Evaluation of the ECMWF System 4 climate forecasts for streamflow forecasting in the Upper Hanjiang River Basin
Fuqiang Tian, Yilu Li, Tongtiegang Zhao, Hongchang Hu, Florian Pappenberger, Yunzhong Jiang and Hui Lu

ABSTRACT
This paper assesses the potential of the European Centre for Medium-Range Weather Forecasts (ECMWF) System 4 forecasts and investigates the post-processing precipitation to enhance the skill of streamflow forecasts. The investigation is based on hydrological modelling and is conducted through the case study of the Upper Hanjiang River Basin (UHRB). A semi-distributed hydrological model, Tsinghua Representative Elementary Watershed (THREW), is implemented to simulate the rainfall–runoff processes, with the help of hydrological ensemble prediction system (HEPS) approach. A post-processing method, quantile mapping method, is applied to bias correct the raw precipitation forecasts. Then we evaluate the performance of raw and post-processed streamflow forecasts for the four hydrological stations along the mainstream of Hanjiang River from 2001 to 2008. The results show that the performance of the streamflow forecasts is greatly enhanced with post-processing precipitation forecasts, especially in pre-dry season (November and December), thus providing useful information for water supply management of the central route of South to North Water Diversion Project (SNWDP). The raw streamflow forecasts tend to overpredict and present similarly to forecast accuracy with the extended streamflow prediction (ESP) approach. Streamflow forecast skill is considerably improved when applying post-processing method to bias correct the ECMWF System 4 precipitation forecasts.

Key words | ECMWF System 4, hydrological ensemble prediction system, post-processing, Upper Hanjiang River Basin

INTRODUCTION
Seasonal streamflow forecasts are crucial for water resource management, including flood control, water supply, hydropower generation, irrigation and drought management (Maurer & Lettenmaier 2004; Graham & Georgakakos 2010; Zhao & Zhao 2014). Xu et al. (2014) proved that the use of forecasts can efficiently improve the overall system performance of cascaded hydropower systems in terms of power generation and reliability. Generally, there are two categories of seasonal streamflow forecasting methods: statistical methods and dynamical methods. Statistical methods formulate the dependence relationship between the given prediction, i.e., future seasonal streamflow, and the selected predictors, including antecedent streamflow and climatic indices and generate forecasts based on the values of predictors (Kalra & Ahmad 2009; Wang et al. 2009; Erdal & Karakurt 2013; Zhang et al. 2015). Zhao et al. (2016) applied Bayesian joint probability modelling approach to generate sub-seasonal to seasonal streamflow forecasts in Australia and found this modelling approach could provide water resource managers with more informative forecasts.
Dynamical methods rely on meteorological forecasts or past observations to drive hydrological models (Mohseni & Stefan 1998; Tucci et al. 2003; Paiva et al. 2013; Zhang et al. 2013; Zhou et al. 2016). Candogan Yossef et al. (2017) used a distributed global hydrological model, PCRaster Global Water Balance model, to produce monthly ensemble streamflow forecasts of 20 large rivers in the world. The extended streamflow prediction (ESP) method proposed by Day (1985) is an early applied dynamical approach that was widely used to seasonal streamflow forecasting. Day’s approach utilises current initial basin conditions and past observed weather data to force the hydrological model and produce ensemble streamflow traces (Hashino et al. 2007; Yang et al. 2015). The underpinning assumption of the ESP approach is that the historical meteorological data (precipitation, temperature, etc.) are substantial representatives of regional conditions that have the same probability of occurrence in the future. In other words, this method does not take into account the forecaster’s knowledge of the climate system. An alternative dynamical method, hydrological ensemble prediction system (HEPS) approach, uses the seasonal forecasts from general circulation models (GCMs) to drive hydrological models. This method has become widely used in seasonal streamflow forecasting, due to increased accuracy and resolution of GCM forecasts (Franz et al. 2003; Gobena & Gan 2010).

Currently, several operational meteorological centres regularly provide GCM seasonal forecasts, for instance, the ECMWF, the United States National Center for Environmental Prediction (NCEP), the UK Meteorological Office (UKMO) and the Australian Bureau of Meteorology. For the ECMWF, the first version of their system (System 1) has been operating since 1997, while the fourth version (System 4) has been operating from November 2011 to October 2017. Several studies have evaluated the precipitation forecast issued by ECMWF System 4 globally as well as in East Asia and China. Weisheimer & Palmer (2014) appraised reliability of the precipitation forecasts from the ECMWF System 4 at a global scale. They defined five reliability categories ranging from ‘perfect’ to ‘dangerous’. In most parts of the world, forecasts fall within the ‘marginally useful’ category. In China, precipitation forecasts fall also in the ‘marginally useful’ category in wet summers and wet winters. Peng et al. (2014) found that ECMWF System 4 generally captures the climatological features of seasonal precipitation over China. Kim et al. (2012) assessed the performance of System 4 winter precipitation forecasts in the northern hemisphere. They showed a positive bias over East Asia. Accordingly, post-processing is a necessary step before GCM outputs can be applied in streamflow forecasts. Quantile mapping method is a popular post-processing method for bias-corrected ensemble GCM forecasts (Wood & Lettenmaier 2006; Rajczak et al. 2016). Its popularity in seasonal forecasting has since grown, because it can enhance forecast skill and reliability by reducing forecast errors (Piani et al. 2009; Teutschbein & Seibert 2012; Mehrotra & Sharma 2016).

With respect to water resources management, it is crucial to assess the potential of ECMWF System 4 forecasts for streamflow forecasts because the raw forecasts from ECMWF System 4 have systematic bias (Kim et al. 2012; Di Giuseppe et al. 2013; Wetterhall et al. 2015). However, this has been paid less attention with just a few exceptions. Trambauer et al. (2015) assessed the ability of ECMWF System 4 forecasts for hydrological drought in the Limpopo River Basin and found that it provides reliable information for reservoir operation and water resource management at the seasonal timescale. Crochemore et al. (2016) evaluated the skill of ECMWF System 4 raw precipitation forecasts for streamflow forecasting at seasonal scale in 16 catchments over France, which showed overprediction in winter season (December to February) and underprediction in autumn season (September to November) for different lead times.

Despite these recent studies, there is a lack of case studies evaluating ECMWF System 4 forecasts for streamflow forecasting in East Asia, albeit its hydrometeorological characteristics. For East Asian monsoon climate regions, including China, the potential of the System 4 dataset for streamflow forecasts should be assessed. The Upper Hanjiang River Basin (UHRB) is a typical subtropical monsoon region in China, which is the drainage basin of the Danjiangkou reservoir. This reservoir serves as the water source for the central route of the SNWDP, which transfers 13 billion m$^3$yr$^{-1}$ of water to the North China Plain since 2014. For the operation of the Danjiangkou reservoir, it is also necessary to enhance the accuracy and reliability of seasonal streamflow forecasts. Therefore, the UHRB is chosen in this
study to evaluate the potential of the ECMWF System 4 forecast dataset for streamflow forecasts up to three months ahead.

The remainder of the paper is organised as follows. Study area is detailed in the next section. Data, HEPS method, hydrological model, post-processing method and forecast evaluation merits are presented in the Materials and methods section. The results of the evaluation of streamflow forecasts are described in the Results and discussion section, and the main findings are summarised in the Conclusion section.

STUDY AREA

The Hanjiang River is the largest tributary of the Yangtze River, originating from the Qinling Mountains in Shaanxi Province. It flows through the Shaanxi, Henan and Hubei provinces of China with a total length of 1,570 km. Our study area is the UHRB draining to the Danjiangkou reservoir, with a drainage area of 95,200 km² (see Figure 1). The mean annual precipitation is approximately 900 mm, of which 70–80% falls in the wet season from May to October (Guo et al. 2013). Figure 2(a) shows the regime curve of precipitation, i.e., monthly precipitation averaged over 2001–2008; Figure 2(b) shows the regime of potential evapotranspiration averaged over 2001–2008; and Figure 2(c) shows the regime curve of streamflow, i.e., the discharge divided by the catchment area. To aid in the seasonal hydrological analysis, we adopted the definition of four seasons following Yang et al. (2014), who divided a water year into four seasons, namely, pre-wet season (May to July), post-wet season (August to October), pre-dry season (November and December) and post-dry season (January to April). This study evaluates the performance of ECMWF System 4 forecast in four sub-basins (Yangxian, Ankang, Baihe and Danjiangkou, ranging from 14,192 km² to 95,200 km²) within the UHRB.

MATERIALS AND METHODS

Data

Daily meteorological forecast data (e.g., precipitation and potential evaporation) are provided by the ECMWF seasonal forecast system. System 4, which initiates every month and integrates for the next seven months at 70-km spatial resolution, has been in operation since November 2011. The reforecast data are available from 1981 to 2010 as

Figure 1  |  Overview of the Upper Hanjiang River Basin (UHRB).
ensembles, including 51 members for the forecasts initiated in February, May, August and November, and 15 members for the remaining months. More details on the ECMWF System 4 can be found in Molteni et al. (2011).

Daily ground gauged data (including precipitation, temperature, wind speed and relative humidity data) at 16 meteorological stations are provided by the China Meteorological Administration. Daily potential evaporation is calculated based on the FAO Penman–Monteith equation (Allen et al. 1998) using gauged meteorological data. We compute the precipitation and potential evaporation for each sub-basin using the Thiessen polygon interpolation approach. Daily streamflow data at Yangxian, Ankang, Baihe and Danjiangkou stations are collected from the Bureau of Hydrology in the Ministry of Water Resources of China. The location of gauging stations is shown in Figure 1, and the time period of gauged data extends from 1970 to 2008.

Ensemble streamflow prediction method

The ESP method usually serves as a reference for validating climate model-based seasonal hydrological predictions (Yuan et al. 2013; Mo & Lettenmaier 2014; Yang et al. 2014). The procedure of the ESP method used in this study is summarised as follows: First, the state of a hydrological model is initialised with observed meteorological forcing through running the model in simulation mode for the year preceding the time of the forecast. Second, driving the model with the initial basin state by an ensemble of raw ECMWF System 4 forecasts and post-processed forecasts. A set of streamflow forecasts is generated with the aim of representing the uncertainty propagated from the meteorological forecasts.

In this work, the hydrological model is run in forecast mode at daily time-step from the beginning of each calendar month until three months later, which corresponds to 0, 1, and 2-month lead time. Lead time is the gap between the time when the forecast is initiated and the verifying month. For example, if we forecast April streamflow on 1st February, then the lead time is counted as two months.

A semi-distributed hydrological model, TsingHua Representative Elementary Watershed (THREW) model, is applied to simulate the rainfall–runoff processes. This model was developed by Tian et al. (2006) and has been successfully applied to the basins under diverse hydrometeorological settings (Li et al. 2012; Liu et al. 2012; He et al. 2015; Tian et al. 2017). For the detail of the THREW model, please refer to Tian et al. (2012) and Mou et al. (2008). Specifically, for the calculation of real evapotranspiration, the THREW model adopts the spatially averaged

Hydrological ensemble prediction system method

The procedure of the HEPS method used in this study is summarised as follows: First, the state of a hydrological model is initialised with observed meteorological forcing through running the model in simulation mode for the year preceding the time of the forecast. Second, driving the model with the initial basin state by an ensemble of raw ECMWF System 4 forecasts and post-processed forecasts. A set of streamflow forecasts is generated with the aim of representing the uncertainty propagated from the meteorological forecasts.

In this work, the hydrological model is run in forecast mode at daily time-step from the beginning of each calendar month until three months later, which corresponds to 0, 1, and 2-month lead time. Lead time is the gap between the time when the forecast is initiated and the verifying month. For example, if we forecast April streamflow on 1st February, then the lead time is counted as two months.

A semi-distributed hydrological model, TsingHua Representative Elementary Watershed (THREW) model, is applied to simulate the rainfall–runoff processes. This model was developed by Tian et al. (2006) and has been successfully applied to the basins under diverse hydrometeorological settings (Li et al. 2012; Liu et al. 2012; He et al. 2015; Tian et al. 2017). For the detail of the THREW model, please refer to Tian et al. (2012) and Mou et al. (2008). Specifically, for the calculation of real evapotranspiration, the THREW model adopts the spatially averaged
evapotranspiration model \citep{Eagleson1978}, in which the real evapotranspiration is determined by the potential evaporation rate as well as soil water supply capacity. The potential evaporation rate is controlled by climatic condition, while the soil and water supply capacity is controlled by soil/plant properties and moisture. The UHRB was divided into 89 representative elementary watersheds for the model setup. According to the ground gauged data of 1970–2000, the THREW model was run at the daily time scale. The model was calibrated at Baihe station and validated at Yangxian, Ankang and Danjiangkou stations. The modelling results showed that the annual Nash–Sutcliffe efficiency (NSE) at four stations (from upstream to downstream: Yangxian, Ankang, Baihe, Danjiangkou) in the calibration period are 0.90, 0.88, 0.88 and 0.91, respectively \citep{Sun2014}.

**Post-processing methods**

Systematic biases in the seasonal precipitation forecasts can cause forecast uncertainty and have unfavourable influence in hydrological forecasting. Quantile mapping (QM), also called distribution mapping, is a popular method for post-processing ensemble GCM forecasts \citep{Wood2002, Yuan2014}. This approach has been widely adopted, because it can enhance forecast skill and reliability by reducing forecast errors \citep{Bennett2014, Crochemore2016}. QM method matches the statistical distribution of precipitation forecasts to the distribution of observations. In the case of ensemble forecasts, the matching is performed for each ensemble member. Madadgar \textit{et al.} \citeyear{Madadgar2014} demonstrated that the QM method matched the statistical distribution of forecasts to the distribution of the observations independently, which induced ‘blind-matching’ and degraded the forecast skill. However, Crochemore \textit{et al.} \citeyear{Crochemore2016} found that QM method could effectively remove bias when the bias was the main deficiency of the raw forecasts, e.g., in the case of ECMWF System 4 precipitation forecasts. Therefore, we apply the QM method to post-process the original System 4 precipitation forecasts in this study.

An approach introduced by Arlot \& Celisse \citeyear{Arlot2010}, namely, leave-one-out cross-validation approach is employed in this study. The leave-one-out cross-validation approach calibrates the QM method in each representative elementary watershed over independent periods within the 2001–2008 period. More specifically, for a given target application year, QM method is trained with observations and forecasts of remaining years. Then, the cross-validation results are applied to the target year for bias correcting. We carry out the calibration of QM method with empirical distribution of daily values of forecasts and observation. In the case of ensemble forecasts, the daily precipitation data of each forecast member are corrected individually.

**Forecast evaluation**

**Deterministic analysis**

Three common metrics are used to assess the accuracy of the streamflow forecasts, including the NSE, the relative mean error (RME) and the correlation coefficient (CC). These metrics have been widely used in previous studies \citep{Demargne2010, Verkade2013, Alfieri2014}. The detail of these metrics is given in the Appendix (available with the online version of this paper).

The skill of monthly streamflow forecasts is evaluated by the mean square error skill score (SS\_MSE) calculated by Equation (1):

\[
SS_{MSE} = 1 - \frac{\sum_{i=1}^{N} [Fst(i) - Obs(i)]^2}{\sum_{i=1}^{N} [Obs(i) - \overline{Obs}]^2} \tag{1}
\]

where $Fst(i)$ and $Obs(i)$ are the forecasted and observed monthly streamflow volume, respectively; $\overline{Obs}$ is the mean value of the observed streamflow volume. $N$ is the total number of forecast years, which is equal to 8 in this study. This skill score is used for evaluating the HEPS performance in individual months. The optimal value of $SS_{MSE}$ is 1.

For deterministic analysis, the ensemble forecasts should integrate into some single values. In this study, the ensemble mean is used to compute these performance metrics in the validation period.

**Probabilistic analysis**

The reliability of the forecast is assessed using the probability integral transform (PIT) diagram \citep{Laio2008}.
which is calculated by Equation (2):

\[ PIT_i = F_t(q_{obs,i}) \]  

(2)

where \( F_t \) is the cumulative distribution function of the forecast, \( q_{obs,i} \) is the corresponding observed data, \( i = 1, 2, \ldots, N \), and \( N \) is the number of observations. The \( PIT_i \) values should follow a uniform distribution between 0 and 1. For a reliable forecast, the observed data \( q_{obs,i} \) should uniformly fall within the prediction distribution and the PIT diagram should align on the 1:1 diagonal line. According to Laio & Tamea (2007), if the scattered points do not lie on the 1:1 line in the PIT diagram, the curve in the PIT diagram generally presents four different shapes, representing four different situations: ‘overprediction’, ‘underprediction’, ‘narrow forecast’ and ‘large forecast’. If the PIT diagram is above (below) the diagonal, the observed values are regularly situated in the lower or upper part of the forecast distribution, which indicates overprediction (underprediction). If the PIT diagram presents a vertical line, the observations fall too frequently at the end of the forecast distribution, which means the predictions are too wide. On the other hand, if the PIT diagram tends to a horizontal line, many observations fall in the mid-range, suggesting the forecasts are too narrow. We also presented the 5% Kolmogorov–Smirnoff confidence interval from the bisector in the following analysis.

The overall performance of the ensemble forecasts is assessed using a ranked probability score (RPS), which has been commonly used to evaluate how well the probability forecasts predict the corresponding observations. Based on percentile determined from the distribution of the observed flows (2001–2008), the thresholds are divided into three categories, including above normal category (>66th percentile), near normal category and below normal category (<33rd percentile). The RPS for one forecast is given by Equation (3) (Wilks 2011):

\[ RPS = \frac{1}{m} \sum_{j=1}^{m} \left( \sum_{i=1}^{m} Y_{fcj} - \sum_{i=1}^{m} O_{obsj} \right)^2 \]  

(3)

where \( J \) is the number of categories (here it is 3), \( Y_{fcj} \) refers to the relative occurrence frequency of ensemble members in the corresponding category, and \( O_{obs} \) represents the observation probability in the category (\( O_{obs} \) equals 1 if the

observed data fall into the \( j \)th category and zero otherwise). For a perfect forecast, the value of RPS is 0.

The forecast skill of the HEPS is evaluated using a ranked probability skill score (RPSS), compared with a reference forecast. For the three categories forecast, the reference forecast probability in the \( j \)th category is 1/3. The RPS for the reference forecast is computed using \( Y_{fcj} = 1/3 \) in Equation (3). The RPSS is defined in Equation (4) (Wilks 2011) as follows:

\[ RPSS = 1 - \frac{\overline{RPS}_f}{\overline{RPS}_{ref}} \]  

(4)

where \( \overline{RPS}_f \) is the average RPS for ensemble forecast and \( \overline{RPS}_{ref} \) is the average RPS for reference forecasts (Wilks 2011). A positive RPSS means that the HEPS forecasts outperform the reference forecasts, and a negative RPSS indicates that the HEPS forecasts produce lower performance than the reference forecasts. For a perfect forecasts, the RPSS equals 1.

As the ECMWF System 4 forecast data are 15 or 51 ensembles members in each month, which induces bias in the computation of RPS skill scores, the correction factor proposed by Ferro et al. (2008) is applied to compute the RPSS in this study.

RESULTS AND DISCUSSION

Forecast illustration

Retrospective HEPS forecasts for the Yangxian, Ankang, Baihe and Danjiangkou hydrological stations were produced using the corresponding ECMWF meteorological forecast data to drive the THREW model. Figure 3 illustrates the raw retrospective HEPS forecasts in January 2001 and July 2001 in Danjiangkou station, which represent the dry and wet seasons, respectively. Each ensemble forecast is drawn with a dotted line. The observation and the post-processed ensemble mean are represented by a solid line. As shown in Figure 3, there is a larger forecast uncertainty at a longer lead time. The 0-month lead forecast is quite accurate and presents narrow forecast range. It also can be seen that the ensemble forecasts present larger overestimation at longer lead time. The hydrograph of the post-processed forecasts captures the observed streamflow more accurately
than raw forecasts, which indicates that post-processed forecasts have positive effect on forecast skill.

**Forecast accuracy**

Three deterministic scores of raw forecasts and post-processed forecasts calculated for the whole period (i.e., 12 months during 2001–2008) are plotted against the scores for ESP forecasts in scatterplots (Figure 4). Deterministic scores for raw forecasts are plotted in the upper panels while the deterministic scores for post-processed forecasts are plotted in the lower panels. Each score is computed for lead times of 0 month, 1 month and 2 month for the four stations.

The figures show that raw forecasts and ESP forecasts have similar forecast skills in different lead times. The scatterplot of NSE, RME and CC for raw forecasts are close to the 1:1 line, while the skill scores for post-processed forecasts tend to be more accurate than ESP forecasts. These results demonstrate that post-processed forecasts have a great impact on forecast accuracy at the four hydrological stations.
stations. For instance, in the case of 0-month lead time, the NSE for post-processed forecasts are 0.46, 0.70, 0.61 and 0.67, and for ESP forecasts are 0.23, 0.25, 0.16 and 0.25 at the four stations, respectively. It should be noted that the ESP forecasts have obvious bias in Yangxian station. This could be attributed to the precipitation series of drainage areas at four gauging stations showing decreasing trends (Sun et al. 2014). Sun et al. (2014) also found that the relative contributions of climate variability to streamflow decreasing at four sub-basins (from upstream to downstream) were 82%, 62%, 54% and 44%, respectively.

**Forecast reliability**

To assess reliability of ensemble streamflow forecasts, we plot the PIT diagrams in Figure 5. There is a jump in the PIT value of 0 for raw streamflow forecasts, which means that observations fall within the lowest interval of the forecast distribution or below it. The forecast reliability increases from 0-month to 2-month lead time, which is generally attributable to the fact that streamflow forecasts at 0-month lead time are more accurate and hence the uncertainty of the forecast is small. Overall, the PIT diagrams indicate that the forecasts have some problems in forecast reliability. This could be attributed to the overestimated precipitation by the ECMWF seasonal forecast System 4, which has been demonstrated by the findings of Kim et al. (2012) in East Asia.

The results of streamflow forecasts obtained from post-processed precipitation forecasts show that significant improvement is achieved after being bias corrected with the QM method for all the lead times. Streamflow forecasts are reliable and close to the 1:1 line in the PIT diagram from 0-month to 2-month lead time. Even though a slight
tendency to overestimate streamflow remains for different lead times, the improvements are noticeable. Our results are consistent with Crochemore et al. (2016), finding that the reliability of seasonal streamflow forecasts is improved after post-processing ECMWF System 4 raw precipitation forecasts.

Figure 5 | PIT diagrams of streamflow forecasts from raw precipitation forecasts (left) and post-processing precipitation (right) for the whole period (2001–2008) with lead times of 0, 1 and 2 months at four hydrological stations. Each line represents the PIT diagram for a station. Dotted lines represent the 5% Kolmogorov–Smirnoff confidence bands.
Performance of monthly forecasts

To evaluate the performance of streamflow forecasts for different calendar months, we calculated and compared SS_MSE for each month in the period 2001–2008. The monthly SS_MSE skill score patterns at four hydrological stations are rather similar. Therefore, the exemplary SS_MSE result in the Danjiangkou station is shown in Figure 6. It is noteworthy that the ESP forecasts are more skillful during August–December. This result is consistent with the previous studies (Wood & Lettenmaier 2008; Yang et al. 2014) that demonstrated the ESP forecast skill is better during the transition period from the wet to dry season. The SS_MSE score of the raw streamflow forecasts is negative during November–April, indicating that the raw streamflow forecast has no skill for this period. This could be attributed to the combined effects of future atmospheric forcing and initial conditions. Yang et al. (2014) found that the future atmospheric forcing is the dominant element on the streamflow forecasts during the transition period from wet to dry season over UHRB. However, the skill of raw ECMWF System 4 precipitation forecasts in dry season is lower than wet seasons (Peng et al. 2014). Therefore, the raw streamflow forecasts produce lower performance during November–April over UHRB. After post-processing precipitation, the streamflow forecasts show remarkable improvement. The SS_MSE score of the post-processed forecasts in all months is consistently positive at 0-month lead time. For 2-month lead time, the post-processed forecasts are also more skillful than ESP forecasts and raw forecasts, but the forecast skill decreases to some extent.

The overall performance of the monthly streamflow forecasts is assessed using RPSS. Figure 7 shows the results of RPSS for raw forecasts and post-processed forecasts with

![Figure 6](https://iwaponline.com/hr/article-pdf/49/6/1864/509238/nh0491864.pdf)
three lead times. It can be seen that raw streamflow forecasts present strong variations at monthly scale. For example, the RPSS value of Yangxian station shows declining trend from January to June. After June, the RPSS increases to the highest positive value in October, and then reduces to the lowest value in December. After post-processing precipitation
forecasts, the ensemble streamflow forecasts have sufficiently improved, which indicates post-processed forecasts with less error than the reference forecasts. Generally, the raw forecasts produced a relatively lower performance in June, November and December and show visible improvements after post-processing precipitation forecasts. At the seasonal scale, the HEPS method from raw forecasts outperformed the reference forecast in the post-wet season, but produced lower performance than the reference forecast in the pre-dry season. It is also apparent from Figure 7 that the post-processed forecasts show better overall performance than reference forecasts in Ankang, Baihe and Danjiangkou stations, especially in pre-dry season, but the streamflow forecasts of Yangxian station show no improvement. This might be because bias correcting precipitation forecast would be more efficient over large basins (Danjiangkou station) and less effective in small basins (Yangxian station). The improved performance during pre-dry season suggests that post-processing ECMWF System 4 precipitation forecast is a useful tool in streamflow forecasts over UHRB. The results of streamflow forecasts from post-processing precipitation could provide useful information for reservoir refill operation in pre-dry season.

Seasonal PIT diagrams of 0-month lead time are shown in Figure 8. Comparing with other seasonal forecasts, the

![Figure 8](https://iwaponline.com/hr/article-pdf/49/6/1864/509238/nh0491864.pdf)

*Figure 8* | PIT diagrams of 0-month lead time streamflow forecasts from raw precipitation (left) and post-processing precipitation (right) in different seasons at four hydrological stations. Each line represents the PIT diagram at a station. Dotted lines represent the 5% Kolmogorov–Smirnov confidence bands. (Continued.)
post-wet season streamflow is more reliable than other seasons. Pre-wet and post-wet season streamflow forecasts from raw precipitation show some reliability, but tend to over-predict streamflow. The PIT diagram of pre-dry season tends to a horizontal line, which indicates the ensemble forecasts are too wide. This finding is consistent with Peng et al. (2014), who claimed that the raw ECMWF System 4 seasonal precipitation forecast performs better in the wet season than in the dry season in China. As precipitation forecasts are post-processed by the QM method, the reliability of streamflow forecasts is considerably enhanced. The improved performance in the post-wet season, pre-dry season and post-dry season provides useful information for water supply management of the central route of the SNWDP. Pre-wet season streamflow forecasts are also more reliable after post-processing precipitation, even though they still have some problems in forecast reliability indicating that there is scope for improvement. Previous studies demonstrated that a simple bias correction method, such as QM method, is not enough to achieve streamflow forecast reliability (Verkade et al. 2013; Roulin & Vannitsem 2015). In addition, post-processing precipitation alone would neglect the uncertainty due to hydrological modeling. Future investigation will deal with the post-processing of the streamflow forecasts in order to
provide the management of the UHRB with improved long-range hydrological predictions.

CONCLUSIONS

This study applies the ECMWF System 4 forecasts for seasonal streamflow forecasting over the UHRB. The performance of the ensemble streamflow forecast is assessed in terms of the forecast accuracy, reliability and spread. The main findings are summarised as follows:

1. During the study period, the raw streamflow forecasts present similar forecast accuracy with ESP forecasts and post-processing precipitation can improve the forecast accuracy substantially. The raw streamflow forecasts show some problems in forecast reliability and a tendency to overpredict streamflow. After post-processing precipitation with the QM method, the reliability of the forecasts is greatly enhanced up to 2-month lead time.

2. The performance of streamflow forecasts presents strong variability at monthly and seasonal scales. The raw streamflow forecasts produce low performance during November–April. As the precipitation forecasts are post-processed by QM method, the results show remarkable improvement over different locations.

3. At the seasonal scale, the improved performance in the post-wet season, pre-dry season and post-dry season provides useful information for water supply management of the central route of the SNWDP, while the pre-dry season streamflow forecasts show poor reliability and require further improvement.

Generally, our results provide valuable information regarding the application of the ECMWF System 4 forecasts for seasonal streamflow prediction in East Asia monsoon climate regions. In future studies, more basins with a diverse climate and landscape properties should be involved in the investigation to strengthen the findings in the present study.

ACKNOWLEDGEMENTS

This research was funded by the Ministry of Science and Technology of P.R. China (2016YFC0402701), the National Science Foundation of China (NSFC 91647205), the foundation of State Key Laboratory of Hydroscience and Engineering of Tsinghua University (2016-KY-03). The authors would like to thank the Information Center of Ministry of Water Resources and China Meteorological Administration for providing the hydrology and meteorological data. Also, special thanks are given to Professor Quanjun Wang from the University of Melbourne for his valuable comments.

REFERENCES


Molteni, F., Stockdale, T., Balsaneda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T. & Vitart, F. 2011 *The new ECMWF Seasonal Forecast System (System 4).* Technical memorandum, ECMWF.


First received 11 November 2017; accepted in revised form 24 March 2018. Available online 17 April 2018.