognized as a critical measure for hydrological hazard early warning by providing more substantial lead times for proactive decision-making in flood and drought mitigation, hydropower optimization and water supply security (Kondal et al., 2024; Lee et al., 2022; Jackson-Blake et al., 2022; Xu et al., 2020).



45 However, constrained by the inherent complexity of hydrological processes and atmospheric  
uncertainties, the accuracy and lead time of current forecasts remain insufficient to fully meet the  
rigorous demands of disaster early warning (Slater et al., 2023; Neri et al., 2020). Consequently,  
enhancing the predictive skill and extending the reliable horizon of streamflow forecasts are  
imperative for strengthening regional disaster resilience and safeguarding water security (Koh and  
50 Galelli, 2024; Sutanto et al., 2020; Pendergrass et al., 2020).

Medium- and long-term streamflow forecasting methods are typically divided into three main  
categories: time series analysis, data-driven methods and dynamical methods. Time series analysis  
forecasts streamflow by extrapolating statistical patterns solely from historical observations (Wang  
55 et al., 2020; Huang et al., 2020; Luo et al., 2019). For instance, Guo et al. (2023) applied a hybrid  
time-series framework based on signal decomposition to forecast medium- and long-term  
streamflow in the lower Yellow River. However, inherently limited to coarser temporal scales and  
devoid of atmospheric drivers, this univariate approach struggles to capture both daily streamflow  
variability and dynamic hydrological responses to climate change (Jamei et al., 2024; Kratzert et al.,  
60 2018). Data-driven methods forecast streamflow by establishing direct quantitative mappings from  
meteorological predictors (Wang et al., 2023; Hunt et al., 2022; Adnan et al., 2020). For example,  
Cheng et al. (2020) utilized long short-term memory (LSTM) networks to map precipitation and  
antecedent flow to future streamflow for long lead-time forecasting. However, lacking explicit  
representation of physical hydrological processes, these data-driven methods are fundamentally  
65 limited in capturing non-stationary environmental impacts thereby hindering the prediction of  
unprecedented hydrological hazards (Reichstein et al., 2019; Shen, 2018). Dynamical methods  
generate streamflow forecasts by driving process-based hydrological models with meteorological  
predictions (Lee et al., 2024; Greuell and Hutjes, 2023; Quedi and Fan, 2020). For instance, Tian et  
al. (2018) employed the THREW model driven by bias-corrected European Centre for Medium-  
70 Range Forecasts (ECMWF) precipitation forecasts for seasonal prediction in the Upper Hanjiang  
River Basin. Among these three categories, dynamical methods leverage their explicit physical  
mechanisms and interpretability to offer superior transferability and the capability to capture climate  
change impacts, thereby making them indispensable for robust hydrological hazard early warning  
(Falck et al., 2025; Zhang et al., 2024; Andrade et al., 2024).

75 As the primary meteorological forcing, precipitation forecasts play a decisive role in determining  
the accuracy of medium- and long-term streamflow forecasts (Ghimire et al., 2021). Currently,  
medium- and long-term precipitation forecasts rely predominantly on numerical weather prediction  
models (Adams and Dymond, 2019; Bauer et al., 2015). However, these raw outputs typically  
80 exhibit large discrepancies and systematic biases (Siqueira et al., 2020; Lin et al., 2019). As the  
forecast lead time extends, their reliability diminishes significantly, making them unsuitable for  
direct application in hydrological hazard assessment (Du et al., 2025; Bogner et al., 2022).  
Consequently, bias correction is essential to render these outputs usable for streamflow forecasting  
(Vernon et al., 2025; Monhart et al., 2019; Anghileri et al., 2019). Traditional correction methods,  
85 such as quantile mapping, have been shown to effectively correct systematic biases in precipitation  
forecasts (Cannon et al., 2015; Teutschbein and Seibert, 2012). Recently, deep learning approaches  
have demonstrated superior performance by leveraging their ability to model complex nonlinear  
relationships within multidimensional datasets (Yin et al., 2023; Lyu et al., 2023). Nie and Sun (2024)



proposed a method combining deep learning and dynamical-statistical projection model to correct  
90 ECMWF sub-seasonal precipitation forecasts, demonstrating significant skill enhancements in  
Southwest China for lead times of up to 30 days. Lyu et al. (2024) utilized a hybrid CSG-UNET  
model to refine ECMWF ensemble precipitation forecasts over the China mainland during summer,  
extending the effective forecast skill to approximately 4 weeks. In general, current research  
predominantly targets ECMWF forecasts due to their superior predictive skill (Falck et al., 2025;  
95 Dong et al., 2025; Andrade et al., 2024; Quedi and Fan, 2020). However, most studies are confined  
to the first 30 days, as results beyond this point are either omitted or unreliable for practical use (Nie  
and Sun, 2024; Lyu et al., 2024; Yin et al., 2023). In contrast, alternative models such as that from  
the UK Met Office (UKMO), which provide 60-day forecasts, remain under-explored due to the  
lower raw accuracy of their precipitation forecasts. To address this gap, this study aims to apply bias  
100 correction to 60-day precipitation forecasts, thereby enhancing reliability and extending the  
effective forecast horizon.

In addition to errors in precipitation forecasts, inherent uncertainties stemming from the  
hydrological model structural generalization and parameter estimation can also propagate into  
105 streamflow predictions (Donegan et al., 2021; Dion et al., 2021), rendering the correction of  
hydrological simulation errors imperative. Common correction strategies range from autoregressive  
methods and data assimilation to deep learning approaches (Tanguy et al., 2025; Sabzipour et al.,  
2023; Siqueira et al., 2021). Among these, the autoregressive with exogenous input (ARX) model  
is particularly suitable for addressing the systematic bias and serial correlation inherent in  
110 hydrological residuals, providing a computationally efficient and effective solution for post-  
processing (Mcinerney et al., 2021; Sharma et al., 2019).

Therefore, this study proposes a 60-day streamflow forecasting framework that integrates a  
convolutional neural network (CNN) to correct precipitation forecasts from the UKMO numerical  
115 weather prediction model, the geomorphology-based eco-hydrological model (GBEHM) to simulate  
hydrological processes, and an ARX model to minimize residual errors. Applying the proposed  
framework in the Upper Yangtze River Basin (UYRB), we expect this research to offer a practical  
and reliable tool for decision support with extended lead times for hydrological hazard early warning.

## 120 **2 Study area and data**

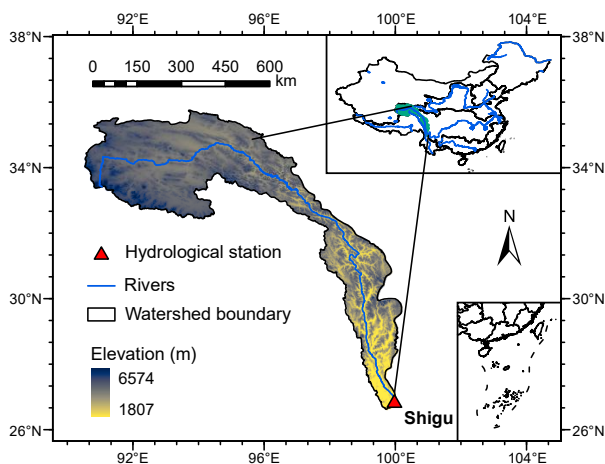
### **2.1 Study area**

The Upper Yangtze River Basin (UYRB) is highly vulnerable to extreme hydrological events driven  
by uneven streamflow distribution, frequently triggering devastating floods and droughts that  
threaten downstream infrastructure and regional water security (Liang et al., 2023; Wang et al.,  
125 2022). In addition, the basin serves as a strategic hub for China's hydropower generation, playing a  
vital role in regional energy security (Zhong et al., 2020). Consequently, improving forecast  
accuracy at medium- and long-range horizons in this region is of paramount importance for  
integrated disaster mitigation and reservoir operation (Su et al., 2017).

130 This study focuses on the UYRB above the Shigu hydrological station, which spans 90° E–101° E  
and 26° N–36° N with a drainage area of approximately  $2.2 \times 10^5$  km<sup>2</sup> (Fig. 1). The terrain is  
complex with elevation gradually decreasing from over 6500 m in the headwaters to about 1800 m



at the outlet. It is a typical alpine cold mountainous zone dominated by grassland, shrubland, and  
barren areas. The region has a mean annual precipitation of around 670 mm and a mean annual  
135 streamflow of around  $1300 \text{ m}^3 \text{ s}^{-1}$  observed at the Shigu station.



**Figure 1.** The upper Yangtze River Basin controlled by the Shigu hydrological station.

## 2.2 Data sources

### 140 2.2.1 Observed meteorological and hydrological data

To support model training and validation, the China Gauge-Based Daily Precipitation Analysis (CGDPA), provided by the National Meteorological Information Center, is used as the reference  
observational dataset. This  $0.25^\circ$  daily precipitation gridded product is derived from over 2,400  
gauge stations across China and covers the period from 1960 to 2019. Its reliability in capturing  
145 precipitation characteristics over complex terrain has been extensively validated (Shen and Xiong,  
2016).

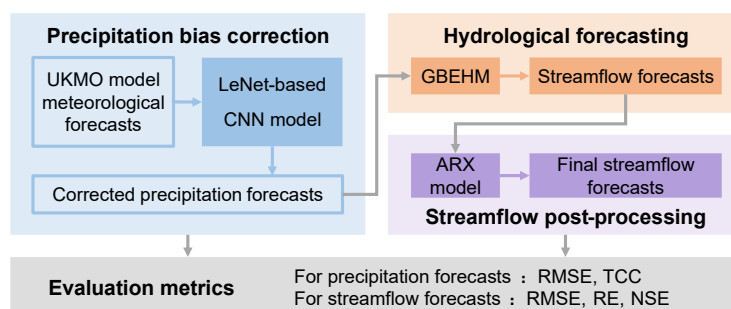
To drive the hydrological model, additional meteorological inputs, including air temperature, wind  
speed, relative humidity, and sunshine duration, are obtained from meteorological stations operated  
150 by the China Meteorological Administration (CMA, <http://data.cma.cn>). To align with the  
hydrological model configuration, these gauge observations are spatially interpolated onto the  
 $8 \text{ km} \times 8 \text{ km}$  simulation grid using the elevation-adjusted angular distance weighting technique  
(Yang et al., 2004). Furthermore, to meet the temporal requirements, the daily data are disaggregated  
into hourly forcing series following the algorithms proposed by Gao et al. (2015).

155

To calibrate the hydrological model and evaluate forecast performance, daily streamflow  
observations from the Shigu hydrological station for 1960 to 2019 are collected from the  
hydrological yearbook.

### 160 2.2.2 Underlying surface data

The distributed model GBEHM requires topographic data and underlying surface properties.  
Topographically, a 90-m resolution digital elevation model (DEM) from the Shuttle Radar  
Topography Mission (SRTM) database (Jarvis et al., 2008) is utilized to delineate the river network



165 **Figure 2.** Flowchart of the proposed forecasting framework.

and basin boundaries. Regarding soil properties, soil texture distributions are derived from Shangguan et al. (2014), while key soil hydraulic parameters, including saturated hydraulic conductivity, saturated and residual water contents, and van Genuchten parameters, are sourced from the China Soil Hydraulic Parameters Dataset (Dai et al., 2013). For land surface characterization, this study uses a 100-m resolution land use map from the Resource and Environment Data Cloud Platform (<http://www.resdc.cn/>). Finally, vegetation dynamics, specifically leaf area index (LAI) and fraction of photosynthetically active radiation (FPAR), are parameterized using the GIMMS NDVI3g-based datasets developed by Zhu et al. (2013).

175

### 2.2.3 Meteorological forecast data

Meteorological forecast data are sourced from the Sub-seasonal to Seasonal Prediction Project (Vitart et al., 2017) which provides medium- and long-range forecasts generated by numerical weather prediction models from various international operational centers. Among the models providing a 60-day lead time (including those from CMA and UKMO), the UKMO model is selected as the source of raw precipitation forecasts due to its superior performance in the study region (Li et al., 2019). The dataset, provided at a 0.25° spatial resolution, covers the hindcast period of 1997–2016 with 44 initialization dates per year. The forecast variables include the target precipitation forecasts together with a suite of predictors for correction: convective precipitation, 2-m dew point temperature, 2-m air temperature, total cloud cover, and multi-level variables (geopotential height, specific humidity, temperature, and wind speed at 200, 500, and 850 hPa). The selection of these specific predictors is based on previous studies which have demonstrated their efficacy in improving forecast accuracy (Zhang et al., 2024; Zhang et al., 2023; Li et al., 2023).

## 190 3 Methods

### 3.1 Overview

In this study, we propose a framework to improve medium- and long-term forecasts of daily precipitation and streamflow with lead times of up to 60 days to enhance hydrological hazard early warning capabilities. The flowchart of the proposed framework is illustrated in Figure 2. First, a LeNet-based CNN model is developed to perform bias correction on the precipitation forecasts by modeling the relationship between local grid precipitation and predictor variables from both the target and surrounding grids. Subsequently, the gridded precipitation forecast serves as input to drive the distributed hydrologic model GBEHM, generating daily streamflow forecasts for the 1–60 day

195



200 horizon. Finally, the simulated streamflow undergoes post-processing via an ARX model to yield the final forecasts. The performance of these forecasts is quantitatively evaluated using metrics including the root mean squared error (RMSE), temporal correlation coefficient (TCC), relative error (RE) and the Nash–Sutcliffe efficiency (NSE).

### 3.2 Convolutional neural networks

205 A modified LeNet-based convolutional neural network (CNN) is designed to model the complex non-linear relationship between local grid precipitation and large-scale atmospheric conditions. The model input consists of a multivariate 3D tensor with dimensions of  $20 \times 9 \times 9$  representing a  $9 \times 9$  spatial neighborhood centered on the target  $0.25^\circ$  grid cell. The input comprises 20 predictor channels, including the DEM, precipitation forecasts and other meteorological variables detailed in Sect. 2.2.3. Structurally, the network integrates three primary components: embedding layers, a convolutional backbone, and fully connected layers. To explicitly capture spatial location dependencies, the latitude and longitude indices of each grid are encoded via two independent embedding layers. The convolutional backbone consists of four cascaded layers (configured with 64, 32, 16, and 8 filters, respectively), each followed by batch normalization to enhance training stability. The extracted spatial feature maps are flattened and concatenated with the spatial embeddings before passing through two fully connected hidden layers.

220 The final layer of the network produces three real numbers ( $z_1, z_2, z_3$ ) that govern the parameters of a censored shifted gamma (CSG) distribution. This probabilistic formulation is specifically adopted to address the mixed discrete-continuous nature of precipitation, effectively characterizing zero-inflation (dry days), skewness, and the transition from light to heavy rainfall (Towler et al., 2025; Zhang et al., 2017). To ensure mathematical validity, the raw network outputs are transformed into the physical distribution parameters of shift ( $\gamma$ ), mean ( $\mu$ ), and standard deviation ( $\sigma$ ) via specific activation functions:  $\gamma = -\sqrt{z_1^2 + \epsilon}$ ,  $\mu = \exp(z_2)$ , and  $\sigma = \sqrt{\exp(z_3) + \epsilon}$ , where  $\epsilon$  is a small constant for numerical stability. Mathematically, the CSG distribution is defined by a latent variable  $X = Z + \gamma$ , where  $Z \sim \text{Gamma}(\alpha, \beta)$  with shape  $\alpha = (\mu/\sigma)^2$  and scale  $\beta = \sigma^2/\mu$ . The precipitation  $Y$  is then obtained by censoring  $X$  at zero, such that  $Y = \max(0, X)$ . The probability of zero precipitation is given by  $P(Y = 0) = F_\Gamma(-\gamma; \alpha, \beta)$ , where  $F_\Gamma$  is the cumulative distribution function of the Gamma distribution. Finally, the model generates a deterministic forecast by constructing a large-scale pseudo-ensemble from the predicted CSG distribution at equal quantiles and calculating the ensemble mean. This approach enables a comprehensive quantification of uncertainty and ensures that low-probability, high-impact extreme events in the tail of distribution are adequately captured, thereby mitigating the smoothing effect inherent in deterministic forecasts.

235 The model is trained using the continuous ranked probability score (CRPS) as the loss function to optimize probabilistic performance:

$$CRPS(F, y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{H}_{\{x \geq y\}})^2 dx \quad (1)$$

where  $F$  is the predicted cumulative distribution function and  $\mathbb{H}$  is the Heaviside step function that equals 1 if  $x \geq y$  and 0 otherwise. To ensure robust generalization, the 20-year dataset is partitioned into training (12 years), validation (4 years), and testing (4 years) subsets within a cross-validation framework. The optimization is performed using the Adam algorithm, with



hyperparameters tuned to minimize overfitting.

### 3.3 Geomorphology-based eco-hydrological model

245 The GBEHM is a physically-based, distributed model designed to simulate hydrological processes  
in topographically complex catchments. To address spatial heterogeneity, the model employs a  
hierarchical sub-grid parameterization scheme (Yang et al., 2015), where the hillslope-valley system  
serves as the fundamental computational unit. Runoff generation is explicitly resolved through three  
primary pathways: surface overland flow, lateral subsurface flow, and groundwater discharge.  
250 Specifically, vertical soil water movement in the unsaturated zone is governed by the one-  
dimensional Richards' equation, while lateral flow in the saturated zone and groundwater-river  
exchange are quantified using Darcy's law and mass balance principles (Cong et al., 2009). To close  
the water and energy budget, the hydrological module is coupled with the Simple Biosphere Model 2  
(SiB2), which estimates evapotranspiration losses, including canopy interception and soil  
255 evaporation, based on energy transfer within the soil–plant–atmosphere continuum (Sellers et al.,  
1996).

A distinctive feature of GBEHM is its enhanced representation of cryosphere hydrology, making it  
particularly robust for cold regions. The model integrates a rigorous coupled heat and water balance  
260 equation (Flerchinger and Saxton, 1989) to simulate soil freeze-thaw cycles, which critically alter  
soil hydraulic conductivity and infiltration capacity. The vertical soil profile is discretized into a  
multi-layer structure extending to a depth of 50 m. Crucially, the active soil layer (top 1–3 m)  
features a refined mesh resolution to accurately capture the dynamics of the active layer thickness  
and the maximum frozen depth (Guo and Wang, 2013).

265 The selection of GBEHM for this study is driven by its proven efficacy in simulating complex  
hydrological processes within the Tibetan Plateau and other high-altitude regions (Shi et al., 2020;  
Gao et al., 2018). The study area is characterized by rugged terrain and a cold climate, where glacier  
and snow melt as well as soil freeze-thaw processes exert a dominant control on the hydrological  
270 regime. Since GBEHM explicitly accounts for phase changes in the soil and the delayed runoff  
response caused by the cryosphere, it is well-suited for accurate runoff forecasting in this catchment.  
In this application, the UYRB is discretized into an 8 km × 8 km grid system and further delineated  
into 479 sub-basins based on the DEM.

### 275 3.4 Autoregressive with exogenous input

While the GBEHM captures the fundamental physical processes of runoff generation, hydrological  
simulations inevitably contain systematic biases and persistent errors due to uncertainties in model  
structure and parameters. To mitigate these discrepancies and improve forecast accuracy, this study  
implements a statistical post-processing technique. Instead of modeling the streamflow directly, an  
280 autoregressive with exogenous input (ARX) model is constructed to simulate and correct the  
hydrological residuals.

Let  $Q_{obs}(t)$  and  $Q_{sim}(t)$  denote the observed and GBEHM-simulated streamflow at time  $t$ ,  
respectively. To stabilize the variance and eliminate seasonal scale effects, both series are first  
285 standardized using the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) derived strictly from the calibration



period. The standardized hydrological error,  $E_{std}(t)$ , is defined as the deviation of the simulated flow from the observations:

$$E_{std}(t) = Q_{obs_{std}}(t) - Q_{sim_{std}}(t) \quad (2)$$

Hydrological errors typically exhibit strong temporal autocorrelation and dependence on flow magnitude. To capture these dynamics, the error at the current time step is modeled as a function of the current simulated state, antecedent simulated states, and antecedent errors. The ARX model formulation is expressed as:

$$E_{std}^*(t) = \beta_0 + \sum_{i=1}^p \phi_i E_{std}(t-i) + \sum_{j=0}^k \gamma_j Q_{sim_{std}}(t-j) + \varepsilon(t) \quad (3)$$

where  $p$  and  $k$  are the lag orders for the autoregressive error term and the exogenous simulated streamflow term respectively, which are optimally determined by identifying the combination that yields the minimum values of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Hipel and McLeod, 1994);  $\phi_i$  represents the autoregressive coefficients describing the persistence of model errors;  $\gamma_j$  are coefficients for the exogenous input, accounting for magnitude-dependent bias;  $\beta_0$  is the intercept standing for the independent bias, and  $\varepsilon(t)$  is the residual white noise. The model parameters are calibrated using the least squares method. The corrected error  $E_{std}^*(t)$  is generated recursively and added to the standardized forecast simulation. Finally, the post-processed streamflow is obtained by applying inverse standardization to the result.

### 3.5 Evaluation metrics

The accuracy of the corrected precipitation forecasts is evaluated using the RMSE to quantify error magnitude and the TCC to measure phase consistency. These metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2} \quad (4)$$

$$TCC = \frac{\sum_{i=1}^n (S_i - \bar{S})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (S_i - \bar{S})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (5)$$

where  $S_i$  and  $O_i$  represent the simulated (or forecasted) and observed values at time step  $i$ , respectively;  $\bar{S}$  and  $\bar{O}$  denote their corresponding means; and  $n$  is the total number of samples.

For streamflow simulations, model performance is appraised using three standard hydrological efficiency criteria: the RMSE (same as above), RE, and NSE. The RE and NSE are calculated as:

$$RE = \frac{\bar{S} - \bar{O}}{\bar{O}} \times 100\% \quad (6)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (7)$$

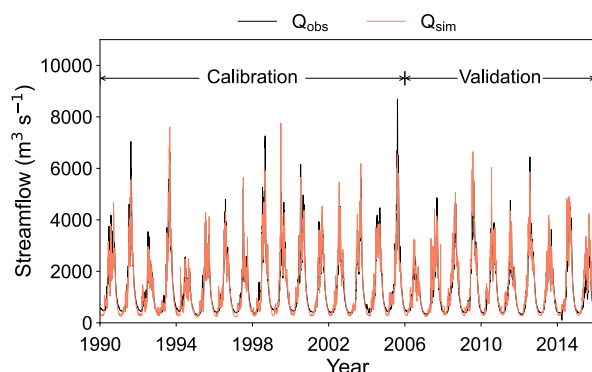
## 4 Results

### 4.1 Calibration and validation of GBEHM and ARX model

The GBEHM is calibrated for the period of 1990–2005 and validated for the period of 2006–2015. Figure 3 illustrates the comparison of the simulated daily streamflow and the observations at the Shigu station. The RMSE, RE and NSE values for the calibration period are  $330 \text{ m}^3 \text{ s}^{-1}$ , 20.2% and 0.93, respectively, and for the validation period are  $347 \text{ m}^3 \text{ s}^{-1}$ , 20.2% and 0.90, which indicating the high reliability of the GBEHM for daily streamflow simulation in the study area.



325 The ARX model is calibrated (2000–2008) and validated (2009–2012) using GBEHM-simulated  
 streamflow driven by CGDPA precipitation data and observed streamflow. After testing different  
 model orders,  $p = 3$  and  $k = 3$  is selected based on validation results (see Sect. S1 in the  
 supplement for details). For the validation period, the raw GBEHM simulation shows the RMSE of  
 351  $\text{m}^3 \text{s}^{-1}$ , RE of 18.9%, and NSE of 0.91. Post-correction with the ARX model, these metrics  
 improve to 267  $\text{m}^3 \text{s}^{-1}$ , 12.5%, and 0.95, respectively. This confirms that the ARX model effectively  
 330 mitigates errors in streamflow simulation.



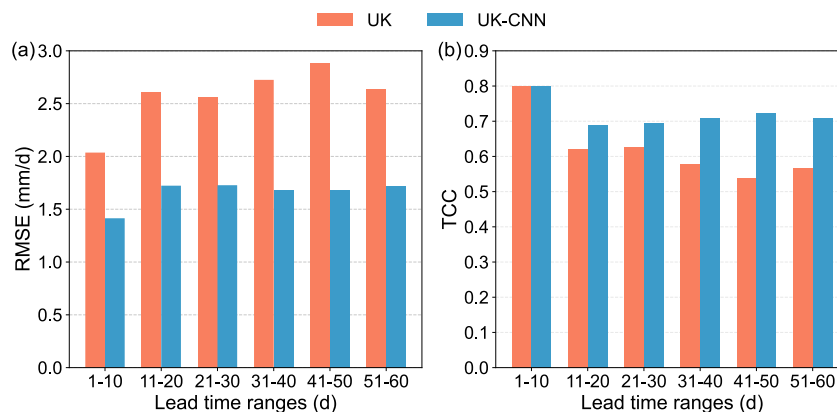
**Figure 3.** Comparison of observed and simulated daily streamflow at the Shigu station during the calibration (1990–2005) and validation (2006–2015) periods.

335

#### 4.2 Evaluation of precipitation forecasts

Figure 4 illustrates the performance of the areal-averaged raw and corrected (denoted as UK and UK-CNN) precipitation forecasts in terms of RMSE and TCC across different lead time ranges for the 2009–2012 period.

340



**Figure 4.** (a) RMSE and (b) TCC of the UK and UK-CNN precipitation forecasts, calculated for aggregated 10-day lead times.



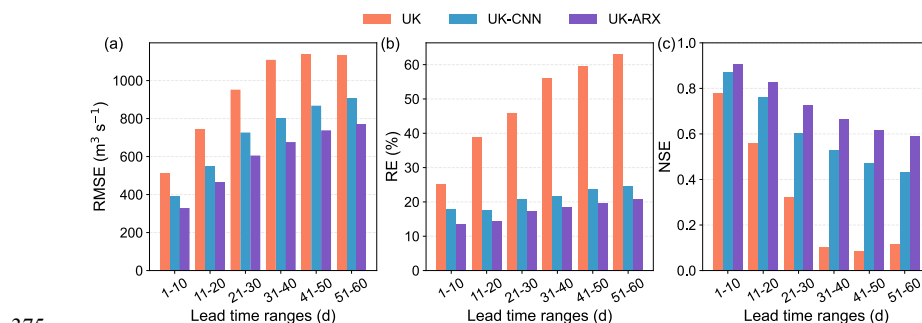
345 In terms of overall performance averaged over all lead times, the areal-averaged UK raw forecasts  
exhibit an RMSE of  $2.6 \text{ mm d}^{-1}$  and a TCC of 0.62. The UK-CNN effectively improve these metrics,  
reducing the average RMSE to  $1.7 \text{ mm d}^{-1}$  (a decrease of 35%) and increasing the TCC to 0.74 (an  
improvement of 18%). Regarding the temporal evolution, the skill of UK raw forecasts naturally  
deteriorates with increasing lead time. The UK-CNN follow a similar declining trend initially but  
350 stabilize at a relatively constant level after approximately 20 days. Consequently, the relative  
improvement gap widens as the lead time extends. For instance, the RMSE reduction increases from  
33% in the first 20 days to 35% and 38% in the middle and last 20-day periods, respectively.  
Similarly, the TCC improvement rises significantly from 5% to 17% and 30% over the same  
intervals.

355

### 4.3 Evaluation of streamflow forecasts

Figure 5 shows the RMSE, RE, and NSE for streamflow forecasts driven by raw UK and corrected  
UK-CNN precipitation, as well as the UK-CNN driven forecasts further post-processed by the ARX  
model (denoted as UK-ARX). In terms of overall performance averaged over all lead times, the  
360 streamflow forecasts driven by raw UK precipitation yield an RMSE of  $930 \text{ m}^3 \text{ s}^{-1}$ , an RE of 48.2%,  
and an NSE of 0.33. Using UK-CNN precipitation significantly improves these metrics to  $706 \text{ m}^3 \text{ s}^{-1}$ ,  
21.0%, and 0.61, representing relative improvements of 24%, 56%, and 87%, respectively. The UK-  
ARX achieves the best performance, further reducing the RMSE and RE to  $596 \text{ m}^3 \text{ s}^{-1}$  and 17.4%,  
and increasing the NSE to 0.72. This corresponds to additional improvements of 16%, 17%, and 18%  
365 over the UK-CNN benchmark.

As the forecast lead time extends, the skill of streamflow forecasts naturally declines with increasing  
lead time; however, the correction methods notably mitigate the rate of this deterioration. For  
instance, the RE of the UK-driven forecasts rises sharply from 25.2% to 63.2% by lead day 51-60,  
370 while the NSE plummets from 0.78 to 0.09. Conversely, the RE of the UK-CNN forecasts increases  
only marginally from 17.6% to 24.5%, with the NSE showing a more gradual decline from 0.87  
to a moderate level of 0.43. The UK-ARX further alleviates this degradation trend, maintaining the RE  
below 20.9% and sustaining the NSE above 0.59 throughout the entire 60-day period.



375

**Figure 5.** (a) RMSE, (b) RE and (c) NSE of streamflow forecasts driven by raw (UK) and corrected (UK-CNN) precipitation, and the streamflow subsequently post-processed by the ARX model (UK-ARX), calculated for aggregated 10-day lead times.



380 Moreover, the extent of improvement exhibits distinct trends. The performance gap between UK  
and UK-CNN becomes more pronounced at longer lead times. For instance, the reduction in RE  
achieved by UK-CNN (compared to UK) increases from around 44% in the first 20 days to 58%  
and 61% in the middle and last 20-day periods, respectively. Conversely, the incremental  
improvement from UK-ARX (relative to UK-CNN) remains stable, providing RE reductions of  
385 around 21%, 16%, and 16% across these lead time ranges.

## 5 Discussion

### 5.1 Efficacy of the CNN-based precipitation bias correction

This study extends the horizon of reliable precipitation forecasting to 60 days by employing a CNN-  
based deep learning model for bias correction. Specifically, the model achieves a TCC of  
390 approximately 0.70 in the extended range (days 21–30), representing a substantial advantage over  
the results reported by Lyu et al. (2023) who achieved a TCC of about 0.35 for the same period  
when correcting summer precipitation over Southern China using ECMWF data with lead times up  
to 30 days. More importantly, the model maintains a TCC above 0.66 even at long lead times of 51–  
395 60 days, with the RMSE stabilizing after the initial 20 days. These findings validate the model's  
robustness in mitigating error accumulation and sustaining forecast reliability up to 60 days.

This substantial enhancement in performance is primarily attributed to the deep learning approach  
and the probabilistic output scheme. First, unlike traditional point-to-point statistical correction  
400 methods (e.g., quantile mapping), the CNN architecture effectively extracts spatial features and  
captures the non-linear dependencies between large-scale atmospheric circulation and local  
precipitation patterns (Baño-Medina et al., 2020; Pan et al., 2019). Second, instead of generating  
deterministic point predictions, this model explicitly corrects the CSG distribution optimized via  
CRPS. This approach not only precisely captures the mixed zero-inflated and heavy-tailed  
405 characteristics of precipitation (Ghazvinian et al., 2022) but also mitigates the smoothing effect  
common in deep learning by deriving forecasts from a pseudo-ensemble (Ravuri et al., 2021; Rasp  
and Lerch, 2018), thereby preserving signal variability and ensuring high reliability even at extended  
lead times.

### 5.2 Added value of ARX-based hydrological post-processing

Our framework extends the valid streamflow forecast horizon to 60 days with the ARX post-  
processing model providing significant added value in mitigating hydrological residuals.  
Specifically, while the skill of UK-CNN forecasts naturally degrades at extended horizons (with  
NSE dropping to 0.43 by day 60), the ARX model effectively alleviates this decline, sustaining a  
415 reliable NSE above 0.57 and keeping RE below 22% throughout the entire period. Furthermore, the  
ARX model achieves a consistent performance boost (about 16–21% reduction in RE) across all  
lead times, demonstrating its unique capability to mitigate intrinsic structural errors within the  
hydrological model and ensuring the validity of forecasts up to 60 days.

420 The efficacy of the ARX model in achieving these results stems from its capacity to mitigate intrinsic  
hydrological modeling uncertainties while remaining a computationally efficient solution. While  
precipitation bias correction improves inputs, hydrological models inevitably introduce intrinsic  
biases due to simplified parameterizations, structural deficiencies, or uncertain initial conditions.



425 The ARX model addresses this by exploiting error autocorrelation to effectively remove systematic  
and temporal discrepancies, ensuring stable correction capabilities across the entire forecast horizon.  
Furthermore, compared to data assimilation approaches which aim to reduce these uncertainties by  
continuously updating internal model states via computationally intensive ensemble simulations  
(Nearing et al., 2022; Liu et al., 2012), the ARX model provides a more efficient alternative by  
bypassing these complex internal adjustments entirely, thereby serving as a straightforward post-  
430 processing solution characterized by interpretability and minimal data requirements.

### 5.3 Attribution of the streamflow forecast error

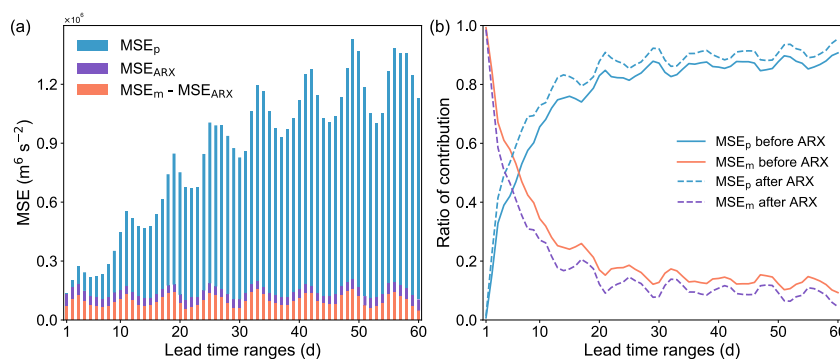
An error decomposition framework is adopted to evaluate the relative contributions of error sources.  
The total streamflow forecast error is decomposed into hydrological model error ( $MSE_m$ ), defined  
435 as the MSE between observed precipitation driven simulations and observed streamflow, and  
precipitation forecast error ( $MSE_p$ ), defined as the MSE between observed and forecast precipitation  
driven simulations. Additionally, the error reduction contribution from the ARX post-processing  
( $MSE_{ARX}$ ) is quantified as the MSE between the streamflow forecasts before and after ARX  
correction. Notably, the total forecast error is generally smaller than the arithmetic sum of  $MSE_m$   
440 and  $MSE_p$ . This non-additivity is attributed to the interaction between error sources, where a  
compensation effect between precipitation biases and hydrological model deficiencies helps  
mitigate the overall error.

Figure 6 depicts the absolute values and relative proportions of  $MSE_m$  and  $MSE_p$  before and after  
445 ARX post-processing. Generally, the magnitudes of  $MSE_m$  and the ARX-induced error reduction  
 $MSE_{ARX}$  show little variation with lead time, while  $MSE_p$  exhibits a steady increase. Prior to ARX  
post-processing,  $MSE_m$  is the dominant error source within the first week of lead time, accounting  
for over 50% of the total. This proportion drops rapidly and stabilizes at around 0.15 after 30 days,  
while the share of  $MSE_p$  inversely rises to stabilize at around 0.85. With the application of ARX,  
450 the dominance of the hydrological model error is significantly curtailed; it ceases to be the primary  
error source after a lead time of only 5 days and stabilizes at a proportion of roughly 0.10 after 20  
days. Notably, the  $MSE$  displays a pronounced 8-day periodicity with peaks occurring at the third  
phase of each cycle. This pattern stems from the systematic alignment of validity dates caused by  
the 8-day forecast interval, where errors from sparse extreme events are systematically projected  
455 onto specific lead times, thus driving the peak formation (details provided in Sect. S2 in the  
Supplement).

The error decomposition analysis reveals a clear shift in dominance: short-term skill is constrained  
by hydrological modeling, while medium-to-long-term skill is limited by precipitation forecasts. In  
460 our study, hydrological model error ( $MSE_m$ ) stabilizes at a notably low contribution of  
approximately 15% at long lead times, outperforming the about 30% reported by Dong et al. (2025)  
who used a hybrid deep learning-conceptual model. This superior performance is attributed to the  
GBEHM's capacity to resolve the spatial heterogeneity of the underlying surface and explicitly  
model the cryosphere hydrological processes, thereby representing the complex runoff generation  
465 mechanisms in this cold mountainous region more accurately. The ARX post-processing further  
compresses this error to about 10% and shortens its dominance period to just 5 days, confirming the  
high reliability of process-based models augmented by post-processing. Nevertheless, the upper



limit of forecast skill remains constrained by meteorological forcing, which dominates the long-term error contribution (>85%). This highlights the indispensability of the proposed framework: effective medium-to-long-range forecasting demands both a high-precision hydrological model to minimize internal uncertainty and advanced precipitation bias correction to mitigate the overwhelming external forcing errors.



475 **Figure 6.** Decomposition of streamflow forecast errors into hydrological model ( $MSE_m$ ) and precipitation forecast ( $MSE_p$ ) components before and after ARX post-processing across lead times.

#### 5.4 Limitations of this study

One limitation of this study lies in the use of an identical set of predictors across the entire 1–60 day forecast horizon. Although separate models are trained for different lead times, allowing the weights of these predictors to adjust dynamically, previous studies indicate that the dominant atmospheric drivers often shift as the lead time extends (Lyu et al., 2023). Therefore, incorporating predictor importance ranking to select lead-time-specific predictor subsets could likely further enhance the model's fitting capability.

Beyond this specific constraint, future research could expand in two aspects. First, while this study yields promising results using UKMO meteorological forecasts, the potential utility of the CMA model, which also supports a 60-day forecast horizon, deserves further evaluation. Second, as the proposed framework is validated at a single hydrological station, broader testing across diverse catchments is essential to verify its generalizability.

#### 6 Summary and conclusions

This study constructs a robust 60-day streamflow forecasting framework by coupling a CNN for correcting UKMO precipitation forecasts, the GBEHM for process-based hydrological simulation, and an ARX model for mitigating hydrological modelling residuals. This framework is applied to the Upper Yangtze River Basin (UYRB).

The results demonstrate a significant extension of the valid streamflow forecast horizon to 60 days. Compared to raw forecasts, the proposed method reduced the RE from a range of 25.2%–63.2% to 13.7%–20.9%, while elevating the NSE from 0.78–0.09 to a reliable 0.91–0.59. In terms of contributions, this performance boost is driven by two key components. First, the CNN-based model



significantly improved meteorological inputs by reducing precipitation RMSE by 35% and elevating TCC from 0.62 to 0.74, particularly at longer lead times; this enhancement in precipitation accounts for approximately two-thirds of the total improvement in streamflow forecasts. Meanwhile, the  
505 ARX post-processing contributed the remaining 33% to the total streamflow error reduction by effectively mitigating intrinsic hydrological residuals.

Our framework highlights the distinct advantages of integrating a physically robust hydrological model with a dual-stage error correction strategy. First, the distributed GBEHM demonstrates  
510 exceptional precision in characterizing complex catchment dynamics, maintaining notably low residual errors (15%) even at extended lead times. Second, by strategically coupling deep learning (CNN) for precipitation input correction with statistical post-processing (ARX) for hydrological output refinement, the framework systematically mitigates both external meteorological biases and internal simulation uncertainties. This synergy yields forecasts that are both volumetrically accurate  
515 and temporally consistent, thereby providing a highly reliable solution with an extended 60-day horizon for hydrological hazard early warning and proactive flood and drought risk mitigation.

520 *Code availability.* The source code for the forecasting framework used in this research is available by contacting the authors upon reasonable request.

*Data availability.* The datasets utilized in this research and their respective sources are detailed in Sect. 2.2. All data supporting the findings of this research are publicly available online: the CGDPA  
525 precipitation data (<http://cdc.nmic.cn/sksj.do?method=ssrjscp>), the meteorological data (<http://data.cma.cn>), the DEM data (<http://srtm.csi.cgiar.org>), the soil data and soil hydraulic parameters (<http://globalchange.bnu.edu.cn>), the land use data (<http://www.resdc.cn/>), the LAI and FPAR data (<https://www.nasa.gov/nasa-earth-exchange-nex/>), the UKMO forecast data (<https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/>), except for the streamflow records  
530 for hydrological gauging stations which are available upon reasonable request.

*Author contribution.* ZL designed the research, developed the code, conducted the data processing, analysis and wrote the original draft of the paper. HY designed the research and edited the manuscript. DY designed the research.  
535

*Competing interests.* The contact author has declared that neither they nor their co-authors have any competing interests.

*Acknowledgements.* This research was supported by the China National Key R&D Program (grant  
540 no. 2021YFC3000202).

## References

Adams, T. E., III and Dymond, R. L.: Possible hydrologic forecasting improvements resulting from  
advancements in precipitation estimation and forecasting for a real-time flood forecast system  
545 in the Ohio River Valley, USA, *J. Hydrol.*, 579, 124138,



- <https://doi.org/10.1016/j.jhydrol.2019.124138>, 2019.
- Adnan, R. M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O., and Li, B.: Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs, *J. Hydrol.*, 586, 124371, 550  
<https://doi.org/10.1016/j.jhydrol.2019.124371>, 2020.
- Andrade, F. S. A., Arsenault, R., Poulin, A., Troin, M., and Armstrong, W.: Application of weather post-processing methods for operational ensemble hydrological forecasting on multiple catchments in Canada, *J. Hydrol.*, 642, 131861, <https://doi.org/10.1016/j.jhydrol.2024.131861>, 2024.
- 555 Anghileri, D., Monhart, S., Zhou, C., Bogner, K., Castelletti, A., Burlando, P., and Zappa, M.: The Value of Subseasonal Hydrometeorological Forecasts to Hydropower Operations: How Much Does Preprocessing Matter?, *Water Resour. Res.*, 55, 10159-10178, <https://doi.org/10.1029/2019wr025280>, 2019.
- Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: Configuration and intercomparison of deep learning neural models for statistical downscaling, *Geosci. Model Dev.*, 13, 2109-2124, 560  
<https://doi.org/10.5194/gmd-13-2109-2020>, 2020.
- Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, *Nature*, 525, 47-55, <https://doi.org/10.1038/nature14956>, 2015.
- Bogner, K., Chang, A. Y. Y., Bernhard, L., Zappa, M., Monhart, S., and Spirig, C.: Tercile Forecasts for Extending the Horizon of Skillful Hydrological Predictions, *J. Hydrometeorol.*, 23, 521-539, 565  
<https://doi.org/10.1175/jhm-d-21-0020.1>, 2022.
- Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?, *J. Clim.*, 28, 6938-6959, <https://doi.org/10.1175/JCLI-D-14-00754.1>, 2015.
- 570 Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., and Pain, C. C.: Long lead-time daily and monthly streamflow forecasting using machine learning methods, *J. Hydrol.*, 590, 125376, <https://doi.org/10.1016/j.jhydrol.2020.125376>, 2020.
- Cong, Z., Yang, D., Gao, B., Yang, H., and Hu, H.: Hydrological trend analysis in the Yellow River basin using a distributed hydrological model, *Water Resour. Res.*, 45, W00A13, 575  
<https://doi.org/10.1029/2008WR006852>, 2009.
- Dai, Y., Shangguan, W., Duan, Q., Liu, B., Fu, S., and Niu, G.: Development of a China dataset of soil hydraulic parameters using pedotransfer functions for land surface modeling, *J. Hydrometeorol.*, 14, 869-887, <https://doi.org/10.1175/JHM-D-12-0149.1>, 2013.
- Dion, P., Martel, J.-L., and Arsenault, R.: Hydrological ensemble forecasting using a multi-model framework, *J. Hydrol.*, 600, 126537, <https://doi.org/10.1016/j.jhydrol.2021.126537>, 2021. 580
- Donegan, S., Murphy, C., Harrigan, S., Broderick, C., Quinn, D. F., Golian, S., Knight, J., Matthews, T., Prudhomme, C., Scaife, A. A., Stringer, N., and Wilby, R. L.: Conditioning ensemble streamflow prediction with the North Atlantic Oscillation improves skill at longer lead times, *Hydrol. Earth Syst. Sci.*, 25, 4159-4183, <https://doi.org/10.5194/hess-25-4159-2021>, 2021.
- 585 Dong, N., Hao, H., Yang, M., Wei, J., Xu, S., and Kunstmann, H.: Deep-learning-based sub-seasonal precipitation and streamflow ensemble forecasting over the source region of the Yangtze River, *Hydrol. Earth Syst. Sci.*, 29, 2023-2042, <https://doi.org/10.5194/hess-29-2023-2025>, 2025.
- Du, Y., Wang, Q. J., Wu, W., and Su, C.-H.: Calibration of precipitation forecasts from NWP models for ungauged locations, *J. Hydrol.*, 661, 133733, <https://doi.org/10.1016/j.jhydrol.2025.133733>,



- 590 2025.
- Falck, A. S., Tomasella, J., Diniz, F. L. R., and Maggioni, V.: Assessment of subseasonal streamflow predictions in a tropical basin, *J. Hydrol.*, 651, 132488, <https://doi.org/10.1016/j.jhydrol.2024.132488>, 2025.
- Flerchinger, G. and Saxton, K.: Simultaneous heat and water model of a freezing snow-residue-soil system I. Theory and development, *Trans. ASAE*, 32, 565-571, <https://doi.org/10.13031/2013.31040>, 1989.
- 595 Gao, B., Qin, Y., Wang, Y., Yang, D., and Zheng, Y.: Modeling ecohydrological processes and spatial patterns in the upper Heihe Basin in China, *Forests*, 7, 10, <https://doi.org/10.3390/f7010010>, 2015.
- 600 Gao, B., Yang, D., Qin, Y., Wang, Y., Li, H., Zhang, Y., and Zhang, T.: Change in frozen soils and its effect on regional hydrology, upper Heihe basin, northeastern Qinghai–Tibetan Plateau, *The Cryosphere*, 12, 657-673, <https://doi.org/10.5194/tc-12-657-2018>, 2018.
- Ghazvinian, M., Zhang, Y., Hamill, T. M., Seo, D.-J., and Fernando, N.: Improving Probabilistic Quantitative Precipitation Forecasts Using Short Training Data through Artificial Neural Networks, *J. Hydrometeorol.*, 23, 1365-1382, <https://doi.org/10.1175/JHM-D-22-0021.1>, 2022.
- 605 Ghimire, G. R., Krajewski, W. F., and Quintero, F.: Scale-Dependent Value of QPF for Real-Time Streamflow Forecasting, *J. Hydrometeorol.*, 22, 1931-1947, <https://doi.org/10.1175/jhm-d-20-0297.1>, 2021.
- Greuell, W. and Hutjes, R. W. A.: Skill and sources of skill in seasonal streamflow hindcasts for South America made with ECMWF's SEAS5 and VIC, *J. Hydrol.*, 617, 128806, <https://doi.org/10.1016/j.jhydrol.2022.128806>, 2023.
- 610 Guo, D. and Wang, H.: Simulation of permafrost and seasonally frozen ground conditions on the Tibetan Plateau, 1981–2010, *J. Geophys. Res. Atmos.*, 118, 5216-5230, <https://doi.org/10.1002/jgrd.50457>, 2013.
- 615 Guo, S. L., Wen, Y. H., Zhang, X. Q., and Chen, H. Y.: Runoff prediction of lower Yellow River based on CEEMDAN-LSSVM-GM(1,1) model, *Sci. Rep.*, 13, 1511, <https://doi.org/10.1038/s41598-023-28662-5>, 2023.
- Hipel, K. W. and McLeod, A. I.: Time series modelling of water resources and environmental systems, Elsevier, ISBN 0080870368, 1994.
- 620 Huang, Z., Zhao, T., Liu, Y., Zhang, Y., Jiang, T., Lin, K., and Chen, X.: Differing roles of base and fast flow in ensemble seasonal streamflow forecasting: An experimental investigation, *J. Hydrol.*, 591, 125272, <https://doi.org/10.1016/j.jhydrol.2020.125272>, 2020.
- Hunt, K. M. R., Matthews, G. R., Pappenberger, F., and Prudhomme, C.: Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States, *Hydrol. Earth Syst. Sci.*, 26, 5449-5472, <https://doi.org/10.5194/hess-26-5449-2022>, 2022.
- 625 Jackson-Blake, L. A., Clayer, F., de Eyto, E., French, A. S., Dolores Frias, M., Mercado-Bettin, D., Moore, T., Puertolas, L., Poole, R., Rinke, K., Shikhani, M., van der Linden, L., and Marce, R.: Opportunities for seasonal forecasting to support water management outside the tropics, *Hydrol. Earth Syst. Sci.*, 26, 1389-1406, <https://doi.org/10.5194/hess-26-1389-2022>, 2022.
- 630 Jamei, M., Jamei, M., Ali, M., Karbasi, M., Farooque, A. A., Malik, A., Cheema, S. J., Esau, T. J., and Yaseen, Z. M.: Quantitative improvement of streamflow forecasting accuracy in the Atlantic zones of Canada based on hydro-meteorological signals: A multi-level advanced



- intelligent expert framework, *Ecol. Inform.*, **80**, 102455,  
635 <https://doi.org/10.1016/j.ecoinf.2023.102455>, 2024.
- Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTN 90m Database, <http://srtm.csi.cgiar.org> (last access: 1 January 2026), 2008.
- Koh, R. and Galelli, S.: Evaluating Streamflow Forecasts in Hydro-Dominated Power Systems-  
640 When and Why They Matter, *Water Resour. Res.*, **60**, e2023WR035825, <https://doi.org/10.1029/2023wr035825>, 2024.
- Kondal, A., Hegewisch, K., Liu, M., Abatzoglou, J. T., Adam, J. C., Nijssen, B., and Rajagopalan, K.: Seasonal forecasts have sufficient skill to inform some agricultural decisions, *Environ. Res. Lett.*, **19**, 124049, <https://doi.org/10.1088/1748-9326/ad8bde>, 2024.
- 645 Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Hernegger, M.: Rainfall-runoff modelling using long short-term memory (LSTM) networks, *Hydrol. Earth Syst. Sci.*, **22**, 6005-6022, <https://doi.org/10.5194/hess-22-6005-2018>, 2018.
- Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., Abeshu, G. W., Agafonova, S., AghaKouchak, A., Aksoy, H., Alvarez-Garreton, C., Aznar, B., Balkhi, L.,  
650 Barendrecht, M. H., Biancamaria, S., Bos-Burgering, L., Bradley, C., Budiyo, Y., Buytaert, W., Capewell, L., Carlson, H., Cavus, Y., Couasnon, A., Coxon, G., Daliakopoulos, I., de Ruyter, M. C., Delus, C., Erfurt, M., Esposito, G., François, D., Frappart, F., Freer, J., Frolova, N., Gain, A. K., Grillakis, M., Grima, J. O., Guzmán, D. A., Huning, L. S., Ionita, M., Kharlamov, M., Khoi, D. N., Kieboom, N., Kireeva, M., Koutroulis, A., Lavado-Casimiro, W., Li, H.-Y., Llasat, M. C., Macdonald, D., Mård, J., Mathew-Richards, H., McKenzie, A., Mejia, A., Mendiondo, E. M., Mens, M., Mobini, S., Mohor, G. S., Nagavciuc, V., Ngo-Duc, T., Thao Nguyen Huynh, T., Nhi, P. T. T., Petrucci, O., Nguyen, H. Q., Quintana-Seguí, P., Razavi, S., Ridolfi, E., Riegel, J., Sadik, M. S., Savelli, E., Sazonov, A., Sharma, S., Sörensen, J., Arguello Souza, F. A., Stahl, K., Steinhausen, M., Stoelzle, M., Szalińska, W., Tang, Q., Tian, F., Tokarczyk, T., Tovar, C.,  
660 Tran, T. V. T., Van Huijgevoort, M. H. J., van Vliet, M. T. H., Vorogushyn, S., Wagener, T., Wang, Y., Wendt, D. E., Wickham, E., Yang, L., Zambrano-Bigiarini, M., Blöschl, G., and Di Baldassarre, G.: The challenge of unprecedented floods and droughts in risk management, *Nature*, **608**, 80-86, <https://doi.org/10.1038/s41586-022-04917-5>, 2022.
- Lee, D., Ng, J. Y., Galelli, S., and Block, P.: Unfolding the relationship between seasonal forecast  
665 skill and value in hydropower production: a global analysis, *Hydrol. Earth Syst. Sci.*, **26**, 2431-2448, <https://doi.org/10.5194/hess-26-2431-2022>, 2022.
- Lee, Y., Pianosi, F., Penuela, A., and Rico-Ramirez, M. A.: Skill of seasonal flow forecasts at catchment scale: an assessment across South Korea, *Hydrol. Earth Syst. Sci.*, **28**, 3261-3279, <https://doi.org/10.5194/hess-28-3261-2024>, 2024.
- 670 Li, W., Chen, J., Li, L., Chen, H., Liu, B., Xu, C.-Y., and Li, X.: Evaluation and Bias Correction of S2S Precipitation for Hydrological Extremes, *J. Hydrometeorol.*, **20**, 1887-1906, <https://doi.org/10.1175/jhm-d-19-0042.1>, 2019.
- Li, Y., Xü, K., Wu, Z., Zhu, Z., and Wang, Q. J.: A statistical-dynamical approach for probabilistic  
675 prediction of sub-seasonal precipitation anomalies over 17 hydroclimatic regions in China, *Hydrol. Earth Syst. Sci.*, **27**, 4187-4203, <https://doi.org/10.5194/hess-27-4187-2023>, 2023.
- Liang, H., Zhang, D., Wang, W., Yu, S., and Nimai, S.: Evaluating future water security in the upper Yangtze River Basin under a changing environment, *Sci. Total Environ.*, **889**, 164101,



- <https://doi.org/10.1016/j.scitotenv.2023.164101>, 2023.
- 680 Lin, R., Zhu, J., and Zheng, F.: The Application of the SVD Method to Reduce Coupled Model Biases in Seasonal Predictions of Rainfall, *J. Geophys. Res. Atmos.*, 124, 11837-11849, <https://doi.org/10.1029/2018jd029927>, 2019.
- 685 Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D. J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci.*, 16, 3863-3887, <https://doi.org/10.5194/hess-16-3863-2012>, 2012.
- Luo, X., Yuan, X., Zhu, S., Xu, Z., Meng, L., and Peng, J.: A hybrid support vector regression framework for streamflow forecast, *J. Hydrol.*, 568, 184-193, <https://doi.org/10.1016/j.jhydrol.2018.10.064>, 2019.
- 690 Lyu, Y., Zhu, S. P., Zhi, X. F., Ji, Y., Fan, Y., and Dong, F.: Improving Subseasonal-To-Seasonal Prediction of Summer Extreme Precipitation Over Southern China Based on a Deep Learning Method, *Geophys. Res. Lett.*, 50, e2023GL106245, <https://doi.org/10.1029/2023GL106245>, 2023.
- 695 Lyu, Y., Zhu, S. P., Zhi, X. F., Wang, J. Y., Ji, Y., Fan, Y., and Dong, F.: Significant advancement in subseasonal-to-seasonal summer precipitation ensemble forecast skills in China mainland through an innovative hybrid CSG-UNET method, *Environ. Res. Lett.*, 19, 074055, <https://doi.org/10.1088/1748-9326/ad5577>, 2024.
- 700 McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Woldemeskel, F., Tuteja, N., and Kuczera, G.: Improving the Reliability of Sub-Seasonal Forecasts of High and Low Flows by Using a Flow-Dependent Nonparametric Model, *Water Resour. Res.*, 57, e2020WR029317, <https://doi.org/10.1029/2020wr029317>, 2021.
- 705 Monhart, S., Zappa, M., Spirig, C., Schar, C., and Bogner, K.: Subseasonal hydrometeorological ensemble predictions in small- and medium-sized mountainous catchments: benefits of the NWP approach, *Hydrol. Earth Syst. Sci.*, 23, 493-513, <https://doi.org/10.5194/hess-23-493-2019>, 2019.
- Nearing, G. S., Klotz, D., Frame, J. M., Gauch, M., Gilon, O., Kratzert, F., Sampson, A. K., Shalev, G., and Nevo, S.: Technical note: Data assimilation and autoregression for using near-real-time streamflow observations in long short-term memory networks, *Hydrol. Earth Syst. Sci.*, 26, 5493-5513, <https://doi.org/10.5194/hess-26-5493-2022>, 2022.
- 710 Neri, A., Villarini, G., and Napolitano, F.: Intraseasonal predictability of the duration of flooding above National Weather Service flood warning levels across the US Midwest, *Hydrol. Process.*, 34, 4505-4511, <https://doi.org/10.1002/hyp.13902>, 2020.
- 715 Nie, Y. and Sun, J.: Improving dynamical-statistical subseasonal precipitation forecasts using deep learning: A case study in Southwest China, *Environ. Res. Lett.*, 19, 074013, <https://doi.org/10.1088/1748-9326/ad5370>, 2024.
- Pan, B., Hsu, K., AghaKouchak, A., and Sorooshian, S.: Improving precipitation estimation using convolutional neural network, *Water Resour. Res.*, 55, 2301-2321, <https://doi.org/10.1029/2018WR024090>, 2019.
- 720 Pendergrass, A. G., Meehl, G. A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A., Bonfils, C. J. W., Gallant, A. J. E., Hoerling, M., Hoffmann, D., Kaatz, L., Lehner, F., Llewellyn, D., Mote, P., Neale, R. B., Overpeck, J. T., Sheffield, A., Stahl, K., Svoboda, M., Wheeler, M. C.,



- Wood, A. W., and Woodhouse, C. A.: Flash droughts present a new challenge for subseasonal-to-seasonal prediction, *Nat. Clim. Chang.*, 10, 191-199, <https://doi.org/10.1038/s41558-020-0709-0>, 2020.
- 725 Quedi, E. S. and Fan, F. M.: Sub seasonal streamflow forecast assessment at large-scale basins, *J. Hydrol.*, 584, 124635, <https://doi.org/10.1016/j.jhydrol.2020.124635>, 2020.
- Rasp, S. and Lerch, S.: Neural networks for postprocessing ensemble weather forecasts, *Mon. Weather Rev.*, 146, 3885-3900, <https://doi.org/10.1175/MWR-D-18-0187.1>, 2018.
- Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., and Madge, S.: Skilful precipitation nowcasting using deep generative models of radar, *Nature*, 597, 672-677, <https://doi.org/10.1038/s41586-021-03854-z>, 2021.
- 730 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, F.: Deep learning and process understanding for data-driven Earth system science, *Nature*, 566, 195-204, <https://doi.org/10.1038/s41586-019-0912-1>, 2019.
- 735 Sabzipour, B., Arsenault, R., Troin, M., and Martel, J.-L.: Sensitivity analysis of the hyperparameters of an ensemble Kalman filter application on a semi-distributed hydrological model for streamflow forecasting, *J. Hydrol.*, 626, 130251, <https://doi.org/10.1016/j.jhydrol.2023.130251>, 2023.
- Sellers, P., Randall, D., Collatz, G., Berry, J., Field, C., Dazlich, D., Zhang, C., Collelo, G., and Bounoua, L.: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, *J. Clim.*, 9, 676-705, [https://doi.org/10.1175/1520-0442\(1996\)009<0676:ARLSPF>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2), 1996.
- 740 Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling, *J. Adv. Model. Earth Syst.*, 6, 249-263, <https://doi.org/10.1002/2013MS000293>, 2014.
- 745 Shao, P., Feng, J., Lu, J., and Tang, Z.: Data-driven and knowledge-guided denoising diffusion probabilistic model for runoff uncertainty prediction, *J. Hydrol.*, 638, 131556, <https://doi.org/10.1016/j.jhydrol.2024.131556>, 2024.
- Sharma, S., Siddique, R., Reed, S., Ahnert, P., and Mejia, A.: Hydrological Model Diversity Enhances Streamflow Forecast Skill at Short- to Medium-Range Timescales, *Water Resour. Res.*, 55, 1510-1530, <https://doi.org/10.1029/2018wr023197>, 2019.
- 750 Shen, C.: A transdisciplinary review of deep learning research and its relevance for water resources scientists, *Water Resour. Res.*, 54, 8558-8593, <https://doi.org/10.1029/2018WR022643>, 2018.
- Shen, Y. and Xiong, A.: Validation and comparison of a new gauge-based precipitation analysis over mainland China, *Int. J. Climatol.*, 36, 252-265, <https://doi.org/10.1002/joc.4341>, 2016.
- 755 Shi, R., Yang, H., and Yang, D.: Spatiotemporal variations in frozen ground and their impacts on hydrological components in the source region of the Yangtze River, *J. Hydrol.*, 590, 125237, <https://doi.org/10.1016/j.jhydrol.2020.125237>, 2020.
- 760 Siqueira, V. A., Fan, F. M., Dias de Paiva, R. C., Ramos, M.-H., and Collischonn, W.: Potential skill of continental-scale, medium-range ensemble streamflow forecasts for flood prediction in South America, *J. Hydrol.*, 590, 125430, <https://doi.org/10.1016/j.jhydrol.2020.125430>, 2020.
- 765 Siqueira, V. A., Weerts, A., Klein, B., Fan, F. M., Dias de Paiva, R. C., and Collischonn, W.: Postprocessing continental-scale, medium-range ensemble streamflow forecasts in South America using Ensemble Model Output Statistics and Ensemble Copula Coupling, *J. Hydrol.*, 600, 126520, <https://doi.org/10.1016/j.jhydrol.2021.126520>, 2021.



- Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y. Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., and Zappa, M.: Hybrid forecasting: blending climate predictions with AI models, *Hydrol. Earth Syst. Sci.*, 27, 1865-1889, <https://doi.org/10.5194/hess-27-1865-2023>, 2023.
- 770 Su, B., Huang, J., Zeng, X., Gao, C., and Jiang, T.: Impacts of climate change on streamflow in the upper Yangtze River basin, *Clim. Chang.*, 141, 533-546, <https://doi.org/10.1007/s10584-016-1852-5>, 2017.
- Sutanto, S. J., Wetterhall, F., and Van Lanen, H. A. J.: Hydrological drought forecasts outperform meteorological drought forecasts, *Environ. Res. Lett.*, 15, 084010, <https://doi.org/10.1088/1748-9326/ab8b13>, 2020.
- 775 Sutanto, S. J., Duku, C., Gülveren, M., Dankers, R., and Paparrizos, S.: Future intensification of compound and consecutive drought and heatwave risks in Europe, *Nat. Hazards Earth Syst. Sci.*, 25, 3879-3895, <https://doi.org/10.5194/nhess-25-3879-2025>, 2025.
- Swain, D. L., Prein, A. F., Abatzoglou, J. T., Albano, C. M., Brunner, M., Diffenbaugh, N. S., Singh, D., Skinner, C. B., and Touma, D.: Hydroclimate volatility on a warming Earth, *Nat. Rev. Earth Environ.*, 6, 35-50, <https://doi.org/10.1038/s43017-024-00624-z>, 2025.
- 780 Tabari, H.: Climate change impact on flood and extreme precipitation increases with water availability, *Sci. Rep.*, 10, 13768, <https://doi.org/10.1038/s41598-020-70816-2>, 2020.
- Tanguy, M., Eastman, M., Chevuturi, A., Magee, E., Cooper, E., Johnson, R. H. B., Facer-Childs, K., and Hannaford, J.: Optimising ensemble streamflow predictions with bias correction and data assimilation techniques, *Hydrol. Earth Syst. Sci.*, 29, 1587-1614, <https://doi.org/10.5194/hess-29-1587-2025>, 2025.
- 785 Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, 456, 12-29, <https://doi.org/10.1016/j.jhydrol.2012.05.052>, 2012.
- 790 Tian, F. Q., Li, Y. L., Zhao, T. T. G., Hu, H. C., Pappenberger, F., Jiang, Y. Z., and Lu, H.: Evaluation of the ECMWF System 4 climate forecasts for streamflow forecasting in the Upper Hanjiang River Basin, *Hydrol. Res.*, 49, 1864-1879, <https://doi.org/10.2166/nh.2018.176>, 2018.
- Towler, E., Stover, D., Acharya, N., Abel, M. R., Currier, W. R., Bellier, J., Cifelli, R., Mahoney, K., Mossel, C., Scheuerer, M., Thorstensen, A., and Viterbo, F.: Implementing and Evaluating National Water Model Ensemble Streamflow Predictions Using Postprocessed Precipitation Forecasts, *J. Hydrometeorol.*, 26, 385-399, <https://doi.org/10.1175/jhm-d-24-0111.1>, 2025.
- 795 Vernon, B., Zhang, W., and Chikamoto, Y.: Improving seasonal precipitation forecasts in the Western United States through statistical downscaling, *Environ. Res. Lett.*, 20, 064008, <https://doi.org/10.1088/1748-9326/add02c>, 2025.
- 800 Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A. W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D.-J., Xiao, H., Zaripov, R., and Zhang, L.: The Subseasonal to Seasonal (S2S) Prediction Project Database, *Bull. Am. Meteorol. Soc.*, 98, 163-173, <https://doi.org/10.1175/BAMS-D-16-0017.1>, 2017.
- 805 Wang, J., Wang, X., Lei, X. H., Wang, H., Zhang, X. H., You, J. J., Tan, Q. F., and Liu, X. L.:



- 810 Teleconnection analysis of monthly streamflow using ensemble empirical mode decomposition, *J. Hydrol.*, 582, 124411, <https://doi.org/10.1016/j.jhydrol.2019.124411>, 2020.
- Wang, M., Wyatt, B. M., and Ochsner, T. E.: Accurate statistical seasonal streamflow forecasts developed by incorporating remote sensing soil moisture and terrestrial water storage anomaly information, *J. Hydrol.*, 626, 130154, <https://doi.org/10.1016/j.jhydrol.2023.130154>, 2023.
- 815 Wang, T., Shi, R., Yang, D., Yang, S., and Fang, B.: Future changes in annual runoff and hydroclimatic extremes in the upper Yangtze River Basin, *J. Hydrol.*, 615, 128738, <https://doi.org/10.1016/j.jhydrol.2022.128738>, 2022.
- Xu, P., Wang, D., Singh, V. P., Lu, H., Wang, Y., Wu, J., Wang, L., Liu, J., and Zhang, J.: Multivariate Hazard Assessment for Nonstationary Seasonal Flood Extremes Considering Climate Change, *J. Geophys. Res. Atmos.*, 125, e2020JD032780, <https://doi.org/10.1029/2020jd032780>, 2020.
- 820 Yang, D., Gao, B., Jiao, Y., Lei, H., Zhang, Y., Yang, H., and Cong, Z.: A distributed scheme developed for eco-hydrological modeling in the upper Heihe River, *Sci. China Earth Sci.*, 58, 36-45, <https://doi.org/10.1007/s11430-014-5029-7>, 2015.
- Yang, D., Li, C., Hu, H., Lei, Z., Yang, S., Kusuda, T., Koike, T., and Musiake, K.: Analysis of water resources variability in the Yellow River of China during the last half century using historical data, *Water Resour. Res.*, 40, W06502, <https://doi.org/10.1029/2003WR002763>, 2004.
- 825 Yin, G. H., Yoshikane, T., Kaneko, R., and Yoshimura, K.: Improving Global Subseasonal to Seasonal Precipitation Forecasts Using a Support Vector Machine-Based Method, *J. Geophys. Res. Atmos.*, 128, e2023JD038929, <https://doi.org/10.1029/2023JD038929>, 2023.
- 830 Zhang, L., Gao, S., and Yang, T.: Adapting subseasonal-to-seasonal (S2S) precipitation forecast at watersheds for hydrologic ensemble streamflow forecasting with a machine learning-based post-processing approach, *J. Hydrol.*, 631, 130643, <https://doi.org/10.1016/j.jhydrol.2024.130643>, 2024.
- Zhang, L. J., Yang, T. T., Gao, S., Hong, Y., Zhang, Q., Wen, X., and Cheng, C. T.: Improving Subseasonal-to-Seasonal forecasts in predicting the occurrence of extreme precipitation events over the contiguous US using machine learning models, *Atmos. Res.*, 281, 106502, <https://doi.org/10.1016/j.atmosres.2022.106502>, 2023.
- 835 Zhang, Y., Wu, L., Scheuerer, M., Schaake, J., and Kongoli, C.: Comparison of Probabilistic Quantitative Precipitation Forecasts from Two Postprocessing Mechanisms, *J. Hydrometeorol.*, 18, 2873-2891, <https://doi.org/10.1175/JHM-D-16-0293.1>, 2017.
- Zhong, W., Guo, J., Chen, L., Zhou, J., Zhang, J., and Wang, D.: Future hydropower generation prediction of large-scale reservoirs in the upper Yangtze River basin under climate change, *J. Hydrol.*, 588, 125013, <https://doi.org/10.1016/j.jhydrol.2020.125013>, 2020.
- 845 Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. R., and Myneni, R. B.: Global data sets of vegetation leaf area index (LAI) 3g and fraction of photosynthetically active radiation (FPAR) 3g derived from global inventory modeling and mapping studies (GIMMS) normalized difference vegetation index (NDVI3g) for the period 1981 to 2011, *Remote Sens.*, 5, 927-948, <https://doi.org/10.3390/rs5020927>, 2013.