https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

# Deep learning based sub-seasonal precipitation and streamflow forecasting over the source region of the Yangtze River

Ningpeng Dong[1], Haoran Hao[2,1], Mingxiang Yang[1], Jianhui Wei[3], Shiqin Xu[4], Harald Kunstmann[3,5,6]

[1]State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, China
[2]State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University, Tianjin 300072, China
[3]Institute of Meteorology and Climate Research (IMKIFU), Karlsruhe Institute of Technology, Campus Alpine, Garmisch-Partenkirchen, Germany
[4]Hydrology, Agriculture and Land Observation (HALO) Laboratory, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[5]Institute of Geography, University of Augsburg, Augsburg, Germany
[6]Centre for Climate Resilience, University of Augsburg, Augsburg, Germany

*Correspondence to*: H. Hao (1020205041@tju.edu.cn) and M. Yang (yangmx@iwhr.com)

**Abstract.** Hydrometeorological forecasting is crucial for managing water resources and mitigating the impacts of extreme hydrologic events. At sub-seasonal scales, readily available hydrometeorological forecast products often exhibit large uncertainties and insufficient accuracies to support decision making. We propose a deep learning based modelling framework for sub-seasonal joint precipitation and streamflow forecasts for a lead time of up to 30 days. This is achieved by coupling (1) a convolutional neural network (CNN) architecture with ResNet blocks for statistically downscaling of the ECMWF raw precipitation forecasts to (2) a hybrid hydrologic model integrating the conceptual Xin'anjiang model (XAJ) and the long-short term memory network (LSTM) for streamflow forecasting. The CNN incorporates a specialized loss function that combines the continuous form of threat score and mean absolute error. Applying the modeling framework to the source region of the Yangtze River Basin, results indicate that the CNN-based downscaling model exhibits ~13% and ~10% less RMSE than the raw ECMWF forecasts and the quantile mapping (QM) forecasts, respectively, averaged over the 30-day lead time. Similarly, the CNN achieves a ~2% and ~5% lower RMSE than raw forecasts and QM for precipitation events above the 90$^{th}$ percentile of historic daily precipitation. Using these precipitation forecasts as meteorological drivers for the hybrid XAJ-LSTM hydrologic model, we found that forecasted streamflow and flood peaks driven by CNN-based precipitation forecasts have 18%-32% lower relative errors and 13%-22% lower RMSE compared to those driven by raw forecasts. However, the standalone XAJ model shows marginal improvements, or in some cases, no improvement at all, with the same enhanced precipitation forecasts. This highlights the importance of understanding the effectiveness of the hydrologic model as part of the sub-seasonal hydrometeorological modeling chain. Our study is expected to provide implications for leveraging advanced AI techniques to enhance sub-seasonal hydrometeorological forecasting accuracy and operational efficiency for effective water resources management and disaster preparedness.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

## 1 Introduction

In past decades, the frequency and intensity of extreme precipitation events are increasing in many areas as global warming continues, thereby amplifying the potential for hazards of extreme weather and hydrologic events (Wei et al., 2018; Yuan et al., 2018; Wang et al., 2019; S. Zhu et al., 2020). Hydrological forecasting has become critically important for managing water resources and mitigating the impacts of these extreme weather and hydrologic events (Robertson and Wang, 2013; Liu et al., 2020; Jiang et al., 2022). Traditional hydrological forecasts, which do not integrate sub-seasonal meteorological forecasts, often provide insufficient lead times for decision-making on flood control, agricultural planning, and ecological preservation efforts (de Andrade et al., 2021; Bierkens, 2015). Integrating both meteorological and hydrological forecasts at sub-seasonal scales is therefore essential to extend lead times, thereby improving water resources management and disaster preparedness over a longer term (Yuan et al., 2016; Cloke and Pappenberger, 2019; Liang et al., 2018; Vigaud et al., 2019).

Advancements in numerical weather prediction (NWP), such as the ECMWF Integrated Forecasting System (IFS) and the NCEP Global Forecast System (GFS) model have greatly improved the accuracy of sub-seasonal weather forecasting (Yuan et al., 2011; Bauer et al., 2015; Brotzge et al., 2023). However, these global models often suffer from relatively coarse resolutions and generalized parameterizations that may not be suitable for regional-scale and local-scale forecasts (Dehshiri et al., 2023; Singhal et al., 2023). Dynamic downscaling, such as that performed by the Weather Research and Forecasting (WRF) model, translates larger-scale atmospheric trends captured by GCMs into fine-scale regional details that reflect local geographic and climatic factors (Merino et al., 2022; Nooni et al., 2022). For instance, recent studies by Gao et al. (2022) and Srivastava et al. (2023) demonstrate the effectiveness of WRF in enhancing the accuracy of precipitation forecasts and capturing the dynamics of severe weather events. Despite these advantages, dynamic downscaling often requires extensive computational resources especially for sub-seasonal scales, and can be sensitive to the quality of input data. Furthermore, the process is constrained by the physical parameterizations that may not always accurately represent localized meteorological conditions, a concern that is increasingly critical under changing climatic conditions (Di Luca et al., 2015; Shi, 2020).

Statistical downscaling techniques, which have been used to relate the larger-scale meteorological patterns to local-scale weather, offer a different approach (Tabari et al., 2021; Zhang et al., 2022; Michalek et al., 2024). Traditional statistical downscaling methods such as quantile mapping have proven effective in reducing the systematic bias of precipitation forecasts with relatively simple inputs (Wilby, R. L., et al., 2004; Vrac, M., & Friederichs, P., 2015). More recently, deep learning models such as convolutional neural networks (CNN) have been reported able to more effectively reduce the total bias of meteorological forecasts due to their ability to learn multi-dimensional representations of data features (Vandal et al., 2019; Sachindra et al., 2018; Ning et al., 2024). Despite general improvements of forecasts, they tend to smooth the extreme precipitation at sub-seasonal scales (Baño-Medina et al., 2021; Kim et al., 2022), likely due to insufficient heavy precipitation samples (Chen et al., 2022). Many studies have since introduced more recent variants of CNNs including U-Net (Ni et al., 2023) and SmaAt-UNet (Li et al., 2024), or coupled standard CNNs with different structures, such as Auto-Encoder (Ling et

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

al., 2022), Transformer (Ling et al., 2024), and in particular ResNet, which shows the potential of mitigating the vanishing gradient issue by introducing the residual paths (Nie et al., 2024). Others have attempted to introduce specialized loss functions to balance heavy and light rains, such as the exponentially weighted mean squared error (Ebert-Uphoff et al., 2020) and Dice loss (You et al., 2022). However, these new developments have not been sufficiently examined for sub-seasonal forecasts.

It is also noteworthy that recent weather prediction models achieved entirely through deep learning has demonstrated the potential to achieve forecast skills comparable to state-of-the-art numerical weather prediction systems. Notable examples include Pangu (Bi et al., 2023) and GraphCast (Lam et al., 2023). These models represent significant advancements in weather forecasting, yet some argue these models are more of a post-processing model rather than a realistic simulator of the atmosphere due to the lack of fidelity and physical consistency of physics-based models (Bonavita et al., 2024).

In addition to meteorological forecasts, sub-seasonal streamflow forecasts are crucial because streamflow at these timescales is directly related to the onset and progression of flooding and drought events. To translate meteorological predictions to streamflow forecasts, both physics-based and data-driven hydrologic forecasting models are widely used. Physics-based models, such as the lumped Xin'anjiang model, HBV model, and the distributed SWAT and VIC models, make predictions by interpreting detailed physical processes (Gassman et al., 2014; Arnold et al., 2015; Dong et al., 2022). Data-driven models were also developed to perform rainfall-runoff modelling and forecasts by learning from big data (Kisi, 2007; Adnan et al., 2019), and have been reported to outperform the well-calibrated physics-based models (Kratzert et al., 2019). It is noteworthy that both models are embedded with uncertainties. Physics-based models may produce inaccurate simulations due to simplified representations of hydrologic processes, and data-driven models may perform less effectively in extrapolation beyond the range of the training data, particularly during extreme events (Addor et al., 2020). By integrating the strengths of both approaches, recent studies have attempted to establish a hybrid model with a higher predictive performance than the physical model alone (Liu et al., 2022). For example, Humphrey et al. (2016) achieved improved streamflow forecast skills by combining Bayesian artificial neural networks with traditional models of GR4J. However, the role of such models as part of the hydrometeorological modelling chain in producing reliable streamflow forecasts has not been well examined at sub-seasonal scales. For example, Crochemore et al. (2016) and Valdez et al. (2022) have suggested that the relationship between the accuracy of precipitation forecasts and the corresponding streamflow forecasts is not necessarily straightforward or linear.

The above considerations are particularly relevant for the source region (SR) of the Yangtze River Basin, which is historically susceptible to extensive flooding and droughts that affect thousands of kilometers downstream (Sun et al., 2016). Aiming at enhancing sub-seasonal hydrometeorological forecasts in this area, we start by addressing the following questions:

(1) How effectively can CNN architectures with recent extensions improve the sub-seasonal precipitation forecasts compared to traditional statistically downscaling models?

(2) How effectively can AI-assisted hydrologic models convert more accurate sub-seasonal precipitation forecasts into more accurate streamflow forecasts compared to traditional conceptual hydrologic models?

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

Specifically, this study investigates sub-seasonal precipitation and streamflow forecast skills for up to 30 days ahead with deep learning models, which integrates a CNN architecture with ResNet blocks and specialized loss functions for post-processing state-of-the-art ECMWF forecasts with a hybrid hydrologic model integrating the Xin'anjiang model (XAJ) and the long-short term memory network (LSTM) for streamflow forecasting. Our approaches and findings are expected to provide implications for operational hydrometeorological forecasts in the SR and also similar basins worldwide.

## 2 Study Area and Data

### 2.1 Study Area

The source region (SR) of the Yangtze River basin is located on the eastern edge of the Tibetan Plateau, between 90°E to 101°E and 26°N to 36°N, as shown in Figure 1. The region of the basin serves as a crucial transitional area from highland mountains to plains in southern China, with the surface elevation decreasing from over 6000 m in the north to just over 2500 m in the south. The climate of the region is subject to both the plateau monsoon climate and the subtropical monsoon climate, with the annual precipitation ranging from 280 mm to 760 mm. The SR has a significant impact on the development and utilization of water resources in the Yangtze River Basin and the Southwest China. The controlling hydrologic station of the SR is Shigu station (Fig. 1), which has a mean annual streamflow of around 1300 $m^3/s$, accounting for around 5% of the total water resources of the Yangtze River Basin.

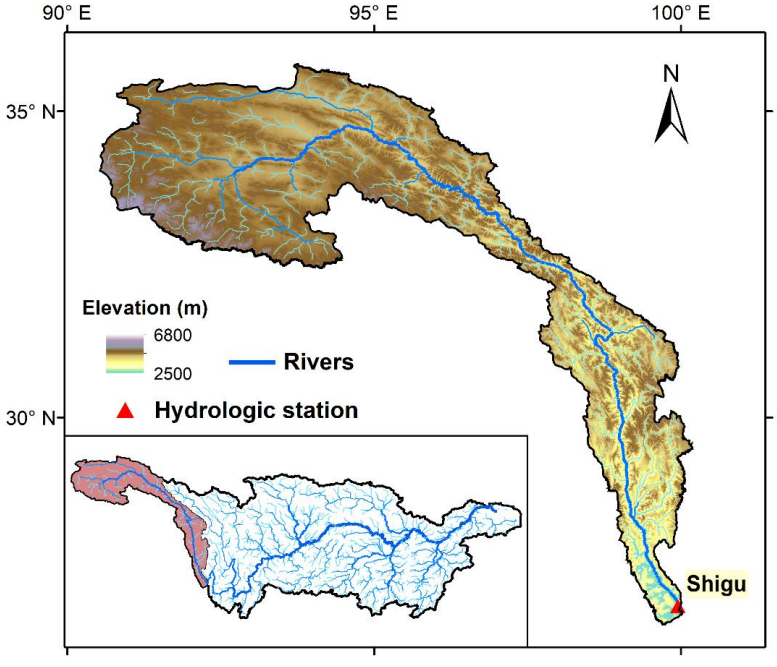



**Figure 1. The source region of the Yangtze River Basin and its location in the Yangtze River Basin.**

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences Discussions

### 2.2 Data sources

*Observed precipitation and temperature.* To train the forecast models and evaluate the accuracy of forecasts, this study employs the 0.25° daily precipitation and temperature grid dataset (CN05.1), released by the National Meteorological Information Center, as the observed data. This dataset is produced by interpolating precipitation and temperature data from over 2,000 meteorological stations across the country and covers the period from 1961 to 2022.

*ECMWF sub-seasonal forecast data.* The European Centre for Medium-Range Weather Forecasts (ECMWF) offers a sub-seasonal forecast service designed to bridge the gap between short-range weather predictions and long-term climate outlooks. This service focuses on predicting atmospheric and oceanic conditions over the next 2 to 6 weeks, providing valuable information for a variety of applications such as water resources management, and disaster preparedness. In this study, we collect the ECMWF Sub-seasonal to Seasonal (S2S) daily forecast data for a lead time of 30 days initialized on 35 dates between May and August per year during 2002-2019. The forecasted variables used in this study include precipitation, convective precipitation at the land surface; and temperature, wind components, geopotential heights, and specific humidity at 200/500/850hPa pressure levels.

*Observed streamflow.* To calibrate the hybrid hydrologic model and evaluate the hydrologic forecasts, the daily streamflow data of Shigu station is collected for 1981-2019.

## 3 Methods

### 3.1 Overview

The presented sub-seasonal hydrometeorological forecasting framework aims to improve the daily precipitation forecasts at the 0.25° grid resolution and the corresponding streamflow forecasts at the Shigu hydrologic station for a lead time up to 30 days. First, we employ the ECMWF S2S gridded sub-seasonal precipitation forecasts for the next 30 days as the raw forecasts. A CNN architecture with ResNet blocks and a specialized loss function is established to statistically downscale and bias correct the ECMWF raw precipitation forecasts to the 0.25° grid resolution. The quantile mapping (QM) also serves as a benchmark to compare with the post-processed precipitation of the CNN.

The post-processed precipitation forecasts, along with ECMWF sub-seasonal daily temperature forecasts corrected by the delta method, are then employed to drive a hybrid hydrologic model to produce the daily streamflow forecasts for lead times of 1-30 days. The hybrid model integrates the conceptual XAJ flood forecast model with snow processes and the LSTM. The streamflow and flood peak forecasts of XAJ-LSTM and standalone XAJ driven by EC-CNN forecasts are then quantitatively evaluated against those driven by raw and QM-based forecasts using a series of metrics. A detailed workflow of this study is presented in Fig. 2.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024

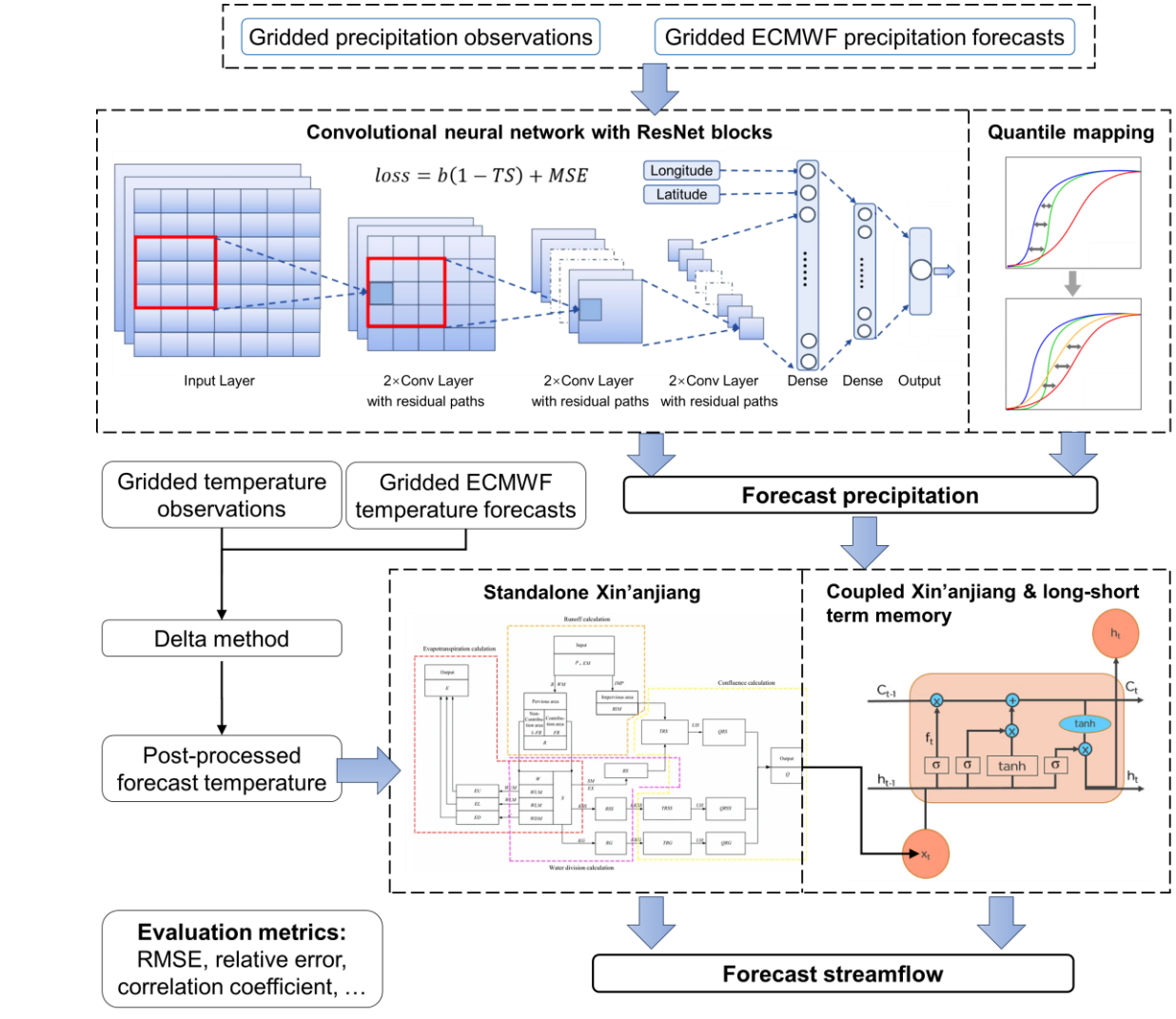




**Figure 2. The workflow of this study.**

### 3.2 Statistically downscaling of precipitation forecasts

#### 3.2.1 Convolutional neural network

A CNN architecture with ResNet blocks is established to learn the spatially dependent relationship between precipitation of local grid cells and predictors from surrounding regions. The model has 19 predictors including the surface elevation, convective precipitation, total precipitation at the surface level, and U/V wind components, specific humidity, temperature, and geopotential height at 200/500/850 hPa pressure levels from ECMWF forecasts. The predictand is the daily precipitation. Figure 2 presents the model structure, which primarily consists of convolutional layers, embedding layers, fully connected layers, and residual paths. The structure is detailed as follows.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

(1) Inputs to the network are predictors from a 7×7 grid patch centered on the target grid cell. This patch includes a total of 19 meteorological variables, resulting in input arrays with sizes of 19×7×7. Here, the ensemble mean forecast is used as predictors, as they typically encompass the majority of information of the ensemble (Li et al., 2022).

(2) The model includes three groups of convolutional layers with residual paths, each group containing two convolutional layers with 3×3 kernels and feature maps of sizes 64, 32, and 16, respectively. Such group, referred to as a ResNet block, mitigates the vanishing gradient problem and improves computational efficiency for a moderately deep learning model as in our study by allowing the gradient to bypass certain layers. For each convolutional layer, the convolution procedure involves moving the kernels along the input spatial fields, with the dot product calculated between the inputs and the kernels to capture spatial features. The $l$ th feature map of the current convolutional layer $X_n^l$ is computed from the previous layer $X_{n-1}$ with $K$ feature maps through the convolutional operation as follows,

$$X_n^l = \mathrm{ELU}\left(b_n^l + \sum_{k=1}^{K} W_n^{k,l} * X_{n-1}^k\right) \tag{1}$$

where $W_n^{k,l}$ are the convolutional kernels; $b_n^l$ is the bias for the $l$ th feature map; the symbol * denotes two-dimensional convolution. Here, we employ the exponential linear units (ELU) as the activation function, i.e.,

$$f(x) = \begin{cases} x & x > 0 \\ a(e^x - 1) & x \leqslant 0 \end{cases} \tag{2}$$

where $a$ is a hyperparameter to be estimated, $x$ is input to ELU function.

(3) To address spatial heterogeneity, embedding layers are introduced to convert the coordinate indices into latitude and longitude decimals (Rasp and Lerch, 2018). The outputs from these embedding layers are merged with the flattened outputs from the ResNet blocks, and this combined data is then fed into two fully connected layers before the output layer.

The Adam optimizer is used to train the CNN model with an early stopping technique to avoid overfitting. Specifically, to account for the small number of extreme precipitation samples, a specialized loss function that combines the Threat Score (TS) and Mean Squared Error (MSE) is used in this study, i.e.,

$$\begin{aligned} loss &= b(1 - TS) + MSE \\ MSE &= \frac{1}{N}\sum_{i=1}^{N} (F_i - O_i)^2 \\ TS &= \frac{H}{H + F + M} \end{aligned} \tag{3}$$

where $b$ represents the weight of extreme precipitation in model training. $F_i$ is the $i$ th forecast data, $O_i$ is the $i$ th observation data, $N$ is the number of data. Meanwhile, $H$, $F$, and $M$ represent the hits, false alarms and misses, respectively. Note that the

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

categorical indices used for calculating TS are discrete, which is not well-suited for training deep learning models. Thus, the differentiable formulations proposed by Larraondo et al (2020) and Lyu et al. (2023) are utilized in this study, i.e.,

$$H = (O > \alpha) \odot sigmoid(F - \alpha) \tag{4}$$

$$F = (O < \alpha) \odot sigmoid(F - \alpha) \tag{5}$$

$$M = (O > \alpha) \odot sigmoid(-F - \alpha) \tag{6}$$

$$sigmoid(x) = \frac{1}{1 + e^{-ax}} \tag{7}$$

in which $\odot$ means element-wise multiplication, the $(O > \alpha)$ and $(O < \alpha)$ are logical operations, which are 1 and 0 when the statement are true and false, respectively. $\alpha$ is the precipitation threshold that corresponds to the 90$^{th}$ percentile of observed precipitation of each grid cell for 2002-2019. The logical operations towards the $F$ (forecast) term are substituted with a sigmoid function, which represents a smooth transition between the Boolean values at the threshold point. In above expressions, $a$ and $b$ are hyperparameters that are determined following Lyu et al. (2023), see Table S1 in the supporting information for detailed values.

Another approach for improving extreme precipitation forecasts is to manually increase the number of heavy precipitation events within the training datasets. This approach is eventually not adopted in our study because it is found to degrade the sub-seasonal forecast accuracy of light precipitation events while not improving the accuracy of heavy precipitation events over the SR region (results not shown). A possible reason is that by doing so artificial disruptions are brought into the distribution of precipitation samples, which could possibly impair the generalization capability of models (You et al., 2023).

### 3.2.2 Quantile mapping

Quantile mapping is a widely used postprocessing technique and is able to effectively enhancing quantitative precipitation forecasts at the sub-seasonal timescale (Baker et al., 2019). Therefore, the current study adopts QM as a benchmark to evaluate the proposed CNN-based model in terms of the sub-seasonal precipitation forecasts. QM estimates the values of the quantile-quantile relation of observations and forecasts for regularly spaced quantiles. Specifically, the period from 2002-2015 is used to estimate the cumulative distribution functions (CDFs) of the forecast precipitation $F$ and observed precipitation $O$ with 100 discretized quantiles. The CDFs are then used to match the corresponding quantiles in the test period of 2016-2019:

$$\tilde{p}_{QM} = O^{-1}[F(p_{EC})] \tag{8}$$

where $\tilde{p}_{QM}$ and $p_{EC}$ are the QM-based precipitation forecasts and the ECMWF raw precipitation forecasts, respectively.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

205 **3.3 Bias correction of temperature forecasts**

In this study, we apply the widely used delta method to correct the ECMWF temperature forecasts for lead times of 1-30 days. We calculate the difference between observed temperature and the forecast temperature (i.e., the delta) for each month and each lead time during the calibration period of 2002-2015, and then apply the monthly delta for each lead time to the forecast temperature during the validation period of 2016-2019. Given that temperature is not the main focus of the paper, plus that
210 temperature forecasts generally have less biases and much less hydrologic impacts than precipitation forecasts (as discussed in Section 5.3), relevant evaluation results are provided in the Text S1 and Fig. S1 in the supporting information.

**3.4 Hybrid hydrologic model of XAJ-LSTM**

**3.4.1 Xin'anjiang model with snow processes**

The XAJ model is a conceptual hydrological model (Zhao, 1992), which has been widely used to generate flood forecasts for
215 humid and semi-humid regions of China. The lumped XAJ model consists of the evapotranspiration module, the runoff generation module, the runoff partition module, and the runoff routing module (Hu et al. 2005). In this study, a modified version of XAJ model with snow accumulation and melting mechanisms is employed to simulate and forecast the daily streamflow of the SR at the sub-seasonal scale, which shows satisfactory accuracies for basins with considerable snow melting runoff in our previous study (Tan et al., 2019).

220 **3.4.2 Long-short term memory network**

In this study, the LSTM model is employed as part of the modelling chain to simulate and predict the sub-seasonal streamflow. LSTMs have memory cells that are analogous to the states of a traditional dynamical system model (Hochreiter and Schmidhuber, 1997), which make them practicable for simulating natural hydrologic systems. Compared with other types of RNNs, LSTMs perform better in coping with exploding and vanishing gradients, which enables them to learn the long-term
225 dependencies between input and output arrays (Zhang et al., 2022). This is particularly desirable for modelling hydrological processes that have relatively long-time dependencies as compared with input-driven processes such as direct surface runoff. For example, Kratzert et al. (2018; 2019) applied LSTMs to hydrologic modelling and show that the internal memory states of the network are highly correlated with observed snow and soil moisture states, even if no snow or soil moisture data were input to the models during training. This model feature allows accurate sub-seasonal hydrologic simulations in the SR where
230 there is snow accumulation around the winter and spring.

**3.4.3 Model integration**

The XAJ model employs daily precipitation and temperature to simulate the daily streamflow at the Shigu station, and the model parameters are calibrated for the period of 1981-2015 and validated for the period of 2016-2019. The particle swarm

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

optimization (PSO) approach is employed to optimize the parameters of the XAJ model, with the Nash-Sutcliffe efficiency (NSE) as the objective function.

The LSTM model takes as input features a time sequence of daily precipitation $\mathbf{p} = \mathbf{p}[1], \mathbf{p}[2], \ldots, \mathbf{p}[N]$, daily temperature $\mathbf{t} = \mathbf{t}[1], \mathbf{t}[2], \ldots, \mathbf{t}[N]$, and XAJ-simulated daily streamflow $\mathbf{q} = \mathbf{q}[1], \mathbf{q}[2], \ldots, \mathbf{q}[N]$ over $N$ time steps. Each element of $\mathbf{p}$, $\mathbf{t}$ and $\mathbf{q}$, namely $\mathbf{p}[n]$, $\mathbf{t}[n]$, and $\mathbf{q}[n]$, is a vector of input features for the past $n\_seq$ days and corresponds to the predictand $\mathbf{o}[n]$, the daily streamflow of Shigu station for timestep $n$. Here $n\_seq$ is an optimized hyperparameter representing the size of input features. In snow-affected regions such as the SR, it is typically set to a larger value to account for the snow accumulation and melting processes, which can span hundreds of days. In our study, LSTM can also be considered a post-processing model of the XAJ model, similar to the CNN model as a post-processing model of the ECMWF forecast model. The LSTM model is trained by the Adam optimizer, and is 5-fold cross-validated using the Randomized Search approach for the period of 1981-2015 (see the Table S2 in the supporting information for details of model hyperparameters). The trained model is then tested for the period of 2016-2019.

### 3.5 Evaluation Metrics

The statistically downscaled precipitation forecasts are evaluated using the root mean squared error (RMSE). Specifically, we classify the daily precipitation less than and greater than the 90$^{th}$ percentile of all historic daily precipitation during 2002-2019 as light rain and heavy rain. The RMSEs are calculated for light and heavy rain events to evaluate the forecasts in predicting common and extreme events, which are both critical for sub-seasonal forecasts that need to inform agricultural planning and flood risk management.

The streamflow forecasts are evaluated using the RMSE, relative error of flow (RE) and Pearson's correlation coefficient (R) for general trends, and the relative error of the maximum daily flow (REF) for extreme events.

## 4 Results

### 4.1 Calibration and validation of the hybrid hydrologic model

The conceptual XAJ model is calibrated for the period of 1981-2015 and validated for the period of 2016-2019. Results in Fig. 3 indicate a satisfactory performance for the standalone XAJ model, as the daily NSE values of simulated streamflow are 0.88 and 0.83 during the calibration and validation period.

The hybrid hydrologic model is calibrated for the period of 1981-2015 (corresponding to the calibration period of XAJ model and the training and cross-validation period of the LSTM model) and validated for the period of 2016-2019 (corresponding to the validation period of XAJ model and the testing period of the LSTM model). Fig. 3 depicts the simulated daily streamflow at Shigu during the validation period, as compared with observations. The results indicate that the daily Nash-Sutcliffe

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

Efficiency (NSE) stands at 0.96 and 0.93 during the calibration and validation period, and the relative error stands at 1.7% and 2.8%, respectively. The mean absolute error and relative error of simulated maximum daily flow are 329 $m^3/s$ and 7.5%, respectively, indicating the model also has a satisfactory ability to simulate large flood events of the basin. By comparing the simulation accuracy of the standalone XAJ model with that of the hybrid model, it is found that the hybrid model can take advantage of the XAJ outputs and improve the streamflow simulations at Shigu station.

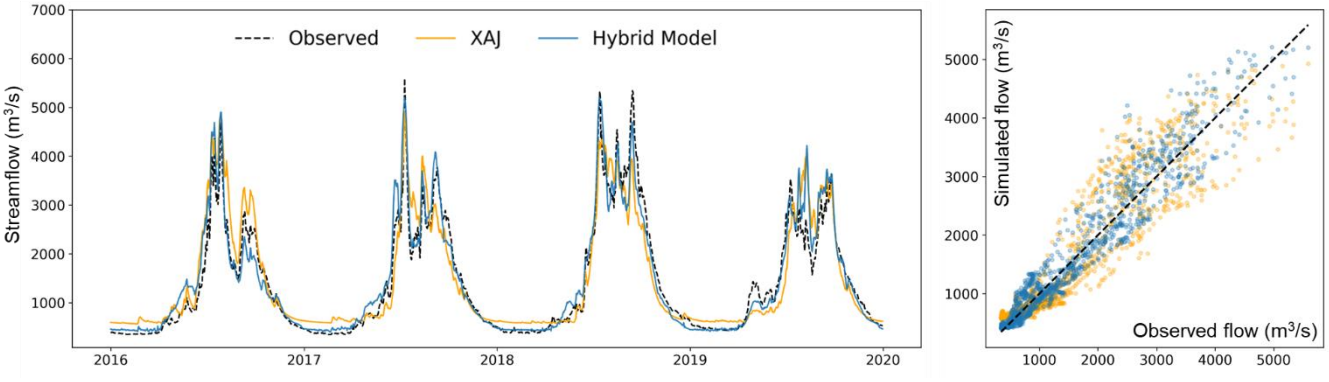



**Figure 3. Streamflow validation of the XAJ model and the hybrid hydrologic model.**

## 4.2 Evaluation of sub-seasonal precipitation forecasts

In this study, the root mean square error (RMSE) is utilized to evaluate the skills of the statistically downscaled precipitation forecasts. The RMSE of EC raw forecasts, EC-QM and EC-CNN averaged over SR at lead times of 1-30 days for the period from 2016 to 2019 are provided in Figure 4, with the error bars representing the $25^{th}$-$75^{th}$ percentile interval.

Generally, the precipitation forecast skills of the raw ECMWF model decrease gradually with the increasing lead times and tend to be constant at a relatively low level for lead times of 15-30 days, which is also observed by Lyu et al. (2023) across Southeast China. The RMSE averaged over all lead times for the raw precipitation forecasts is 2.76 mm/day. The QM effectively reduces RMSE at all lead times by an average of 0.1 mm/day (~5%), indicating the effectiveness of QM in improving precipitation at sub-seasonal scales. By contrast, the CNN model exhibits ~10% less RMSE compared to QM for all lead times, which reduces the RMSE of raw forecasts by 0.35 mm/day (~13%). In particular, the RMSE of EC-CNN forecasts are reduced more significantly for longer lead times, as the RMSE is reduced by 0.20 mm/day, 0.45 mm/day, and 0.43 mm/day for lead times of 1-10 days, 11-20 days, and 21-30 days, while the RMSE of EC-QM forecasts sees a relatively steady reduction over all lead times. In addition, the EC-CNN forecasts exhibit narrower $25^{th}$-$75^{th}$ percentile intervals across different initial forecast dates than raw EC forecasts and EC-QM forecasts, suggesting that the CNN model tends to produce precipitation forecasts with more stable skill metrics across different initial dates. These results preliminarily demonstrate the superiority of proposed CNN method to the raw ECMWF model and the QM.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024

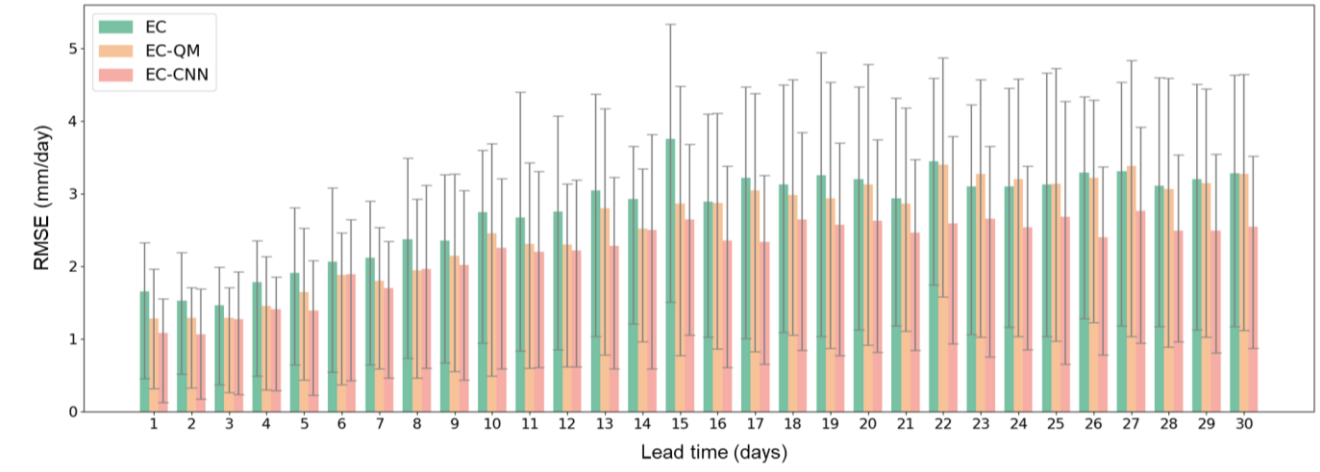
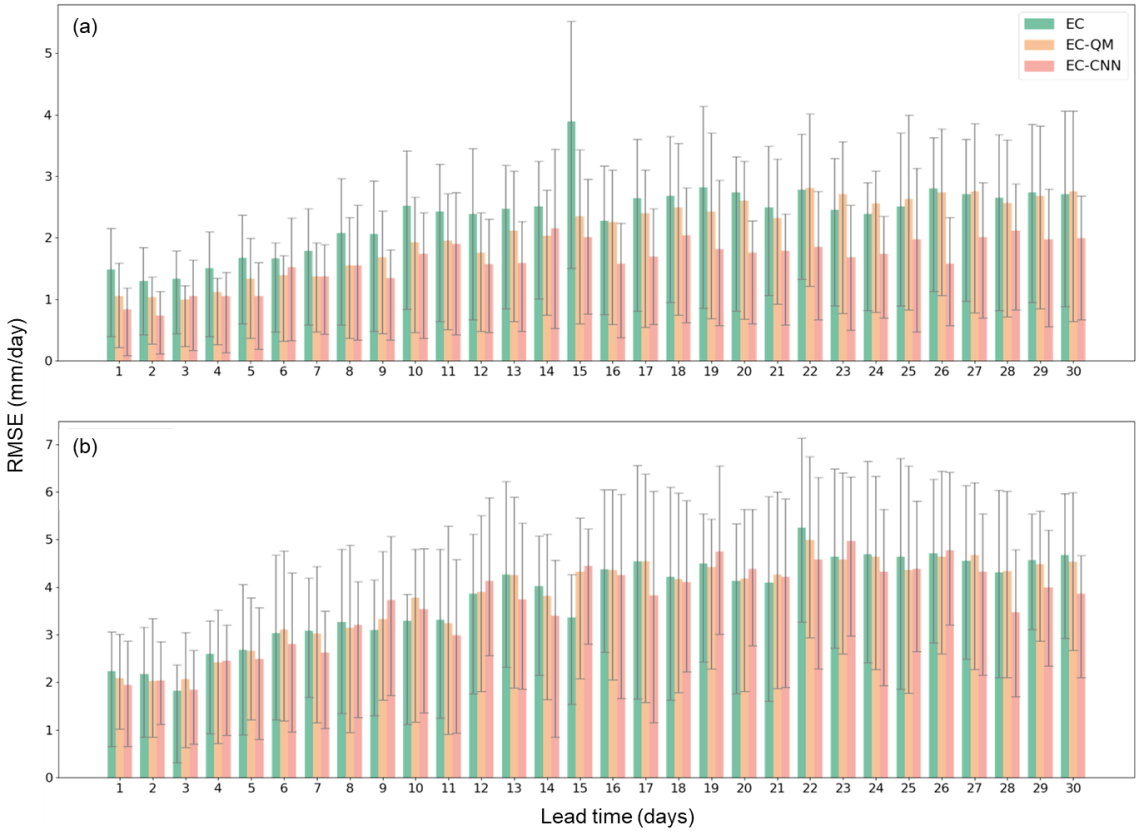




**Figure 4. Root mean square error (RMSE) of ECMWF raw precipitation forecasts (EC), QM precipitation forecasts (EC-QM), and CNN precipitation forecasts (EC-CNN) for lead times of 1-30 days during the test period. Error bar represents the 25$^{th}$-75$^{th}$ percentile interval.**

290 To investigate the performance of EC-CNN forecasts in extreme precipitations, we classify the daily precipitation less than and greater than the 90$^{th}$ percentile of all historic daily precipitation during 2002-2019 as light rain and heavy rain, respectively. Figure 5 displays the variations of RMSE for light rain and heavy rain averaged over SR at lead times of 1-30 days, with the error bars representing the 25$^{th}$-75$^{th}$ percentile interval.

Generally, the RMSE of EC forecasts increases with the increasing lead times for both light and heavy rains. For light rain,
295 QM improves the raw EC forecast by reducing the RMSE from 2.4 mm/day to 2.2 mm/day, whereas the CNN model can further improve the forecast by reducing to RMSE to 1.8 mm/day. In terms of the heavy rain, QM shows no improvements as compared to the raw forecasts, with the RMSE increasing from 3.8 mm/day to 3.9 mm/day, suggesting that QM has a limited ability to improve the forecast skills for extreme precipitation events in the SR. The CNN generally shows slight improvements (~2%) in RMSE as compared to raw forecasts, and the RMSE of EC-CNN forecasts is smaller than that of EC-QM for most
300 of the lead times, suggesting CNN exhibits advantages over QM for extreme events. This is particularly the case for lead times of 21-30 days, where the RMSE of EC-CNN forecasts is 0.16 mm/day lower than that of EC-QM and is 0.12 mm/day lower than that of the raw forecasts. Overall, these results imply that, for extreme precipitation events, the EC-CNN forecast has an advantage over the EC-QM forecast, and also shows a slightly better accuracy than the raw EC forecasts.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


305 **Figure 5. RMSE of (a) light rain and (b) heavy rain for ECMWF raw precipitation forecasts (EC), QM precipitation forecasts (EC-QM), and CNN precipitation forecasts (EC-CNN) at lead times of 1-10 days, 11-20 days, and 21-30 days. Error bar represents the 25$^{th}$-75$^{th}$ percentile interval.**

To investigate the spatial characteristics of precipitation forecasts, Figure 6 describes the precipitation distributions of EC raw forecasts, EC-CNN forecasts and observed precipitations averaged over all lead times and all initial forecast dates. In terms of
310 the raw ECMWF forecasts, overestimated precipitation can be observed for the southern part of the basin where the mean bias can go up to ~3 mm/day. By contrast, the CNN improves the precipitation forecast over the majority of SR, with most of the basin experiencing a mean bias of less than 1 mm/day.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

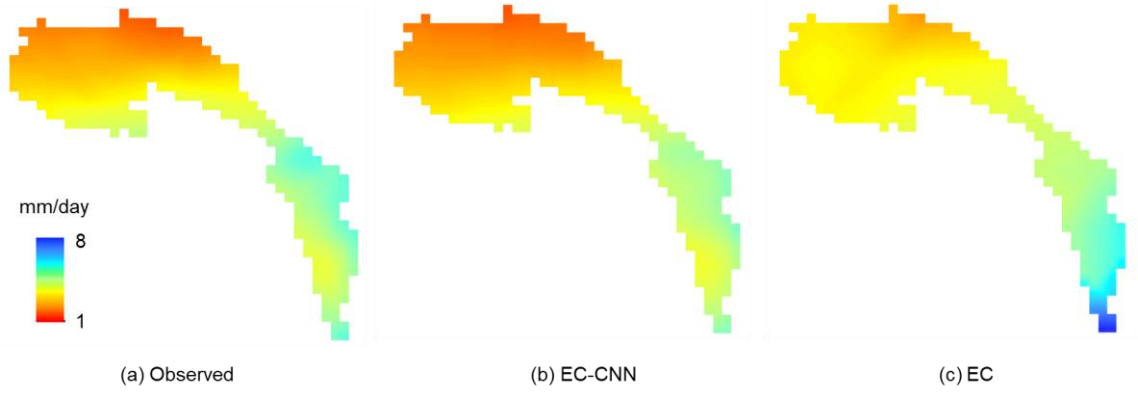



**Figure 6. The spatial distribution of (a) observed precipitation, (b) EC-CNN forecasts, and (c) EC raw forecasts averaged over all lead times and all initial forecast dates during the test period.**

### 4.3 Evaluation of sub-seasonal streamflow forecasts

Figure 7 shows the relative error of daily streamflow (RE), relative error of maximum daily streamflow (REF), and correlation coefficient (R) of the XAJ-LSTM and XAJ streamflow forecasts driven by EC raw precipitation forecasts, QM forecasts, and CNN forecasts, respectively.

Results indicate that, for the XAJ-LSTM hybrid model, the accuracy metrics of streamflow forecasts decrease with the increase in the lead times for all precipitation forecasts. For example, the relative error of streamflow forecasts driven by EC raw forecasts increases from around 12.5% for 1-10 days to over 25% and 29% for 11-20 days and 21-30 days, respectively. The EC-QM (EC-CNN) forecasts reduce the relative error of the raw forecasts by 12% (20%), 16% (24%) and 9% (21%), respectively, and reduces the relative error of maximum daily flow by 27% (29%), 16% (32%) and 11% (18%), respectively. For the standalone XAJ model, the relative error of streamflow forecasts driven by EC raw forecasts increases from around 16% for 1-10 days to over 26.5% and 28% for 11-20 days and 21-30 days, respectively. The EC-QM (EC-CNN) forecasts reduce the relative error of the raw forecasts by 17% (17%), 12% (17%) and 1% (4%), respectively, and reduces the relative error of maximum daily flow by 9% (4%), 11% (14%) and 2% (11%), respectively.

Comparing the XAJ-LSTM model and the XAJ model shows that improvements in CNN-based precipitation forecast lead to significantly greater enhancements in streamflow forecasts for the XAJ-LSTM model than for the XAJ model. Notably, the EC-CNN driven XAJ-LSTM streamflow forecast sees 13%-22% (5%-11%) less RMSE than that driven by raw EC forecasts (EC-QM forecasts) over different lead time periods. On the other hand, the EC-CNN driven XAJ streamflow forecasts generally show no improvements (and could be even worse) with respect to the RMSE compared to that driven by EC and EC-QM, despite that the EC-CNN sees much less error than EC and EC-QM (Figs. 4 & 5). This result suggests that improving sub-seasonal precipitation forecasts may not necessarily translate to a streamflow improvement, because it is related to not only the skill of precipitation forecasts but are also to a large extent the hydrologic model.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024

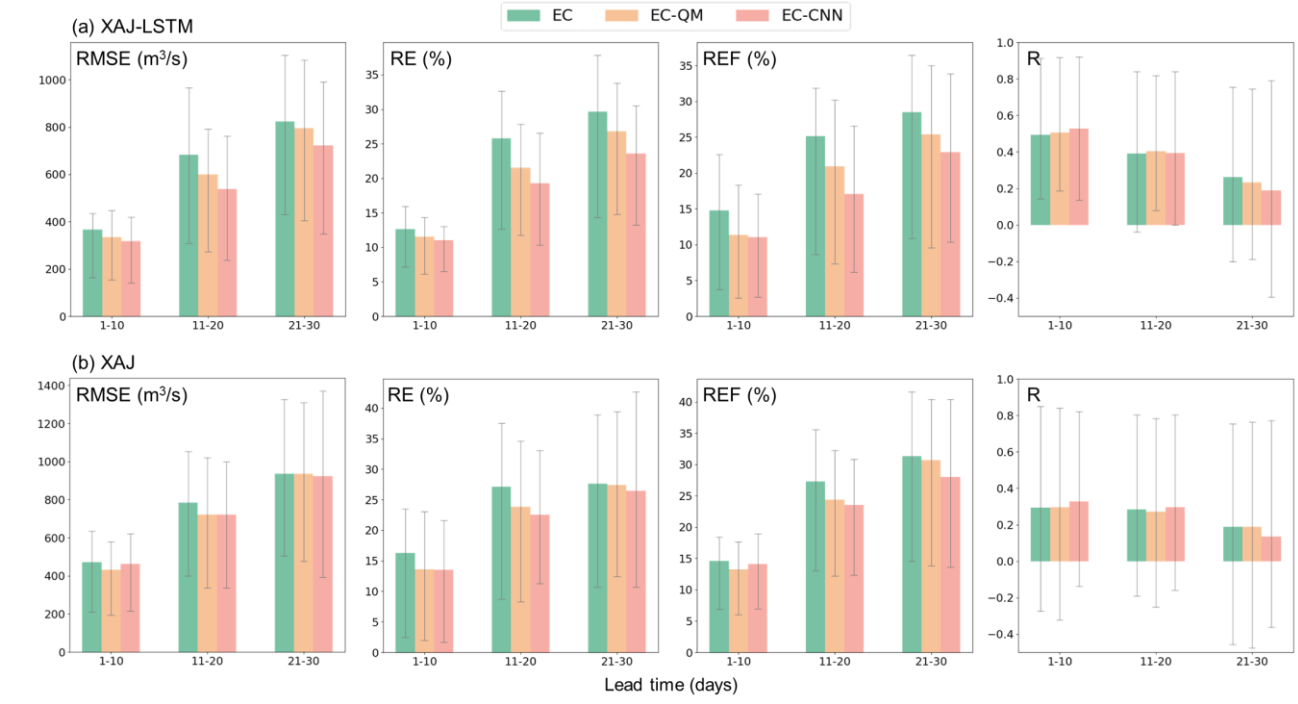




**Figure 7. The root mean square error (RMSE), relative error of streamflow (RE), relative error of maximum daily flow (REF) and correlation coefficient (R) for the (a) XAJ-LSTM and (b) XAJ streamflow forecasts driven by EC forecasts, EC-QM forecasts and EC-CNN forecasts for lead times of 1-10 days, 11-20 days, and 21-30 days. Error bar represents the 25$^{th}$-75$^{th}$ percentile interval.**

Figure 8 presents several examples of XAJ-LSTM streamflow forecasts initialized on different dates. In most cases, the EC-CNN forecasts can reduce the streamflow bias in a more flexible manner than the EC-QM forecasts, resulting in more accurate overall streamflow predictions. For example, the CNN model increases (decreases) the EC precipitation and hence the forecast streamflow initialized on May 02, 2017 (August 01, 2017), which improves the forecast skills in both cases. On the other hand, the QM reduces the precipitation and hence the streamflow forecasts for both dates, which improves the forecast on August but degrades the accuracy on May. Similarly, for the forecast initialized on Aug 11, 2017, the CNN model decreases the EC raw precipitation for lead times of 1-20 days and increases it for lead times of 21-30 days, which alleviates the streamflow overestimation on late August and underestimation on early September. However, the QM model consistently predicts lower precipitation and hence streamflow compared to raw EC forecasts for all lead times, worsening the underestimation in early September.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024

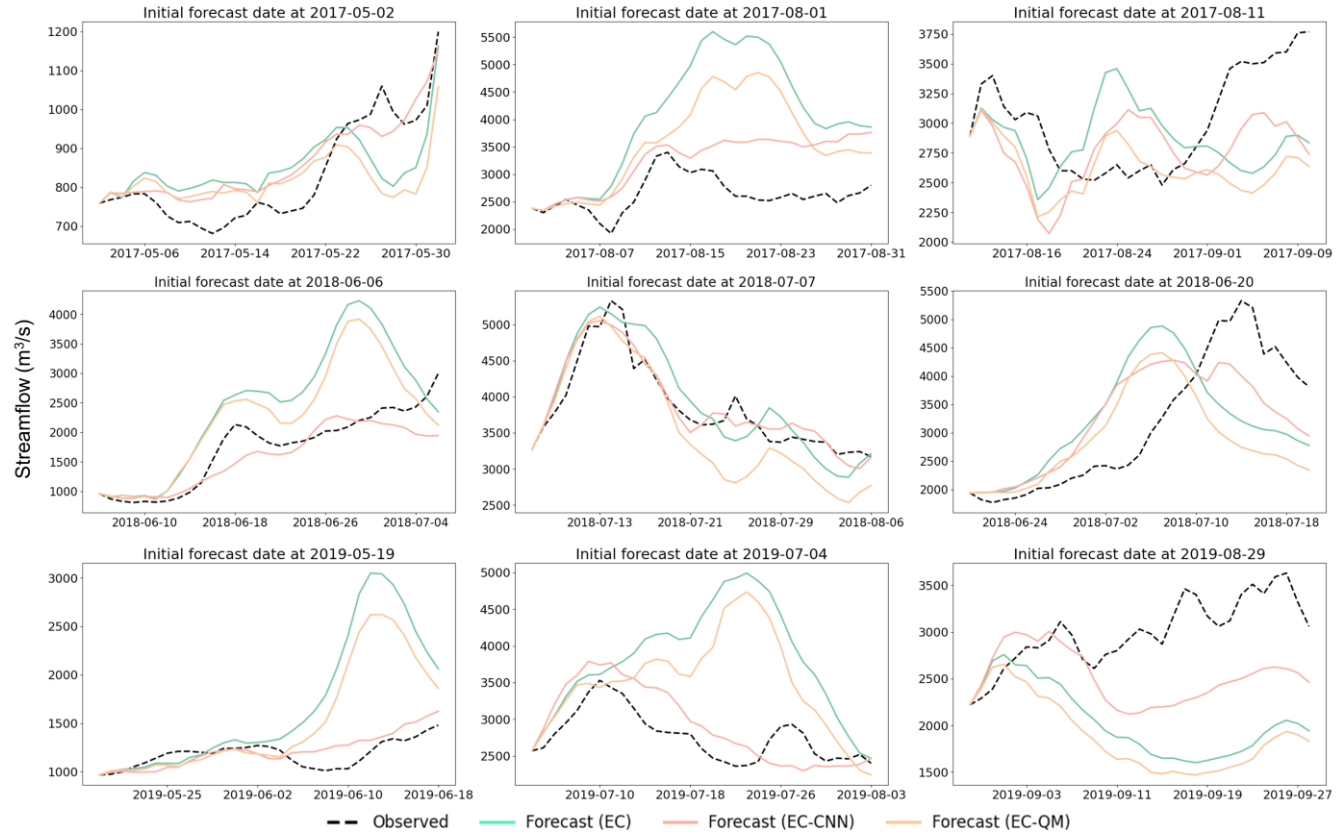




**Figure 8. Examples of sub-seasonal XAJ-LSTM streamflow forecasts for a lead time of 30 days driven by EC forecasts, EC-QM forecasts and EC-CNN forecasts.**

## 5 Discussion

### 5.1 Deep learning models can outperform traditional statistical downscaling methods in both mean and extremes

Traditional post-processing methods for precipitation forecasts often rely on local precipitation forecasts as the sole predictor, which can limit their ability to fully utilize the spatial information embedded in raw forecasts (Sun et al., 2021). In this study, a CNN post-processing model with ResNet blocks and a weighted loss function specialized on extreme events is established to investigate its potential to overcome these limitations by establishing multi-dimensional relationships between atmospheric circulation predictors and local precipitation.

We compare the CNN model with the commonly used quantile mapping (QM) bias correction method. The result reveals that the CNN model outperforms QM in improving sub-seasonal precipitation forecast skills, both in terms of capturing the general trends and predicting extreme precipitation events. This superior performance is likely due to the specialized loss function that balances the prediction of light rain events and extreme events by incorporating the mean absolute error and the threat score.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

This balance is crucial for sub-seasonal forecasts, where both event types impact water resource management. In general, the CNN structure used in our study is not only effective and easy to implement but also more computationally efficient than more complex CNN variations like SmaAt-UNet. Nevertheless, newer variants may better leverage multi-scale spatial information and incorporate multiple auxiliary predictors relevant to local weather conditions (Rasp and Lerch, 2018; Peng et al., 2020; Baño-Medina et al., 2020). Future research will focus on integrating these variants with new loss functions to achieve more desirable forecast outcomes.

The rapid development of AI-based weather prediction models in recent years, such as Pangu and GraphCast, has also demonstrated the potential of these models to achieve forecast skills comparable to state-of-the-art physics-based models (Bi et al., 2023; Lam et al., 2023). In comparison, our CNN-based statistically downscaled model of ECMWF precipitation forecasts offers improved sub-seasonal forecast skills with significantly lower computational resources, which could make it a practical and efficient tool for operational use in local meteorological or water agencies to provide high-quality forecasts and issue early warmings. Our results also underscore the potential of combining advanced AI techniques with physics-based forecasting methods to achieve superior performance and operational efficiency in weather prediction.

### 5.2 Better sub-seasonal precipitation forecasts may not guarantee better streamflow forecasts

The evaluation of sub-seasonal streamflow forecasts in Section 4.3 reveals a complex relationship between precipitation forecast accuracy and streamflow forecast performance. For example, the results presented in Figure 7 demonstrate that while improvements in precipitation forecasts generally lead to better streamflow forecasts, this relationship is not straightforward and can be influenced significantly by the choice of hydrologic model.

For example, a notable finding is that the hybrid XAJ-LSTM model shows much more substantial streamflow improvements with better precipitation forecasts compared to the standalone XAJ model. Specifically, the XAJ-LSTM model, which combines the strengths of LSTM networks and the XAJ hydrologic model, benefits significantly from the enhanced accuracy of EC-CNN forecasts. This model demonstrates a considerable reduction in RMSE for streamflow predictions over various lead times. On the other hand, the standalone XAJ model exhibits marginal improvements, or in some cases, no improvement at all, when driven by the same enhanced precipitation forecasts.

This disparity suggests that while advanced precipitation forecasts provide more accurate inputs, the ability of hydrologic models to effectively utilize these inputs is crucial. Similar findings are also reported by Valdez et al. (2022) for lead times of 7 days, who attributes the potential degradation of streamflow forecasts to other sources of uncertainties that may cancel out the added values of precipitation forecast improvements. The integration of machine learning (LSTM) and physical process representations (XAJ) allows it to better capture the long-term dependencies in hydrological processes, making it more responsive to the quality of precipitation forecasts at sub-seasonal scales. This synergy between the two models enables to

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

leverage the strengths of both conceptual understanding and data-driven prediction, which can also be extended to other basins with similar hydrological characteristics for addressing the sub-seasonal forecasting challenges.

### 5.3 Attribution of the XAJ-LSTM streamflow forecast error

To identify the possible sources of error for the XAJ-LSTM streamflow forecasts, an error decomposition method is employed to break down the total forecast error into its constituent parts. The specific contributions of each error source are isolated by calculating the RMSE of the streamflow simulations driven by observed precipitation and temperature (i.e., the hydrologic modelling error, $E_m$), the RMSE of the streamflow forecasts driven by forecast precipitation and observed temperature (i.e., the precipitation forecast error, $E_p$), and the RMSE of the streamflow forecasts driven by observed precipitation and forecast temperature (i.e., the temperature forecast error, $E_t$). Note that the total error between the observed streamflow and the forecast streamflow driven by forecast precipitation and temperature may not equal to the sum of $E_m$, $E_p$, and $E_t$, due to the interacting effects between multiple sources of error. This is manifested by the analysis result that the individual contributions of $E_p$, $E_m$, and $E_t$ to the total error add up to a value greater than 100%, indicating that there is a compensatory effect between multiple sources of error that reduces the total error.

Figure 9 depicts the individual contribution of $E_p$, $E_m$, and $E_t$ to their combined error. In general, the hydrologic modelling error $E_m$ dominates for lead times of 1-3 days, accounting for over 50% of the three sources of error combined. The ratio of $E_m$ decreases rapidly with the increase in lead times, and reaches a steady value of around 0.3 after the lead time of 15 days. The contribution ratio of precipitation forecasts error $E_h$ rise rapidly for lead times of 1-7 days, and stands at a steady value of around 0.6 after the lead time of 15 days. The temperature forecast error $E_t$, while present, has a less pronounced impact compared to $E_m$ and $E_p$, accounting for 5%-10% of the combined error. This is an expected result as precipitation generally impacts the streamflow more significantly than temperature.

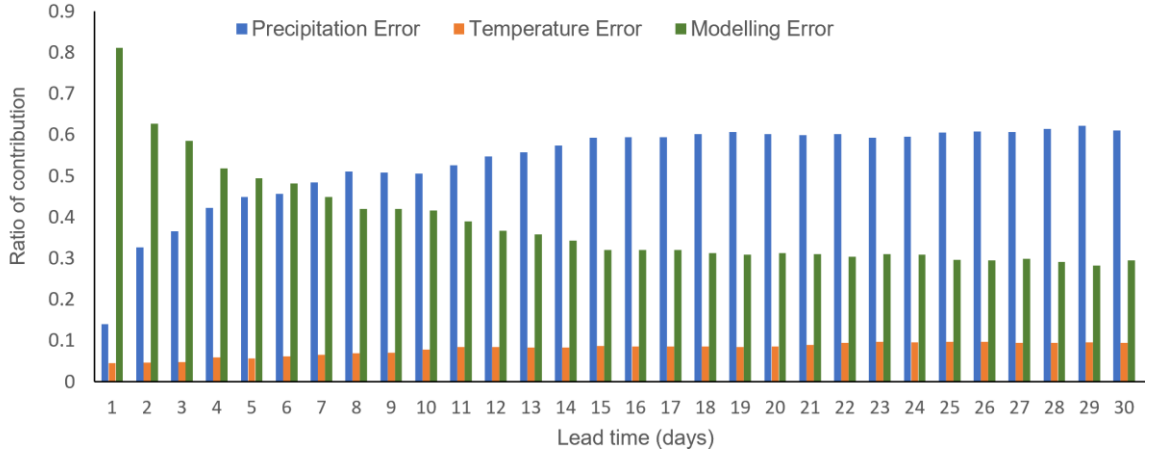



**Figure 9. Contribution of precipitation forecast errors, temperature forecast errors, and hydrologic modelling errors to their combined error.**

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

## 6 Summary and conclusions

This study proposes a deep learning based modelling framework for sub-seasonal hydrometeorological forecasts (i.e., precipitation and streamflow) for a lead time up to 30 days. The framework couples a CNN architecture with ResNet blocks for spatiotemporal downscaling of ECMWF raw precipitation forecasts to a hybrid hydrologic model integrating the conceptual XAJ model and LSTM for streamflow forecasting. The CNN model incorporates a specialized loss function that combines the continuous form of TS and MAE.

By applying the modelling framework to the source region of the Yangtze River Basin, we show that the CNN-based downscaling model exhibits advantages over quantile mapping in improving the precipitation forecasts in terms of both the general trends and the extreme events. The CNN-based model consistently outperforms the raw ECMWF forecasts and the traditional QM approach across all lead times, achieving an average RMSE value that are over 10% lower than both two forecasts. This improvement is also noted in extreme precipitation events, as demonstrated by a ~2% and ~5% lower RMSE of the CNN for heavy rain events as compared to raw forecasts and QM forecasts, especially for longer lead times.

With these precipitation forecasts serving as meteorological drivers of a hybrid XAJ-LSTM hydrologic model, it is found that CNN-based models can reduce the relative error of streamflow forecasts by 20%-24% compared to raw precipitation forecasts, particularly for longer lead times. This outperforms QM, which reduces the relative error of streamflow by 9%-16% compared to raw precipitation forecasts. The CNN-based precipitation forecasts also prove effective in deriving more reliable streamflow forecasts during extreme hydrological events (such as floods) for the XAJ-LSTM model. However, for the standalone XAJ model, the streamflow forecasts show marginal improvements, or in some cases, no improvement at all, with the same CNN enhanced precipitation forecasts. This highlights the importance of understanding the effectiveness of the hydrologic model as part of the sub-seasonal hydrometeorological modeling chain.

From a practical perspective, the proposed modelling framework is computationally efficient, requiring lower computational resources compared to fully AI models, traditional dynamic downscaling methods, and distributed hydrologic models. This makes it a viable tool for operational use in local meteorological and water management agencies to provide more accurate forecasts and issue early warnings. This study also shows the potential of combining advanced AI techniques with traditional hydrologic modelling approaches to achieve superior performance in sub-seasonal hydrometeorological forecasting, offering a robust and adaptable solution for effective water resources management and disaster preparedness.

## Code Availability Statement

The CNN model for statistically downscaling and bias correcting the ECMWF raw forecasts, and the XAJ hydrologic model with snow processes are both deposited at Zenodo repository (DOI: 10.5281/zenodo.12664798) and will be made publicly


available upon acceptance of this paper. The LSTM and the hybrid hydrologic model are developed and configured using the NeuralHydrology package (Kratzert et al., 2022), available at https://neuralhydrology.readthedocs.io/en/latest/index.html.

## Data Availability Statement

The ECMWF forecast data and observed precipitation and temperature data are all deposited at Zenodo repository (DOI: 10.5281/zenodo.12664851) and will be made publicly available upon acceptance of this paper.

## Author Contributions

ND contributed to the research design, compiled the data set, conducted the data processing, analysis, and manuscript preparation. HH contributed to the research design and code development. MY, JW, LC, SX, and HK contributed to the research design and manuscript preparation and editing.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## Acknowledgements

This research was funded by the National Key Research and Development Project (No. 2021YFC3000202; No. 2023YFC3081000). Jianhui Wei is supported financially by the German Federal Ministry of Science of Education (BMBF) through funding of the KARE_II project (01LR2006D1).

## References

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E., and Wilby, R. L.: Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting, Prog. Phys. Geogr., 36, 480–513, https://doi.org/10.1177/0309133312444943, 2012.

Adnan, R. M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B., and Kisi, O.: Daily streamflow prediction using optimally pruned extreme learning machine, J. Hydrol., 577, 123981, https://doi.org/10.1016/j.jhydrol.2019.123981, 2019.

Arnold, J. G., et al.: SWAT: Model use, calibration, and validation, Trans. ASABE, 55, 1491–1508, https://doi.org/10.13031/trans.55.10799, 2015.

Balint, G., Csik, A., Bartha, P., Gauzer, B., and Bonta, I.: Application of meteorological ensembles for Danube flood forecasting and warning, in: Transboundary Floods: Reducing Risks through Flood Management, edited by: Marsalek, J.,


Stancalie, G., and Balint, G., NATO Sci. Ser., Springer, Dordrecht, the Netherlands, 57–68, https://doi.org/10.1007/1-4020-4902-1_6, 2006.

Baño-Medina, J., Manzanas, R., & Gutiérrez, J. M. Configuration and intercomparison of deep learning neural models for statistical downscaling. Geoscientific Model Development, 13(4), 2109–2124, https://doi.org/10.5194/gmd-13-2109-2020, 2020.

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections, Clim. Dyn., 57, 2941–2951, https://doi.org/10.1007/s00382-021-05843-8, 2021.

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.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/10.1038/s41586-023-06191-0, 2023.

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resour. Res., 51, 4923–4947, https://doi.org/10.1002/2015WR017173, 2015.

Bonavita, M.: On some limitations of current machine learning weather prediction models, Geophys. Res. Lett., 51, e2023GL107377, https://doi.org/10.1029/2023GL107377, 2024.

Brotzge, J. A., Berchoff, D., Carlis, D. L., Carr, F. H., Carr, R. H., Gerth, J. J., et al.: Challenges and opportunities in numerical weather prediction, Bull. Am. Meteorol. Soc., 104, E698–E705, https://doi.org/10.1175/BAMS-D-22-0121.1, 2023.

Cloke, H. L., and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Chen, G., and Wang, W.-C.: Short-term precipitation prediction for contiguous United States using deep learning, Geophys. Res. Lett., 49, e2022GL097904, https://doi.org/10.1029/2022GL097904, 2022.

Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, Hydrol. Earth Syst. Sci., 20, 3601–3618, https://doi.org/10.5194/hess-20-3601-2016, 2016.

de Andrade, F.M., Young, M.P., MacLeod, D., Hirons, L.C., Woolnough, S.J. and Black, E. Subseasonal precipitation prediction for Africa: Forecast evaluation and sources of predictability. Weather and Forecasting, 36(1), pp.265-284, 2021.

Dehshiri, S. S. H., and Firoozabadi, B.: A multi-objective framework to select numerical options in air quality prediction models: A case study on dust storm modeling, Sci. Total Environ., 863, 160681, https://doi.org/10.1016/j.scitotenv.2022.160681, 2023.

Dong, N., Wei, J., Yang, M., Yan, D., Yang, C., Gao, H., Arnault, J., Laux, P., Zhang, X., Liu, Y., and Niu, J.: Model estimates of China's terrestrial water storage variation due to reservoir operation, Water Resour. Res., 58, e2021WR031787, https://doi.org/10.1029/2021WR031787, 2022.


505 Di Luca, A., de Elía, R., and Laprise, R.: Potential for added value in precipitation simulated by high-resolution nested regional climate models and observations, Clim. Dyn., 44, 2519–2537, https://doi.org/10.1007/s00382-014-2327-0, 2015.

Ebert-Uphoff, I., and Hilburn, K.: Evaluation, tuning and interpretation of neural networks for working with images in meteorological applications, Bull. Am. Meteorol. Soc., 1–47, https://doi.org/10.1175/BAMS-D-19-0324.1, 2020.

Gao, S., Huang, D., Du, N., Ren, C., and Yu, H. WRF ensemble dynamical downscaling of precipitation over China using
510 different cumulus convective schemes. Atmospheric Research, 271, 106116, https://doi.org/10.1016/j.atmosres.2022.106116, 2022.

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G.: The Soil and Water Assessment Tool: Historical development, applications, and future research directions, Trans. ASABE, 57, 1211–1250, https://doi.org/10.13031/trans.57.10791, 2014.

Guo, C., and Berkhahn, F.: Entity embeddings of categorical variables, arXiv preprint, arXiv:1604.06737, 2016.

515 Hu, C. H., Guo, S. L., Xiong, L. H., and Peng, D. Z.: A modified Xin'anjiang model and its application in northern China, Hydrol. Res., 36, 175–192, https://doi.org/10.2166/nh.2005.0013, 2005.

Humphrey, G. B., Gibbs, M. S., Dandy, G. C., and Maier, H. R.: A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network, J. Hydrol., 540, 623–640, https://doi.org/10.1016/j.jhydrol.2016.06.026, 2016.

520 Jaun, S., Ahrens, B., Walser, A., Ewen, T., and Schär, C.: A probabilistic view on the August 2005 floods in the upper Rhine catchment, Nat. Hazards Earth Syst. Sci., 8, 281–291, https://doi.org/10.5194/nhess-8-281-2008, 2008.

Jiang, Z., Yang, S., Liu, Z., Xu, Y., Xiong, Y., Qi, S., et al.: Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin, Environ. Model. Softw., 155, 105436, https://doi.org/10.1016/j.envsoft.2022.105436, 2022.

525 Jiang, M., Weng, B., Chen, J., Huang, T., Ye, F., and You, L.: Transformer-enhanced spatiotemporal neural network for post-processing of precipitation forecasts, J. Hydrol., 630, 130720, https://doi.org/10.1016/j.jhydrol.2024.130720, 2024.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, Bull. Am. Meteorol. Soc., 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

530 Kim, T., Yang, T., Zhang, L., & Hong, Y. Near real-time hurricane rainfall forecasting using convolutional neural network models with Integrated Multi-satellitE Retrievals for GPM (IMERG) product. Atmospheric Research, 270, 106037. https://doi.org/10.1016/j.atmosres.2022.106037, 2022.

Kisi, O.: Streamflow forecasting using different artificial neural network algorithms, J. Hydrol. Eng., 12, 532–539, https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532), 2007.

535 Kratzert, F., Gauch, M., Nearing, G., and Klotz, D. NeuralHydrology---A Python library for Deep Learning research in hydrology. Journal of Open Source Software, 7(71), 4050, https://doi.org/10.21105/joss.04050, 2022.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Hydrology and Earth System Sciences
Discussions

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, 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.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resour. Res., 55, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... and Battaglia, P.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, https://doi.org/10.1126/science.adh7887, 2023.

Larraondo, P. R., Renzullo, L. J., Van Dijk, A. I., Inza, I., and Lozano, J. A.: Optimization of deep learning precipitation models using categorical binary metrics, J. Adv. Model. Earth Syst., 12, e2019MS001909, https://doi.org/10.1029/2019MS001909, 2020.

Li, J., Li, L., Zhang, T., Xing, H., Shi, Y., Li, Z., Wang, C., and Liu, J.: Flood forecasting based on radar precipitation nowcasting using U-net and its improved models, J. Hydrol., 632, 130871, https://doi.org/10.1016/j.jhydrol.2023.130871, 2024.

Li, W., Pan, B., Xia, J., and Duan, Q. Convolutional neural network-based statistical post-processing of ensemble precipitation forecasts. Journal of hydrology, 605, 127301, https://doi.org/10.1016/j.jhydrol.2021.127301, 2022.

Liang, P., Lin, H., and Ding, Y.: Dominant modes of subseasonal variability of East Asian summertime surface air temperature and their predictions, J. Climate, 31, 2729–2743, https://doi.org/10.1175/JCLI-D-17-0368.1, 2018.

Ling, F., Li, Y., Luo, J.-J., Zhong, X., and Wang, Z.: Two deep learning-based bias-correction pathways improve summer precipitation prediction over China, Environ. Res. Lett., 17, 124025, https://doi.org/10.1088/1748-9326/acaea0, 2022.

Liu, J., Yuan, X., Zeng, J., Jiao, Y., Li, Y., Zhong, L., and Yao, L.: Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning, Hydrol. Earth Syst. Sci., 26, 265–278, https://doi.org/10.5194/hess-26-265-2022, 2022.

Liu, D., Jiang, W., Mu, L., and Wang, S.: Streamflow prediction using deep learning neural network: case study of Yangtze River, IEEE Access, 8, 90069–90086, https://doi.org/10.1109/ACCESS.2020.2993751, 2020.

Lyu, Y., Zhu, S., Zhi, X., 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.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, https://doi.org/10.1029/2009RG000314, 2010.

Merino, A., García-Ortega, E., Navarro, A., Sánchez, J. L., and Tapiador, F. J.: WRF hourly evaluation for extreme precipitation events, Atmos. Res., 274, 106215, https://doi.org/10.1016/j.atmosres.2022.106215, 2022.


570 Michalek, A. T., Villarini, G., and Kim, T.: Understanding the impact of precipitation bias-correction and statistical downscaling methods on projected changes in flood extremes, Earth's Future, 12, e2023EF004179, https://doi.org/10.1029/2023EF004179, 2024.

Nooni, I. K., Tan, G., Hongming, Y., Saidou Chaibou, A. A., Habtemicheal, B. A., Gnitou, G. T., and Lim Kam Sian, K. T.: Assessing the performance of WRF Model in simulating heavy precipitation events over East Africa using satellite-based
575 precipitation product, Remote Sens., 14, 1964, https://doi.org/10.3390/rs14091964, 2022.

Ni, L., Wang, D., Singh, V. P., Wu, J., Chen, X., Tao, Y., Zhu, X., Jiang, J., and Zeng, X.: Monthly precipitation prediction at regional scale using deep convolutional neural networks, Hydrol. Process., 37, e14954, https://doi.org/10.1002/hyp.14954, 2023.

Nie, Y., and Sun, J.: Improving dynamical-statistical subseasonal precipitation forecasts using deep learning: A case study in
580 Southwest China, Environ. Res. Lett., https://doi.org/ 10.1088/1748-9326/ad5370, 2024.

Peng, T., Zhi, X., Ji, Y., Ji, L., and Tian, Y.: Prediction skill of extended range 2-m maximum air temperature probabilistic forecasts using machine learning postprocessing methods, Atmosphere, 11, 1–17, https://doi.org/10.3390/atmos11080805, 2020.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian model averaging to calibrate forecast
585 ensembles, Mon. Weather Rev., 133, 1155–1174, https://doi.org/10.1175/MWR2906.1, 2005.

Rasp, S., Lerch, S., 2018. Neural networks for postprocessing ensemble weather forecasts. Mon. Weather. Rev. 146(11), 3885–3900.

Robertson, D. E., and Wang, Q. J.: Seasonal forecasts of unregulated inflows into the Murray River, Australia, Water Resour. Manag., 27, 2747–2769, https://doi.org/10.1007/s11269-013-0313-4, 2013.

590 Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., and Perera, B. J. C. Statistical downscaling of precipitation using machine learning techniques. Atmospheric research, 212, 240-258, https://doi.org/10.1016/j.atmosres.2018.05.022, 2018

Saha, S., et al.: The NCEP Climate Forecast System Version 2, J. Clim., 27, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

Scheuerer, M., and Hamill, T. M.: Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted
595 gamma distributions, Mon. Weather Rev., 143, 4578–4596, https://doi.org/10.1175/MWR-D-15-0068.1, 2015.

Shi, X.: Enabling smart dynamical downscaling of extreme precipitation events with machine learning, Geophys. Res. Lett., 47, e2020GL090309, https://doi.org/10.1029/2020GL090309, 2020.

Singhal, A., Jaseem, M., and Jha, S. K.: Spatial connections in extreme precipitation events obtained from NWP forecasts: A complex network approach, Atmos. Res., 282, 106538, https://doi.org/10.1016/j.atmosres.2022.106538, 2023.

600 Srivastava, A. K., Ullrich, P. A., Rastogi, D., Vahmani, P., Jones, A., and Grotjahn, R. Assessment of WRF (v 4.2. 1) dynamically downscaled precipitation on subdaily and daily timescales over CONUS. Geoscientific Model Development, 16(13), 3699-3722, https://doi.org/10.5194/gmd-16-3699-2023, 2023.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024




Sun, L., and Lan, Y. Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods. International Journal of Climatology, 41(2), 1128-1147, https://doi.org/10.1002/joc.6769, 2021.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons, Rev. Geophys., 56, 79–107, https://doi.org/10.1002/2014RG000477, 2016.

Tabari, H., Paz, S. M., Buekenhout, D., and Willems, P.: Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought, Hydrol. Earth Syst. Sci., 25, 3493–3517, https://doi.org/10.5194/hess-25-3493-2021, 2021.

Taillardat, M., Mestre, O., Zamo, M., and Naveau, P.: Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics, Mon. Weather Rev., 144, 2375–2393, https://doi.org/10.1175/MWR-D-15-0260.1, 2016.

Tan, Y., Dong, N., Hou, A., and Yan, W.: An improved Xin'anjiang hydrological model for flood simulation coupling snowmelt runoff module in Northwestern China, Water, 15, 3401, https://doi.org/10.3390/w15193401, 2023.

Valdez, E. S., Anctil, F., and Ramos, M.-H.: Choosing between post-processing precipitation forecasts or chaining several uncertainty quantification tools in hydrological forecasting systems, Hydrol. Earth Syst. Sci., 26, 197–220, https://doi.org/10.5194/hess-26-197-2022, 2022.

Vandal, T., Kodra, E., & Ganguly, A. R. Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. Theoretical and Applied Climatology, 137, 557-570, https://doi.org/10.1007/s00704-018-2613-3, 2019

Vigaud, N., Tippett, M. K., and Robertson, A. W.: Deterministic skill of subseasonal precipitation forecasts for the East Africa-West Asia sector from September to May, J. Geophys. Res.-Atmos., 124, 11887–11896, https://doi.org/10.1029/2019JD030747, 2019.

Vrac, M., and Friederichs, P.: Multivariate-intervariable, spatial, and temporal bias correction, J. Clim., 28, 218–237, https://doi.org/10.1175/JCLI-D-14-00059.1, 2015.

Wang, R., Zhang, J., Guo, E., Zhao, C., and Cao, T.: Spatial and temporal variations of precipitation concentration and their relationships with large-scale atmospheric circulations across Northeast China, Atmos. Res., 222, 62–73, https://doi.org/10.1016/j.atmosres.2019.02.008, 2019.

Wei, L., Hu, K.-H., and Hu, X.-D.: Rainfall occurrence and its relation to flood damage in China from 2000 to 2015, J. Mt. Sci., 15, 2492–2504, https://doi.org/10.1007/s11629-018-4931-4, 2018.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., and Mearns, L. O.: Guidelines for use of climate scenarios developed from statistical downscaling methods, Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA, 2004.

https://doi.org/10.5194/hess-2024-212
Preprint. Discussion started: 29 July 2024


Xu, Y. P., Gao, X., Zhu, Q., and Zhang, Y.: Coupling a regional climate model and distributed hydrological model to assess future water resources in Jinhua River Basin, East China, ASCE J. Hydrol. Eng., 20, 04014054, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001007, 2015.

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., and Zhao, B.: A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data, J. Hydrol., 590, 125206, https://doi.org/10.1016/j.jhydrol.2020.125206, 2020.

You, X. X., Liang, Z. M., Wang, Y. Q., and Zhang, H.: A study on loss function against data imbalance in deep learning correction of precipitation forecasts, Atmos. Res., 281, 106500, https://doi.org/10.1016/j.atmosres.2023.106500, 2023.

Yuan, X., Ma, F., Wang, L., Zheng, Z., Ma, Z., Ye, A., and Peng, S.: An experimental seasonal hydrological forecasting system over the Yellow River basin – Part 1: Understanding the role of initial hydrological conditions, Hydrol. Earth Syst. Sci., 20, 2437–2451, https://doi.org/10.5194/hess-20-2437-2016, 2016.

Yuan, X., Wang, S., and Hu, Z.-Z.: Do climate change and El Niño increase likelihood of Yangtze River extreme rainfall?, Bull. Am. Meteorol. Soc., 99, S113–S117, https://doi.org/10.1175/BAMS-D-17-0089.1, 2018.

Yuan, X., Wood, E. F., Luo, L., and Pan, M.: A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, Geophys. Res. Lett., 38, L13401, https://doi.org/10.1029/2011GL047792, 2011.

Zhang, Q., Li, Y. P., Huang, G. H., Wang, H., Li, Y. F., Liu, Y. R., and Shen, Z. Y.: A novel statistical downscaling approach for analyzing daily precipitation and extremes under the impact of climate change: Application to an arid region, J. Hydrol., 615, 128730, https://doi.org/10.1016/j.jhydrol.2022.128730, 2022.

Zhang, Y., Ragettli, S., Molnar, P., Fink, O., and Peleg, N.: Generalization of an Encoder-Decoder LSTM model for flood prediction in ungauged catchments, J. Hydrol., 614, 128577, https://doi.org/10.1016/j.jhydrol.2022.128577, 2022.

Zhao, R. J.: The Xin'anjiang model applied in China, J. Hydrol., 135, 371–381, https://doi.org/10.1016/0022-1694(92)90096-E, 1992.

Zhu, E., Yuan, X., and Wood, A.: Benchmark decadal forecast skill for terrestrial water storage estimated by an elasticity framework, Nat. Commun., 10, 1237, https://doi.org/10.1038/s41467-019-09245-3, 2019.

Zhu, S., Remedio, A. R. C., Sein, D. V., Sielmann, F., Ge, F., Xu, J., et al.: Added value of the regionally coupled model ROM in the East Asian summer monsoon modeling, Theor. Appl. Climatol., 140, 375–387, https://doi.org/10.1007/s00704-020-03093-8, 2020.