Extended-Range Runoff Forecasting Using a One-Way Coupled Climate–Hydrological Model: Case Studies of the Yiluo and Beijiang Rivers in China

Lüliu Liu [1,*], Chan Xiao [1], Liangmin Du [2], Peiqun Zhang [1] and Guofu Wang [1]

[1] National Climate Center, China Meteorological Administration, Beijing 100081, China; xiaochan@cma.gov.cn (C.X.); zhangpq@cma.gov.cn (P.Z.); wanggf@cma.gov.cn (G.W.)

[2] Wuhan Regional Climate Center, Wuhan 430074, China; duliangmin@outlook.com

[*] Correspondence: liull@cma.gov.cn; Tel.: +86-10-5899-5906

Abstract: Extended-range runoff forecasting is important for water resources management and energy planning. Experimental extended-range runoff was hindcasted, based on an extended-range climate model, developed by National Climate Center of the China Meteorological Administration, and semi-distributed hydrological model HBV-D. The skill of the runoff forecasts was explored using mean square skill score (MSSS), anomaly correlation coefficient (ACC), and areas under the relative operating characteristics curve (AUC) for three terciles for three experimental 51-day periods during flood season (June 1 to July 21, July 1 to August 20 and August 1 to September 20) for two rivers in China. The results revealed decreasing trends of the five indices, and varying length of the continuous longest skilful time slice from 3 days to 6 weeks depending on index, period and river location. In most cases, skilful abnormal terciles forecast occurred more often or with similar frequency to deterministic forecasts. It suggests that ensemble probability forecasting is a method with potential for extended-range river runoff forecast. Further, abnormal terciles are more skillful than normal terciles, and above normal are more skillful than below normal. In terms of a temporal mean of the MSSS and ACC, deterministic forecasts are skillful for both rivers in all three periods, but more skillful for the Beijiang River than for the Yiluo River in most cases.

Keywords: extended-range runoff forecast; DERF2.0; HBV; Yiluo River; Beijiang River

1. Introduction

Reliable and timely runoff forecasts have potential to provide critical information with lead times ranging from minutes to years. Such information is vital for numerous users, e.g., the emergency services, hydropower generators, irrigators, and rural and urban water supply authorities, as well as environmental managers in their role of mitigating disasters, controlling hydropower activities, determining industrial, domestic and agricultural water allocations and protecting the environment [1–4]. To meet the requirements of various users, in past half decades, couples of hydrological forecast models and forecasting systems have been proposed and developed. For example, a hydrometeorolgical ensemble predictions system (HEPS) was launched operationally in 2008 in Switzerland, which coupled a hydraulic model and a semi-distributed hydrological model with climate forcing from two climate models [5]. The Australian Bureau of Meteorology regularly issues seasonal stream flow forecasts for various river basins based on the Bayesian joint probability model, including probability distribution, terciles and exceedance probability forecasts (http://www.bom.gov.au/water/ssf/). The National Weather Service Advance Hydrologic Prediction Service provides hydrologic forecasts including the chances of flood or drought for almost 4000
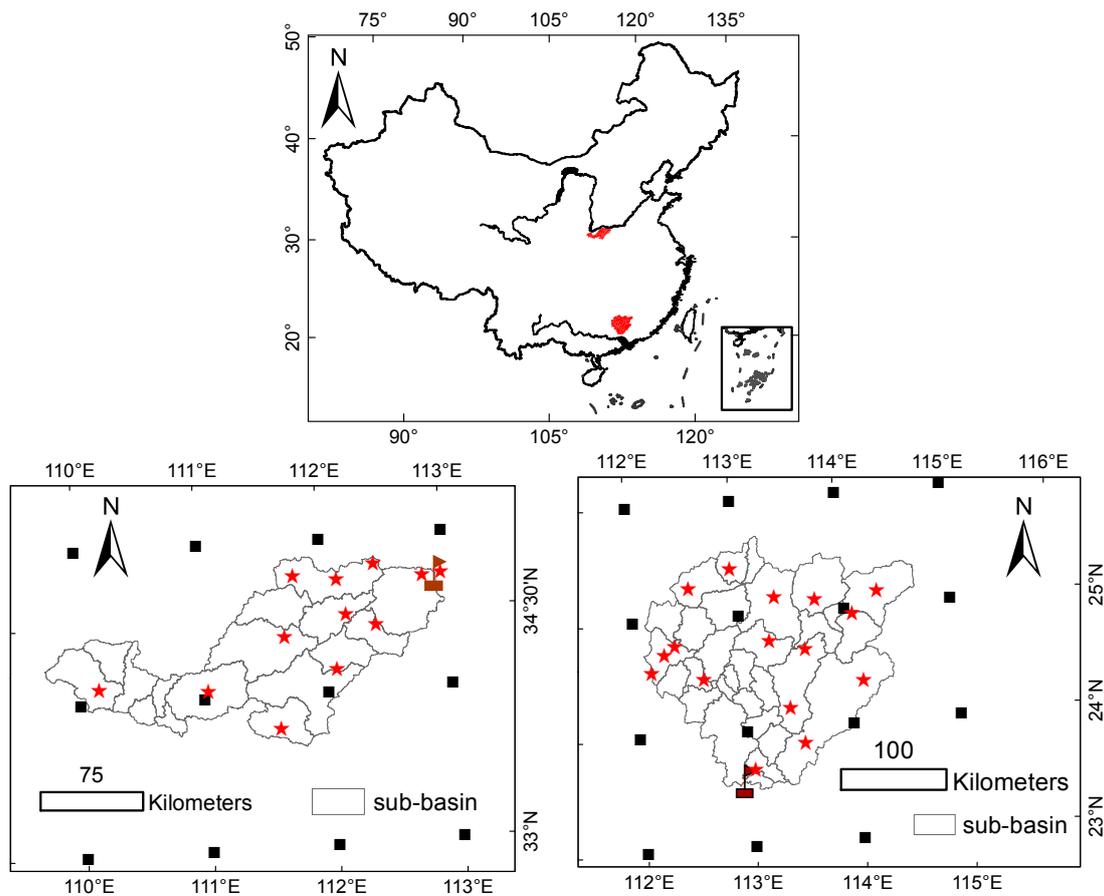
locations across the United States with forecast periods ranging from hours to months, using various models (https://water.weather.gov/ahps/forecasts.php). A hydrological outlook across the UK is published monthly, including terciles forecast for river flow and groundwater by expert merging of a statistical analogue and persistence method, Ensemble Streamflow Prediction method, and a hydrologic modeling system. (http://www.hydoutuk.net/latest-outlook/).

Data-driven models and process-based models represent two general approaches used in hydrological forecasting. Date-driven models do not explicitly incorporate hydrological processes [6] and therefore should be used with caution for conditions outside the observed domain. No definitive conclusion has been drawn regarding identification of the model capable of producing optimum runoff forecasts [7]. Process-based hydrological forecasts often contain large uncertainties that are attributable primarily to the hydrological models, initial conditions of the basin and atmospheric forcing [8]. In recent years, mixed approaches have emerged, which merges the merits of statistical and dynamical models i.e., [4,9]. Thus, numerous approaches or models have been investigated to improve hydrological forecasting. Among them, ensemble climate-model-based hydrological forecasting as an emerging field of interest since the early 21st century [10], is effective in achieving hydrological forecasts with longer lead times in comparison with the traditional Ensemble Streamflow Prediction method [11]. It has been used successfully both for cascading uncertainties and for improving the skill of hydrological forecasting [12–14]. However, obtaining accurate runoff forecasts remains a challenge, especially in relation to the medium and the long term during the flood season [7]. For example, errors or biases induced by coarse grid scales of outputs of climate models are still issues, although different post-processing such as bias correction and downscaling were proposed [4,9]. Process-based forecast uncertainties need further investigation using different methods such as ensemble techniques [15–19]. In addition, flood frequency distribution and runoff generation are worth analyzing to gain skillful runoff, especially flood forecasting [20]. Assessing the impacts of model selection and climate change is important [3,21].

The aim of this study was to explore the operational potential of extended-range runoff forecasts by one-way coupling an operational extended-range climate model and a hydrological model. It was achieved by assessing the assets and shortcoming of the approach for two rivers in China with different locations and different climate. Such research is fundamental regarding hydrological forecasting and for improving forecasting skills. The remainder of the paper is arranged as follows: Section 2, Study area, gives a brief introduction of the studied river basin. Section 3, Data and Method, introduces available climate and hydrological data, the hydrological model and climate model, experimental design of runoff forecasting, and verification methods. Section 4, Results, assesses both the performance of the hydrological model in relation to the two rivers and the skill of the deterministic and tercile runoff forecasts for the two river basins during three different periods in the flood season. Finally, Section 5, Discussion, and Section 6, Conclusions, are presented.

2. Study Area

The Yiluo River in China is the largest tributary between the middle and lower reaches of the Yellow River below the Sanmenxia Dam (33.5°–35° N, 109.5°–113.5° E). The basin has a drainage area of 18,881 km^2 and it supported a population of 7.7 million in 2010 [22]. The annual average mean temperature is 10–13 °C and the mean annual precipitation is about 660 mm. Dominated by the East Asian summer monsoon, 60% of the annual precipitation falls during the flood season (June–September) [23], which can easily cause flooding in summer. The annual runoff is approximately 2.74 billion m^3 [24]. Modulated by the seasonal precipitation pattern, the runoff during the flood season accounts for approximately 46.5% of the annual total [25]. In this study, the runoff through the Heishiguan hydrological station was forecasted. The area of the watershed above this station is 18,563 km^2 [23], which accounts for 98.3% of the total area of the Yiluo River basin. The locations of the Yiluo River and Heishiguan station are shown in Figure 1.

Figure 1. (**Upper**) Locations of the Yiluo and Beijiang rivers in China. Hydrological stations (dark red flag), climate stations (red star), grid notes of the dynamic extended-range forecast (DERF2.0) (black square) as well as sub-basins division of the (**lower left**) Yiluo River and (**lower right**) Beijiang River.

The Beijiang River is the second largest tributary of the Pearl River in southern China (23.1°–25.5° N, 111.9°–114.8° E. The basin has a drainage area of 46,710 km^2 [26] and it supported a population of 8.0 million in 2010, which is a value derived from gridded population data of China [27]. The annual mean temperature of the basin is 18–21 °C [28] and the mean annual precipitation varies within the range 712–4040 mm [29]. Overall, 75% of the annual precipitation falls during the flood season (April–September), which is when 70–80% of the annual runoff occurs [29,30]. In this study, the runoff through the Shijiao hydrological station was forecasted. The area of the watershed above the station constitutes 82.1% of the total area of the Beijiang River basin. The locations of the Beijiang River and Shijiao station are shown in Figure 1.

3. Data and Method

3.1. Available Datasets

This study used a digital elevation model (DEM) with scale 1:250,000, developed by the National Geomatics Center of China [31] and a digital land use map with scale 1:500,000, developed by the Ministry of Land and Resources of China [32]. Spatial soil data with 30 arc-second resolution and soil properties were generated from a harmonized world soil database [33]. Their spatial resolution was to resample to 1 km when setting up the hydrological model.

Climate data included observed and simulated daily precipitation and daily mean air temperatures. The observed meteorological records at 12 climate stations (Figure 1) obtained during 1968–2016 were acquired from the National Meteorological Information Center of the China Meteorological

Administration, and were used to calibrate and validate the hydrological model. Simulated climate data during 1983–2016 were hindcasted using the dynamic extended-range forecast (DERF2.0) operational system, developed by National Climate Center of the China Meteorological Administration. They were used to reforecast runoff as climate forcing of the hydrological model.

Digital daily flow records during 1970–2000 at the Heishiguan hydrological station were obtained from the Yellow River Resources Bureau. Digital daily flow records during 1970–1990 and monthly flow records during 1991–2000 at the Shijiao hydrological station were obtained from the Pearl River Resources Bureau. The Heishiguan and Shijiao data were used to calibrate and validate the HBV-D model for the Yiluo River and the Beijiang River, respectively. To evaluate the performance of the hydrological model for daily flow through Shijiao station during 1991–2000, daily values were estimated using the recorded monthly flow data.

3.2. Methods

3.2.1. Hydrological Model

HBV-D is a semi-distributed basin-scale hydrological model. It divides a basin into small hydrological units based on certain parameters, e.g., elevation, land cover type and climate conditions. It is a derivative of the Nordic HBV model. Compared with the original HBV, HBV-D improves the description of land cover characteristics and has more physically sound evapotranspiration schemes [34]. HBV-D has been applied to studies in China across the scales from small watersheds to large basins covering various climatic conditions [22,35–37], because of its simple structure and climate inputs, which only asks for daily temperature and precipitation.

The parameters of the HBV-D model for the two river basins were calibrated and validated using observed daily climate and streamflow data during two periods, i.e., 1970–1980 and 1981–2000, respectively. The performance of the model was evaluated using the Nash–Sutcliffe efficiency (NSE), coefficient of determination (R^2) and percentage bias (PBIAS). Generally, performance of a model can be considered satisfactory if the value of both NSE and R^2 is >0.50 and that of PBIAS is <±25% for flow simulation [38]. Although the natural hydrology of both studied rivers has been disturbed by human activities, the human alteration of the natural flow was not considered in the hydrological model because of the difficulty in recreating the undisturbed conditions.

3.2.2. Climate Model

DERF2.0, the dynamic extended-range forecast operational system version 2.0, was developed by National Climate Center of the China Meteorological Administration. It was based on the atmospheric circulation model BCC_AGCM2.2, which is a spectral model with horizontal T106 truncation resolution (~1.0° × 1.0°) and 26 vertical levels. DERF2.0 has operated daily since June 2014, and produces 57-day daily climate forecasts for four ensemble members using a lagged average forecasting strategy with 6-h intervals of atmospheric initial conditions. It has higher skill for precipitation and temperature forecast in China than its previous version DERF1.0, and ever skillfully forecasted flood in 1998 and drought in 2016 in China [39].

3.2.3. Runoff Forecasting

In general, runoff forecasting can be divided into four processing steps: (1) the hydrological model setup, (2) climate prediction processing, (3) runoff prediction, and (4) post-processing activities. The hydrological model comprises a process of selecting, parameterizing and linking a suite of models to simulate the hydrologic system to be forecasted. Its aim is to reduce the uncertainty associated with the hydrological model. The aim of climate prediction processing is to remove bias from the raw output of the climate model, or to interpolate the climate prediction of the model depending on the requirements of the hydrological model. Post-processing activities are intended to provide a probability prediction of runoff by combining the ensemble forecast and the reforecast information.

In this study, semi-distributed hydrological model HBV-D was selected, because it only requires daily temperature and precipitation as climate input. It can simulate runoff of the study area well [22] using the climate forcing from Water and Global Change Program (WATCH) meteorological forcing dataset [40]. However, the model was parameterized once more using observed climate forcing. Following this, the forecasted daily temperature and precipitation from DERF2.0 were interpolated to the center of each sub-basin using the inverse distance weighting method, and were corrected using the percentile method [41] to reduce bias and variance. Thirdly, runoff was forecasted using calibrated HBV-D forced by corrected climate with the hydrological initial conditions obtained by spinning up the HBV-D model for two years using observed climate data. Following these steps, ensemble runoff for three periods: June 1 to July 21 (Period1), July 1 to August 20 (Period2) and August 1 to September 20 (Period3) during 1991–2016 was reforecasted by the HBV-D simulation using the climate forcing of 24 members at six integrations of lead times. The dates of the DERF2.0 initial conditions, forecasted periods and leads are listed in Table 1. Finally, the ensemble mean and three terciles forecasts for the three periods in every year during 1991–2016 were estimated based on the hindcast dataset.

Table 1. DERF2 operation date/forecasted periods and ensemble members lead times.

Date of DERF2.0 Initial Conditions/Forecasted Period			Lead Time
Period1	Period2	Period3	
May 26/Jun 1 to Jul 21	Jun 25/Jul 1 to Aug 20	Jul 26/Aug 1 to Sep 20	7–57 days
May 27/Jun 1 to Jul 21	Jun 26/Jul 1 to Aug 20	Jul 27/Aug 1 to Sep 20	6–56 days
May 28/Jun 1 to Jul 21	Jun 27/Jul 1 to Aug 20	Jul 28/Aug 1 to Sep 20	5–55 days
May 29/Jun 1 to Jul 21	Jun 28/Jul 1 to Aug 20	Jul 29/Aug 1 to Sep 20	4–54 days
May 30/Jun 1 to Jul 21	Jun 29/Jul 1 to Aug 20	Jul 30/Aug 1 to Sep 20	3–53 days
May 31/Jun 1 to Jul 21	Jun 30/Jul 1 to Aug 20	Jul 31/Aug 1 to Sep 20	2–52 days

3.2.4. Runoff Reforecast Verification

The runoff reforecast during 1983–2016 was verified for Period1, Period2 and Period3 using indices of the mean square skill score (MSSS), anomaly correlation coefficient (ACC), and relative operating characteristics (ROC) for three terciles.

The MSSS, which is applicable to non-categorical deterministic forecasts only, provides a comparison of forecast performance relative to forecasts of climatology. The MSSS was estimated using Equations (1)–(3). Further details regarding the MSSS are available in the literature [42]. In this study, the MSSS was used to compare the ensemble mean of the runoff reforecast with the reference runoff, which was estimated using the multi-year mean of simulated runoff driven by the observed climate. A forecast can be considered perfect if the MSSS = 1, whereas it is determined worse than the reference forecast if the MSSS < 0.

$$\mathrm{MSSS} = 1 - MSE/MSE^* \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(R_i^{sim} - R_i^{obs}\right)^2 \tag{2}$$

$$MSE^* = \left(\frac{n}{n-1}\right)^2 \sum_{i=1}^{n}\left(\overline{R^{obs}} - R_i^{obs}\right)^2 \tag{3}$$

Here, MSE and MSE^* are the mean square error of the ensemble mean of the runoff forecast and the climatology forecast, respectively, R_i^{sim} and R_i^{obs} are the ensemble mean of the runoff forecast and the observed runoff, respectively, in year i and $\overline{R^{obs}}$ is the mean of R_i^{obs} during 1983–2016.

The ACC is another index with which to evaluate climate-hydrological forecasts. It is regarded as a measure of the strength and direction of the linear relationship between observations and the forecast. A forecast can be considered perfect if the ACC = 1, whereas it is determined worse than the climatology if the ACC < 0.

The ROC can also be used for verification of probability forecasts [42]. In this study, the standardized areas under ROC curves (AUCs) for three tercile categories of daily runoff were calculated. A forecast system can be considered perfect if the AUC = 1, whereas a forecast system can be considered to have no valid information if the AUC = 0.5. A forecast system for which the value of the AUC is >0.5 can be considered to have certain skill. The areas for above normal, near normal and below normal are denoted AUC_A, AUC_N and AUC_B, respectively.
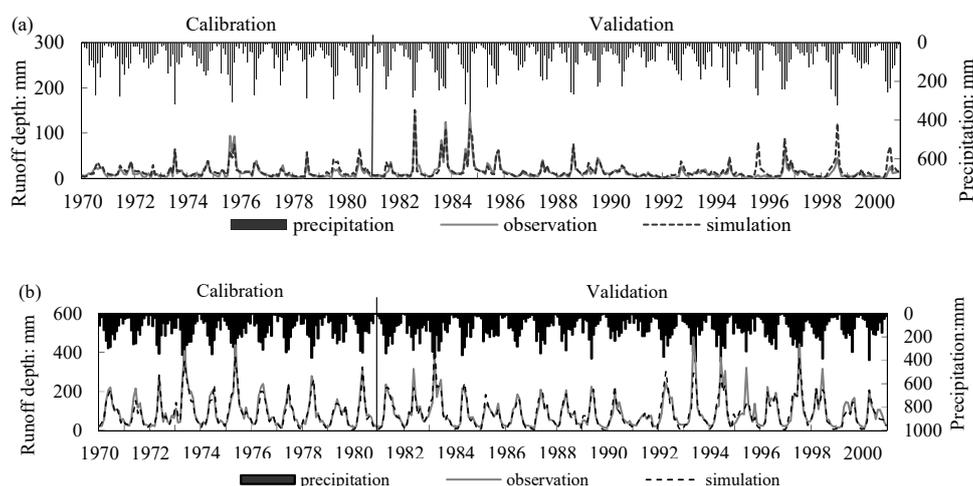
4. Results

4.1. Calibration and Validation of the HBV-D Model

For the calibration and validation periods, the model performance statistics are all considered satisfactory in relation to the Yiluo and Beijiang rivers (Table 2), i.e., both NSE and R^2 are >0.5 and PBAIS is <±25% for both periods. Additionally, the HBV-D model is also shown well able to simulate monthly runoff using gridded climate forcing (NSE > 0.73, R^2 > 0.80 and PBIAS < ±15%) during 1971–2000 for monthly flow through the two hydrological stations [16]. From the hydrographs shown in Figure 2, it can be seen that the monthly and seasonal patterns of observed runoff are reproduced well. This, together with the successful application of the HBV-D model in different climatic regions, verified that the model could be used to predict the runoff anomaly when forced with accurate precipitation and temperature forecasts and proper initial conditions.

Table 2. Performance of the HBV-D model in daily runoff simulation.

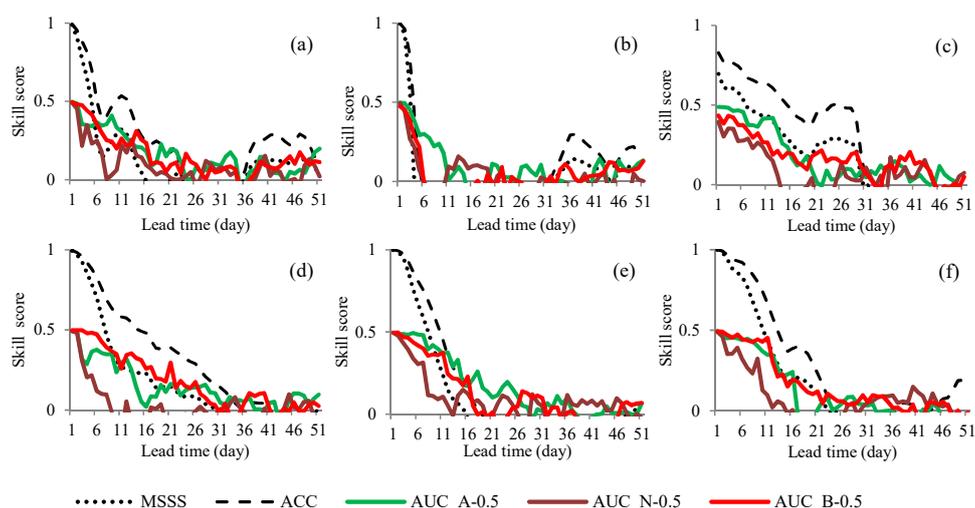| Hydrological Station (Calibration/Validation Periods)	Series	NSE	R^2	PBIAS (%)
Heishiguan (1970–1980/1981–2000)	Daily	0.64/0.56	0.81/0.76	18.0/17.6
	Monthly	0.77/0.71	0.90/0.88	
Shijiao (1970–1980/1981–2000)	Daily	0.74/0.81	0.87/0.90	−5.9/−0.9
	Monthly	0.94/0.93	0.97/0.97	

Figure 2. Monthly precipitation and observed/simulated runoff at (**a**) Heishiguan station on the Yiluo River and (**b**) Shijiao station on the Beijiang River for periods of calibration (1970–1980) and validation (1981–2000).

4.2. Runoff Forecast Skill

As shown in Figure 3, the daily MSSS shows a declining trend. The length of the skilful time slice varies with forecasted period and river location. Among the three periods, the MSSS reduces most rapidly for both rivers in Period2. As listed in Table 3, the shortest time slice lasts only three days for the Yiluo River in Period2, whereas the longest time slice lasts one month for the Yiluo River in

Period3 and for the Beijiang River in Period1. Thus, the highest mean MSSS for the Yiluo River (0.19) is in Period3 and that for the Beijiang River (0.16) is in Period1, whereas the corresponding lowest values for both rivers (0 and 0.07, respectively) are in Period2 (Table 3). The intermediate skilful time slice for the Yiluo and Beijiang rivers lasts two or three weeks with mean MSSS values of 0.13 and 0.12, respectively. Additionally, the daily skilful forecasts on other days can be observed in Figure 3 and they can be inferred from Table 3. Skill is evidently higher (lower) for the Beiijiang River than for the Yiluo River for Period1 and Period2 (Period3) in terms of mean MSSS and length of the first skilful time slice.

Figure 3. Skilful daily indices for (**a**–**c**) the Yiluo River and (**d**–**f**) the Beijiang River for Period1, Period2 and Period3, respectively (Only positive values of the mean square skill score (MSSS) and anomaly correlation coefficient (ACC), and AUC-0.5 are illustrated to show skilful days. **a** and **d**, **b** and **e**, **c** and **f** are for Period1, Period2 and Period3 respectively.).

Table 3. Temporal mean of MSSS, ACC and AUC during the 51-day forecasted periods.

River	Index	Jun 1 to Jul 21 Period1	Jul 1 to Aug 20 Period2	Aug 1 to Sep 20 Period3
	MSSS	0.13/15/33	0.0/3/22	0.19/30/30
	ACC	0.26/23/39	0.07/5/27	0.25/31/31
Yiluo River	AUC_A	0.82/34/47	0.58/14/25	0.67/21/47
	AUC_N	0.66/21/43	0.53/5/24	0.57/12/40
	AUC_B	0.78/27/48	0.53/5/26	0.67/17/35
	MSSS	0.16/30/30	0.07/15/18	0.12/23/23
	ACC	0.30/41/41	0.14/18/22	0.28/24/38
Beijiang River	AUC_A	0.65/39/48	0.90/28/44	0.61/15/36
	AUC_N	0.52/8/26	0.65/12/40	0.56/14/33
	AUC_B	0.67/30/42	0.84/18/35	0.65/37/42

Note: The three numbers (left to right) in each row of the columns for each period represent the value of the index, number of days of the first continuous skilful time slice and number of all skilful days, respectively.

The variation of the ACC is similar to that of the MSSS. The lowest mean ACC value is also observed for Period2, but the highest observed value for both rivers is in Period1, i.e., 0.07 and 0.26 respectively for the Yiluo River, and 0.14 and 0.30 respectively for the Beijiang River. Compared with the MSSS, the ACC score is obviously higher, i.e., the first continuous skilful time slice lasts longer and the total skilful days are more, as illustrated in Figure 3 and listed in Table 3. The longest daily skilful time slice lasts about six weeks for the Bejiang River in Period1, which is 11 days longer than the MSSS. The shortest daily skilful time slice lasts five days for the Yiluo River in Period2, which is two days longer than the MSSS. The most days with positive daily ACCs is seen in Period1, i.e., 6 (11) days more
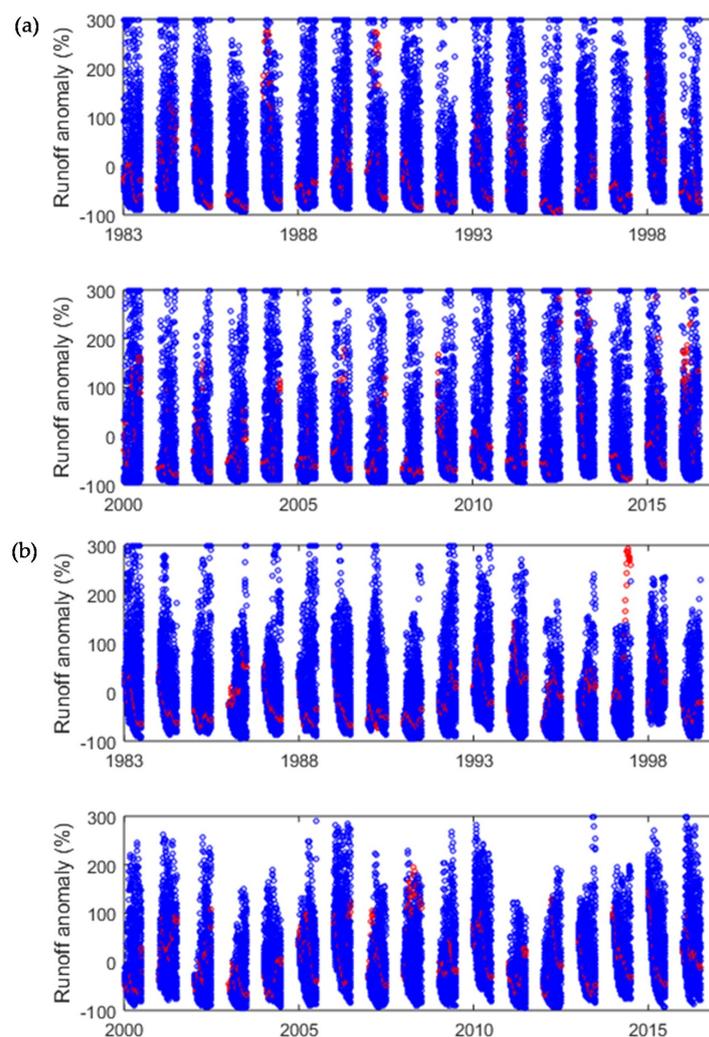
for the Yiluo River (Beijiang River) than the corresponding number for the MSSS. Moreover, the least skilful days of the ACC are 22 days, which is 5 days more than the MSSS. Compared with the Yiluo River, the skill is higher in terms of the mean ACC in three periods and the length of the first skilful time slice of the ACC in the first two periods, while it is lower for Period3 for the Beijiang River.

Trends that first decline and then become flat can be observed in the AUC curves for the three terciles in Figure 3. The continuous daily skilful time slices for the abnormal terciles are longer than the normal terciles, and those for above normal terciles are longer than the below normal terciles in five out of six cases. The longest slice lasts 39 days, while the shortest lasts 5 days. However, the most skilful days are 48, while the least skilful days are 26 among 9 cases. The maximum and minimum values of the mean AUC_A are 0.82 and 0.58 (0.90 and 0.61), respectively, for the Yiluo River (Beijiang River).

In most cases, skilful AUC days for abnormal terciles occur more often than or with similar frequency to deterministic forecasts.

4.3. Reproducing Ability of Ensemble Runoff Forecast

Figure 4 illustrates the reproduction of the ensemble forecast for the two rivers during Period1. Some daily observations fall outside the range of the ensemble forecasts on certain days for both rivers (Figure 4), as well as for other periods (not shown), which partially explains the negative indices on somedays.

Figure 4. Daily ensemble runoff reforecast (blue dots) and observations (red dots) for (**a**) the Yiluo River and (**b**) the Beijiang River during Period1 1983–2016.

5. Discussion

Decreasing trends of the five indices were observed during the three 51-day periods, and the length of the skilful time slices varied with forecast period and geographic location. This might be attributable to a decrease of the skill of climate forcing with lead time and to seasonal and regional variations. Previous studies have indicated that the skill of precipitation and air temperature forecasts drops markedly beyond a period of one month [43,44]. However, it is interesting that skilful daily forecasts were detected in this study after some negative values were derived. It could be a result of the combined influence of initial hydrological conditions and climate forcing, together with the variation of their relative contributions with lead time [11], season, land surface condition and river location.

Runoff forecast skill was found higher for the Beijiang River than the Yiluo River in most deterministic forecast cases explored. This could be attributed partially to better performance of the HBV-D model for the Beijiang River than the Yiluo River (Table 2). The opposite case found for Period3 might be attributable partially to the slightly lower precipitation skill of DERF2.0 for the Beijiang River than the Yiluo River in terms of the MSSS and ACC in August and September (see Figures 21 and 23 of [45]). However, it will be necessary to conduct further studies to investigate the reasons.

It could be expected that runoff forecast skill would be improved when using a multi-model ensemble, which outperforms a single-model ensemble in terms of probabilistic forecast skill and reliability [3,44], and when conducting post-processing of the coupled global climate model and/or hydrological outputs [4,14,46–48]. However, only a slight skill improvement was detected for the runoff forecasts driven by corrected climate data in comparison with those using raw DERF2.0 output (results not shown). Moreover, it would be potentially more beneficial to use multiple climate models rather than multiple hydrological models for a given river basin. This is because the uncertainties associated with climate model forecasts are more likely to be the main reason for the low skill in seasonal hydroclimate forecasting in comparison with those associated with the initial conditions and the hydrological models [14].

This study assessed the potential of the proposed extended-range hydrological model in two rivers targeting three periods. It is vital for a possible operational system to assess its skill in more river basins for more periods, and to improve skill by using various approaches such as multi-model ensemble, model parameters calibrating, and post-processing.

6. Conclusions

All five daily indices examined in this study showed decreasing trends for both rivers during the three periods considered. The highest daily value derived was almost equal to 1. Moreover, the length of the continuous skilful time slice was found to depend on the index, period and river location. The longest continuous skilful time slice was about six weeks, whereas the shortest was only three days for the deterministic forecasts and five days for the probabilistic forecasts.

The mean of the deterministic forecast skill during the 51-day periods was found higher for the Beijiang River than for the Yiluo River in all three periods, except for in the terms of MSSS in Period3. In terms of the mean MSSS, the lowest skill (<0.8) was observed in Period2 for both rivers among the three periods, but the highest skill was 0.19 in Periods3 for the Yiluo River and was 0.16 in Period1 for the Beijiang River. In terms of the mean ACC, the lowest skill was 0.07 and 0.14 for Periods2, but the highest skill was 0.26 and 0.30 for Periods1 for the Yiluo River and Beijiang River respectively, among the three periods.

In producing the ensemble probabilistic runoff forecasts, in most cases, the continuous skilful time slices for the abnormal terciles were found longer than for the normal terciles, and those for the above normal terciles were found longer than for the below normal terciles.

Ensemble probability forecasting is a method with potential for extended-range river runoff forecasting, compared with deterministic methods. In most cases, when using the DERF2.0 climate model and the HBV-D hydrological model, it was found that skilful AUC days for abnormal terciles occurred more often or with similar frequency to deterministic forecasts. This study confirms

that ensemble runoff forecasting is superior to deterministic forecasting even when applying a single-model ensemble.

References

1. Fan, Y.R.; Huang, G.H.; Li, Y.P.; Wang, X.Q. Probabilistic prediction for monthly streamflow through coupling stepwise cluster analysis and quantile regression methods. *Water Resour. Manag.* **2016**, *30*, 5313–5331. [CrossRef]
2. Hopson, T.M.; Webster, P.J. A 1–10-day ensemble forecasting scheme for the major river basins of bangladesh: Forecasting severe floods of 2003–07. *J. Hydrometeorol.* **2010**, *11*, 618–641. [CrossRef]
3. Yuan, X. An experimental seasonal hydrological forecasting system over the Yellow River basin—Part 2: The added value from climate forecast model. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2453–2466. [CrossRef]
4. Crochemore, L.; Ramos, M.H.; Pappenberger, F. Bias correcting precipitation forecast to improve the skill of seasonal streamflow forecast. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3601–3618. [CrossRef]
5. Addor, N.; Jaun, S.; Fundel, F.; Zappa, M. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): Skill, case studies and scenarios. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2327–2347. [CrossRef]
6. Shortridge, J.E.; Guikema, S.D.; Zaitchik, B.F. Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrol. Earth Syst. Sci.* **2016**, *7*, 2611. [CrossRef]
7. Tan, Q.F.; Lei, X.H.; Wang, X.; Wang, H.; Wen, X.; Ji, Y.; Kang, A.Q. An adaptive middle and long-term runoff forecast model using EEME-ANN hybrid approach. *J. Hydrol.* **2018**, *567*, 767–780. [CrossRef]
8. Yang, L.; Tian, F.; Sun, Y.; Yuan, X.; Hu, H. Attribution of hydrologic forecast uncertainty within scalable forecast windows. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 775786. [CrossRef]
9. Yuan, X.; Wood, E.F.; Ma, Z. A review on climate-model based seasonal hydrologic forecasting: Physical understanding and system development. *Wiley Interdiscip. Rev. Water* **2015**, *2*, 523–536. [CrossRef]
10. Wood, A.W.; Mauer, E.P.; Kumar, A.; Lettenmaier, D.P. Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.* **2002**, *107*, 449. [CrossRef]
11. Yuan, X. An experimental seasonal hydrological forecasting system over the Yellow River basin—Part 1: Understanding the role of initial hydrological conditions. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2437–2451. [CrossRef]
12. Pappenberger, F.; Beven, K.J.; Hunter, N.M.; Bates, P.B.; Gouweleeuw, B.T.; Thielen, J.; de Roo, A.P.J. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecast in System (EFFS). *Hydrol. Earth Syst. Sci.* **2005**, *9*, 381–393. [CrossRef]
13. Cloke, H.L.; Pappenberger, F. Ensemble flood forecasting: Review. *J. Hydrol.* **2009**, *375*, 613–626. [CrossRef]
14. Yuan, X.; Wood, E.F.; Roundy, J.K.; Pan, M. CFSv2-based seasonal hydroclimatic forecast over the conterminous United States. *J. Clim.* **2013**, *26*, 4828–4847. [CrossRef]
15. Stensrud, D.J.; Brooks, H.E.; Du, J.; Tracton, M.S.; Rogers, E. Using ensembles for short-range forecasting. *Mon. Weather Rev.* **1999**, *127*, 433–446. [CrossRef]
16. Brown, J.D.; Heuvelink, G. Assessing uncertainty propagation through physically based models of soil water flow and solute transport. In *Encyclopedia of Hydrological Sciences*; Anderson, M., Ed.; John Wiley and Sons: Hoboken, NJ, USA, 2005; pp. 1181–1195.
17. Olsson, J.; Lindström, G. Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. *J. Hydrol.* **2008**, *350*, 14–24. [CrossRef]

18. Brown, J.D.; Seo, D.J. A noparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *J. Hydrometeorol.* **2010**, *11*, 642–665. [CrossRef]

19. Najafi, M.R.; Moradkhani, H.; Piechota, T.C. Ensemble streamflow prediction: Climate signal weighting methods vs. climate forecast system reanalysis. *J. Hydrol.* **2012**, *442*, 105–116. [CrossRef]

20. Iacobellis, V.; Fiorentino, M.; Gioia, A.; Manfreda, S. Best fit and selection of theoretical flood frequency distributions based on different runoff generation mechanisms. *Water* **2010**, *2*, 239–256. [CrossRef]

21. Najafi, M.R.; Moradkhani, H.; Jung, I.W. Assessing the uncertainties of hydrologic model selection in climate change impact studies. *Hydrol. Process.* **2011**, *25*, 2814–2826. [CrossRef]

22. Liu, L.; Xu, H.; Wang, Y.; Jiang, T. Impacts of 1.5 and 2 °C global warming on water availability and extreme hydrological events in Yiluo and Beijiang River catchments in China. *Clim. Chang.* **2017**, *145*, 145–158. [CrossRef]

23. He, R.M.; Wang, G.Q.; Zhang, J.Y. Impacts of environmental change on runoff in the Yiluohe River basin of the middle Yellow River. *Res. Soil Water Conserv.* **2007**, *2*, 297–301. (In Chinese)

24. Wang, W.T.; Qian, L.X. Spatial distribution and seasonal variation of evapotranspiration in Yiluo River basin on MODIS data. *Resour. Sci.* **2012**, *34*, 1582–1590. (In Chinese)

25. Guo, W.N.; Zhao, W.J. Variation characteristics of runoff in the Yiluo River. *Water Resour. S. N. Water Diversion* **2015**, *21*, 61–62. (In Chinese)

26. Cheng, Z.L. Analysis on characteristic of floods and floods combining in the upstream of the Bejiang River. *Guangong Water Resour. Hydropower* **2006**, *19*, 46–49. (In Chinese)

27. Fu, J.Y.; Jiang, D.; Huang, Y.H. Population Grid_China. *Acta Geogr. Sinica* **2014**, *69*, 41–44. (In Chinese)

28. Gao, C.Y.; Zhao, Q.H.; Liu, Q. Fractal characteristic of soil particle size under different vegetation types in riparian zone of the main stream of Beijiang River. Res. *Soil Water Conserv.* **2016**, *23*, 37–42.

29. He, L.; Chen, W.Y. An analysis of water regime in Beijiang's lower reaches. *J. Foshan Univ.* **2003**, *21*, 67–69. (In Chinese)

30. Huang, G.R.; Wu, C.H.; Liu, Z.Y.; Chen, Z.J.; Hu, J.W.; Yin, Z.J. Precipitation of extreme floods of the Feilaixia reservoir in the Beijiang River basin under climate change. *Adv. Water Sci.* **2015**, *26*, 10–19. (In Chinese)

31. Shang, Y.L.; Wang, D.H.; Li, L. Development and updation of national dataset at a scale of 1:250,000. *Sci. Surv. Mapp.* **2004**, *29*, 92–95. (In Chinese)

32. Zha, Z.X.; Wu, M.H.; Ren, X.Y. Dataset of national land-use at a scale of 1:500,000. *Land Resour. Informatiz.* **2001**, *2*, 17–21. (In Chinese)

33. FAO; IIASA; ISRIC; ISS-CAS; JRC. *Harmonized World Soil Database*; Version 1.0; FAO: Rome, Italy; IIASA: Laxenburg, Austria, 2008.

34. Krysanova, V.; Bronstert, A.; Müller-Wohlferl, D.I. Modelling river discharge for large drainage basins: From lumped to distributed approach. *Hydrol. Sci. J.* **1999**, *44*, 313–333. [CrossRef]

35. Fan, J.; Jiang, Y.A.; Chen, Y.; Chen, P.X.; Bai, S.Q.; Yu, X.J. The critical rainfall calculation in Kaidu River based on HBV hydrological model. *Desert Oasis Meteorol.* **2014**, *8*, 31–35. (In Chinese)

36. Su, B.; Huang, J.; Zeng, X.; Gao, C.; Jiang, T. Impacts of climate change on streamflow in upper Yangtze River basin. *Clim. Chang.* **2017**, *141*, 533–546. [CrossRef]

37. Liu, L.; Jiang, T.; Xu, H.; Wang, Y. Potential threats from variations of hydrological parameters to the Yellow River and Pearl River Basins in China over the next 30 years. *Water* **2018**, *10*, 883. [CrossRef]

38. Moriasi, D.N.; Arnold, J.G.; van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]

39. He, H.G.; Li, Q.P.; Wu, T.W.; Tang, H.Y.; Hu, Z.Y. Chinese temperature and precipitation evaluation of monthly dynamic extended range forecast operational system DERF2.0 in China. *J. Atmos. Sci.* **2014**, *38*, 950–964. (In Chinese)

40. Weedon, G.P.; Gomes, S.; Viterbo, P.; Österle, H.; Adam, J.C.; Bellouin, N.; Boucher, Q.; Best, M. *The WATCH Forcing Data 1958–2001: A Meteorological Forcing Dataset for Land Surface- and Hydrological-Models*. WATCH Technical Report No. 22. 2010, pp. 1–41. Available online: www.eu-watch.org (accessed on 1 February 2017).

41. Liu, L.L.; Ren, G.Y. Percentile statistical downscaling method and its application in the correction of GCMs daily precipitation in China. *Plateau Meteorol.* **2012**, *31*, 715–722. (In Chinese)

42. WMO. *Standardised Verification System (SVS) for Long-Range Forecasts (LRF): New Attachment II-8 to the Manual on the GDPFS (WMO-No. 485)*; WMO: Geneva, Switzerland, 2006; Volume I. Available online: http://www.bom.gov.au/wmo/lrfvs/users.shtml (accessed on 4 February 2017).

43. Lavers, D.; Luo, L.; Wood, E.F. A multiple model assessment of seasonal climate forecast skill for applications. *Geophys. Res. Lett.* **2009**, *3*, L23711. [CrossRef]

44. Yuan, X.; Wood, E.F.; Luo, L.; Pan, M. A first look at Climate Forecast System version 2(CFSV2) for hydrological seasonal prediction. *Geophys. Res. Lett.* **2011**, *38*, L13402. [CrossRef]

45. Wu, J.; Zhao, C.B.; Ren, H.L. *Report to Temperature, Precipitation, and Circulation Evaluation of Monthly Dynamic Extended Range Forecast Operational System DERF.0 over the World*; Laboratory for Climate Studies, China Meteorological Administration: Beijing, China, 2014. (In Chinese)

46. Luo, L.; Wood, E.F. Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the eastern United States. *J. Hydrometeorol.* **2008**, *9*, 866–884. [CrossRef]

47. Yuan, X.; Wood, E.F. Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resour. Res.* **2012**, *48*, W12519. [CrossRef]

48. Liu, L.L.; Du, L.M.; Liao, Y.M.; Li, Y.; Liang, X.Y.; Tang, J.Y.; Zhao, Y.H. Probability prediction of monthly precipitation in summer over Huaihe River in China on spatial-temporal statistical downscaling. *Meteorol. Mon.* **2018**, *44*, 1490–1496. (In Chinese)