Deep learning based sub-seasonal precipitation and streamflow ensemble 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: N. Dong (dongnp@iwhr.com) and 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 ensemble forecasts for a lead time of up to 30 days. This is achieved by coupling (1) an ensemble of enhanced convolutional neural network (CNN) models architecture with ResNet blocks and a specialized loss function for statistically downscaling of the ECMWF raw precipitation ensemble precipitation forecasts to (2) a hybrid hydrologic model integrating the conceptual Xin'anjiang model (XAJ) and the long-short term memory network (LSTM) for ensemble streamflow forecasting. The CNN incorporates a specialized loss function that combines the continuous form of threat score and mean squared error. Applying the modeling framework to the source region of the Yangtze River Basin, results indicate that the CNN-based downscaling model exhibits ~34 13% and ~26 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 approximately a ~6 2% and ~10 5% lower RMSE than raw forecasts and QM for heavy precipitation events. 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 16 8%-33 2% lower relative errors and 20%-31% 13%-22% lower RMSE compared to those driven by raw forecasts. However, the standalone XAJ model shows only 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.

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) models, 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). On the other hand, forecasting weather and predicting climate using machine learning, especially deep learning (DL), has recently become a hot topic. A common approach for this purpose is to use preceding predictors from observational or reanalysis data to forecast subsequent predictands (Weyn et al., 2021; Xie et al., 2023; Ham et al., 2019). An alternative method involves postprocessing dynamical

65 forecasts. For instance, Cho et al. (2020) applied machine learning techniques, including random forests and support vector machines, to develop statistical relationships for temperature adjustments. Similarly, Kim et al. (2021) utilized Long Short-Term Memory (LSTM) networks to correct bias in the amplitude and phase of the Madden–Julian Oscillation. 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

70 al., 2019; Sachindra et al., 2018; Ning et al., 2024). For example, Lagerquist et al. (2019) used a CNN to identify fronts in gridded data for spatially explicit prediction of synoptic-scale fronts.

~~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 in~~of~~ forecasts, the~~se~~ DL-based

75 models~~y~~ 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 the U-shaped U-Net (Han et al, 2021; Horat and Lerch, 2024; Ni et al., 2023) and SmaAt-UNet (Li et al., 2024), or coupled standard CNNs with different structures, such as Auto-Encoder (Ling et al., 2022~~),)~~ and Transformer (Ling et al., 2024)~~., and I~~in particular, the residual network, ResNet, has been introduced in sub-seasonal forecast correction, which

80 shows the potential of mitigating the vanishing gradient issue by introducing the residual paths (Jin et al 2022; 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.

Other state-of-the-art forms of deep learning for weather forecasts include fully DL-based models, such as Pangu (Bi et al.,

85 2023) and GraphCast (Lam et al., 2023), which are reported able to achieve forecast skills comparable to numerical weather prediction systems. While these models may appear quite different from statistical post-processing deep learning models, some argue that these models act more as post-processing tools rather than realistic simulators of the atmosphere due to the lack of physical fidelity and consistency (Bonavita et al., 2024). Although not the primary focus of this paper, this calls attention to the scientific community to critically evaluate and differentiate between the capabilities and applications of fully DL-based

90 models and DL models designed for post-processing.~~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~~

95 ~~(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 CLHMS~~SWAT~~ and VIC models, make predictions by interpreting detailed physical processes (Gassman et al., 2014; Arnold et al., 2015; Dong et al., 2022; Dong et al., 2023). 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) combined a Bayesian neural network (BNN) with the traditional GR4J model and achieved improved forecast accuracy compared to using either the BNN or GR4J alone. ~~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~~. Ensemble approach with DL models also shows promising results (Ferranti et al., 2018; Bremnes, 2020) that require further investigation.

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 for the wet season 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?

Specifically, this study investigates sub-seasonal precipitation and streamflow ensemble forecast skills for up to 30 days ahead with deep learning models, which integrates ~~a~~ enhanced CNN ~~architecture~~ models with ResNet blocks and specialized loss functions for post-processing ~~state-of-the-art~~the ensemble ECMWF forecasts with a hybrid hydrologic model ~~integrating~~ of 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~~and subtropical monsoon climates, 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 (Hao et al., 2024). 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.

The SR spans a large north-south range, with the northern region deep in the Tibetan Plateau and dominated by a plateau monsoon climate, and the southern region characterized by low hills and a subtropical monsoon climate. These contrasting environments suggest different runoff generation mechanisms between the two regions. To ensure the accuracy of streamflow simulations at Shigu, we therefore divided the SR into two sub-basins (i.e., northern and southern basins in Figure 1) for hydrologic modeling in Section 3.4.

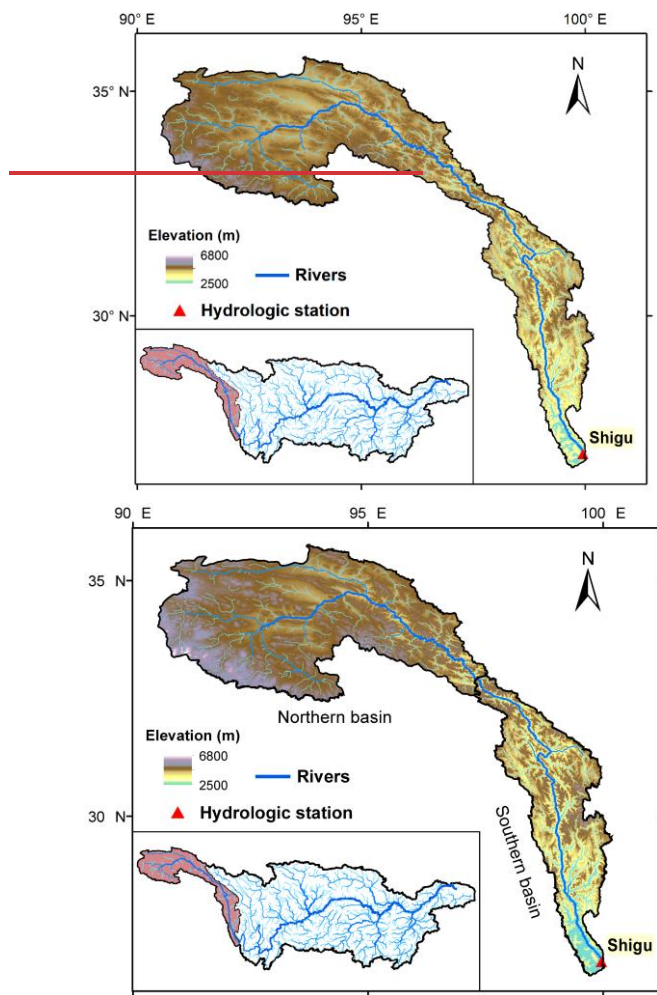



Figure 1. The source region of the Yangtze River Basin and its location in the Yangtze River Basin.

2.2 Data sources

150 *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~~ reference 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 reforecast data. The European Centre for Medium-Range Weather Forecasts (ECMWF) offers a sub-155 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 reforecast data for a lead time of 30 days initialized on 35 dates during the wet season (between May and August) per year during 2002-2019. The forecasted variables used in this study 160 include precipitation, convective precipitation at the land surface~~;~~, and temperature, wind components, geopotential heights, and specific humidity at 200/500/850hPa pressure levels. All of these variables are collected at a spatial resolution of 1.5°.

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

165 **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 during the wet season (May to August) for a lead time up to 30 days. We first employ all 10 ensemble members from the ECMWF S2S gridded sub-seasonal precipitation reforecast dataset for the next 30 days as raw forecasts, denoted as EC. ~~First, we employ the ECMWF 170 S2S gridded sub-seasonal precipitation forecasts for the next 30 days as the raw forecasts.~~ An ensemble of~~-~~ enhanced CNN ~~models~~architecture with ResNet blocks and a specialized loss function is established to statistically downscale and bias correct each ensemble member of the 1.5° ~~ECMWF~~ EC raw precipitation forecasts to ~~the~~ 0.25° grid resolution, with its post-processed forecast denoted EC-CNN (Section 3.2.1). The quantile mapping (QM) serves~~also serves~~ as a benchmark ~~to compare with the post-processed precipitation of the CNN~~for comparison, with its post-processed forecast denoted EC-QM (Section 3.2.2).~~.~~

175 The~~se post-processed~~ three gridded ensemble precipitation forecasts (EC, EC-QM and EC-CNN), along with ECMWF gridded sub-seasonal daily temperature forecasts corrected by the delta method (Section 3.3), are ~~then~~ employed to drive two~~a hybrid~~ lumped hydrologic model~~s~~ to produce the daily streamflow forecasts for lead times of 1-30 days. All these gridded forecasts are areal-averaged over the two sub-basins of the SR (Figure 1) before being input to the lumped hydrologic models. The first

hydrologic model is a standalone XAJ model (Section 3.4.1), and the second model is a~~The~~ hybrid model that integrates the
conceptual XAJ ~~flood forecast~~ model ~~with snow processes~~ and the LSTM (hereinafter XAJ-LSTM) (Sections 3.4.2 and 3.4.3).
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 EC~~raw~~ and EC-QM ~~based~~ forecasts.

The evaluation metrics include ~~using~~ deterministic ~~a series of~~ metrics of Root Mean Squared Error (RMSE), Relative Error
(RE) and the Nash-Sutcliffe efficiency (NSE) for the ensemble mean, and probabilistic metrics of Continuous Ranked
Probability Score (CRPS) for a total of 10 ensemble members (Section 3.5). A detailed workflow of this study is presented in
Fig. 2.

Figure 2. The workflow of this study.

190 **3.2 Statistically downscaling of ensemble precipitation forecasts**

3.2.1 Enhanced Cconvolutional neural network

An ensemble of enhanced CNN models with ResNet blocks and a specialized loss function is established to learn the spatially dependent relationship between fine-resolution local precipitation and coarse-resolution predictors from surrounding regions. Specifically, it downscales a total of 10 ensemble members of the ECMWF S2S reforecasts from a 1.5° resolution to a 0.25° resolution, using the CN05.1 reference precipitation dataset. 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.~~ The model takes spatially distributed inputs of 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. These inputs cover a 3×3 area of coarse grid cells at 1.5° resolution, centered around the target fine grid cell at 0.25° resolution. Due to the square-shaped input structure of the CNN model, some ECMWF data from outside the basin boundary are included in the input. For the outputs, the predictand is the daily precipitation at a spatial resolution of 0.25°, and the CNN loops over each fine-resolution grid cell (0.25°) within the basin boundary, thereby generating a high-resolution precipitation forecast for the entire SR. Figure 2 presents the model structure, which primarily consists of convolutional layers, embedding layers, fully connected layers, and residual paths. More details on ~~T~~the model structure ~~is~~ are ~~detailed~~ described as follows.

(1) Inputs to the network are predictors from a 37×37 grid patch (1.5° resolution) centered on the target grid cell (0.25° resolution). This patch includes a total of 19 meteorological variables, resulting in input arrays with sizes of 19×37×37. ~~Here, the ensemble mean forecast is used as predictors, as they typically encompass the majority of information of the ensemble~~ The spatial dimension of ~~3×3~~ is selected because it shows the best performance among four candidates of 1×1, 3×3, 5×5, and 7×7 (Li et al., 2022).

(2) The model includes three ResNet blocks, with each block ~~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,~~ blocks mitigate~~s~~ 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 (Li et al., 2022). 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,~~

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

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.,

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

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

225 (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)
230 and Mean Squared Error (MSE) is used in this study, i.e.,

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

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

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

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

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

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

240 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. α is the precipitation threshold that corresponds to the 90th 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
245 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 EC-CNN ~~based~~ model ~~in terms of the sub-seasonal precipitation forecasts~~.

We implement QM using a non-parametric approach that adjusts the quantiles of the forecasted and observed data without assuming a specific distribution. Specifically, the empirical cumulative distribution functions (CDFs) of observed and forecasted daily precipitation are built respectively, and each percentile of the forecasted data is adjusted to match the corresponding percentile in the observed data. Dry days with a precipitation amount less than 0.1 mm are excluded from the derivation of CDFs (Gudmundsson et al., 2012). To match the 1.5° forecast resolution with the 0.25° reference dataset resolution, the empirical CDFs are established separately for each 0.25° grid from the corresponding 1.5° forecast grid cell. Manzanas et al (2018), Cannon et al. (2015) and other studies have indicated the effectiveness of this implementation in improving the overall precipitation forecasts.

~~QM estimates the values of the quantile-quantile relation of observations and forecasts for regularly spaced quantiles. Specifically~~Here, the period from 2002-2015 is used to estimate the ~~cumulative distribution functions (~~empirical ~~CDFs) of the forecast precipitation *F* and observed precipitation *O* with 100 discretized quantiles.~~CDFs, and these ~~The~~ CDFs are then used to ~~correct the~~~~m~~a EC forecasts~~tch 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. The QM is constructed separately for each lead time to account for forecast bias variations across different lead times. For each lead time, a single model is applied across all months, which is aligned with the structure of the CNN model built in this study.

设置了格式: 字体: (中文) +中文正文 (宋体), (中文) 简体中文 (中国大陆)

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 lead time during ~~the~~ May and August~~calibration period~~ of 2002-2015 as a calibration period, and then apply ~~the monthly~~ a single delta model for each lead time to the forecast temperature during May and August ~~the validation period~~ of 2016-2019 as a validation period. Given that temperature is not the main focus of the paper, plus that 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

280 **3.4.1 Xin'anjiang model**

The XAJ model is a conceptual hydrological model (Zhao, 1992), which has been widely used to generate flood forecasts for 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

285 streamflow of the SR at the sub-seasonal scale, which shows satisfactory accuracies for basins with snow melting runoff in our previous study (Tan et al., 2019).

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

290 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 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

295 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 there is snow accumulation around the winter and spring.

3.4.3 Model integration

The XAJ model employs daily precipitation and temperature of the two sub-basins to simulate the daily streamflow at the

300 Shigu station., and Tthe model parameters are calibrated for the period of 1981-2015 and validated for the period of 2016-2019. The particle swarm optimization (PSO) approach is employed to optimize the parameters of the XAJ model, with the Nash-Sutcliffe efficiency (NSE) NSE as the objective function.

The LSTM model takes as input features a time sequence of daily precipitation $\mathbf{p}_i = \mathbf{p}_i[1], \mathbf{p}_i[2], ..., \mathbf{p}_i[N]$, daily temperature $\mathbf{t}_i = \mathbf{t}_i[1], \mathbf{t}_i[2], ..., \mathbf{t}_i[N]$, $(i = 1, 2)$ of two sub-basins, and XAJ-simulated daily streamflow $\mathbf{q} = \mathbf{q}[1], \mathbf{q}[2], ..., \mathbf{q}[N]$ over N

305 time steps. Each element of $\mathbf{p}_i$, $\mathbf{t}_i$ and $\mathbf{q}_i$, namely $\mathbf{p}_i[n]$, $\mathbf{t}_i[n]$, and $\mathbf{q}_i[n]$, 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-root (~~root mean squared error (RMSE)~~RMSE) and relative error (RE) for the ensemble mean and using the CRPS for the total 10 ensemble members. Specifically, we classify the 5-day daily precipitation less than and greater than the 90th percentile of all historic 5-day~~daily~~ precipitation during 2002-2019 as light rain events and heavy rain events. The RMSEs are calculated for all rain events ~~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)~~RE, ~~and Pearson's correlation coefficient (R)~~ and NSE 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 Figure 3 indicate a satisfactory performance for the standalone XAJ model. T~~, as t~~he daily NSE values of simulated streamflow are 0.88 and 0.83 during the calibration and validation period. The relative error of streamflow is 1.0% and 2.6% during the calibration and validation period. The mean absolute error and relative error of simulated maximum daily flow are 844 m^3/s and 17.0% during the calibration period, and are 379 m^3/s and 7.9% during the validation period, respectively.

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 Efficiency (NSE)~~ 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% during the validation period, 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.

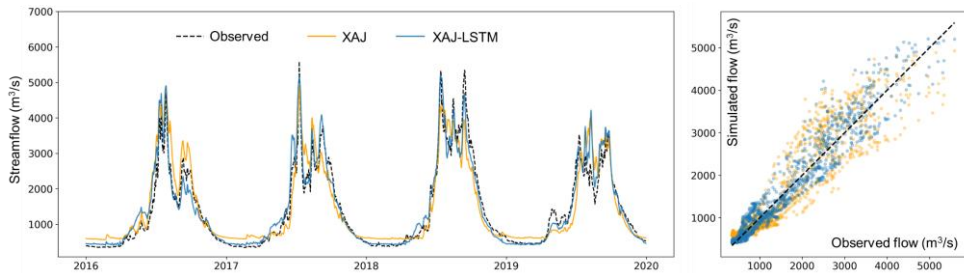



Figure 3. Streamflow validation of the XAJ model and the ~~hybrid~~ XAH-LSTM hydrologic model.

340 **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 and RE of areal-averaged~~EC~~ EC ~~raw forecasts~~, EC-QM and EC-CNN precipitation ~~averaged over SR~~ at different lead time ~~ranges of 1-30 days~~ for the period from 2016 to 2019 are provided in Figure 4, with the error bars representing the 25th-75th percentile interval.

345 Generally, the EC raw 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 areal-averaged~~raw~~ EC precipitation ~~forecasts~~ is ~~2.76~~1.13 mm/day. The EC-QM effectively reduces RMSE at all lead times by an average of 0.1~~2~~ mm/day (~~~5~~11%), indicating the effectiveness of QM in improving precipitation at sub-seasonal scales. By contrast, the EC-CNN ~~model~~ exhibits ~~~10~~26% less

350 RMSE compared to EC-QM for all lead times, which reduces the RMSE of ~~raw~~ EC forecasts by 0.~~35~~38 mm/day (~~~13~~34%). The RE shows a similar trend to RMSE, and the RE of EC, EC-QM, and EC-CNN are 27%, 36%, and 42% averaged over all lead times, respectively.

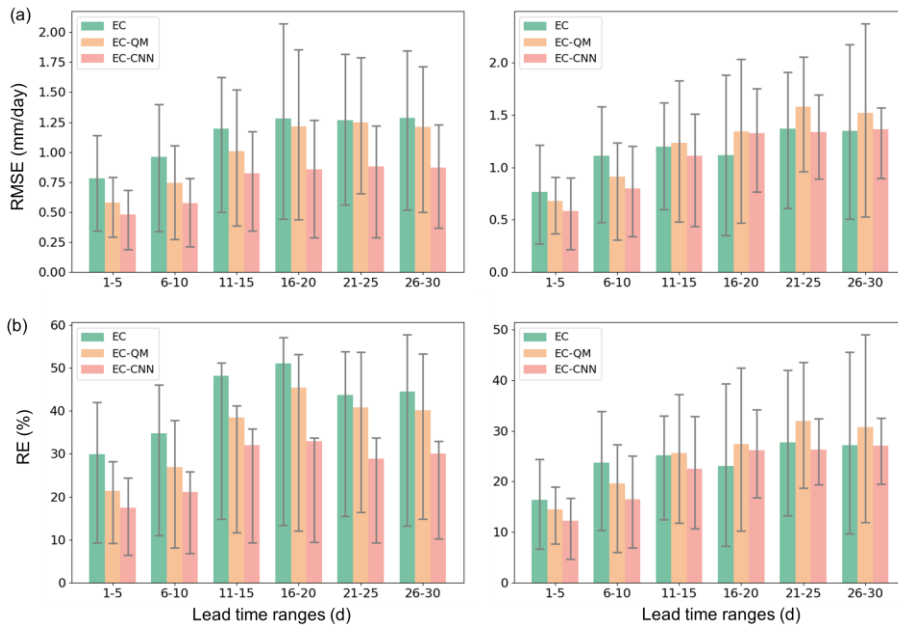
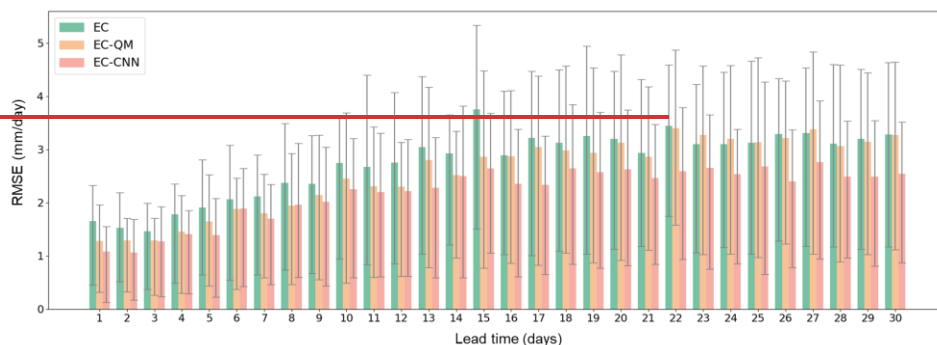
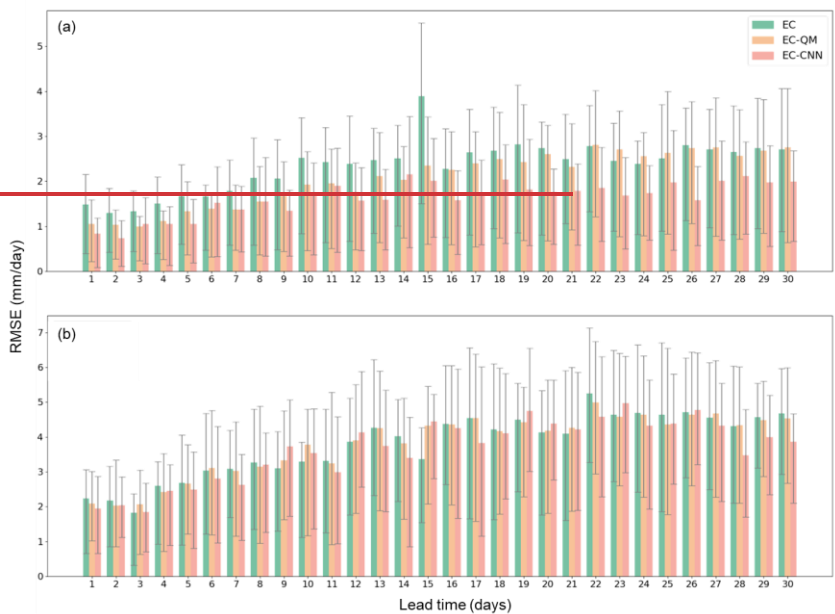



Figure 4. (a) RMSE and (b) RE of all rain events (left) and heavy rain events (right) for the ensemble means of EC, EC-QM and EC-CNN 5-day precipitation forecasts at lead times of 1-5 days, 6-10 days, 11-15 days, 16-20 days, 21-25 days, 26-30 days. Error bar represents the 25th-75th percentile interval.

In particular, the ~~forecast accuracy~~ RMSE ~~of EC-CNN forecasts are~~is ~~reduced~~ improved by EC-CNN forecasts ~~more significantly~~to a relatively steady extent at all~~for longer~~ lead times, as the RMSE is reduced by 33%-34%~~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~~ for different lead time ranges~~, while~~. On the other hand, the RMSE improvements of ~~RMSE of~~ EC-QM forecasts decrease rapidly with the increase in lead times. For example, the EC-QM reduces the RMSE of EC forecasts ~~sees a~~ by 24% for the first 10 days, by 11% for the middle 10 days, and by 4% for the last 10 days.~~relatively steady reduction over all lead times~~. In addition, ~~the~~ EC-CNN forecasts exhibit narrower 25th-75th percentile intervals of RMSE and RE across different ~~initial forecast~~ initialized 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 initialized dates. These results preliminarily demonstrate the superiority of proposed CNN method to the raw EC forecasts~~MWF model~~ and the EC-~~QM~~ forecasts.

Right panels in ~~To investigate the performance of EC-CNN forecasts in extreme precipitations, we classify the daily precipitation less than and greater than the 90th percentile of all historic daily precipitation during 2002-2019 as light rain and heavy rain, respectively.~~ Figure ~~4~~5 display~~s~~ the variations of RMSE for ~~light rain and~~ heavy rain events averaged over SR at lead times of 1-30 days, with the error bars representing the 25th-75th percentile interval.

Generally, the RMSE of EC forecasts increases with the increasing lead times for both light and heavy rains. ~~For light rain, 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 events, ~~EC-~~QM shows no improvements as compared to the ~~raw~~ EC forecasts averaged over all lead times, with the RMSE increasing ~~from 3.8 mm/day to 3.9 mm/day~~by ~5%, suggesting that QM has a limited ability to improve the forecast skills for extreme precipitation events in the SR. The ~~EC-~~CNN generally shows slight improvements (~~~2~~6%) in RMSE as compared to raw forecasts, and the RMSE of EC-CNN forecasts is smaller than that of EC-QM for ~~most of the~~all lead times, suggesting CNN exhibits advantages over QM for extreme events. This is particularly the case for lead times of ~~2~~11-~~30~~10 days, where the RMSE of EC-CNN (EC-QM) forecasts is ~~0.16~~26% (14%) ~~mm/day~~lower than that of ~~EC forecasts~~EC-QM ~~and is 0.12 mm/day lower than that of the raw forecasts~~. Overall, these results imply that, for ~~extreme~~heavy precipitation ~~rainfall~~ 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~~—~~. The RE shows a similar trend to RMSE, as the RE of EC, EC-QM, and EC-CNN are 22%, 25%, and 24% averaged over all lead times. The EC-CNN effectively reduces the bias of heavy rain events at all lead times except for the 16-20 day range.

395   To investigate the spatial characteristics of precipitation forecasts, Figure 5 presents the spatial distribution on the RMSE of EC and EC-CNN forecast for lead times of 1-10, 11-20 and 21-30 days. It is clear that the EC-CNN improves the forecast skill of the raw ECMWF forecasts over the majority of the SR for all lead times. For example, the RMSE is reduced from 3-5 mm/day for EC forecasts to 1-2 mm/day for EC-CNN forecasts at the northern sub-basin for all lead times. Similar improvements can also be seen around the southern part of the basin, for example the southernmost part of SR sees RMSE

400   over 10 mm/day for EC forecasts at lead times of 11-30 days but reduces to 6-7 mm/days for EC-CNN forecasts. In addition, by comparing Fig. R1b and Fig. R1c it can be seen that the EC-CNN shows larger improvements than EC-QM across the SR for all lead times. The above results indicate that EC-CNN not only improves the raw forecasts temporally, but also enhances their spatial accuracy across various regions of the SR. This basin-wise improvement allows for more reliable predictions across diverse hydrological zones within the SR, which could further benefit the hydrologic modelling. Figure 6 describes the

405   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 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.

410

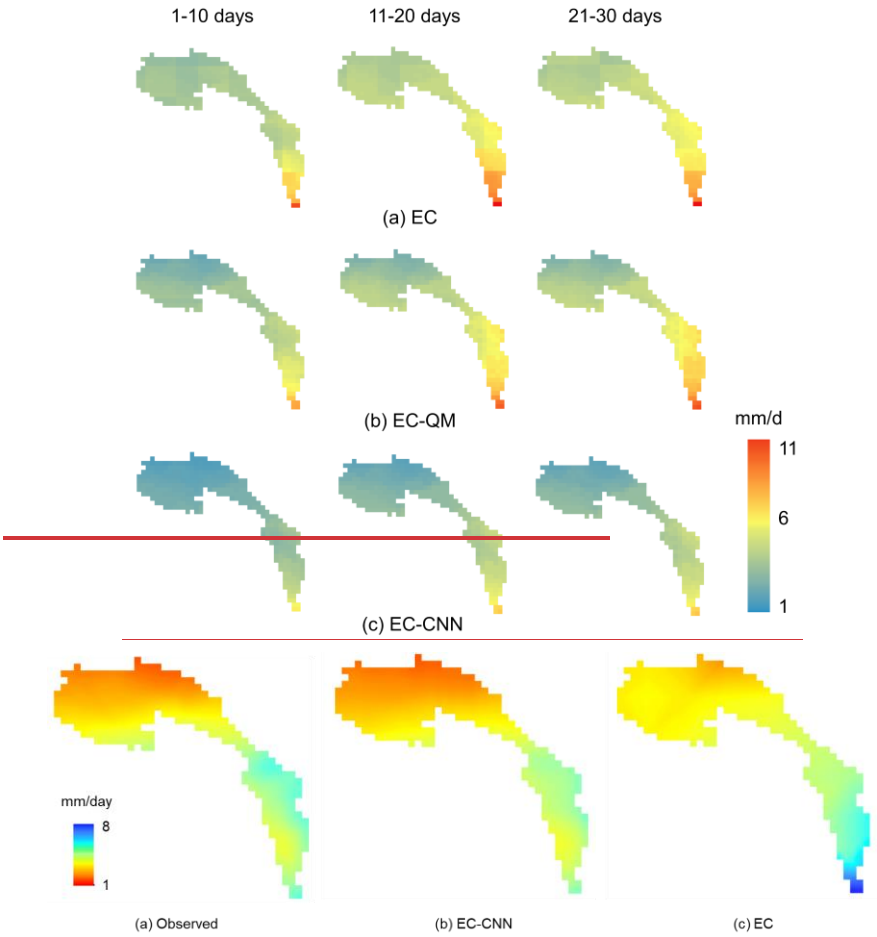



Figure 65. The spatial distribution of RMSE for the ensemble means of (a) EC forecasts, (b) EC-QM, and (c) EC-CNN forecasts averaged over lead times of 1-10 days, 11-20 days, and 21-30 days during the test period. The spatial distribution of (a) observed

precipitation, (b) EC-CNN forecasts, and (c) EC raw forecasts averaged over all lead times and all initial forecast dates during the test period.

Figure 6 presents cumulative distribution functions (CDFs) of EC, EC-QM and EC-CNN forecasted precipitation averaged over the SR across all lead times. Notably, the EC-QM forecast consistently aligns well with the observed CDF across all lead times, which reflects the designed purpose of QM to match the empirical distribution through quantile mapping. This result demonstrates EC-QM is good at correcting the raw EC forecast to follow the observed distribution closely, despite an overall large RMSE and RE compared to EC-CNN.

The EC-CNN forecast shows improvement over the EC forecasts and EC-QM forecasts by better approximating the observed CDF for the first 15 days. However, as compared to EC and EC-QM, the EC-CNN begins to deviate more significantly from the observed CDF as lead times increase. Specifically, the EC-CNN forecast appears to concentrate around medium precipitation values for 16-30 days, which underestimates the frequency of both lighter and heavier precipitation events. This pattern suggests that, while EC-CNN improves the overall accuracy of light and heavy rains of the raw EC forecasts at all lead times (Figure 4), the distributional accuracy, particularly for extreme precipitation events, may be compromised over extended lead times. A discussion on the possible cause of biases in CDF for different forecasts is provided in Section 5.4.

Figure 6. **The cumulative distribution function (CDF) of areal-averaged precipitation for observed precipitation, the ensemble means of EC, EC-QM, and EC-CNN forecasts at different lead times during the test period.**

Figure 7 presents the CRPS of EC, EC-QM and EC-CNN forecasted precipitation averaged over the SR across the lead times. CRPS evaluates how close the ensemble forecast distribution is to the observed value, and a value close to zero means a better ensemble forecast. As can be seen in the figure, the CRPS for EC is around 1.1 mm/day at lead times of 1-10 days and increases to around 1.4 mm/day at lead times of 11-30 days. The EC-QM reduces the CRPS by an average of around 0.1 mm/day at all lead times, indicating an improvement in the probabilistic calibration and sharpness of the ensemble forecasts. The EC-CNN further reduces the CRPS for most of the lead times as compared to the EC-QM, especially for the first 5 days where the CRPS is 0.4 mm/day lower than EC forecasts and 0.2 mm/day lower than EC-QM forecasts. This shows that the EC-CNN has an enhanced capability of representing the range of possible outcomes and improving the overall reliability in probabilistic forecasting. Such an advantage also offers better decision-making insights under uncertainty, which is favorable for risk management and planning across various time horizons.

440

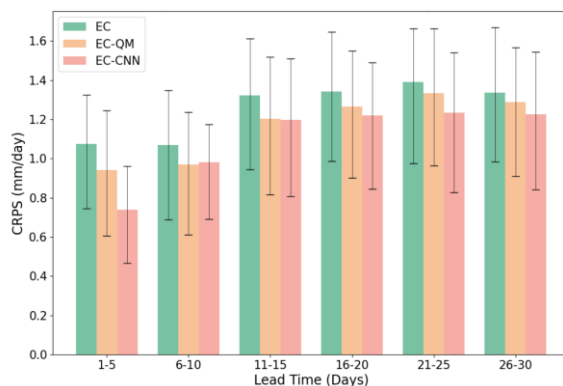



Figure 7. The CRPS of areal-averaged precipitation for the EC, EC-QM, and EC-CNN ensemble forecasts at different lead times during the test period. Error bar represents the 25th-75th percentile interval.

4.3 Evaluation of sub-seasonal streamflow forecasts

445   Figure 7 8 shows the ~~relative error~~RE, REF, RMSE and NSE ~~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~~, EC-QM ~~forecasts~~, and EC-CNN precipitation forecasts, respectively.

    Results indicate that, for the XAJ-LSTM hybrid model, the accuracy metrics of streamflow forecasts decrease ~~with~~as the lead times increase ~~in the lead times~~ for all precipitation forecasts. ~~For example, the relative error of streamflow forecasts driven~~

450 ~~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 example, the RE (REF) of streamflow forecasts driven by EC precipitation increases from around 11.6% (14%) for 1-10 days to 24% (23%) and 28% (27%) for 11-20 days and 21-30 days, respectively. The EC-QM (EC-CNN)

455 forecasts reduce the RE of the EC forecasts by approximately 6.9% (16.4%), 13.8% (32.5%), and 12.5% (26.8%), respectively, and reduce the REF of the EC forecasts by approximately 14.3% (28.6%), 10.4% (35.7%), and 13.3% (25.6%), respectively. It is noted that the improvements brought by EC-QM and EC-CNN are larger in lead times of 11-20 days than in 1-10 days, which was also observed by Zhang et al. (2023), Lyu et al. (2023), and Li et al. (2024) in specific cases.

~~For the standalone XAJ model, the relative error of streamflow forecasts driven by EC raw forecasts increases from around~~

460 ~~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. For the standalone XAJ model, the RE (REF) of streamflow forecasts driven by EC raw forecasts increases from around 18% (21%) for 1-10 days to 31% (33%) and 34% (31%) for 11-20 days and 21-30 days, respectively. The EC-QM (EC-CNN) forecasts reduce the RE of the EC forecasts by approximately 5.6% (9.4%), 8.1% (18.4%), and 0.6% (12.4%) for the 1-10, 11-20, and 21-30 day lead times, respectively. For the relative error of maximum daily flow, the EC-QM (EC-CNN) forecasts reduce it by approximately 9.4% (26.7%), 11.6% (24.8%), and 0.9% (5.8%) for each corresponding lead time range.

The above results indicate that (1) the streamflow and flood biases are smaller for XAJ-LSTM than for XAJ for all lead times, and (2)

Comparing the XAJ-LSTM model and the XAJ model shows that improvements in EC-CNN and EC-QM-based precipitation forecast lead to significantly greater enhancementsenhance the in streamflow forecast accuracy more effectivelys for the XAJ-LSTM model than for the XAJ model. Notably, the EC-CNN (EC-QM) driven XAJ-LSTM streamflow forecast sees 1320%-2231% (56%-1112%) less RMSE than that driven by raw EC forecasts (EC-QM forecasts) over different lead time periods. On the other hand, the EC-CNN (EC-QM) driven XAJ streamflow forecasts only see 7%-11% (1%-7%) less RMSE generally show no improvements (and could be even worse) with respect to the RMSE compared to that driven by EC and EC-QM, reflecting a much less improvement compared to those driven by XAJ-LSTM. A similar trend can also be observed for NSE, as XAJ-LSTM shows more improvement than XAJ when EC precipitation forecasts are replaced by EC-CNN and EC-QM precipitation forecasts. 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 isare also to a large extent the hydrologic model.

It is also noted that, despite the NSE values improve with the EC-CNN precipitation forecasts, they are mostly negative for both hydrologic models. This suggests further improvements may be required to achieve a more accurate hydrologic forecast at sub-seasonal scales. A discussion on this aspect is provided in Section 5.4.

带格式的: 居中

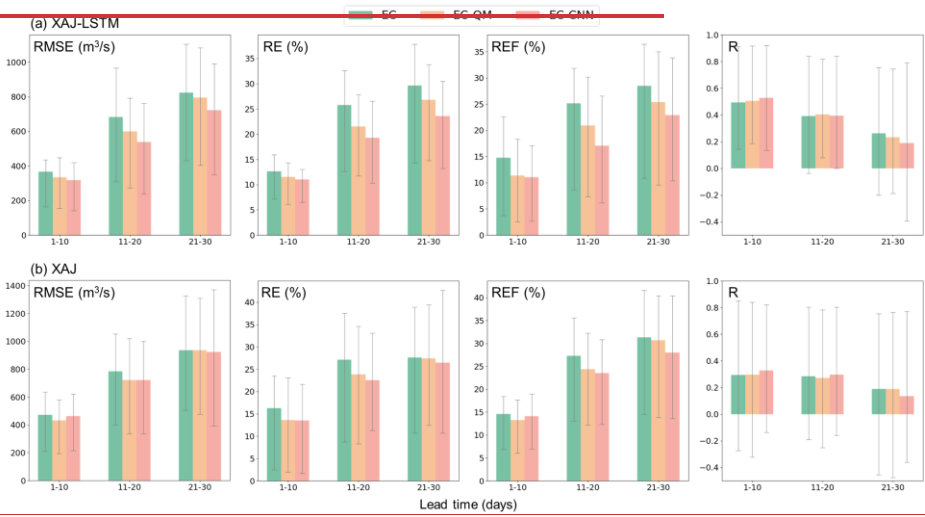
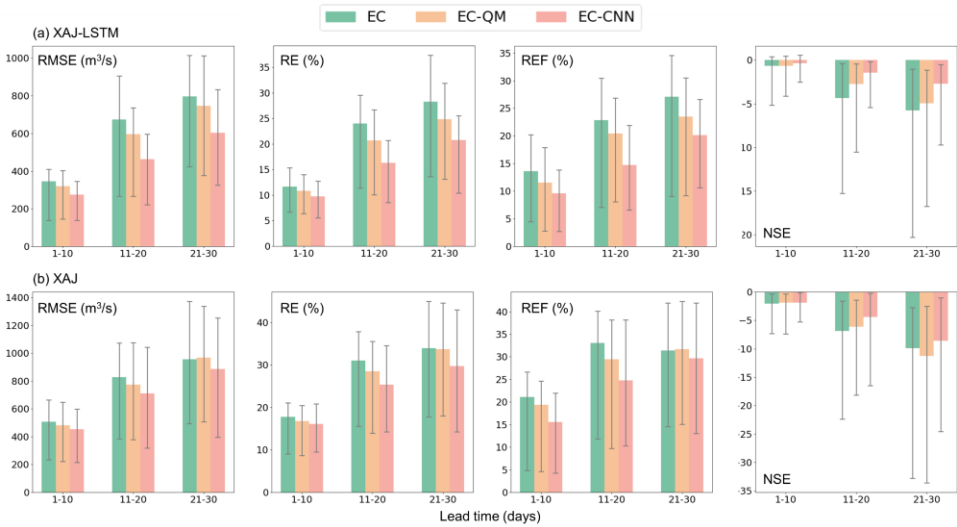



Figure ~~7~~8. The ~~root mean square error (RMSE), relative error of streamflow (RE),~~RMSE, RE, REF, and ~~relative error of maximum daily flow (REF) and NSE~~correlation coefficient (R) **for the (a) XAJ-LSTM and (b) XAJ streamflow forecasts driven by** the ensemble means of ~~EC forecasts~~, EC-QM ~~forecasts,~~ **and EC-CNN forecasts for lead times of 1-10 days, 11-20 days, and 21-30 days. Error bar represents the 25th-75th percentile interval.**

Figure ~~8~~9 presents ~~several~~ examples of XAJ-LSTM streamflow ensemble forecasts ~~initialized~~ issued on different dates, with the RE of total forecast flow presented in each subplot. 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 across different dates. For example, the CNN model ~~increases~~ decreases (in~~de~~creases) the EC precipitation and hence the forecast streamflow ~~initialized~~ issued on ~~May~~ June 0~~6~~2, 201~~7~~7 (August 0~~8~~1, 201~~9~~7), which improves the forecast skills in both cases. The EC-CNN reduces the relative error from 22.1% of raw EC forecasts to -2.5% for the 30-day streamflow forecast issued on June 06, 2017, and reduces the relative error from -21.5% of raw EC forecasts to -9.2% for the 30-day streamflow forecast issued on August 08, 2019. On the other hand, the QM reduces the precipitation and hence the streamflow forecasts for both dates, which ~~reduce~~ improves the ~~forecast~~ relative error to 11.0% on ~~August~~ June 2017 but ~~degrades~~ increases the relative error to -27.3%~~the accuracy~~ on ~~May~~August 2019. Similarly, for the forecast ~~issued~~nitialized on Aug 11, 2017, ~~the~~ EC-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~~ EC-QM ~~model~~ consistently predicts lower precipitation and hence streamflow compared to ~~raw~~ EC forecasts ~~for~~ at all lead times, worsening the underestimation in early September.

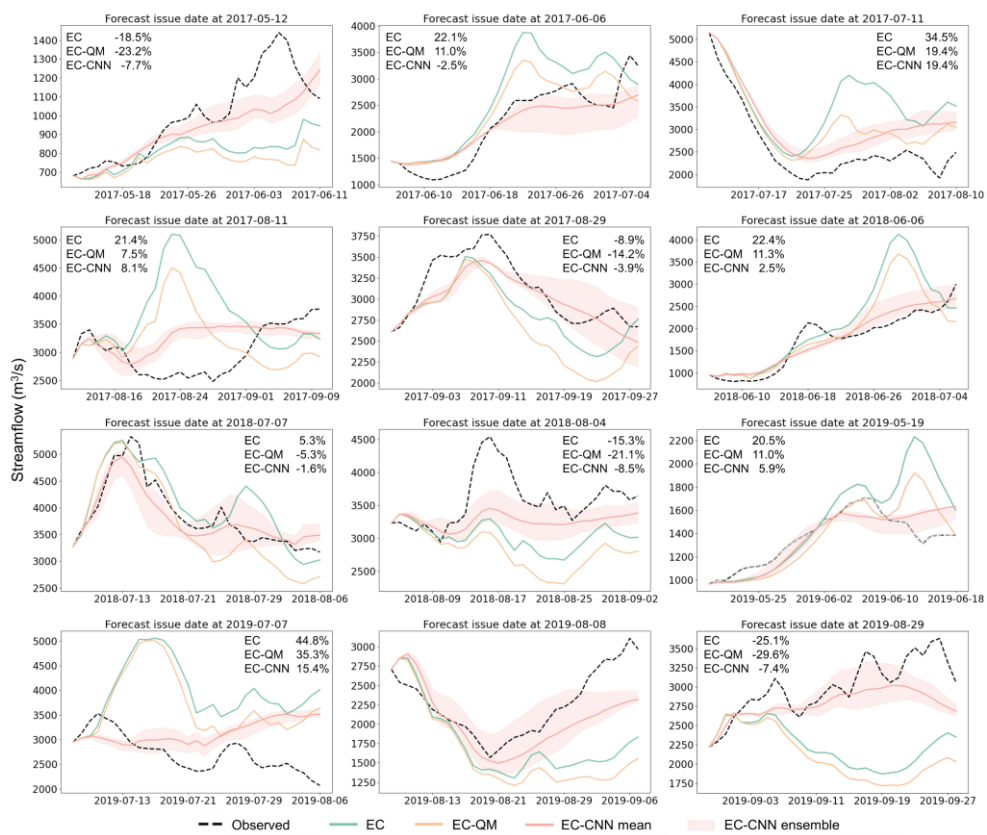
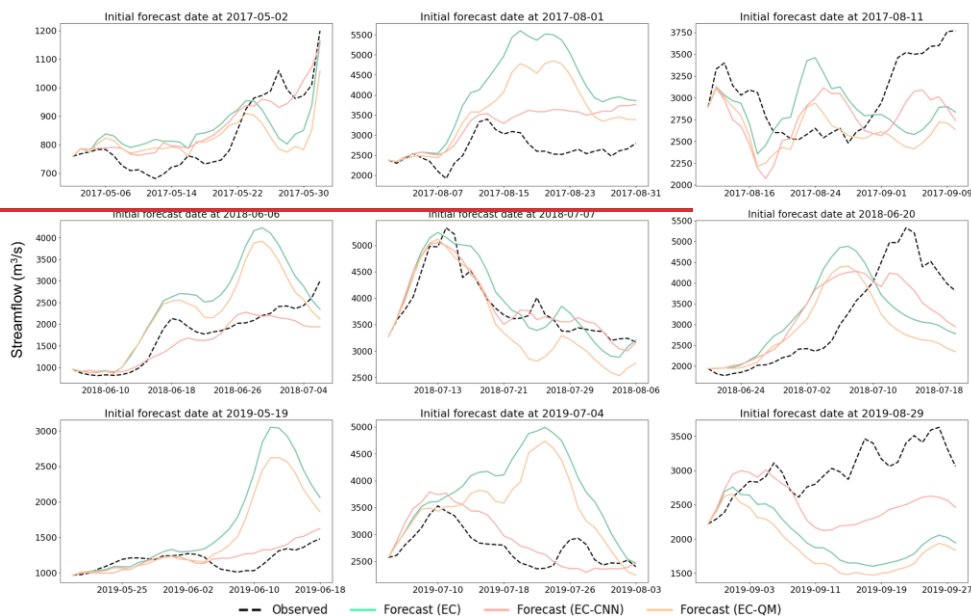



Figure ~~8~~9. Examples of sub-seasonal XAJ-LSTM streamflow forecasts for a lead time of 30 days driven by ~~EC forecasts~~EC, EC-QM, ~~forecasts and ensemble and~~ EC-CNN precipitation forecasts. Shaded areas represent the 25th-75th percentile. The relative error of total forecast flow is shown in each subplot.

Figure 10 presents the CRPS of EC, EC-QM and EC-CNN driven XAJ-LSTM streamflow forecasts at Shigu across the lead times. The CRPS for EC is around 120 m^3/s at lead times of 1-5 days and increases rapidly to around 300 m^3/s at lead times of 6-10 days, and further to 500 m^3/s at lead times of 21-30 days. The EC-QM reduces the CRPS by an average of around 65 m^3/s at all lead times, and the EC-CNN further reduces the CRPS for most of the lead times as compared to the EC-QM, especially for the lead times of 6-20 days where the CRPS is about 60 m^3/s lower than EC forecasts and 25 m^3/s lower than EC-QM forecasts. However, for lead times of 26-30 days, the CRPS of EC-CNN is slightly larger than that of EC-QM, indicating the advantage of EC-CNN in ensemble forecasting is not evident for extended forecast lead times. Nevertheless, the EC-CNN improves the overall reliability in probabilistic streamflow forecasting for all lead times as compared to EC and for most lead times as compared to EC-QM, which can benefit the downstream water resources management under uncertainty.

520

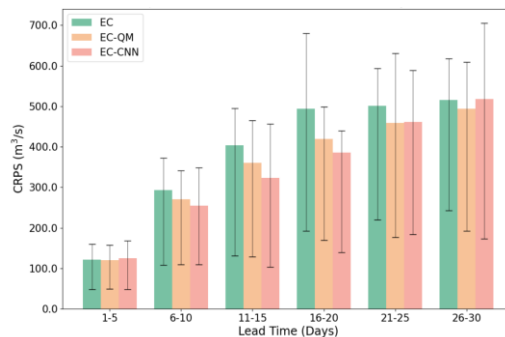



Figure 10. The CRPS of streamflow forecasts driven by the EC, EC-QM, and EC-CNN ensemble forecasts at different lead times during the test period. Error bar represents the 25th-75th percentile interval.

5 Discussion

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

525 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, an ensemble of enhanced CNN post-processing models with ResNet blocks and a weighted loss function specialized on extreme events is is established to investigate its potential to overcome these limitations by establishing multi-dimensional relationships between atmospheric circulation predictors and local precipitation.

530 We compare the CNN model with the commonly used quantile mapping (QM) bias correction method. It is noted that, for several lead time ranges, EC-QM forecasts show no improvements in RMSEs compared to those of raw EC forecasts. This was also observed by some of the recent studies using QM for statistical downscaling precipitation. For example, Li et al. (2023), Huang et al. (2022), and Mao et al. (2015) show that, while QM is generally effective in adjusting model bias towards observations, it does not always lead to improvements of the forecast accuracy. One plausible reason could be the limited

535 applicability of the relatively simple QM method. Specifically, QM primarily adjusts the distribution of forecasted values towards the distribution of historical observations, with no account of the atmospheric conditions associated with those forecasts. However, physics-based numerical weather predictions involve complex and nonlinear errors that QM may not be able to fully correct.

On the other hand, The result reveals that the CNN model improves the RMSE and RE of forecast precipitation at all lead times,
540 and outperforms QM in improving sub-seasonal precipitation forecast skills, both in terms of capturing the general trends, and predicting extreme precipitation events, and approximating the probabilistic distribution at sub-seasonal scales. 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 squared error and the threat score. 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 8 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

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 ensemble-mean streamflow simulations forecast driven by observed precipitation and temperature (i.e., the hydrologic modelling error, E_m), the RMSE of the ensemble-mean streamflow forecasts driven by forecast precipitation and observed temperature (i.e., the precipitation forecast error, E_p), and the RMSE of the ensemble-mean 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 11 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_p 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 ratio of 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 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 than 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.

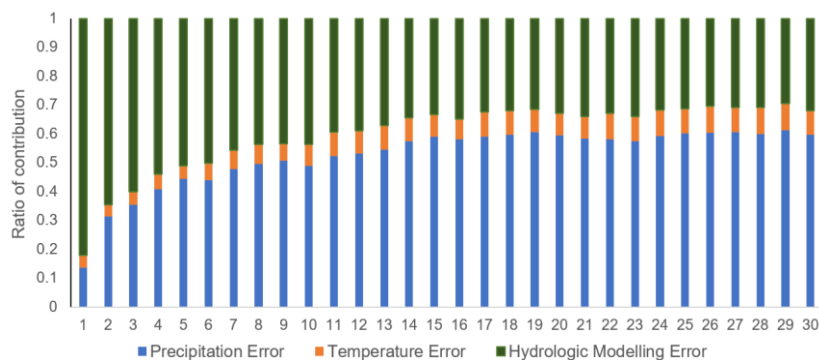



Figure 9̶11. Contribution of precipitation forecast errors, temperature forecast errors, and hydrologic modelling errors to their combined error.

5.4 Limitations of this study

One limitation of the proposed EC-CNN model is its reduced accuracy in capturing the CDF over lead times extending beyond 15 days. While EC-CNN outperforms in terms of RMSE and RE for both light and heavy rainfall events (Figure 4), its CDF deviates from observed patterns for these longer lead times. In contrast, EC and EC-QM tend to align more closely with observed CDFs at these lead times. We attribute this to the following possible reason.

Over extended lead times (>15 days), extreme precipitation values of EC and EC-QM forecasts tend to be more incorrectly assigned to specific times, leading to, for instance, storm-level precipitation predictions on dry days or nearly zero precipitation predictions on storm days. While these misplaced extremes inflate daily errors, they still allow the overall CDF of the forecast to retain both high and low ends of the precipitation distribution. The EC-CNN model, however, is designed to minimize daily errors by adjusting outliers and bringing exaggerated values closer to moderate levels, thereby reducing large forecast errors. For example, when EC forecasts a high precipitation event on a dry day, EC-CNN mitigates this by lowering the extreme to a more typical value. While this adjustment helps decrease RMSE and RE, it inadvertently compresses the distribution toward the center by reducing the frequency of both extreme high and low precipitation values. This approach limits the model ability to capture the full distribution, resulting in a CDF that is overly concentrated around moderate values.

Another limitation is the relatively low Nash-Sutcliffe Efficiency (NSE) values in streamflow forecasts, especially over extended lead times. Despite improvements in streamflow forecast accuracy due to EC-CNN and the XAJ-LSTM, NSE values are found predominantly negative. This can be primarily due to the inaccuracies in the precipitation forecasts that drive these hydrologic models, as high NSE values require precipitation inputs to be accurate both spatially and temporally. Improving the hydrologic model alone is unlikely to address this issue substantially, as it depends heavily on input accuracy (Section 5.3).

带格式的: 正文, 定义网格后不调整右缩进, 段落间距段前: 6 磅, 不调整西文与中文之间的空格, 不调整中文和数字之间的空格

Nevertheless, with our proposed coupled EC-CNN and XAJ-LSTM framework, the overall relative error of forecast flow can be reduced to ~10% for the next 10 days and ~20% for the next 30 days (Figure 7), which can provide implications for water management and disaster prevention. Future improvements in NSE could focus on refining precipitation forecasting with more advanced AI models. Additionally, LSTM model training on multiple basins and fine-tuning on specific target basins could further enhance the streamflow accuracy over extended lead times (Kratzert et al., 2019).

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 of up to 30 days. The framework couples (1) an ensemble of enhanced CNN models~~architecture~~ with ResNet blocks for ~~spatiotemporal~~ statistically downscaling ~~of~~ ECMWF raw precipitation ensemble forecasts to (2) a hybrid hydrologic model integrating the conceptual XAJ model and LSTM for streamflow forecasting. The CNN models incorporate~~s~~ 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, and ensemble distribution. 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~~ around ~~10~~30% lower than ~~both two forecasts~~both forecasts. This improvement is also noted in extreme precipitation events, as demonstrated by ~~a approximately~~ ~~~6~~2% and ~~~10~~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~~16% ~~-33%~~ ~~24~~% compared to raw precipitation forecasts, particularly for longer lead times. This outperforms QM, which reduces the relative error of streamflow by ~~9~~7%-14%~~%~~ ~~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, with the average relative error of maximum daily flow reduced by 26%-36%. 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
650 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 is 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
655 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

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

Conflict of Interest

The authors declare that they have no conflict of interest.

665 **Acknowledgements**

Ningpeng Dong received funding for this study through the National Key Research and Development Program of China (2023YFC3081000; 2021YFC3000202) and the National Natural Science Foundation of China (42401053). This research was funded by the National Key Research and Development Project (No. 2021YFC3000202; No. 2023YFC3081000). Jianhui Wei is financially supported financially by the German Federal Ministry of Science of Education (BMBF) through funding of the
670 KARE_II project (01LR2006D1). This study is also funded by China Power Construction Corporation Technology Project (DJ-HXGG-2021-04) and Key R&D Plan Project in Yunnan Province (202203AA080010).

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.

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges, Hydrol. Sci. J., 65, 712–725, https://doi.org/10.1080/02626667.2019.1683182, 2020.

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.

Bremnes, J. B.: Ensemble postprocessing using quantile function regression based on neural networks and Bernstein polynomials, Mon. Weather Rev., 148, 403–414, 2020.

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.

705 Cannon, A. J., Sobie, S. R., & 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, 2015.

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.

Cho, D., C. Yoo, J. Im, and D. H. Cha, 2020: Comparative assessment of various machine learning-based bias correction
710 methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. Earth and Space Science, 7, e2019EA000740, https://doi.org/10.1029/2019EA000740.

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.

Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal
715 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,
720 https://doi.org/10.1016/j.scitotenv.2022.160681, 2023.

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.

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,
725 https://doi.org/10.1029/2021WR031787, 2022.

Dong, N., Yang, M., Wei, J., Arnault, J., Laux, P., Xu, S., Wang, H., Yu, Z., and Kunstmann, H.: Toward improved parameterizations of reservoir operation in ungauged basins: A synergistic framework coupling satellite remote sensing, hydrologic modeling, and conceptual operation schemes, Water Resour. Res., 59(3), e2022WR033026, https://doi.org/10.1029/2022WR033026, 2023.

730 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.

Ferranti, L., Corti, S., & Janousek, M.: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector, Q. J. R. Meteorol. Soc., 144, 317–326, 2018.

Gao, S., Huang, D., Du, N., Ren, C., and Yu, H. WRF ensemble dynamical downscaling of precipitation over China using
735 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.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Downscaling RCM precipitation to the station scale using statistical transformations – A comparison of methods. Hydrology and Earth System Sciences, 16(9), 3383-3390. doi:10.5194/hess-16-3383-2012

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

Ham, Y.-G., Kim, J.-H., and Luo, J.-J. (2019). Deep learning for multi-year ENSO forecasts. Nature, 573, 568–572.

Han, L., Chen, M., Chen, K., Chen, H., Zhang, Y., Lu, B., Song, L., and Qin, R. (2021). A deep learning method for bias correction of ECMWF 24–240 h forecasts. Adv. Atmos. Sci., 38, 1444–1459.

Hao, H., Dong, N., Yang, M., Wei, J., Zhang, X., Xu, S., Yan, D., Ren, L., Leng, G., Chen, L., and Zhou, X.: The changing hydrology of an irrigated and dammed Yangtze River: Streamflow, extremes, and lake hydrodynamics, Water Resour. Res., 60(10), e2024WR037841, https://doi.org/10.1029/2024WR037841, 2024.

Horat, N. and Lerch, S., 2024. Deep Learning for Postprocessing Global Probabilistic Forecasts on Subseasonal Time Scales. Monthly Weather Review, 152(3), pp.667-687.

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.

Huang, Z., Zhao, T., Xu, W., Cai, H., Wang, J., Zhang, Y., Liu, Z., Tian, Y., Yan, D., & Chen, X. (2022). A Seven-Parameter Bernoulli-Gamma-Gaussian Model to Calibrate Subseasonal to Seasonal Precipitation Forecasts. Journal of Hydrology, 610, 127896. https://doi.org/10.1016/j.jhydrol.2022.127896

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.

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, 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.

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.

Jin, W., Zhang, W., Hu, J., Weng, B., Huang, T., and Chen, J. (2022). Using the residual network module to correct the sub-seasonal high temperature forecast. Front. Earth Sci., 9, 760766.

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.

Kim, H., Ham, Y.-G., Joo, Y.-S., and Son, S.-W. (2021). Deep learning for bias correction of MJO prediction. Nat. Commun., 12, 3087.

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.

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.

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.

Lagerquist, R., A. McGovern, and D. J. Gagne II, 2019: Deep learning for spatially explicit prediction of synoptic-scale Fronts. Wea. Forecasting, 34, 1137−1160, https://doi.org/10.1175/WAF-D-18-0183.1.

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, L., Yun, Z., Liu, Y., Wang, Y., Zhao, W., Kang, Y., & Gao, R. (2024). Improving Categorical and Continuous Accuracy of Precipitation Forecasts by Integrating Empirical Quantile Mapping and Bernoulli-Gamma-Gaussian Distribution. Atmospheric Research, 298, 107133. https://doi.org/10.1016/j.atmosres.2024.107133

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.

800   Li, X., Wu, H., Nanding, N., Chen, S., Hu, Y., & Li, L. (2023). Statistical Bias Correction of Precipitation Forecasts Based on Quantile Mapping on the Sub-Seasonal to Seasonal Scale. Remote Sensing, 15(7), 1743. https://doi.org/10.3390/rs15071743

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
805   precipitation prediction over China, Environ. Res. Lett., 17, 124025, https://doi.org/10.1088/1748-9326/acaea0, 2022.

Ling, F., Luo, J.-J., Li, Y., Tang, T., Bai, L., Ouyang, W., and Yamagata, T. (2022). Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole. Nat. Commun., 13, 7681.

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.

810   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.

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,
815   https://doi.org/10.1029/2023GL106245, 2023.

Manzanas, R., Lucero, A., Weisheimer, A., & Gutiérrez, J. M.: Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts?, Clim. Dyn., 50, 1161–1176, 2018.

Mao, G., Vogl, S., Laux, P., Wagner, S., & Kunstmann, H. (2015). Stochastic Bias Correction of Dynamically Downscaled Precipitation Fields for Germany Through Copula-Based Integration of Gridded Observation Data. Hydrology and Earth
820   System Sciences, 19(4), 1787-1806. https://doi.org/10.5194/hess-19-1787-2015

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.

825   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.

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.

830   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 Southwest China, Environ. Res. Lett., https://doi.org/ 10.1088/1748-9326/ad5370, 2024.

835 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 precipitation product, Remote Sens., 14, 1964, https://doi.org/10.3390/rs14091964, 2022.

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,
840 2020.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian model averaging to calibrate forecast 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.

845 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.

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-
850 00823.1, 2014.

Scheuerer, M., and Hamill, T. M.: Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted 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.

855 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.

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.

860 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.

865 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.

870 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.

875 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, 880 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, 885 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.

Weyn, J. A., Durran, D. R., Caruana, R., and Cresswell-Clay, N. (2021). Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. J. Adv. Modeling Earth Syst., 13, e2021MS002502.

890 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.

Xie, J., Hsu, P.-C., Hu, Y., Ye, M., and Yu, J. (2023). Skillful extended-range forecast of rainfall and extreme events in East China based on deep learning. Weather Forecast., 38, 467–486.

895 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, T., Liang, Z., Li, W., Wang, J., Hu, Y., & Li, B. (2023). Statistical Post-Processing of Precipitation Forecasts Using Circulation Classifications and Spatiotemporal Deep Neural Networks. Hydrology and Earth System Sciences, 27(10), 1945-1962. https://doi.org/10.5194/hess-27-1945-2023

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.

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.

930 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-
935 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-
940 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
945 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.

950 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.
955 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.

960 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.

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 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.

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.

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.

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.

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.

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.

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.

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 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 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 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.

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 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.

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.

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.

1095 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,
1100 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,
1105 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.

1110 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.

1115 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
1120 in the East Asian summer monsoon modeling, Theor. Appl. Climatol., 140, 375–387, https://doi.org/10.1007/s00704-020-03093-8, 2020.