1 **Ensemble streamflow forecasting over a cascade reservoir catchment with**

2 **integrated hydrometeorological modeling and machine learning**

3

4 Junjiang Liu[1], Xing Yuan[1,2*], Junhan Zeng[1], Yang Jiao[1], Yong Li[3], Lihua Zhong[3], Ling

5 Yao[4]

6

7 [1]School of Hydrology and Water Resources, Nanjing University of Information

8 Science and Technology, Nanjing 210044, China

9 [2]Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute

10 of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

11 [3]Guangxi Meteorological Disaster Prevention Center, Nanning 530022, China

12 [4]Guangxi Guiguan Electric Power Co., Ltd., Nanning 530029, China

13

14

15

16 *Hydrology and Earth System Sciences*

17 October 12, 2021

18

[*]*Corresponding author address:* Xing Yuan, School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China E-mail: xyuan@nuist.edu.cn

Abstract. A popular way to forecast streamflow is to use bias-corrected meteorological forecast to drive a calibrated hydrological model, but these hydrometeorological approaches have deficiency 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 of 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 the meteorological forecast, land surface hydrological model simulation and machine learning to forecast hourly streamflow over the Yantan catchment, where the streamflow is influenced both by 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% as compared with the traditional ensemble streamflow prediction (ESP) approach during the first 7 days. The deterministic forecast error can be further reduced by 6% in the first 72 hours when combining the hydrometeorological forecast 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 hours. This study implies the potential of improving flood forecast over a cascade reservoir catchment by integrating meteorological forecast, hydrological modeling and machine learning.

43

1. Introduction

Flood events are the most destructive ones among the natural disasters, causing huge damages to human society. Reservoirs are massively constructed to regulate river flows, which has significantly reduced flood risks or damages (Ji et al., 2020). However, the number and intensity of precipitation extreme events are increasing in many areas as the global warming continues, thus amplify the potential of flood hazards (Hao et al., 2013; Shao et al., 2016; Wei et al., 2018; Yuan et al., 2018a; Wang et al., 2019). Accurate streamflow forecast is thus needed to provide guidelines for reservoir operations (Robertson et al., 2013), especially when the flood risk is increasing under global warming.

A common approach of streamflow forecast is to use hydrological models, where the first attempt could be traced back to 1850s, using simple regression-type approaches to predict discharge from observed precipitation (Mulvaney, 1850). Since then, model concepts have been further augmented by designing new data networks, addressing heterogeneity of hydrological processes, capturing the nonlinear characteristics of hydrologic system and parameterizing models (Hornberger and Boyer, 1995; Kirchner, 2006). With the advancements of computer technology and high-resolution observation, a well-parameterized hydrological model can simulate streamflow with high accuracy (Kollet et al., 2010; Ye et al., 2014; Graaf et al., 2015; Yuan et al., 2018b).

64   Streamflow simulations from hydrological models heavily rely on

65   meteorological forcing inputs, especially precipitation, which can be measured at

66   in-situ gauges or retrieved from satellites and radars. However, for medium-range (2–

67   15 days ahead) streamflow forecasts, precipitation forecast is needed (Hopson et al.,

68   2002). To improve the forecast, ensemble techniques that can give a deterministic

69   estimate as well as its uncertainty became popular. Ensemble weather forecasting can

70   be traced back to 1963 when Leith transferred a deterministic forecast into an

71   ensemble using the Monte-Carlo method to describe the atmospheric uncertainty

72   (Leith, 1963). In the 1990s, ensemble forecasting was developed into an integral part

73   of numerical weather prediction, which showed higher skill than the deterministic

74   forecast even with higher model resolution (Toth et al., 2001). Due to its rapid

75   development, ensemble weather forecasts and climate predictions are applied to

76   hydrological forecasting studies by combining with hydrological models (Jasper et al.,

77   2002; Balint et al., 2006; Jaun et al., 2008; Xu et al., 2015; Yuan et al., 2016; Zhu et

78   al., 2019). Provided with streamflow variability, a reservoir can maintain a reliable

79   utility from natural streamflow better than provided with a deterministic streamflow

80   forecast (Zhao et al., 2011). However, the streamflow prediction skill depends on

81   whether the precipitation forecasts introduced into the hydrological model are skillful

82   (Alfieri et al., 2013). When assessing the skill of this hydrometeorological forecast

83   approach, a benchmark is needed. Using ensembles of historical climatology data

84   (Day, 1985) as meteorological forecast inputs, which is known as ensemble

85   streamflow prediction (ESP), is often selected as the benchmark approach.

86   Evaluations of hydrological forecasts indicated that forecast skill has a close

87   relationship with catchment size, geographical locations and resolutions (Alfieri et al.,

88   2013; Pappenberger et al., 2015), which means there is a necessity to compare with

89   the ESP to show the skill of the hydrometeorological forecast approach.

90   Although physically based hydrological models are widely used, it is still hard

91   to apply a hyper-resolution distributed model for streamflow forecasting due to its

92   demand for observation data, complex model structures and computational resources

93   requirements for calibration and application (Wood et al., 2011; Kratzert et al., 2018;

94   Yaseen et al., 2018). In cascade reservoir systems, there are two sources of streamflow,

95   one is from the rainfall within the interval basin and the other is from the upstream

96   reservoir discharge. While the rainfall-runoff relationship is well studied, it is

97   challenging to reproduce the reservoir operating rules in a physical model (Gao et al.,

98   2010; Zhang et al., 2016; Dang et al., 2020).

99   Machine learning methods can recognize patterns hidden in input data and can

100   simulate or predict streamflow without explicit descriptions of the underlying physical

101   processes (Kisi et al., 2007; Adnan et al., 2019). Neural networks are suitable for

102   streamflow forecasting among machine learning models, some of them can even

103   outperform physically based hydrological models. For example, Humphrey et al.

104   (2016) showed that their combined Bayesian artificial neural network with the modèle

105   du Génie Rural à 4 paramètres Journalier (GR4J) approach outperforms the GR4J

106   model in monthly streamflow forecasting. Kratzert et al. (2019) showed that the long

107   short-term memory (LSTM)-based approach outperforms a well-calibrated

108   Sacramento Soil Moisture Accounting Model (SAC-SMA). Yang et al. (2020) used

109   the geomorphology-based hydrological model (GBHM) combined with traditional

110   ANN model to simulate daily streamflow, which can provide enough physical

111   evidence and can run with less observation data. Although neural network models are

112   criticized with little physical evidence (Abrahart et al., 2012), their potential in

113   hydrological forecasting is yet to be explored.

114   In this study, we combine the machine learning with hydrometeorological

115   approach for hourly streamflow forecast over a cascade reservoir catchment located in

116   southwestern China. We use the meteorological hindcast data from European Centre

117   for Medium-Range Weather Forecasts (ECMWF) model that participated in the

118   THORPEX Interactive Grand Global Ensemble (TIGGE) project to drive a newly

119   developed high-resolution land surface model, named as the Conjunctive

120   Surface-Subsurface Process model version 2 (CSSPv2, Yuan et al., 2018b), to provide

121   runoff and streamflow forecasts, and correct the forecasts via LSTM model. We aim

122   to improve flood forecast over the cascade reservoir catchment by integrating

123   meteorological forecast, hydrological modeling and machine learning. So we strive to

124   (1) calibrate the hydrological model, (2) bias correct the meteorological forecasts, (3)

125   evaluate the streamflow forecast skill and (4) test the physical-statistical combined

126   approach.

127   **2. Study Area, Data, Model and Method**

128   *2.1 Study Area*

129   The Yantan Hydropower Station is in the middle reaches of Hongshui River in

130   Dahua Yao Autonomous County, Guangxi Province. The Yantan Hydropower Station

131   is the fifth level in the 10-level development of Hongshuihe hydropower base in

132   Nanpanjiang River, connected with upstream Longtan Hydropower Station and the

133   downstream Dahua Hydropower Station. The drainage area between the Longtan

134   Hydropower Station and Yantan Hydropower Station is 8,900 km^2. The annual mean

135   streamflow at Yantan gauge is 55.5 billion m^3. The river passes through karst

136   mountain area, with narrow valley, steep slope and scattered cultivated land, and the

137   average slope is 0.036%. Figure 1 shows the locations of 4 hydrological gauges, with

138   detailed information listed in Table 1.

139   *2.2 Data and Method*

140   *2.2.1 Hydrometeorological observations*

141   There are 97 meteorological observation stations within the catchment (Figure

142   1). Here, observed hourly 2m-temperature, 10m-wind speed, relative humidity,

143   accumulated precipitation and surface pressure data were interpolated into a 5km

144   gridded observation dataset via inverse distance weight method. The hourly surface

145   downward solar radiation data from China Meteorological Administration Land Data

146   Assimilation System (CLDAS) was also interpolated into 5km via bilinear

147   interpolation method. The hourly surface downward thermal radiation   was estimated

8

148   by specific humidity, pressure, temperature. This dataset was used to drive the

149   CSSPv2 land surface hydrological model.

150   The monthly runoff for each 5km grid was estimated by disaggregating control

151   streamflow station observations with the ratio of observed grid monthly precipitation

152   and catchment mean precipitation. The gridded runoff was used to calibrate the

153   CSSPv2 model at each grid (Yuan et al., 2016). The calibrated runoff parameters can

154   better represent the heterogeneity of the rainfall-runoff processes and make precise

155   runoff simulations.

156   *2.2.2 Ensemble Meteorological hindcast data and ESP hindcasts*

157   The TIGGE dataset consists of ensemble forecast data from 10 global Numerical

158   Weather Prediction centers started from October 2006, which has been made available

159   for scientific research, via data archive portals at ECMWF and the Chine

160   Meteorological Administration (CMA). TIGGE has become a focal point for a range

161   of research projects, including research on ensemble forecasting, predictability, and

162   the development of products to improve the prediction of severe weather (Bougeault

163   et al., 2010). In this paper, TIGGE data from April to September during 2013-2017

164   from ECMWF were used as meteorological hindcast data. The 3-hourly

165   meteorological hindcasts for 7-day lead time from 51 ensemble members (including

166   control forecast) were interpolated into 5km resolution via bilinear interpolation. The

167   forecast precipitation and temperature were corrected to match the observational

168   means to remove the biases.

169   The ESP was accomplished by applying historical meteorological forcings (Day,

170   1985). In this paper, the meteorological forcings from the same date as the forecast

171   start date to the next 9 days of each year (excluding the target year) were selected as

172   the ESP forcings. Take April 1st, 2013 as example, the 7-day observations started from

173   April 1st to April 10th (i.e., April 1st-April 7th, April 2nd-April 8th, …, April 10th-April

174   16th) in the year of 2014, 2015, 2016 and 2017 were selected as the forecast ensemble

175   forcings of the issue date (April 1st), with a total of 40 ensemble members. The

176   detailed information about the raw datasets are listed in Table 2

177   *2.2.3 CSSPv2 streamflow hindcasts*

178   The physical hydrological model used in this paper is the Conjunctive

179   Surface-Subsurface Process model version 2 (CSSPv2; Yuan et al., 2018). The

180   CSSPv2 model is a distributed, grid-based land surface hydrological model, which

181   was developed from the Common Land Model (Dai et al., 2003, 2004), but with better

182   representations in lateral surface and subsurface hydrological processes and their

183   interactions. The routing model used here employs the kinetic wave equation as

184   covariance function, which is solved via a Newton algorithm. A main reason for

185   adopting this covariance function is that it suits the basin with mountainous terrain.

186   The CSSPv2 model was successfully used to perform a high-resolution (3 km) land

187   surface simulation over the Sanjiangyuan region, which is the headwater of major

188   Chinese rivers (Ji and Yuan, 2018). In this paper, we calibrated CSSPv2 model against

189   monthly estimated runoff to simulate the natural hydrological processes by using the

190   Shuffled Complex Evolution (SCE-UA) approach (Duan et al., 1994). The calibrated

191   parameters include maximum velocity of baseflow, variable infiltration curve

192   parameter, fraction of maximum soil moisture where non-linear baseflow occurs and

193   fraction of maximum velocity of baseflow where non-linear baseflow begins. The

194   hourly observed streamflow at Yantan hydrological gauge was used to calibrate the

195   CSSPv2 routing model manually, including slope, river density, roughness, width and

196   depth. The observed streamflow at Longtan hydrological gauge were added into the

197   corresponding grid to provide upstream streamflow information. We used a

198   high-resolution elevation database (hereafter referred to as DEM90) for sub-grid

199   parameterization and figured out the initial values of these river channel parameters.

200   We first extracted the slope angle and the natural river flow path from DEM90, and

201   then identified the accurate river network using a drainage area threshold of 0.18 km^2.

202   River density and bed slope values for each 5km grid were calculated as:

203    $$rivden = \sum l / A, \hspace{2cm} (1)$$

204    $$bedslp = mean(\tan(\beta)), \hspace{2cm} (2)$$

205   where rivden is the river density (km/km^2), bedslp is the river channel bed slope

206   (unitless), A is the area of a 5km grid (km^2), $\sum l$ is the total river channel length (m)

207   within the grid, β is the slope angle (radian) for each river segment located in the grid.

208   Other river channel parameters were estimated by empirical formulas (Getirana

209   et al., 2012; Luo et al., 2017) as follows:

$$W = 1.956 \times A_{acc}^{0.413}, \tag{3}$$

$$H = 0.245 \times A_{acc}^{0.342}, \tag{4}$$

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

where W, H and n are river width (m), depth (m) and roughness (unitless) for each

5km grid; Aacc means the upstream drainage area (km2); Hmax and Hmin refer to the

maximum and minimum values of river depth calculated by Eq. (4).

Through a trial-and-error procedure, we calibrated these river channel parameters

to match the simulated streamflow with observed hourly records at 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 range are listed in Table 3

After calibration, we drove the CSSPv2 model using 5km regridded and

bias-corrected TIGGE-ECMWF forecast forcing during 2013-2017 to provide a set of

7-day hindcasts. Streamflow hindcasts both from the ESP and the hydrometerological

approach (TIGGE-ECMWF/CSSPv2) were corrected by matching monthly mean

streamflow observations to remove the biases, and the hindcast experiments were

termed as ESP-Hydro and Meteo-Hydro (Table 4). Figure 2 shows the procession of

the CSSPv2 hindcasts: the calibrated CSSPv2 model was first driven with observation

dataset to generate initial hydrological conditions (soil moisture, surface water, etc.)

for each forecast issue date, then CSSPv2 model was driven with forecast data

230   (TIGGE-ECMWF or ESP) at every forecast issue date with the generated initial

231   conditions to perform a 7-day hindcast.

232   *2.2.4 LSTM streamflow forecast*

233   LSTM is a type of recurrent neural network model which learns from sequential

234   data. The input of the LSTM model includes forecast interval streamflow at the

235   specified forecast step obtained from TIGGE-ECMWF/CSSPv2, historical upstream

236   streamflow observation, and historical streamflow observation at Yantan hydrological

237   gauge. The network was trained on sequences of April to September in 2013-2017,

238   with six historical streamflow observations and one forecast interval streamflow to

239   predict the total streamflow at each forecast time step (Figure 2). The LSTM was

240   calibrated through a cross validation method, by leaving the target year out.

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

242
$$\mathbf{q_0} = \frac{(\mathbf{q}-\mathbf{q_{min}})}{(\mathbf{q_{max}}-\mathbf{q_{min}})}, \tag{6}$$

243   Where $\mathbf{q_0}$, $\mathbf{q}$, $\mathbf{q_{max}}$ and $\mathbf{q_{min}}$ are the normalized variable, input variable, the

244   maximum and minimum of the sequence of the variable, respectively. The hindcast

245   experiment was termed as Meteo-Hydro-LSTM (Table 2). In addition, we also tried

246   an LSTM streamflow forecast approach which only uses 6-hr historical streamflow

247   data as inputs, and the experiment was termed as LSTM (Table 2). The process of

248   LSTM is similar to Meteo-Hydro-LSTM but without the forecast interval streamflow,

249   which is also shown in Figure 2.

250 *2.3 Evaluation Method*

251   The root-mean squared error (RMSE) was used to evaluate the deterministic

252 forecast, i.e., the ensemble means of 51 (ECMWF) or 40 (ESP) forecast members. To

253 evaluate probabilistic forecasts, the Continuous Ranked Probability Score (CRPS)

254 was calculated as follows:

255
$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2, \qquad (7)$$

256 where

257
$$F_o(y) = \begin{cases} 0, & y < observed\ value \\ 1, & y \geq observed\ value \end{cases} \qquad (8)$$

258 is a cumulative-probability step function that jumps from 0 to 1 at the point where the

259 forecast variable y equals the observation and $F(y)$ is a cumulative-probability

260 distribution curve formed by the forecast ensembles. The CRPS has a negative

261 orientation (smaller values are better), and it rewards concentration of probability

262 around the step function located at the observed value (Wilks, 2005). The skill score

263 for deterministic forecast was calculated as

264
$$SS_{RMSE} = \frac{RMSE - RMSE_{ref}}{0 - RMSE_{ref}} = 1 - \frac{RMSE}{RMSE_{ref}} \quad . \qquad (9)$$

265 The skill score for probabilistic forecast (CRPSS) could be calculated similarly based

266 the CRPS.

267 **3. Results**

268 *3.1 Evaluation of CSSP calibration*

269   The employed CSSPv2 model is a fully distributed hydrological model and the

270   streamflow is calculated through a process of converting gridded rainfall into runoff

271   and a process of runoff routing. Figure 3 shows the runoff calibration results by

272   calculating the NSE of monthly runoff simulations compared with observed gridded

273   monthly runoff. After calibrating the CSSPv2 runoff model, the NSE of all grids are

274   above 0, which indicates that the runoff simulation results in all grids are more

275   reliable than the climatology method. In addition, grids distributed in the downstream

276   region have better NSE than the upstream grids. The NSE values of the grids in the

277   southern part are greater than 0.5, which accounts for two thirds of the interval basin

278   area. Higher NSE in the upstream part of Jiazhuan station (Figure 1) is due to more

279   humid climate (not shown), where hydrological models usually have better

280   performance over wetter areas. For the downstream areas with less precipitation, the

281   higher NSE is related to the higher percentage of sand in the soil (not shown). Under

282   the same meteorological conditions, there is higher hydraulic conductivity with higher

283   sand content (Wang et al., 2016), and it yields less runoff under infiltration excess,

284   which is more suitable for the saturation excess-based runoff generation for the

285   CSSPv2 model (Yuan et al., 2018b).

286   Figures 4 and 5 show the results after the calibration of the routing model, where

287   CSSPv2 is driven by observed meteorological forcings to provide streamflow

288   simulations and compare against observed streamflow at Yantan hydrological gauge.

289   Figure 4 shows the daily and monthly streamflow simulation results. The monthly

290   result (Fig. 4f) shows that the simulated streamflow closely follows the observed

291 streamflow, and the NSE is 0.96. The daily streamflow simulations during flood

292 seasons (Figs. 4a-4e) also show a good performance, and the NSE is 0.92. During

293 June and July in years of 2014, 2015 and 2017, the CSSPv2 model underestimated the

294 daily streamflow with a maximum of 1104 m^3/s and an average of 334 m^3/s (Figs. 4b,

295 4c, 4e). In years of 2013 and 2016, the difference between observed and simulated

296 streamflow is relatively small, and the average difference is 96 m^3/s (Figs. 4a, 4d).

297   Figure 5 shows the hourly streamflow simulation results for a few flood events.

298 Figure 5a shows that the CSSPv2 model can accurately simulate the streamflow

299 response to a rainfall event after a dry period. Figures 5b-5d show that for

300 instantaneous heavy rainfall events, the CSSPv2 model over-predicted the water loss

301 during recession period. Figures 5e-5f show that for continuous rainfall events, the

302 simulated streamflow has a larger fluctuation than observation. The simulated

303 streamflow is also smoother than observation. Nevertheless, the NSE for the hourly

304 streamflow simulation is 0.61, which suggests that CSSPv2 has an acceptable

305 performance at hourly time scale.

306 *3.2 Bias correction of TIGGE-ECMWF meteorological forecasts*

307   The resolution of TIGGE-ECMWF grid data is 0.25°, so the data was

308 interpolated to 5km grid to drive the CSSPv2 model. We calculated both observations'

309 and TIGGE-ECMWF's yearly average precipitation and temperature, then performed

310 a bias correction by adding back the difference (for precipitation) or multiplying back

311 the ratio (for temperature) to match the observations' averages. Figure 6 shows the

312 correlation coefficient and RMSE of TIGGE-ECMWF precipitation and temperature

313 forecasts as compared against observations, either before or after bias correction. The

314 51-ensemble mean precipitation and temperature (the red dashed lines) shows better

315 performance than the best ensemble members (the green dashed lines), with an

316 average RMSE reduction of 3.66 mm/day and average correlation increase of 0.04 for

317 precipitation, and average RMSE reduction of 0.1K and average correlation increase

318 of 0.03 for temperature. After bias correction, the 51-ensemble means still perform

319 better than best ensemble members. Compared with ensemble mean results before

320 bias correction, the RMSE reduced by 0.23 mm/day for the bias-corrected

321 precipitation, and reduced by 1K for the bias-corrected surface air temperature. For

322 the bias-corrected ensemble mean results, the average RMSE and correlation are 14.6

323 mm/day and 0.44 for precipitation, and 1.25 K and 0.87 for surface air temperature.

324 *3.3 Comparison between ESP-Hydro and Meteo-Hydro streamflow forecast*

325 Figure 7 presents the variations of RMSE and CRPS for ESP-Hydro and

326 Meteo-Hydro hourly streamflow forecast at Yantan hydrological gauge. For

327 probabilistic forecast, Figure 7a shows that the CRPS for Meteo-Hydro streamflow

328 forecast ranges from 165 to 225 m^3/s while the CRPS for ESP-Hydro streamflow

329 forecast ranges from 170 to 230 m^3/s. The Meteo-Hydro approach performs better

330 than ESP-Hydro with lower CRPS at all lead times, with an average of 6%

331 improvement in CRPSS (Figure 7c). For deterministic forecast, Figure 7b shows that

332 the RMSE for Meteo-Hydro streamflow forecast ranges from 250 to 350 m^3/s, while

333 the RMSE for ESP-Hydro streamflow forecast ranges from 250 to 390 m^3/s. The

334 Meteo-Hydro approach also performs better than ESP-Hydro with lower RMSE at all

335   lead times especially after 3 days, with the average reduction of RMSE reaching 6%

336   (Figure 7d).

337   Figure 7 also shows that both forecast skills have a similar diurnal cycle, where

338   RMSE and CRPS reach their peaks around 00UTC and drop to their lows at 06UTC.

339   Figure 8 shows the diurnal cycle of model employed variables, which are observed

340   catchment mean rainfall, observed streamflow at Yantan and Longtan hydrological

341   gauges, to explain the diurnal cycle of ESP-Hydro and Meteo-Hydro forecasting skills.

342   These three input variables show different diurnal patterns. The observed rainfall

343   starts to rise at 00UTC and reaches its maximum at 06UTC. The observed streamflow

344   at Yantan hydrological gauge drops to its minimum at 12UTC and rises to its

345   maximum at 00UTC. The streamflow from upstream Longtan hydrological gauge

346   starts to drop at 00UTC and reaches its minimum at 06UTC. After comparing these

347   diurnal cycles with the cycle of forecast skill, it is found that the forecast skill

348   decreases when the upstream Longtan outflow starts to decrease, and the precipitation

349   starts to increase. When the upstream Longtan outflow increases and the precipitation

350   starts to decrease (after 06UTC), the forecast skill rises. Such information indicates

351   that the hydrological model performs worse in the case of heavier rainfall event, and

352   the decrease of upstream outflow may amplify such degradation when the portion of

353   interval rainfall-runoff increased.

354   *3.4 Meteo-Hydro-LSTM streamflow forecast*

355   Machine learning methods can recognize patterns hidden in input data and can

356   simulate or predict streamflow without explicit descriptions of the underlying physical

357   processes. Figure 9 shows the RMSE of Meteo-Hydro-LSTM streamflow forecast

358   using the ensemble mean hydrological forecast as described in the section above, and

359   the past 6-hour observed streamflow of Yantan hydrological gauge as input.

360   Compared with Meteo-Hydro and ESP-Hydro approach, applying LSTM model can

361   further decrease the RMSE within the first 72 hours. The RMSE of

362   Meteo-Hydro-LSTM approach ranges from 205 to 363 m^3/s during these three days,

363   suggesting an average of 6% improvement against Meteo-Hydro approach.

364   Figure 9 also shows the RMSE of LSTM streamflow forecast only using the past

365   6-hour observed streamflow of Yantan hydrological gauge as input. Without using the

366   physical model forecast, RMSE is improved only when the lead time is less than 1 day.

367   And the performance of LSTM is far worse than Meteo-Hydro streamflow forecast

368   when lead time is more than 2 days.

369   Figure 10 shows several examples of streamflow forecasts by

370   Meteo-Hydro-LSTM approach and Meteo-Hydro approaches to show the forecast

371   improvements in details. The Meteo-Hydro-LSTM approach reduced the flood peak

372   value and the water loss during flood recession period compared with Meteo-Hydro

373   streamflow forecast approach, which improves the streamflow prediction for most

374   cases (Figs. 10b-10f). However, when the upstream reservoir's flood operation is

375   triggered by continuous heavy rain, the Meteo-Hydro may underpredict the

376   streamflow. With the LSTM model further decreases the streamflow, the

377   Meteo-Hydro-LSTM method can end up with worsening the streamflow forecast,

378 which means the machine learning method may improve forecasts when trained in

379 different flood operating situations (Figure 10a).

4. Conclusions

380

381 In this study, we developed and evaluated a streamflow forecasting framework

382 by coupling meteorological forecasts with a land surface hydrological model (CSSPv2)

383 and a machine learning method (LSTM) over a cascade reservoir catchment using

384 hindcast data from 2013 to 2017. The monthly observed runoff was used to calibrate

385 the runoff generation module of the CSSPv2 model grid by grid, and the hourly

386 observed streamflow at Yantan hydrological gauge was used to calibrate the routing

387 module of the CSSPv2 model. Then, the bias-corrected TIGGE-ECMWF ensemble

388 forecasts were used to drive the CSSPv2 for streamflow forecasts, and the LSTM

389 model was used to correct the streamflow forecasts, resulted in an integrated

390 meteorological-hydrological-machine learning forecast framework.

391 With automatic offline calibration of the CSSPv2 model, and the NSE values are

392 0.96, 0.92 and 0.61 for streamflow simulations at the Yantan gauge at monthly, daily

393 and hourly time scales, respectively. The bias-corrected ensemble mean

394 TIGGE-ECMWF forcings which perform the best among all ensemble members,

395 show average RMSE and correlation of 14.6 mm/day and a 0.44 for precipitation

396 forecasts, and 1.3 K and 0.87 for surface air temperature forecasts. By comparing with

397 the hourly observed streamflow, the integrated hydrometeorological forecast approach

398   (Meteo-Hydro) increases the probabilistic and deterministic forecast skill against the

399   initial condition-based approach (ESP-Hydro) by 6%.

400   Adding LSTM model to the hydrometeorological forecast (Meteo-Hydro-LSTM)

401   can further reduce the forecast error. Within the first 72 hours, LSTM can improve the

402   forecast skill with a maximum of 25% and an average of 6%. However, if we do not

403   use the streamflow predicted by Meteo-Hydro, the error from the LSTM increases

404   rapidly after 24 hours, and the historical data-based LSTM method performs worse

405   than the Meteo-Hydro method. Most cascade reservoirs yet cannot forecast

406   streamflow beyond 6 hours, and the integrated Meteo-Hydro-LSTM approach has

407   potential to improve the forecasts at long leads. This study mainly focused on

408   exploring the added values of meteorology-hydrology coupled forecast and LSTM

409   forecast in a non-closed catchment, so the forecast uncertainty from upstream outflow

410   was ignored by using the observed outflow. In the future, the upstream outflow

411   forecast is planned to include, but this requires the development of upstream

412   hydrometeorological forecast capability, as well as the reservoir regulation forecast

413   that is very challenging. The artificial intelligence (AI) techniques are expected to

414   complement the physical model for reservoir regulation forecast.

415

422 **Data availability**

423 The TIGGE-ECMWF hindcast data can be downloaded from

424 https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=pf/ (Parsons et al., 2017),

425 the in-situ observations and simulation data are available upon request.

426 **Competing interests**

427 The contact author has declared that neither they nor their co-authors have any

428 competing interests.

429 **Author contributions**

430 X.Y. conceived and designed the study. J.L. performed the analyses and wrote the

431 initial draft, X.Y. revised the paper, with substantial contributions from all authors.

432

References

433 **References**

434 Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., et

435   al.: Two decades of anarchy? emerging themes and outstanding challenges for

436   neural network river forecasting. Prog. Phys. Geogr. 36(4), 480-513.

437   https://doi.org/10.1177/0309133312444943, 2012.

438 Adnan, R.M., et al.: Daily streamflow prediction using optimally pruned extreme

439   learning machine. J. Hydrol. 577. https://doi.org/10.1016/j.jhydrol.2019.123981,

440   2019.

441 Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., & Pappenberger, F.: GloFAS-global

442   ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst.

443   Sci. 17 (3), 1161–1175. https://doi.org/10.5194/hess17-1161-2013, 2013.

444 Balint, G., Csik, A., Bartha, P., Gauzer, B., Bonta, I.: Application of meterological

445   ensembles for Danube flood forecasting and warning. In: Marsalek, J., Stancalie,

446   G., Balint, G. (Eds.), Transboundary Floods: Reducing Risks through Flood

447   Management. Springer, NATO Science Series, Dordecht, The Netherlands, pp.

448   57–68. https://doi.org/10.1007/1-4020-4902-1_6 , 2006.

449 Bougeault, P., et al.: The THORPEX interactive grand global ensemble, Bull. Am.

450   Meteorol. Soc., 91, 1059–1072. http://dx.doi.org/10.1175/2010BAMS2853.1,

451   2010.

452 Dai, Y., Dickinson, R. E., Wang, Y. P.: A two-big-leaf model for canopy temperature,

453   photosynthesis, and stomatal conductance. J. Clim. 17(12),2281–2299.

454   https://doi.org/10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2 , 2004.

455   Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al.:

456   The Common Land Model. Bull. Am. Meteorol. Soc. 84(8), 1013–1024.

457   https://doi.org/10.1175/BAMS-84-8-1013, 2003.

458   Dang, T. D., Chowdhury, A. K., Galelli, S.: On the representation of water reservoir

459   storage and operations in large-scale hydrological models: implications on

460   model parameterization and climate change impact assessments. Hydrol. Earth

461   Syst. Sci., 24, 397–416. https://doi.org/10.5194/hess-24-397-2020, 2020.

462   Day, G.N.: Extended Streamflow Forecasting Using NWSRFS. J. Water Resour. Plan

463   Manag. 111 (2): 157-170, 1985.

464   Duan, Q., Sorooshian, S., Gupta, V. K.: Optimal use of SCEUA global optimization

465   method for calibrating watershed models, J. Hydrol., 158, 265–284,

466   https://doi.org/10.1016/0022-1694(94)90057-4, 1994.

467   Gao, X., Zeng, Y., Wang, J., Liu, H.: Immediate impacts of the second impoundment

468   on fish communities in the Three Gorges Reservoir, Environ. Biol. Fish., 87,

469   163–173. https://doi.org/10.1007/s10641-009-9577-1., 2010.

470   Getirana, A. C. V., Boone, A., Yamazaki, D., Decharme, B., Papa, F., and Mognard, N.:

471   The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the

472   Amazon   Basin,   J.   Hydrometeorol.,   13,   1641–1665,

473   https://doi.org/10.1175/JHM-D-12-021.1, 2012.

474   Graaf, I. D., Sutanudjaja, E. H., Beek, L. V., et al.: A high-resolution global-scale

475   groundwater   model.   Hydrol.   Earth   Syst.   Sci.   19(2):823-837.

476   https://doi.org/10.5194/hess-19-823-2015, 2015.

Hao, Z., Aghakouchak, A., Phillips, T. J.: Changes in concurrent monthly precipitation and temperature extremes. Environ. Res. Lett. 8(3), 1402-1416. https://doi.org/10.1088/1748-9326/8/3/034014, 2013.

Hopson, T., 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(3), 618-641. https://doi.org/10.1175/2009JHM1006.1, 2010.

Hornberger, G. M., E. W. Boyer.: Recent advances in watershed modeling, U.S. Natl. Rep. Int. Union Geod. Geophys. 1991 – 1994, Rev. Geophys., 33, 949 – 957. https://doi.org/10.1029/95RG00288, 1995.

Humphrey, G.B., Gibbs, M.S., Dandy, G.C., 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., Lang, H.: Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. J. Hydrol. 267 (1–2), 40–52. https://doi.org/10.1016/S0022-1694(02)00138-5, 2002.

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

Ji, P., X. Yuan., Y. Jiao., C. Wang., S. Han., C. Shi.: Anthropogenic contributions to the 2018 extreme flooding over the upper Yellow River basin in China. Bull. Am.

499 Meteorol. Soc. 101(1), S89-S94, https://doi.org/10.1175/BAMS-D-19-0105.1,

500 2020.

501 Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements,

502 analyses, and models to advance the science of hydrology, Water Resour. Res.

503 42, 1–5. https://doi.org/10.1029/2005WR004362, 2006.

504 Kisi, O.: Streamflow forecasting using different artificial neural network algorithms. J.

505 Hydrol. Eng. 12 (5), 532–539.

506 https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532), 2007.

507 Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., &

508 Vereecken, H., et al.: Proof of concept of regional scale hydrologic simulations

509 at hydrologic resolution utilizing massively parallel computer resources. Water

510 Resour. Res. 46(4). https://doi.org/10.1029/2009WR008730, 2010.

511 Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M.: Rainfall–runoff

512 modelling using long short-term memory (LSTM) networks. Hydrol. Earth Syst.

513 Sci. 22 (11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018, 2018.

514 Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., Nearing, G. S.:

515 Towards Improved Predictions in Ungauged Basins: Exploiting the Power of

516 Machine Learning. Water Resour. Res. 55.

517 https://doi.org/10.1029/2019wr026065, 2019.

518 Leith, C. E.: Theoretical skill of monte carlo forecasts. Mon. Weather Rev. 102(6),

519 409-418. https://doi.org/10.1175/1520-0493(1974)1022.0.CO;2, 1974.

520 Luo, X., Li, H. Y., Ruby, L. L., Tesfa, T. K., Augusto, G., & Fabrice, P., et al.:

26

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(3), 1-42. 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 of flood discharges in a given

catchment, in: Proceedings Institution of Civil Engineers, Dublin, Vol. 4, 18–31,

1850.

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K.,

et al.: How do I know if my forecasts are better? Using benchmarks in

hydrological ensemble prediction. J. Hydrol. 522, 697–713.

https://doi.org/10.1016/j.jhydrol.2015.01.024, 2015.

Parsons, D. B., Beland, M., Burridge, D., Bougeault, P., Brunet, G., Caughey, J., et al.:

Thorpex research and the science of prediction. Bull. Am. Meteorol. Soc., 98,

807-830, https://doi.org/10.1175/BAMS-D-14-00025.1 , 2017.

Robertson, D. E., Wang, Q. J.: Seasonal Forecasts of Unregulated Inflows into the

Murray River, Australia. Water. Resour. Manag. 27(8):2747–2769.

https://doi.org/10.1007/s11269-013-0313-4, 2013.

Shao, J., Wang, J., Lv, S., Bing, J.: Spatial and temporal variability of seasonal

precipitation in Poyang Lake basin and possible links with climate indices.

Hydrol. Res. 47(S1):51–68. https://doi.org/10.2166/nh.2016.249, 2016.

Toth, Z., Zhu, Y., Marchok, T.: The use of ensembles to identify forecasts with small

and large uncertainty. Weather Forecast 16(4), 463-463. https://doi.org/10.1175/1520-0434(2001)0162.0.CO;2, 2001.

Wang, R., Zhang, J., Guo, E., Zhao, C., 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.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Volume 91, Second Edition (International Geophysics), 2005.

Wood, E. F., et al.: 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., Zhang, Y., Kang, L.: 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, 2015, https://doi.org/10.1061/(ASCE) HE.1943-5584.0001007, 2015.

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., 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., 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., Schaake, J.: Hydrologic post-processing of MOPEX streamflow simulations. J. Hydrol. 508, 147-156, doi:10.1016/j.jhydrol.2013.10.055, 2014

Yuan, X., Ma, F., Wang, L., Zheng, Z., Ma, Z., Ye A., 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., S. Wang, and Z.-Z. Hu, 2018a: 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, 2018a.

Yuan, X., Ji, P., Wang, L., Liang, X. Z., Yang, K., Ye, A., et al.: 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. https://doi.org/10.1029/2018MS001412, 2018b.

Zhang, Y., Erkyihum, S. T., 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., Yang, D.W.: Effect of streamflow forecast uncertainty on real-time reservoir operation. Adv. Water Resour. 34 (4), 495–504, https://doi.org/10.1016/j.advwatres.2011.01.004, 2011.

Zhu, E., X. Yuan., A. Wood.: Benchmark Decadal Forecast Skill for Terrestrial Water

587   Storage Estimated by an Elasticity Framework. Nat. Commun. 10, 1237,

588   https://doi.org/10.1038/s41467-019-09245-3, 2019.

589

590 **Table 1.** Information of 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)

591

592   **Table 2.** Information of hydrological datasets

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

593

594   **Table 3.** Descriptions of calibrated parameters

Parameters	Range
Maximum velocity of baseflow (mm/day)	0.00000116 ~ 0.000579
Fraction of maximum velocity of baseflow where non-linear baseflow begins	0.001 ~ 0.99
Fraction of maximum soil moisture where non-linear 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

595

596   **Table 4.** Experimental design in this study.

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

597

598

599

600
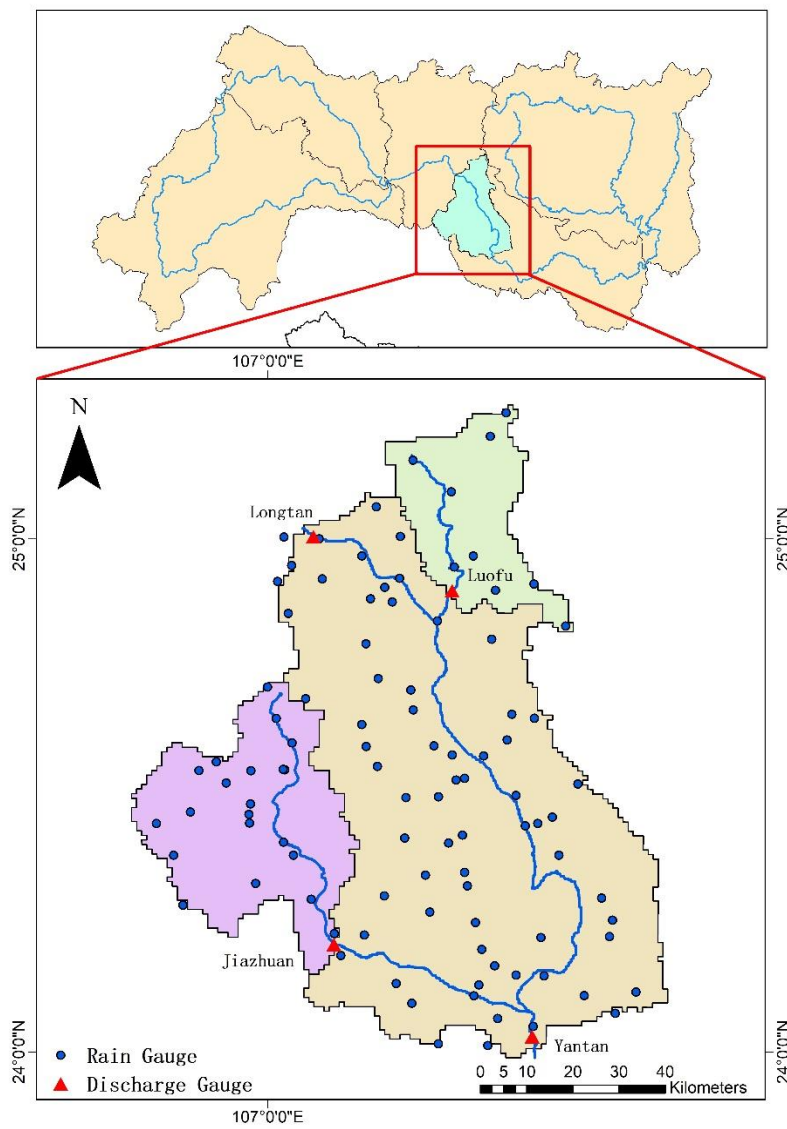
Figure 1. Locations of discharge gauges and rain gauges over the Yantan basin.

602
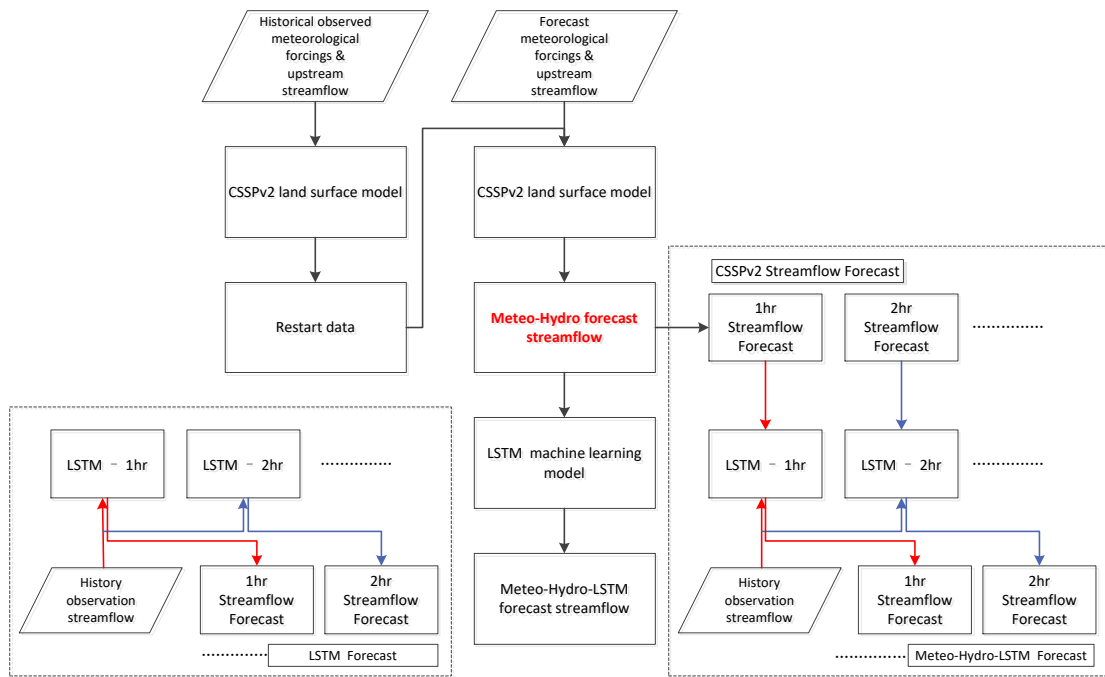
Figure 2. A diagram for the integrated hydrometeorological and machine learning streamflow prediction.

603

604

605

606

607
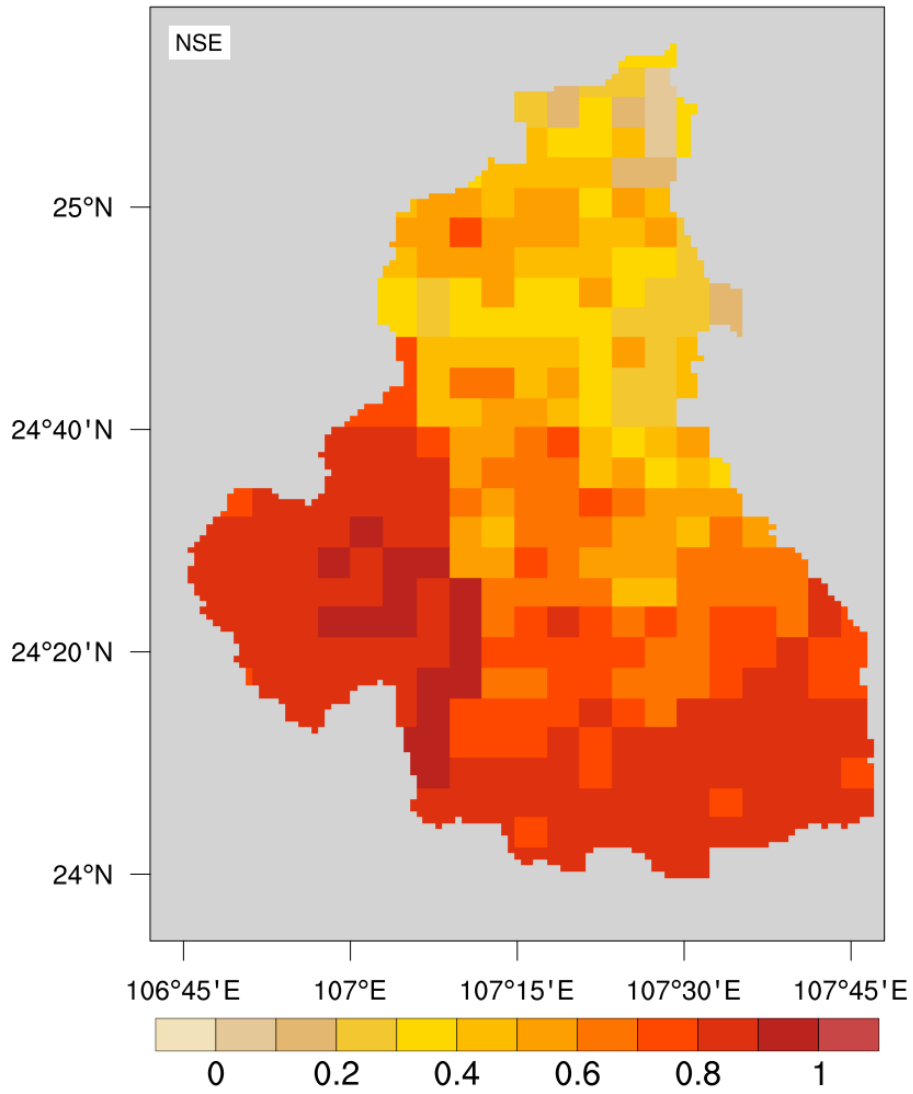
Figure 3. Nash-Sutcliff efficiency coefficients for the calibrated grid runoff simulation

from CSSPv2.

610

611
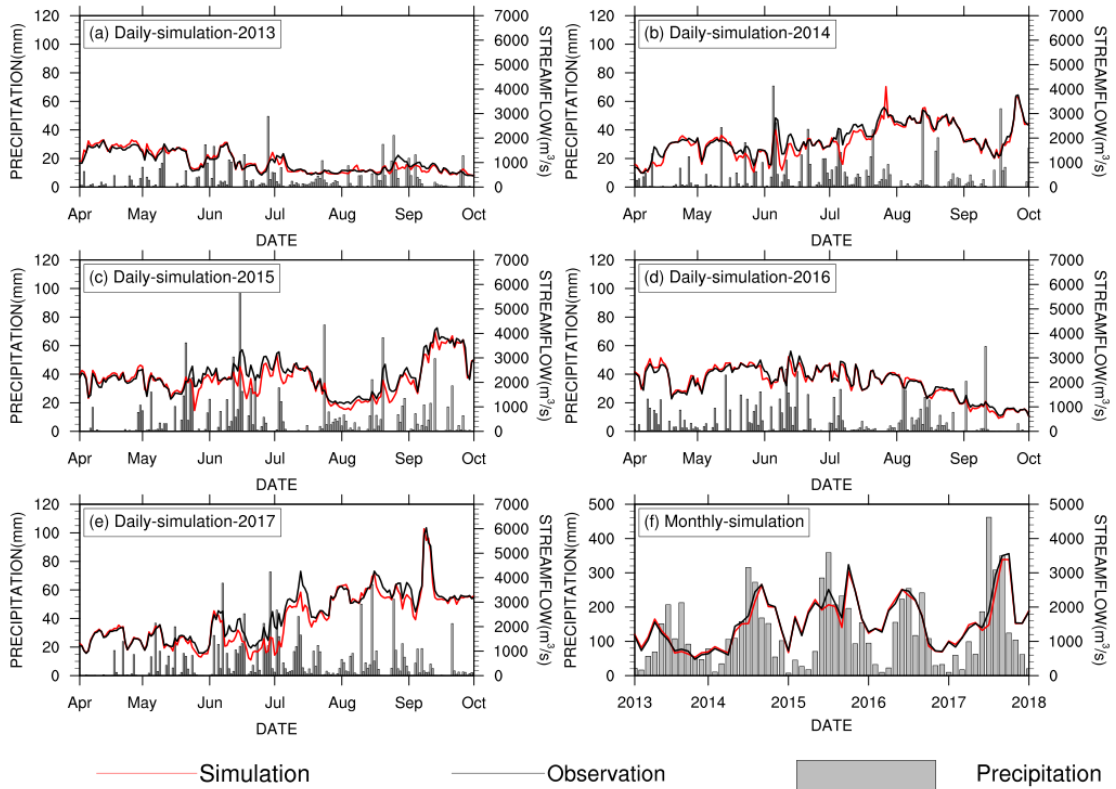
Figure 4. Evaluation of streamflow simulations at Yantan gauge. The black and red

lines are observed and simulated streamflow. (a)-(e) are for daily streamflow, and (f)

is for monthly streamflow. The gray bars represent daily (or monthly) precipitation.

615

616
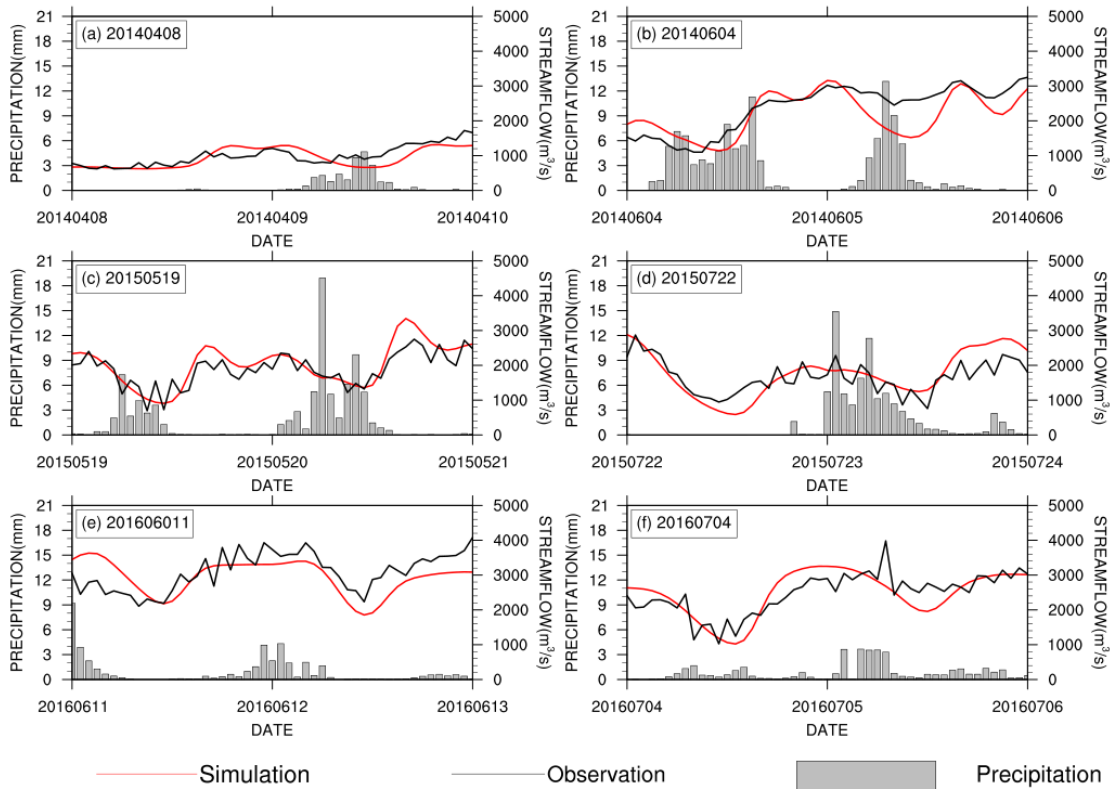
Figure 5. The same as Figure 4, but for the evaluation of hourly streamflow simulations at Yantan gauge.

619

620
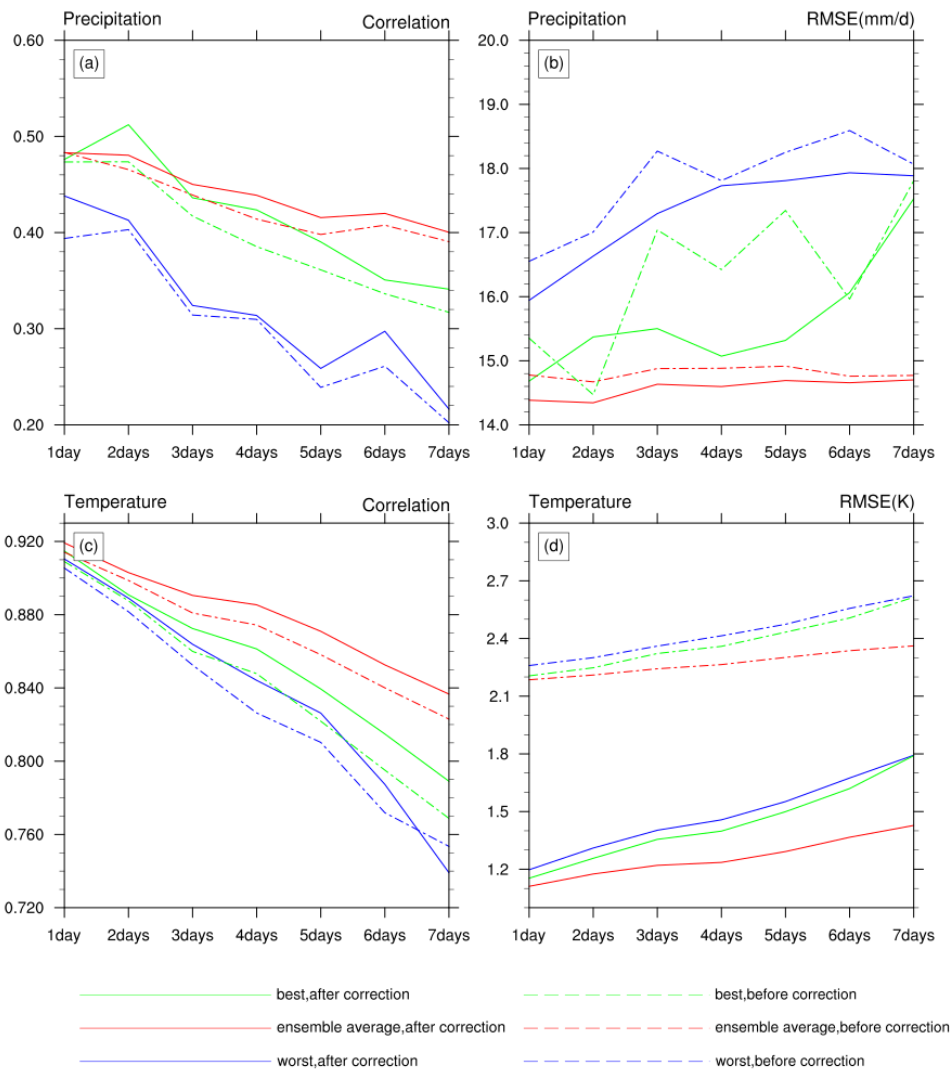
621

622 **Figure 6.** Evaluation of precipitation and temperature hindcasts from

623 TIGGE-ECMWF. The red and blue lines represent the best and worst results among 51

624 TIGGE-ECMWF ensemble members respectively, and the green lines represent the

625 results for the ensemble means of 51 members. Solid and dashed lines represent the

626 results after and before bias corrections, respectively.

627

628

629
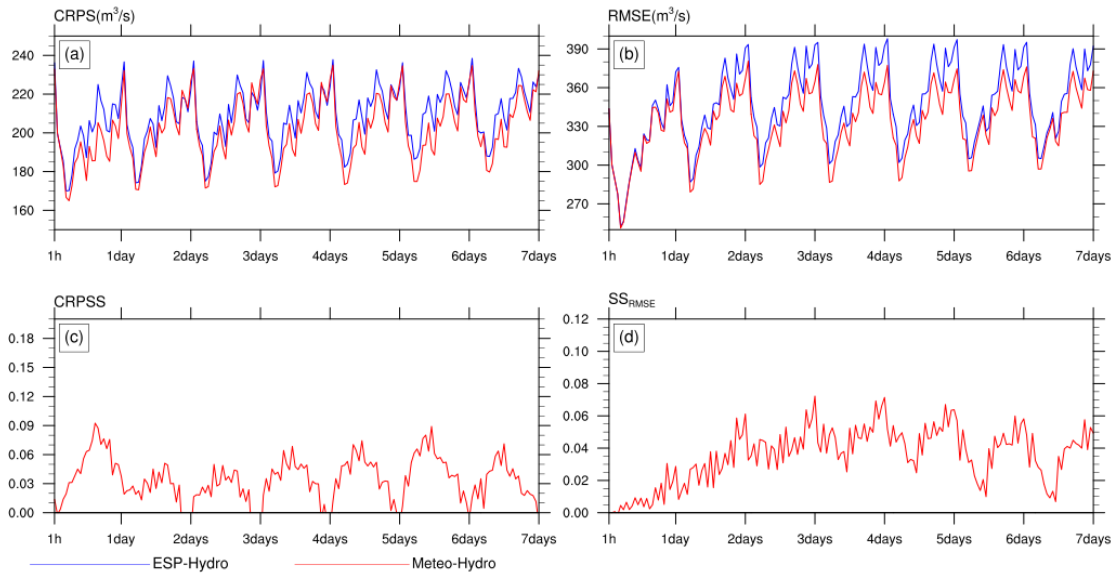
Figure 7. (a) Continuous Ranked Probability Score (CRPS) and (b) Root Mean Squared Error (RMSE) for daily streamflow ensemble forecasts at Yantan gauge. (c) and (d) are the skill score in terms of CRPS and RMSE for Meteo+Hydro, where ESP+Hydro is used as reference forecast.
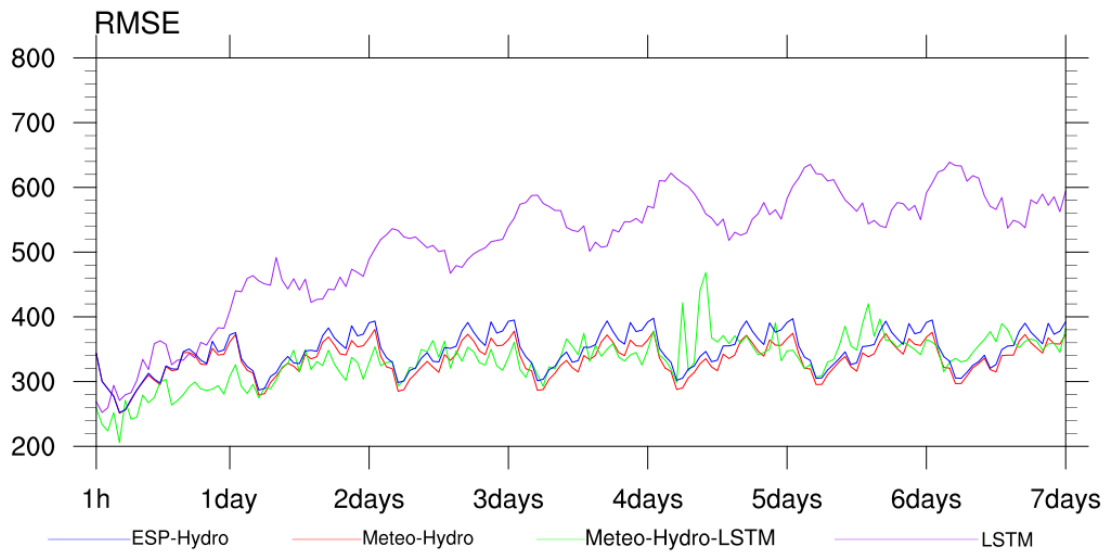
634

635

636
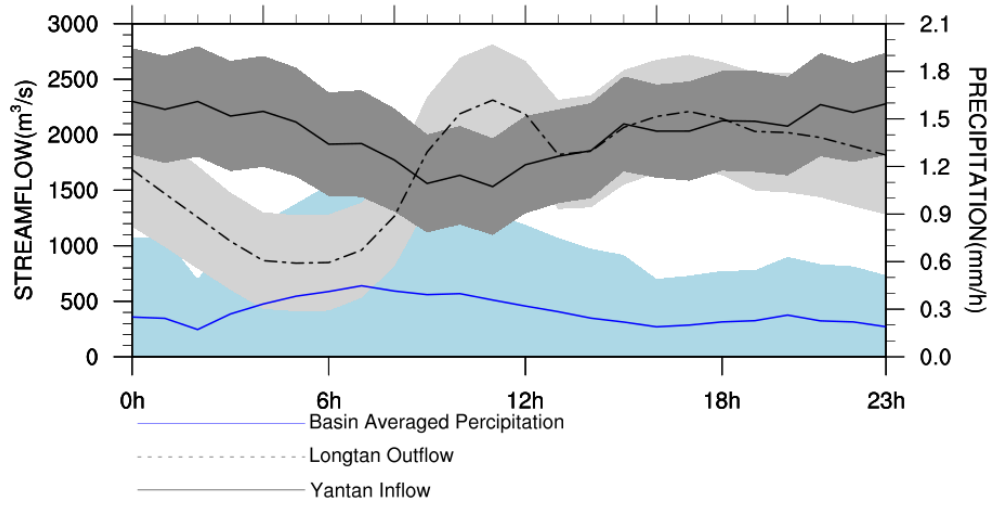
**Figure 8**. Diurnal cycle of Longtan outflow (m^3/s; dashed black line), Yantan inflow

(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.

640

641

642 **Figure 9.** RMSE (m^3/s) for hourly streamflow hindcasts from four forecast

643 approaches. The green line represents the Meteo+Hydro+LSTM forecast, the red line

644 represents the Meteo+Hydro forecast, the blue line represent the ESP+Hydro forecast,

645 and the purple line represents the LSTM forecast based on historical streamflow

646 observation alone.

647
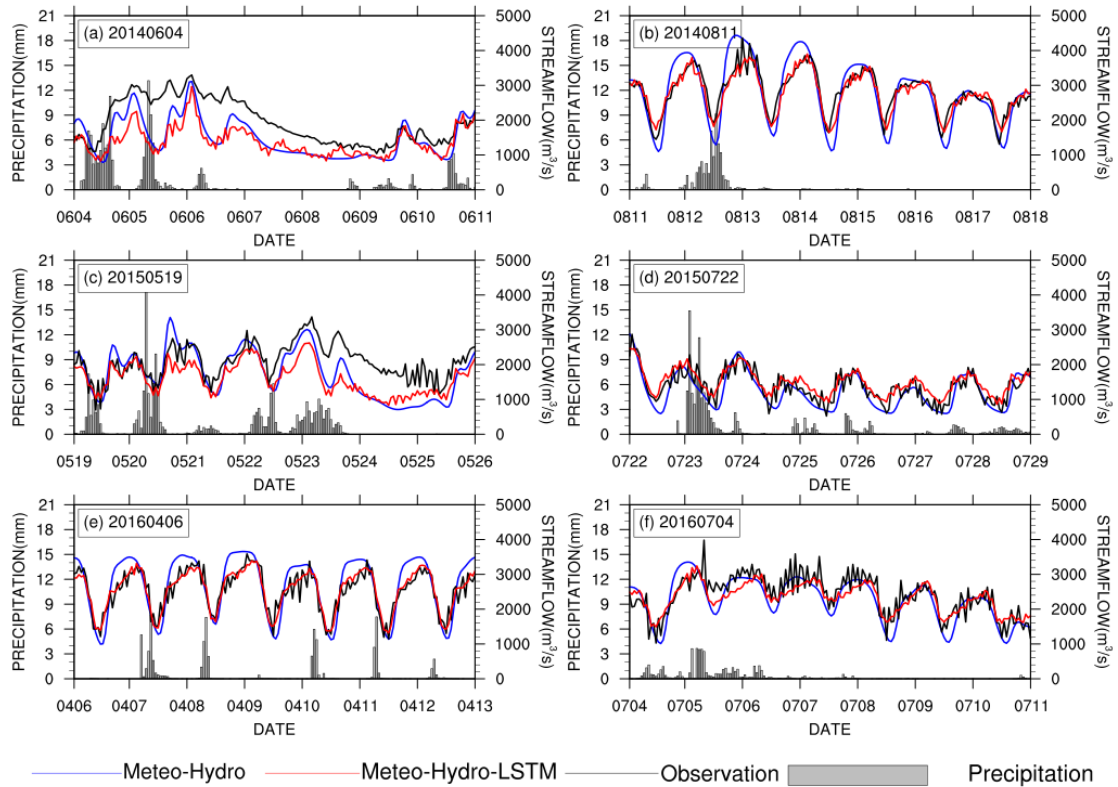
Figure 10. Evaluation of the forecast approaches for a few flooding events. The black lines are observed streamflow from Yantan hydrological gauge, the blue lines are the Meteo+Hydro ensemble mean streamflow forecast, and the red lines are the Meteo+Hydro+LSTM forecast streamflow by using Meteo+Hydro ensemble mean forecast with LSTM. The gray bars represent hourly precipitation averaged over the basin.