



Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning

Junjiang Liu¹, Xing Yuan^{1,2}, Junhan Zeng¹, Yang Jiao¹, Yong Li³, Lihua Zhong³, and Ling Yao⁴

¹School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China

²Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

³Guangxi Meteorological Disaster Prevention Center, Nanning 530022, China

⁴Guangxi Guiguan Electric Power Co., Ltd., Nanning 530029, China

Correspondence: Xing Yuan (xyuan@nuist.edu.cn)

Received: 23 July 2021 – Discussion started: 30 July 2021

Revised: 21 November 2021 – Accepted: 6 December 2021 – Published: 18 January 2022

Abstract. A popular way to forecast streamflow is to use bias-corrected meteorological forecasts to drive a calibrated hydrological model, but these hydrometeorological approaches suffer from deficiencies over small catchments due to uncertainty in meteorological forecasts and errors from hydrological models, especially over catchments that are regulated by dams and reservoirs. For a cascade reservoir catchment, the discharge from the upstream reservoir contributes to an important part of the streamflow over the downstream areas, which makes it tremendously hard to explore the added value of meteorological forecasts. Here, we integrate meteorological forecasts, land surface hydrological model simulations and machine learning to forecast hourly streamflow over the Yantan catchment, where the streamflow is influenced by both the upstream reservoir water release and the rainfall–runoff processes within the catchment. Evaluation of the hourly streamflow hindcasts during the rainy seasons of 2013–2017 shows that the hydrometeorological ensemble forecast approach reduces probabilistic and deterministic forecast errors by 6 % compared with the traditional ensemble streamflow prediction (ESP) approach during the first 7 d. The deterministic forecast error can be further reduced by 6 % in the first 72 h when combining the hydrometeorological forecasts with the long short-term memory (LSTM) deep learning method. However, the forecast skill for LSTM using only historical observations drops

sharply after the first 24 h. This study implies the potential of improving flood forecasts over a cascade reservoir catchment by integrating meteorological forecasts, hydrological modeling and machine learning.

1 Introduction

Floods are the most destructive events among natural disasters, causing huge amounts of damage to human society. Reservoirs are constructed to regulate river flows and have significantly reduced flood risks and damage (Ji et al., 2020). However, the number and intensity of extreme precipitation events are increasing in many areas as global warming continues, thereby amplifying the potential for flood hazards (Hao et al., 2013; Shao et al., 2016; Wei et al., 2018; Yuan et al., 2018a; Wang et al., 2019). Thus, accurate streamflow forecasts are needed to provide guidelines for reservoir operations (Robertson and Wang, 2013).

A common approach to streamflow forecasting is to use hydrological models; the first attempt at this kind streamflow forecasting can be traced back to the 1850s and involved simple regression-type approaches to predict discharge from observed precipitation (Mulvaney, 1851). Since then, model concepts have been further augmented by designing new data networks, addressing the heterogeneity of hydrologi-

cal processes, capturing the nonlinear characteristics of hydrologic system and parameterizing models (Hornberger and Boyer, 1995; Kirchner, 2006). With advancements in computer technology and high-resolution observation, a well-parameterized hydrological model can now simulate streamflow with high accuracy (Kollet et al., 2010; Ye et al., 2014; Graaf et al., 2015; Yuan et al., 2018b).

Streamflow simulations from hydrological models heavily rely on meteorological forcing inputs, especially precipitation, which can be measured at in situ gauges or retrieved from satellites and radars. However, for medium-range (2–15 d ahead) streamflow forecasts, precipitation forecasts are needed (Hopson and Webster, 2010). To improve the forecasts, ensemble techniques that can give a deterministic estimate as well as the estimate's uncertainty have become popular. Ensemble weather forecasting can be traced back to 1963 (Lorenz, 1963). Later, Leith (1974) transferred a deterministic forecast into an ensemble using the Monte Carlo method in order to describe the atmospheric uncertainty. In the 1990s, ensemble forecasting was developed into an integral part of numerical weather prediction that showed higher skill than the deterministic forecast, even with higher model resolution (Toth et al., 2001). Due to the rapid development of this technique, ensemble weather forecasting and climate predictions are applied to hydrological forecasting studies by combining them with hydrological models (Jasper et al., 2002; Balint et al., 2006; Jaun et al., 2008; Xu et al., 2015; Yuan et al., 2016; Zhu et al., 2019). Provided with an ensemble of streamflow forecasts and their forecast variability, a reservoir can maintain a reliable utility from natural streamflow better than that provided with a deterministic streamflow forecast only (Zhao et al., 2011). However, the streamflow prediction skill depends on whether the precipitation forecasts introduced into the hydrological model are skillful (Alfieri et al., 2013). When assessing the skill of this hydrometeorological forecasting approach, a benchmark is needed. Using ensembles of historical climatology data (Day, 1985) as meteorological forecast inputs, which is known as ensemble streamflow prediction (ESP), is often selected as the benchmark approach. Evaluations of hydrological forecasts have indicated that forecast skill has a close relationship with the catchment size, geographical location and resolution (Alfieri et al., 2013; Pappenberger et al., 2015); thus, there is a necessity to compare these forecasts with the ESP in order to establish the skill of the hydrometeorological forecasting approach.

Although physically based hydrological models are widely used, it is still hard to apply a hyper-resolution distributed model to streamflow forecasting due to its demand for observation data, its complex model structures, and the computational resource requirements for calibration and application (Wood et al., 2011; Kratzert et al., 2018; Yaseen et al., 2018). In cascade reservoir systems, there are two sources of streamflow: the rainfall within the interval basin and the upstream reservoir discharge. While the rainfall–runoff relationship is

well studied, it is challenging to reproduce the reservoir operating rules in a physical model (Gao et al., 2010; Zhang et al., 2016; Dang et al., 2020).

Machine learning methods can recognize patterns hidden in input data and can simulate or predict streamflow without explicit descriptions of the underlying physical processes (Kisi, 2007; Adnan et al., 2019). Neural networks are suitable for streamflow forecasting among machine learning models, and some of them can even outperform physically based hydrological models. For example, Humphrey et al. (2016) showed that their combined Bayesian artificial neural network (ANN) with the modèle du Génie Rural à 4 paramètres Journalier (GR4J) approach outperforms the GR4J model with respect to monthly streamflow forecasting. Kratzert et al. (2019) showed that an approach based on the long short-term memory (LSTM) technique outperforms a well-calibrated Sacramento Soil Moisture Accounting Model (SAC-SMA). Yang et al. (2020) used a geomorphology-based hydrological model (GBHM) combined with a traditional ANN model to simulate daily streamflow, which can provide enough physical evidence and can run with less observation data. Although neural network models are criticized with little physical evidence (Abrahart et al., 2012), their potential in hydrological forecasting is yet to be explored.

In this study, we combine machine learning with a hydrometeorological approach for hourly streamflow forecasting over a cascade reservoir catchment located in southwestern China. We use the meteorological hindcast data from the European Centre for Medium-Range Weather Forecasts (ECMWF) model that participated in the THORPEX (The Observing-system Research and Predictability Experiment) Interactive Grand Global Ensemble (TIGGE) project to drive a newly developed high-resolution land surface model, named “CSSPv2” (Conjunctive Surface-Subsurface Process, version 2; Yuan et al., 2018b), to provide runoff and streamflow forecasts, and we corrected the forecasts using the LSTM model. We aim to improve flood forecasting over the cascade reservoir catchment by integrating meteorological forecasts, hydrological modeling and machine learning. So we strive to (1) calibrate the hydrological model, (2) bias correct the meteorological forecasts, (3) evaluate the streamflow forecast skill and (4) test the combined physical–statistical approach.

2 Study area, data, model and method

2.1 Study area

The Yantan Hydropower Station is in the middle reaches of the Hongshui River in Dahua Yao Autonomous County, Guangxi Province. This station is the fifth level in the 10-level development of the Hongshuihe hydropower base in the Nanpanjiang River, connected with the upstream Longtan

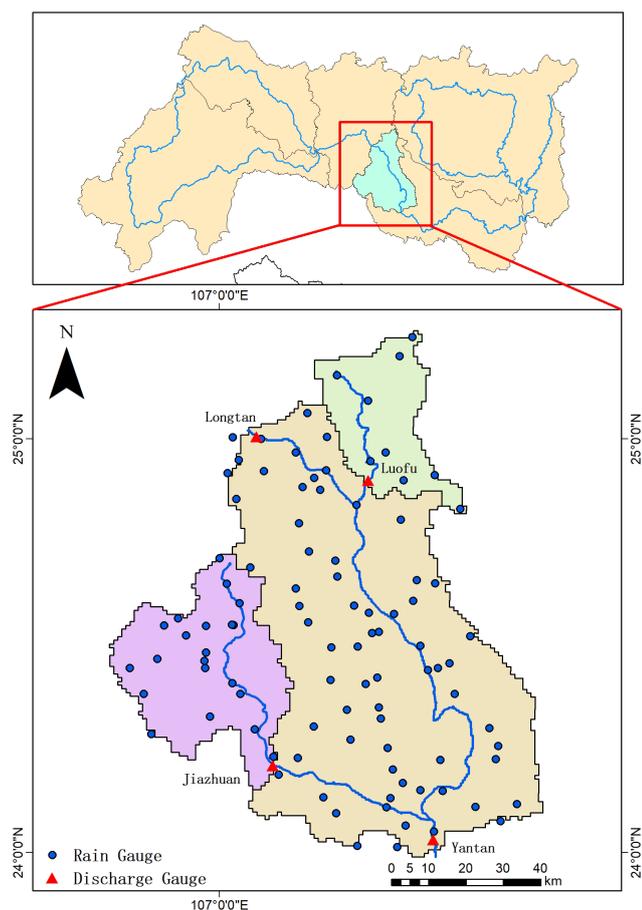


Figure 1. Locations of discharge gauges and rain gauges over the Yantan Basin.

Table 1. Information on hydrological gauges.

Gauge	Longitude (° E)	Latitude (° N)	Drainage area (km ²)
Longtan	107.09	25.00	–
Yantan	107.50	24.11	5950 (orange area in Fig. 1)
Luofu	107.36	24.90	800 (green area in Fig. 1)
Jiazhuan	107.12	24.21	2150 (purple area in Fig. 1)

Hydropower Station and the downstream Dahua Hydropower Station. The drainage area between the Longtan Hydropower Station and Yantan Hydropower Station is 8900 km². The annual mean streamflow at the Yantan hydrological gauge is 55.5×10^9 m³. The river passes through a karst mountain area, with a narrow valley, steep slope and scattered cultivated land, and the average slope is 0.036 %. Figure 1 shows the locations of four hydrological gauges, with detailed information listed in Table 1.

2.2 Data and method

2.2.1 Hydrometeorological observations

There are 97 meteorological observation stations within the catchment (Fig. 1). Here, observed hourly 2 m temperature, 10 m wind speed, relative humidity, accumulated precipitation and surface pressure data were interpolated onto a 5 km gridded observation dataset using the inverse distance weight method. The hourly surface downward solar radiation data from the China Meteorological Administration Land Data Assimilation System (CLDAS) were also interpolated onto a 5 km dataset using the bilinear interpolation method. The hourly surface downward thermal radiation was estimated by specific humidity, pressure and temperature. This dataset was used to drive the CSSPv2 land surface hydrological model.

The monthly runoff for each 5 km grid was estimated by disaggregating control streamflow station observations with the ratio of observed grid monthly precipitation and catchment mean precipitation. The gridded runoff was used to calibrate the CSSPv2 model at each grid (Yuan et al., 2016). The calibrated runoff parameters can be used to better represent the heterogeneity of the rainfall–runoff processes and make precise runoff simulations.

2.2.2 Ensemble meteorological hindcast data and ESP hindcasts

The TIGGE dataset consists of ensemble forecast data from 10 global numerical weather prediction centers starting from October 2006; the dataset has been made available for scientific research via data archive portals at ECMWF and the China Meteorological Administration (CMA). TIGGE has become the focal point for a range of research projects, including research on ensemble forecasting, predictability and the development of products to improve the prediction of severe weather (Bougeault et al., 2010). In this paper, TIGGE data from April to September during 2013–2017 from ECMWF were used as meteorological hindcast data. The 3-hourly meteorological hindcasts for a 7 d lead time from 51 ensemble members (including a control forecast) were interpolated to a 5 km resolution via bilinear interpolation. The forecast precipitation and temperature were corrected to match the observational means in order to remove the biases.

The ESP was accomplished by applying historical meteorological forcings (Day, 1985). In this paper, the meteorological forcings from the same date as the forecast start date to the next 9 d of each year (excluding the target year) were selected as the ESP forcings. Take 1 April 2013 as an example, the 7 d observation periods starting from 1 to 10 April (i.e., 1–7 April, 2–8 April, ..., 10–16 April) in the years 2014, 2015, 2016 and 2017 were selected as the forecast ensemble forcings of the issue date (1 April), resulting in a total of

40 ensemble members. The detailed information on the raw datasets is given in Table 2.

2.2.3 CSSPv2 streamflow hindcasts

The physical hydrological model used in this paper is the Conjunctive Surface-Subsurface Process model, version 2 (CSSPv2; Yuan et al., 2018b). The CSSPv2 model is a distributed, grid-based land surface hydrological model that was developed from the Common Land Model (Dai et al., 2003, 2004), but it has better representations of lateral surface and subsurface hydrological processes and their interactions. The routing model used here employs the kinetic wave equation as a covariance function, which is solved via a Newton algorithm. A main reason for adopting this covariance function is that it suits basins with mountainous terrain. The CSSPv2 model was successfully used to perform a high-resolution (3 km) land surface simulation over the Sanjiangyuan region, which is the headwater of major Chinese rivers (Ji and Yuan, 2018). In this paper, we calibrated the CSSPv2 model against monthly estimated runoff to simulate the natural hydrological processes using the shuffled complex evolution (SCE-UA) approach (Duan et al., 1994). The calibrated parameters include the maximum velocity of baseflow, the variable infiltration curve parameter, the fraction of maximum soil moisture where nonlinear baseflow occurs and the fraction of maximum velocity of baseflow where nonlinear baseflow begins. The hourly observed streamflow at the Yantan hydrological gauge was used to manually calibrate the CSSPv2 routing model, including the slope, river density, roughness, width and depth. The observed streamflow values at the Longtan hydrological gauge were added into the corresponding grid to provide upstream streamflow information. We used a high-resolution elevation database (hereafter referred to as DEM90) for sub-grid parameterization and then calculated the initial values of these river channel parameters. We first extracted the slope angle and the natural river flow path from DEM90 and then identified the accurate river network using a drainage area threshold of 0.18 km². River density and bed slope values for each 5 km grid were calculated as follows:

$$\text{rivden} = \sum l/A, \quad (1)$$

$$\text{bedslp} = \text{mean}(\tan(\beta)), \quad (2)$$

where rivden is the river density (km/km²), bedslp is the river channel bed slope (unitless), A is the area of a 5 km grid (km²), $\sum l$ is the total river channel length (m) within the grid and β is the slope angle (radian) for each river segment located in the grid.

Other river channel parameters were estimated using empirical formulas (Getirana et al., 2012; Luo et al., 2017) as follows:

$$W = 1.956 \times A_{\text{acc}}^{0.413}, \quad (3)$$

$$H = 0.245 \times A_{\text{acc}}^{0.342}, \quad (4)$$

$$n = 0.03 + (0.05 - 0.03) \frac{H_{\text{max}} - H}{H_{\text{max}} - H_{\text{min}}}, \quad (5)$$

where W , H and n are river width (m), depth (m) and roughness (unitless) for each 5 km grid; A_{acc} is the upstream drainage area (km²); and H_{max} and H_{min} refer to the maximum and minimum values of river depth calculated by Eq. (4).

Using a trial-and-error procedure, we calibrated these river channel parameters to match the simulated streamflow with observed hourly records at the Yantan hydrological gauge. The simulation results were evaluated by calculating the Nash–Sutcliffe efficiency (NSE) with corresponding observation data. The descriptions of the calibrated parameters and their ranges are given in Table 3.

After calibration, we drove the CSSPv2 model using 5 km regridded and bias-corrected TIGGE-ECMWF forecast forcing during 2013–2017 to provide a set of 7 d hindcasts. Streamflow hindcasts from both the ESP and the hydrometeorological approach (TIGGE-ECMWF/CSSPv2, where CSSPv2 was driven by TIGGE-ECMWF) were corrected by matching monthly mean streamflow observations to remove the biases, and the hindcast experiments were termed “ESP-Hydro” and “Meteo-Hydro” (Table 4). Figure 2 shows the procession of the CSSPv2 hindcasts: the calibrated CSSPv2 model was first driven with the observation dataset to generate initial hydrological conditions (e.g., soil moisture and surface water) for each forecast issue date, and the CSSPv2 model was then driven with forecast data (TIGGE-ECMWF or ESP) at every forecast issue date with the generated initial conditions to perform a 7 d hindcast.

2.2.4 LSTM streamflow forecast

Long short-term memory (LSTM) is a type of recurrent neural network model that learns from sequential data. The input of the LSTM model includes the forecast interval streamflow at the specified forecast step obtained from TIGGE-ECMWF/CSSPv2, historical upstream streamflow observations and historical streamflow observations at the Yantan hydrological gauge. The network was trained on sequences of April to September data from 2013 to 2017, with six historical streamflow observations and one forecast interval streamflow to predict the total streamflow at each forecast time step (Fig. 2). The LSTM was calibrated using a cross validation method by leaving the target year out.

Before calibration, all input and output variables were normalized as follows:

$$q_0 = \frac{(q - q_{\text{min}})}{(q_{\text{max}} - q_{\text{min}})}, \quad (6)$$

Table 2. Information on hydrological datasets. (Please note that dates are given in the following format in this table: yyyy/mm/dd.)

Dataset	Time range	Time step
Rain gauge observation forcing	2013/1/1–2017/12/31	Hourly
Longtan & Yantan discharge gauge streamflow data	2013/1/1–2017/12/31	Hourly
Jiazhuan & Luofu discharge gauge streamflow data	2013/4/1–2017/9/30	Daily
TIGGE-ECMWF forecast forcing	2013/4/1–2017/9/30	Hourly

Table 3. Descriptions of calibrated parameters.

Parameters	Range
Maximum velocity of baseflow (mm/d)	0.00000116–0.000579
Fraction of maximum velocity of baseflow where nonlinear baseflow begins	0.001–0.99
Fraction of maximum soil moisture where nonlinear baseflow occurs	0.2–0.99
Variable infiltration curve parameter	0.001–1
River width (m)	0–101.16
River depth (m)	0–6.46
River density (km/km ²)	0.049–1.03
River roughness	0.033–0.05
River slope	0.015–0.47

where q_0 , q , q_{\max} and q_{\min} are the normalized variable, the input variable, and the maximum and minimum of the sequence of the variable, respectively. The hindcast experiment was termed “Meteo-Hydro-LSTM” (Table 2). In addition, we also tried an LSTM streamflow forecasting approach that only used 6 h historical streamflow data as inputs; this experiment was termed “LSTM” (Table 2). The process of LSTM is similar to Meteo-Hydro-LSTM but without the forecast interval streamflow, which is also shown in Fig. 2.

2.3 Evaluation method

The root-mean-squared error (RMSE) was used to evaluate the deterministic forecast, i.e., the ensemble means of 51 (ECMWF) or 40 (ESP) forecast members. To evaluate probabilistic forecasts, the continuous ranked probability score (CRPS) was calculated as follows:

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2, \tag{7}$$

where

$$F_o(y) = \begin{cases} 0, & y < \text{observed value} \\ 1, & y \geq \text{observed value} \end{cases} \tag{8}$$

is a cumulative probability step function that jumps from zero to one at the point where the forecast variable y equals the

observation, and $F(y)$ is a cumulative probability distribution curve formed by the forecast ensembles. The CRPS has a negative orientation (smaller values are better), and it rewards the concentration of probability around the step function located at the observed value (Wilks et al., 2005). The skill score for deterministic forecast was calculated as

$$SS_{RMSE} = \frac{RMSE - RMSE_{ref}}{0 - RMSE_{ref}} = 1 - \frac{RMSE}{RMSE_{ref}}. \tag{9}$$

The skill score for a probabilistic forecast (CRPSS) could be calculated similarly to the SS_{RMSE} .

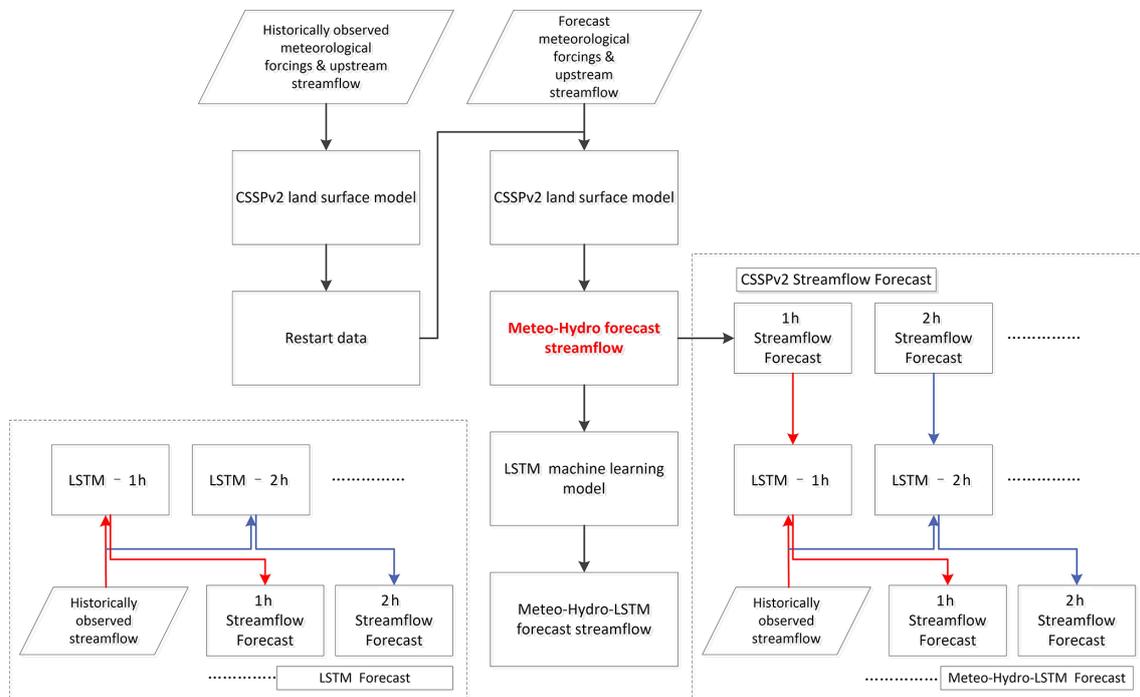
3 Results

3.1 Evaluation of CSSP calibration

The employed CSSPv2 model is a fully distributed hydrological model, and the streamflow is calculated through a process of converting gridded rainfall into runoff and a process of runoff routing. Figure 3 shows the runoff calibration results by calculating the NSE of monthly runoff simulations compared with observed gridded monthly runoff. After calibrating the CSSPv2 runoff model, the NSE of all grids are above zero, which indicates that the runoff simulation results in all grids are more reliable than the climatology method. In addition, grids distributed in the downstream region have

Table 4. Experimental design in this study.

Experiments	Description
ESP-Hydro	Using the CSSPv2 land surface hydrological model driven by randomly sampled historical meteorological forcings
Meteo-Hydro	Using the CSSPv2 model driven by bias-corrected TIGGE-ECMWF hindcast meteorological forcings
Meteo-Hydro-LSTM	Using the LSTM model to correct streamflow from the Meteo-Hydro hindcast
LSTM	Using the LSTM model to forecast streamflow based on observations only

**Figure 2.** A diagram for the integrated hydrometeorological and machine learning streamflow prediction.

better NSE than the upstream grids. The NSE values of the grids in the southern part are greater than 0.5, which accounts for two-thirds of the interval basin area. Higher NSE values in the upstream part of Jiazhan station (Fig. 1) are due to the more humid climate (not shown), as hydrological models usually have better performance over wetter areas. For the downstream areas with less precipitation, the higher NSE values are related to the higher percentage of sand in the soil (not shown). Under the same meteorological conditions, there is higher hydraulic conductivity with higher sand content (Wang et al., 2016), and it yields less runoff under infiltration excess, which is more suitable for the saturation-excess-based runoff generation for the CSSPv2 model (Yuan et al., 2018b).

Figures 4 and 5 show the results after the calibration of the routing model, where CSSPv2 is driven by observed meteorological forcings to provide streamflow simulations and compared against observed streamflow at the Yantan hydrological gauge. Figure 4 shows the daily and monthly streamflow simulation results. The monthly result (Fig. 4f) shows that the simulated streamflow closely follows the observed streamflow, and the NSE is 0.96. The daily streamflow simulations during flood seasons (Fig. 4a–e) also show good performance, and the NSE is 0.92. During June and July in 2014, 2015 and 2017, the CSSPv2 model underestimated the daily streamflow with a maximum of $1104 \text{ m}^3/\text{s}$ and an average of $334 \text{ m}^3/\text{s}$ (Fig. 4b, c, e). In 2013 and 2016, the difference between the observed and simulated streamflow is relatively small, and the average difference is $96 \text{ m}^3/\text{s}$ (Fig. 4a, d).

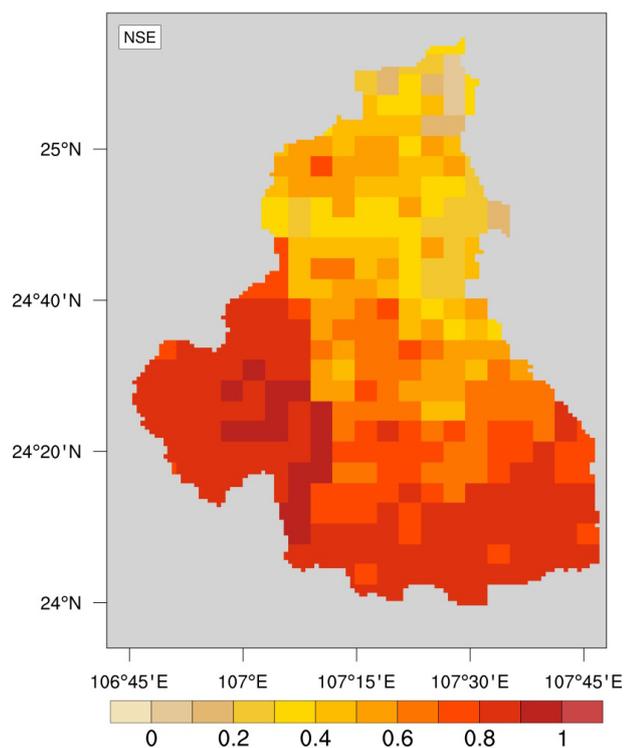


Figure 3. Nash–Sutcliffe efficiency coefficients for the calibrated grid runoff simulation from CSSPv2.

Figure 5 shows the hourly streamflow simulation results for a few flood events. Figure 5a shows that the CSSPv2 model can accurately simulate the streamflow response to a rainfall event after a dry period. Figure 5b–d show that the CSSPv2 model overpredicted water loss during the recession period for instantaneous heavy rainfall events. Figure 5e–f show that the simulated streamflow has a larger fluctuation than the observations for continuous rainfall events. The simulated streamflow is also smoother than the observations. Nevertheless, the NSE for the hourly streamflow simulation is 0.61, which suggests that CSSPv2 has acceptable performance on an hourly timescale.

3.2 Bias correction of the TIGGE-ECMWF meteorological forecasts

The resolution of TIGGE-ECMWF grid data is 0.25° , so the data were interpolated onto a 5 km grid to drive the CSSPv2 model. We calculated the annual average precipitation and temperature for both the observations and TIGGE-ECMWF and then performed a bias correction by adding back the difference (for precipitation) or multiplying back the ratio (for temperature) to match the observations' averages. Figure 6 shows the correlation coefficient and RMSE of TIGGE-ECMWF precipitation and temperature forecasts compared against the observations, either before or after bias correction. The 51-ensemble mean shows better performance for

precipitation and temperature (the red dashed lines) than the best ensemble members (the green dashed lines), with an average RMSE reduction of 3.66 mm/d and an average correlation increase of 0.04 for precipitation as well as an average RMSE reduction of 0.1 K and an average correlation increase of 0.03 for temperature. After bias correction, the 51-ensemble means still perform better than best ensemble members. Compared with the ensemble mean results before bias correction, the RMSE decreased by 0.23 mm/d for the bias-corrected precipitation and decreased by 1 K for the bias-corrected surface air temperature. For the bias-corrected ensemble mean results, the average RMSE and correlation are 14.6 mm/d and 0.44 for precipitation, and they are 1.25 K and 0.87 for surface air temperature.

3.3 Comparison between the ESP-Hydro and Meteo-Hydro streamflow forecasts

Figure 7 presents the variations in RMSE and CRPS for the ESP-Hydro and Meteo-Hydro hourly streamflow forecasts at the Yantan hydrological gauge. For the probabilistic forecast, Fig. 7a shows that the CRPS for the Meteo-Hydro streamflow forecast ranges from 165 to 225 m^3/s , while the CRPS for the ESP-Hydro streamflow forecast ranges from 170 to 230 m^3/s . The Meteo-Hydro approach performs better than ESP-Hydro, with a lower CRPS at all lead times and an average 6 % improvement in the CRPSS (Fig. 7c). For the deterministic forecast, Fig. 7b shows that the RMSE for the Meteo-Hydro streamflow forecast ranges from 250 to 350 m^3/s , while the RMSE for the ESP-Hydro streamflow forecast ranges from 250 to 390 m^3/s . The Meteo-Hydro approach also performs better than ESP-Hydro with a lower RMSE at all lead times, especially after 3 d, with the average reduction in the RMSE reaching 6 % (Fig. 7d).

Figure 7 also shows that the forecast skill of both metrics have a similar diurnal cycle, where the RMSE and CRPS reach their peaks at around 00:00 UTC and drop to their lows at 06:00 UTC. Figure 8 shows the diurnal cycle of the variables employed in the model, namely observed catchment mean rainfall and observed streamflow at the Yantan and Longtan hydrological gauges, to explain the diurnal cycle of the ESP-Hydro and Meteo-Hydro forecasting skill. These three input variables show different diurnal patterns: the observed rainfall starts to rise at 00:00 UTC and reaches its maximum at 06:00 UTC; the observed streamflow at the Yantan hydrological gauge drops to its minimum at 12:00 UTC and rises to its maximum at 00:00 UTC; and the streamflow from upstream of the Longtan hydrological gauge starts to drop at 00:00 UTC and reaches its minimum at 06:00 UTC. After comparing these diurnal cycles with the cycle of forecast skill, it is found that the forecast skill decreases when the upstream Longtan outflow starts to decrease and the precipitation starts to increase. When the upstream Longtan outflow increases and the precipitation starts to decrease (after 06:00 UTC), the forecast skill rises.

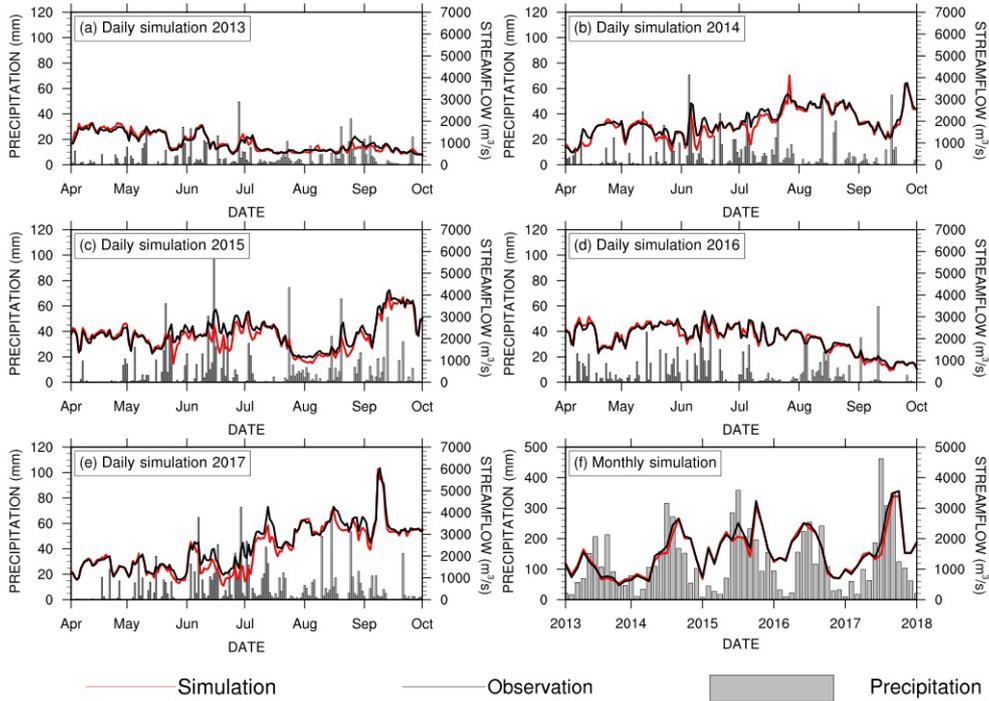


Figure 4. Evaluation of streamflow simulations at the Yantan hydrological gauge. The black and red lines are the observed and simulated streamflow. Panels (a)–(e) show daily streamflow, and panel (f) shows monthly streamflow. The gray bars represent daily (or monthly) precipitation.

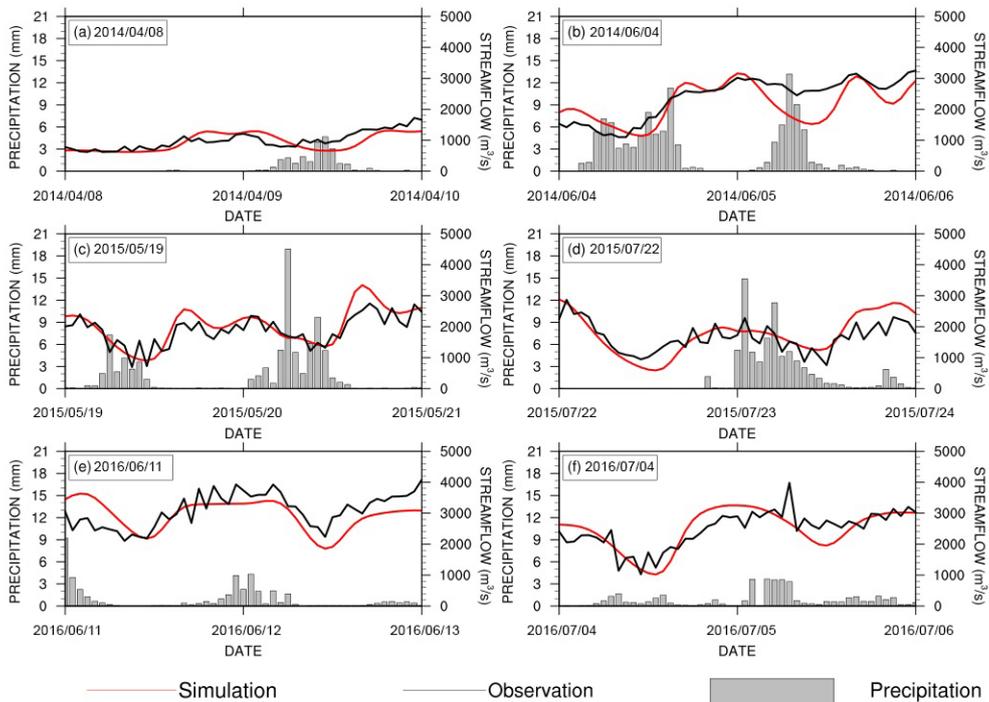


Figure 5. The same as Fig. 4 but for the evaluation of hourly streamflow simulations at the Yantan hydrological gauge. (Please note that dates are given in the following format in this figure: yyyy/mm/dd.)

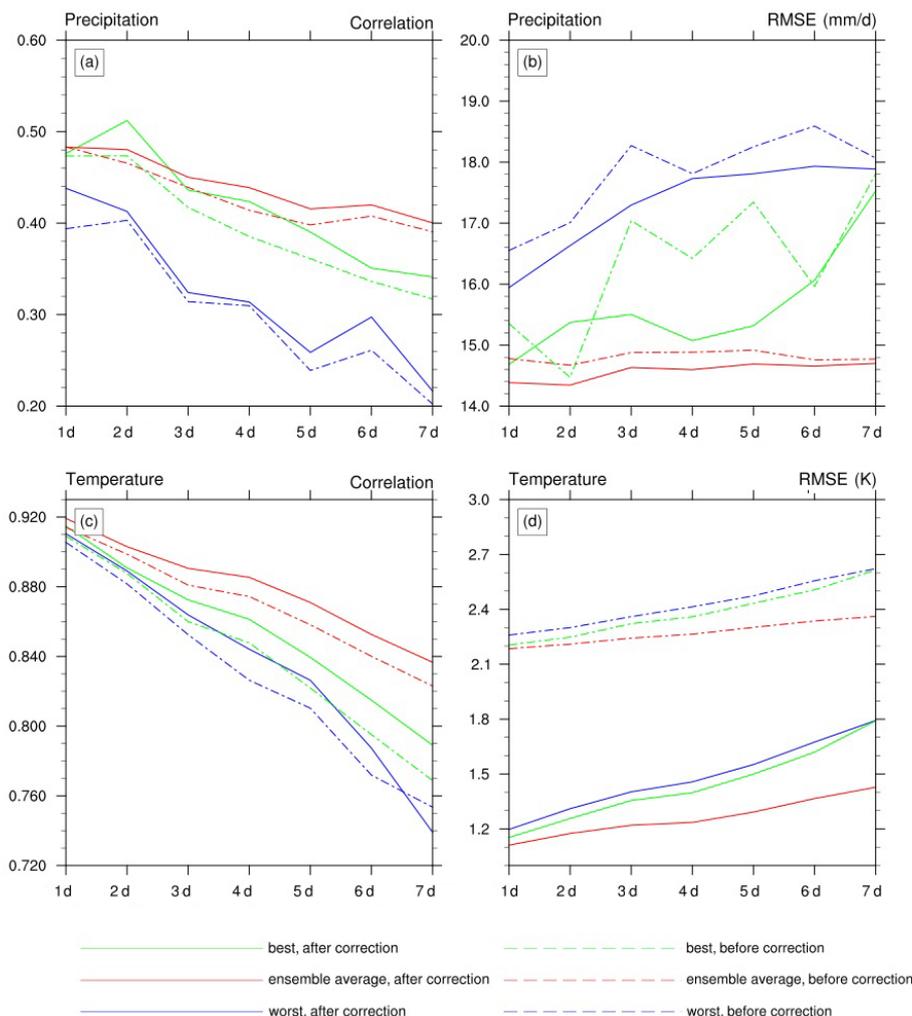


Figure 6. Evaluation of precipitation and temperature hindcasts from TIGGE-ECMWF. The red and blue lines represent the best and worst results among the 51 TIGGE-ECMWF ensemble members, respectively, and the green lines represent the results for the ensemble means of 51 members. Solid and dashed lines represent the results after and before bias corrections, respectively.

3.4 The Meteo-Hydro-LSTM streamflow forecast

Machine learning methods can recognize patterns hidden in input data and can simulate or predict streamflow without explicit descriptions of the underlying physical processes. Figure 9 shows the RMSE of the Meteo-Hydro-LSTM streamflow forecast using the ensemble mean hydrological forecast as described in the section above and using the past 6 h observed streamflow of the Yantan hydrological gauge as input. Compared with the Meteo-Hydro and ESP-Hydro approach, applying the LSTM model can further decrease the RMSE within the first 72 h. The RMSE of the Meteo-Hydro-LSTM approach ranges from 205 to 363 m³/s during these 3 days, suggesting an average 6% improvement compared with the Meteo-Hydro approach.

Figure 9 also shows the RMSE of the LSTM streamflow forecast using only the past 6 h observed streamflow of the

Yantan hydrological gauge as input. Without using the physical model forecast, the RMSE is improved only when the lead time is less than 1 d. Moreover, the performance of LSTM is far worse than the Meteo-Hydro streamflow forecast when the lead time is more than 2 d.

Figure 10 presents several examples of streamflow forecasts using the Meteo-Hydro-LSTM and Meteo-Hydro approaches to show the forecast improvements in detail. The Meteo-Hydro-LSTM approach reduced the flood peak value and the water loss during flood recession period compared with the Meteo-Hydro streamflow forecast approach, which improves the streamflow prediction for most cases (Fig. 10b–f). However, when the upstream reservoir’s flood operation is triggered by continuous heavy rain, Meteo-Hydro may underpredict the streamflow. As the LSTM model further decreases the streamflow, the Meteo-Hydro-LSTM method can end up worsening the streamflow forecast, which means that

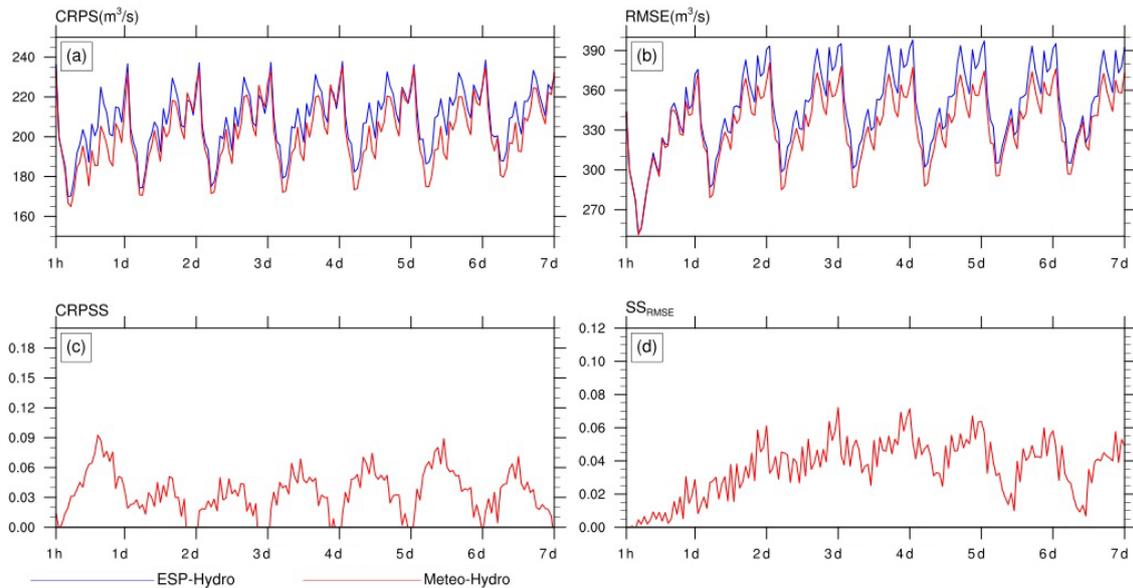


Figure 7. (a) The continuous ranked probability score (CRPS) and (b) root-mean-squared error (RMSE) for the daily streamflow ensemble forecasts at the Yantan hydrological gauge. Panels (c) and (d) show the skill score in terms of the CRPS and RMSE for Meteo-Hydro, where ESP-Hydro is used as the reference forecast.

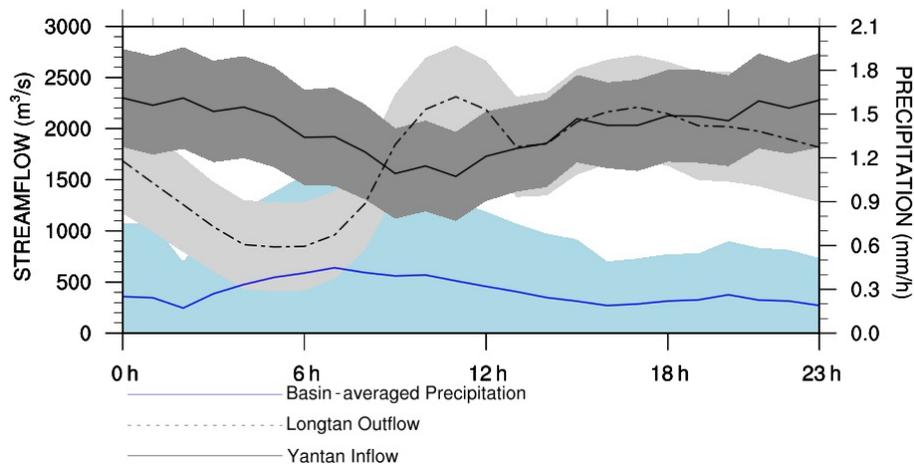


Figure 8. Diurnal cycle of Longtan outflow (m^3/s ; dashed black line), Yantan inflow (m^3/s ; solid black line) and basin-averaged precipitation (mm/h ; blue line) as well as their ranges. The time shown in this figure is universal time.

the machine learning method may not always improve the forecasts (Fig. 10a).

4 Conclusions

In this study, we developed and evaluated a streamflow forecasting framework by coupling meteorological forecasts with a land surface hydrological model (CSSPv2) and a machine learning method (LSTM) over a cascade reservoir catchment using hindcast data from 2013 to 2017. The monthly observed runoff was used to calibrate the runoff generation module of the CSSPv2 model grid by grid, and the hourly ob-

served streamflow at the Yantan hydrological gauge was used to calibrate the routing module of the CSSPv2 model. The bias-corrected TIGGE-ECMWF ensemble forecasts were then used to drive the CSSPv2 for streamflow forecasts, and the LSTM model was used to correct the streamflow forecasts, resulting in an integrated meteorological–hydrological–machine learning forecast framework.

With automatic offline calibration of the CSSPv2 model, the NSE values are 0.96, 0.92 and 0.61 for streamflow simulations at the Yantan hydrological gauge at monthly, daily and hourly timescales, respectively. The bias-corrected ensemble mean TIGGE-ECMWF forcings, which perform the

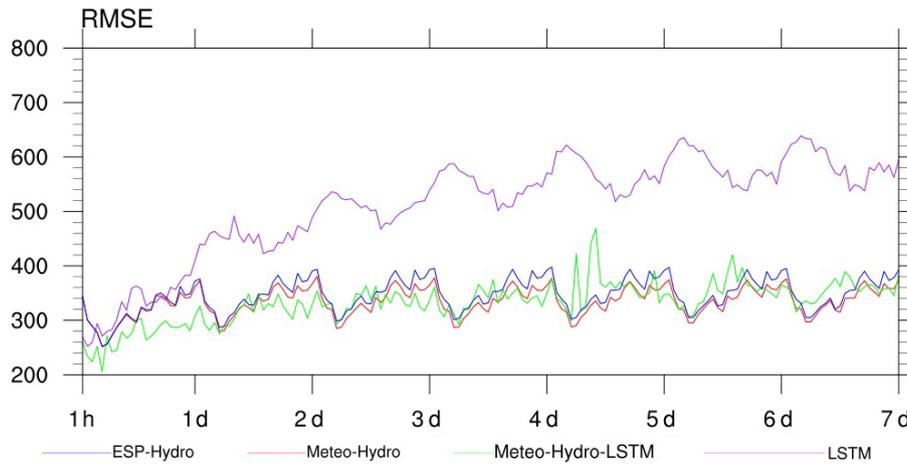


Figure 9. The RMSE (m^3/s) for the hourly streamflow hindcasts from four forecasting approaches. The green line represents the Meteo-Hydro-LSTM forecast, the red line represents the Meteo-Hydro forecast, the blue line represent the ESP-Hydro forecast and the purple line represents the LSTM forecast based on historical streamflow observations alone.

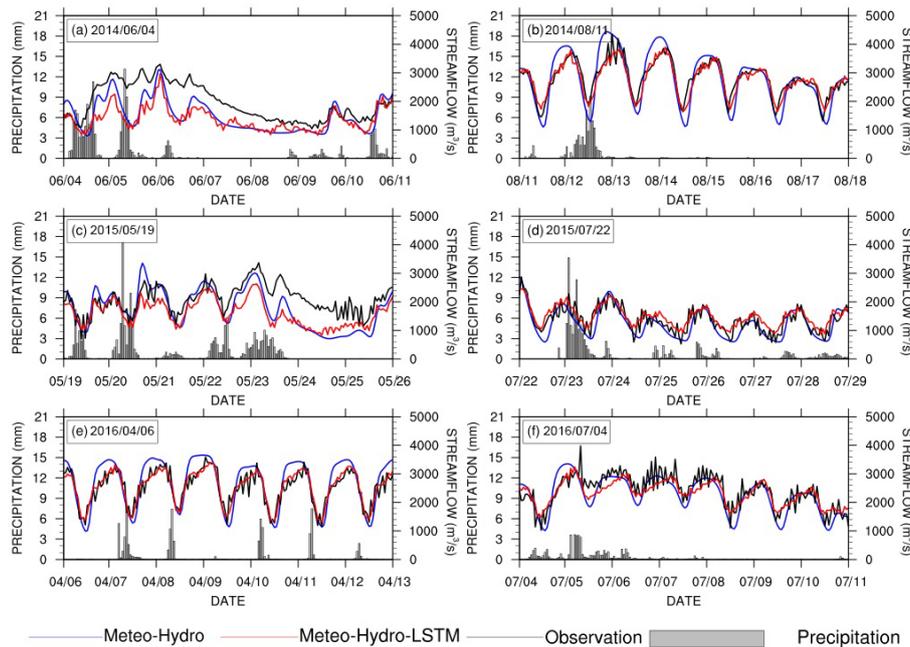


Figure 10. Evaluation of the forecast approaches for a few flooding events. The black lines are observed streamflow from the Yantan hydrological gauge, the blue lines are the Meteo-Hydro ensemble mean streamflow forecast and the red lines are the Meteo-Hydro-LSTM forecast streamflow using the Meteo-Hydro ensemble mean forecast with LSTM. The gray bars represent hourly precipitation averaged over the basin. (Please note that dates are given in the following format in this figure: mm/dd.)

best among all ensemble members, show average respective RMSE and correlation values of 14.6 mm/d and 0.44 for precipitation forecasts and 1.3 K and 0.87 for surface air temperature forecasts. By comparing these results with the hourly observed streamflow, it is found that the integrated hydrometeorological forecast approach (Meteo-Hydro) increases the probabilistic and deterministic forecast skill against the initial condition-based approach (ESP-Hydro) by 6 %.

Adding the LSTM model to the hydrometeorological forecast (Meteo-Hydro-LSTM) can further reduce the forecast error. Within the first 72 h, LSTM can improve the forecast skill by a maximum of 25 % and an average of 6 %. However, if we do not use the streamflow predicted by Meteo-Hydro, the error from the LSTM increases rapidly after 24 h, and the historical-data-based LSTM method performs worse than the Meteo-Hydro method. Most cascade reservoirs cannot

currently forecast streamflow beyond 6 h, and the integrated Meteo-Hydro-LSTM approach has the potential to improve the forecasts at long lead times. This study mainly focused on exploring the added value of meteorology–hydrology coupled forecast and LSTM forecasts in a non-closed catchment; therefore, the forecast uncertainty from upstream outflow was ignored by using the observed outflow. In the future, it is planned to include the upstream outflow forecast; however, this will be very challenging, as it requires the development of an upstream hydrometeorological forecast capability as well as reservoir regulation forecasts. Artificial intelligence (AI) techniques are expected to complement the physical model for reservoir regulation forecasts.

Data availability. The TIGGE-ECMWF hindcast data can be downloaded from <https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=pf/> (Parsons et al., 2017). The in situ observations and simulation data are available from the authors upon request.

Author contributions. XY conceived and designed the study. JL performed the analyses and wrote the initial draft of the paper. XY revised the paper with substantial contributions from all authors. JZ and YJ provided JL with modeling technical support. YL and LZ provided the observed meteorology data used in this study. LY provided the observed hydrology data used in this study.

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

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We would like to thank the reviewers for their constructive comments and suggestions.

Financial support. This research has been supported by the National Key R&D Program of China (grant no. 2018YFA0606002) and the National Natural Science Foundation of China (grant nos. 41875105 and 41901035).

Review statement. This paper was edited by Bob Su and reviewed by Tongtiengang Zhao and three anonymous referees.

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 anar-

- chy? 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.
- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth Syst. Sci.*, 17, 1161–1175, <https://doi.org/10.5194/hess-17-1161-2013>, 2013.
- 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 Science Series, Springer, Dordrecht, the Netherlands, 57–68, https://doi.org/10.1007/1-4020-4902-1_6, 2006.
- Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y., Parsons, D., Raoult, B., Schuster, D., Dias, P. S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., and Worley, S.: The THORPEX interactive grand global ensemble, *B. Am. Meteorol. Soc.*, 91, 1059–1072, <https://doi.org/10.1175/2010BAMS2853.1>, 2010.
- Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.: The Common Land Model, *B. Am. Meteorol. Soc.*, 84, 1013–1024, <https://doi.org/10.1175/BAMS-84-8-1013>, 2003.
- Dai, Y., Dickinson, R. E., and Wang, Y. P.: A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance, *J. Climate*, 17, 2281–2299, [https://doi.org/10.1175/1520-0442\(2004\)017<2281:ATMFACT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2281:ATMFACT>2.0.CO;2), 2004.
- Dang, T. D., Chowdhury, A. F. M. K., and Galelli, S.: On the representation of water reservoir storage and operations in large-scale hydrological models: implications on model parameterization and climate change impact assessments, *Hydrol. Earth Syst. Sci.*, 24, 397–416, <https://doi.org/10.5194/hess-24-397-2020>, 2020.
- Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, *J. Water Resour. Plan Manag.*, 111, 157–170, 1985.
- de Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model, *Hydrol. Earth Syst. Sci.*, 19, 823–837, <https://doi.org/10.5194/hess-19-823-2015>, 2015.
- Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of SCEUA global optimization method for calibrating watershed models, *J. Hydrol.*, 158, 265–284, [https://doi.org/10.1016/0022-1694\(94\)90057-4](https://doi.org/10.1016/0022-1694(94)90057-4), 1994.
- Gao, X., Zeng, Y., Wang, J., and Liu, H.: Immediate impacts of the second impoundment on fish communities in the Three Gorges Reservoir, *Environ. Biol. Fish.*, 87, 163–173, <https://doi.org/10.1007/s10641-009-9577-1>, 2010.
- Getirana, A. C. V., Boone, A., Yamazaki, D., Decharme, B., Papa, F., and Mognard, N.: The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon Basin, *J. Hydrometeorol.*, 13, 1641–1665, <https://doi.org/10.1175/JHM-D-12-021.1>, 2012.

- Hao, Z., Aghakouchak, A., and Phillips, T. J.: Changes in concurrent monthly precipitation and temperature extremes, *Environ. Res. Lett.*, 8, 1402–1416, <https://doi.org/10.1088/1748-9326/8/3/034014>, 2013.
- Hopson, T. and Webster, P.: A 1–10 day ensemble forecasting scheme for the major river basins of Bangladesh: forecasting severe floods of 2003–2007, *J. Hydrometeorol.*, 11, 618–641, <https://doi.org/10.1175/2009JHM1006.1>, 2010.
- Hornberger, G. M., and Boyer, E. W.: Recent advances in watershed modeling, *Rev. Geophys.*, 33, 949–957, <https://doi.org/10.1029/95RG00288>, 1995.
- 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.
- Jasper, K., Gurtz, J., and Lang, H.: Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model, *J. Hydrol.*, 267, 40–52, [https://doi.org/10.1016/S0022-1694\(02\)00138-5](https://doi.org/10.1016/S0022-1694(02)00138-5), 2002.
- 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.
- Ji, P., and Yuan, X.: High-resolution land surface modeling of hydrological changes over the Sanjiangyuan region in the eastern Tibetan Plateau: 2. Impact of climate and land cover change, *J. Adv. Model. Earth Syst.*, 10, 2829–2843, <https://doi.org/10.1029/2018MS001413>, 2018.
- Ji, P., Yuan, X., Jiao, Y., Wang, C., Han, S., and Shi, C.: Anthropogenic contributions to the 2018 extreme flooding over the upper Yellow River basin in China, *B. Am. Meteorol. Soc.*, 101, S89–S94, <https://doi.org/10.1175/BAMS-D-19-0105.1>, 2020.
- Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, W03S04, <https://doi.org/10.1029/2005WR004362>, 2006.
- 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\(532\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532)), 2007.
- Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., and Simmer, C.: Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources, *Water Resour. Res.*, 46, W04201, <https://doi.org/10.1029/2009WR008730>, 2010.
- 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.
- Leith, C. E.: Theoretical Skill of Monte Carlo Forecasts, *Mon. Weather Rev.*, 102, 409–418, [https://doi.org/10.1175/1520-0493\(1974\)102<0409:TSOMCF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2), 1974.
- Lorenz, E. N.: Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, 20, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2), 1963.
- Luo, X., Li, H.-Y., Leung, L. R., Tesfa, T. K., Getirana, A., Papa, F., and Hess, L. L.: Modeling surface water dynamics in the Amazon Basin using MOSART-Inundation v1.0: impacts of geomorphological parameters and river flow representation, *Geosci. Model Dev.*, 10, 1233–1259, <https://doi.org/10.5194/gmd-10-1233-2017>, 2017.
- Mulvaney, T. J.: On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment, *Trans. Inst. Civil Eng. Ireland*, 4, 18–33, 1851.
- Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., Salamon, P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, *J. Hydrol.*, 522, 697–713, 65 <https://doi.org/10.1016/j.jhydrol.2015.01.024>, 2015.
- Parsons, D. B., Beland, M., Burridge, D., Bougeault, P., Brunet, G., Caughey, J., Cavallo, S. M., Charron, M., Davies, H. C., Niang, A. D., Ducrocq, V., Gauthier, P., Hamill, T. M., Harr, P. A., Jones, S. C., Langland, R. H., Majumdar, S. J., Mills, B. N., Moncrieff, M., Nakazawa, T., Paccagnella, T., Rabier, F., Redelsperger, J.-L., Riedel, C., Saunders, R. W., Shapiro, M. A., Swinbank, R., Szunyogh, I., Thorncroft, C., Thorpe, A. J., Wang, X., Waliser, D., Wernli, H., and Toth, Z.: Thorpex research and the science of prediction, *B. Am. Meteorol. Soc.*, 98, 807–830, <https://doi.org/10.1175/BAMS-D-14-00025.1>, 2017 (data available at: <https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=pf/>, last access: 12 January 2022).
- 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.
- Shao, J., Wang, J., Lv, S., and Bing, J.: Spatial and temporal variability of seasonal precipitation in Poyang Lake basin and possible links with climate indices, *Hydrol. Res.*, 47, 51–68, <https://doi.org/10.2166/nh.2016.249>, 2016.
- Toth, Z., Zhu, Y., and Marchok, T.: The use of ensembles to identify forecasts with small and large uncertainty, *Weather Forecast.*, 16, 463–477, [https://doi.org/10.1175/1520-0434\(2001\)016<0463:TUOETI>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0463:TUOETI>2.0.CO;2), 2001.
- 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.
- Wang, Y., Fan, J., Cao, L., and Liang, Y.: Infiltration and Runoff Generation Under Various Cropping Patterns in the Red Soil Region of China, *Land Degrad. Dev.*, 27, 83–91, <https://doi.org/10.1002/ldr.2460>, 2016.
- 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.
- Wilks, D. S., Dmowska, R., Hartmann, D., and Rossby, T. H.: *Statistical Methods in the Atmospheric Sciences*, second edn., International Geophysics Series, volume 91, Academic Press, ISBN 9780080456225, 2005.

- Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, 47, W05301, <https://doi.org/10.1029/2010WR010090>, 2011.
- 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](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.
- Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., and Chau, K.-W.: An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction, *J. Hydrol.*, 569, 387–408, <https://doi.org/10.1016/j.jhydrol.2018.11.069>, 2018.
- Ye, A., Duan, Q., Yuan, X., Wood, E. F., and Schaake, J.: Hydrologic post-processing of MOPEX streamflow simulations, *J. Hydrol.*, 508, 147–156, <https://doi.org/10.1016/j.jhydrol.2013.10.055>, 2014.
- 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?, *B. Am. Meteorol. Soc.*, 99, S113–S117, <https://doi.org/10.1175/BAMS-D-17-0089.1>, 2018a.
- Yuan, X., Ji, P., Wang, L., Liang, X. Z., Yang, K., Ye, A., Su, Z., and Wen, J.: High-resolution land surface modeling of hydrological changes over the Sanjiangyuan region in the eastern Tibetan plateau: 1. Model development and evaluation, *J. Adv. Model Earth Syst.*, 10, 2806–2828, <https://doi.org/10.1029/2018MS001412>, 2018b.
- Zhang, Y., Erkyihum, S. T., and Block, P.: Filling the GERD: evaluating hydroclimatic variability and impoundment strategies for Blue Nile riparian countries, *Water Int.*, 41, 593–610, <https://doi.org/10.1080/02508060.2016.1178467>, 2016.
- Zhao, T. T. G., Cai, X. M., and Yang, D. W.: Effect of streamflow forecast uncertainty on real-time reservoir operation, *Adv. Water Resour.*, 34, 495–504, <https://doi.org/10.1016/j.advwatres.2011.01.004>, 2011.
- 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.