



The role of probabilistic precipitation forecasts in hydrologic predictability

Seung Beom Seo¹ · Jang Hyun Sung²

Received: 13 March 2019 / Accepted: 14 May 2020 / Published online: 28 May 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

Accurate streamflow forecasts enable the appropriate management of water resources. Although there is a general consensus that climate information can enhance hydrological predictability, this might not be the case if the accuracy of the given climate information is unreliable. Hence, this study has developed a modeling framework to estimate the role of climate information in forecasting accurate streamflow. Ensemble streamflow prediction (ESP) technology was adopted as a dynamic hydrologic forecast method to 35 watersheds in South Korea. The probabilistic precipitation forecast (PPF), issued by the Korea Meteorological Administration, was used as climate information for updating the probabilities of climate scenarios. First, we found that the current PPF is not accurate enough for significantly enhancing the streamflow forecasting accuracy. Subsequently, multiple sets of PPF were synthetically generated to evaluate the role of climate information. Given the perfect categorical climate forecasts, we found that there is much potential for the enhancement of streamflow forecast skill especially in the seasons that exhibit greater streamflow variability. However, there is less potential for increasing the streamflow forecasting skill under below-normal conditions. The proposed modeling framework is capable of quantifying the magnitude of potential improvement in hydrological predictability under the assumption that better climate information will be available in the future. We expect that this modeling framework can be effectively applied to other regions across a wide range of climate regimes.

Keywords Streamflow forecast · Probabilistic precipitation forecast · ESP · Climate information · Hydrological predictability

1 Introduction

Hydrologic forecasts on multiple timescales facilitate management decisions across various temporal horizons (Labadie 2004). Accurate streamflow forecasts enable the appropriate management of water resources, such as drought mitigation, flood prevention, and reservoir operation. Over the past few decades, the ensemble streamflow prediction (ESP) technique (Day 1985) has been widely used for streamflow forecasting (Franz et al. 2003; Jeong and Kim 2005; Kim et al. 2006; Renner et al. 2009; Gobena and Gan 2010; Olsson and Lindströma 2010; Hwang et al. 2011; Najafi et al. 2012; Liu et al. 2015). The ESP uses a hydrological model with weather

scenarios sampled from historical observations. The soil moisture conditions are initialized at the time of the forecast. Then, the hydrological model runs using sampled alternate weather inputs to generate the ensemble forecasts of streamflow. This single model ESP can be extended to multi-model ESP by using multiple hydrological models. Since each weather input then is inserted into each hydrologic model, the number of forecast samples is then increased in proportion to the number of hydrologic models. Furthermore, if weather forecasts driven by climate models (e.g., general circulation models) are available, the number of streamflow forecasts also would be increased. In order to address possible uncertainty arises in multiple models, in general, it is preferred using as many models as possible.

The basic assumption behind the ESP technique is that the likelihood of occurrence of each member of ensemble forecasts is equal. However, if additional climate forecast information is given, the probabilities assigned to individual forecasts can be updated. Kelman et al. (1990) and Faber and Stedinger (2001) illustrated how the probabilities of individual climate series can be updated using the Bayes theorem. Additionally, the conditional distribution of future streamflow

✉ Jang Hyun Sung
jhsung1@korea.kr

¹ International School of Urban Sciences, University of Seoul, Seoul 02504, South Korea

² Han River Flood Control Office, Ministry of Environment, Seoul 06501, South Korea

can be derived using the Bayesian methods and climate forecast information (Krzysztofowicz 2001; Herr and Krzysztofowicz 2010; Bradley et al. 2015; Seo et al. 2019).

Climate and weather forecasts are the application of contemporary technologies to predict the future state of the atmosphere for a given location. Over the past decades, information regarding large-scale climate phenomena, such as the El Niño–Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) pattern, has been widely used to discuss linkages between large-scale climate information and regional climatology (Dettinger et al. 1998; Gershunov 1998; Clark et al. 2001; Li and Chen 2014). Subsequently, these teleconnection patterns have been applied to enhance ESP skills relating to seasonal and sub-seasonal forecasting (Kalra et al. 2013; Beckers et al. 2016). However, teleconnection patterns that are generally dominant on a continental scale often fail to improve forecast skill on a watershed scale (Grantz et al. 2005).

The National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS)—an end-to-end Hydrologic Ensemble Forecast Service—extends the lead time of hydrologic ensemble forecasts from 6 h to 1 year and includes additional weather and climate information as well as improved quantification of major uncertainties (Demargne et al. 2014). Due to the chaotic nature of the atmospheric processes and the incomplete modeling of these processes, forecasts become less accurate as the forecasting lead time increases. The accuracy of climate forecasts cannot be assured due to huge uncertainty in forecasts, particularly in regions with high rainfall variability. The updated probabilities on the individual scenarios might not be able to improve hydrologic forecast accuracy if the given climate information's accuracy is not assured. Thus, it is a fundamental premise that accurate climate forecasts must be used when updating the probabilities of individual streamflow forecasts with ESP.

Though the ensemble approach holds great potential for hydrologic prediction systems, obtaining accurate climate forecasts is an essential prerequisite for efficient operation of hydrologic forecasting for water resources management such as flood prevention, drought mitigation, and dam operation. Under circumstances where accurate climate forecasts are not guaranteed, the role of climate forecasts in enhancing hydrologic ensemble forecast accuracy needs to be quantified in order to identify areas where investments in forecast systems and processes will significantly benefit hydrologic prediction. This is because the accuracy of the hydrologic ensemble forecasts can be enhanced if the accuracy of the weather forecasts is increased through future developments on atmospheric and climate modeling technologies. Nevertheless, there have been little efforts on discussion about the impacts of to-be-improved climate forecast on the accuracy of the streamflow forecasts.

Therefore, the present study develops a modeling framework to evaluate the role of climate forecast on ESP accuracy. The rest of the paper is organized as follows: the following

section describes methodological backgrounds; “Section 3” provides a case study along with data sets; “Section 4” highlights the results of the case study; and “Section 5” presents the conclusions.

2 Methodological backgrounds

2.1 Probabilistic forecast in Korea

The Korea Meteorological Administration (KMA) has been reporting mid-term and long-range forecasts to the public for 12 provinces of the Korean Peninsula since June 2014. The mid-term forecast predicts 1-week ahead precipitation and temperature probabilistic forecasts on daily basis. On the other hand, long-range forecast reports the 1-month and 3-month outlook for average temperature and precipitation on a weekly and monthly basis, respectively, as a form of probability for each tercile interval: below-normal, normal, and above-normal. Figure 1 illustrates an actual example of the 3-month outlook for the average temperature and precipitation probability, reported on a monthly basis. As shown in Fig. 1, the probability for each tercile interval is forecasted for both precipitation and average temperature. Based on the values of these probabilities, forecast statement is categorized into three stages—above-normal, normal, and below-normal.

KMA generates the various short-, mid-, and long-range forecasts based on statistical techniques and climate models using numerical techniques. Nonetheless, only long-range outlook information was used in this study because the objective of this paper is evaluating the role of probabilistic forecast on long-term (e.g., monthly or seasonal) streamflow forecast accuracy. The methodology proposed in this paper can be extended to shorter lead-time forecast applications.

2.2 Ensemble Streamflow prediction (ESP)

ESP has been a widely used method for probabilistic forecasting in operational hydrology. The ESP runs a rainfall-runoff model with observed meteorological inputs to generate an ensemble of possible streamflow hydrographs (Kim et al. 2001; Gobena and Gan 2010; Najafi et al. 2012). In the ESP, all the meteorological scenarios are input into the rainfall-runoff model under the basic assumption that they are equally likely to occur in the future. Since initial conditions of the rainfall-runoff model vary depending on the timing of the forecast, the ensemble of streamflow forecast varies in accordance with the initial conditions—even for the same meteorological scenarios. Therefore, the generated ensemble of streamflow is also a function of the concurrent hydrological states estimated in the rainfall-runoff model. Thus, this technique is sometimes referred to as a conditional Monte Carlo simulation approach (Day 1985).

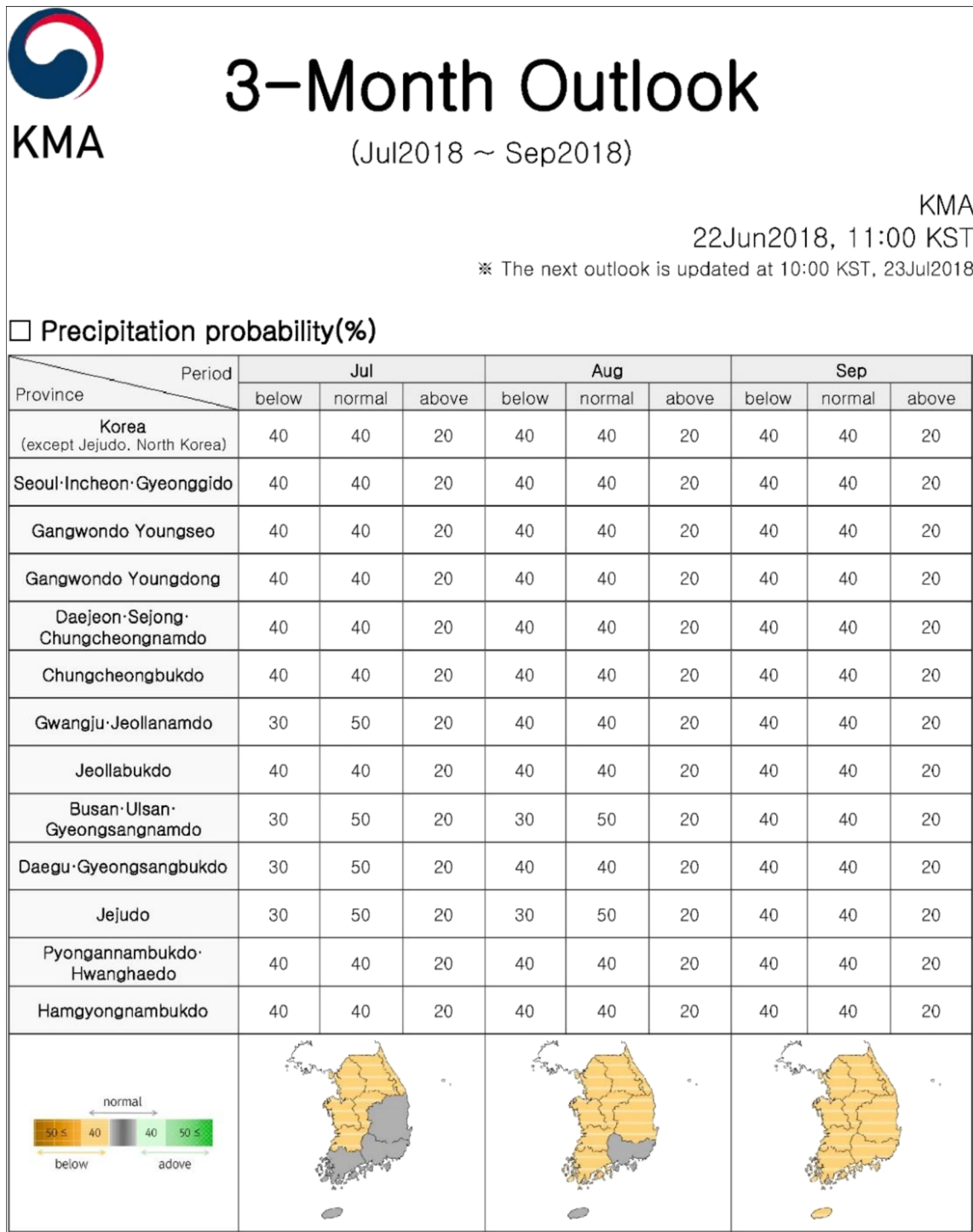


Fig. 1 An example of the long-range forecast of South Korea: 3-month outlook reported on June 22, 2018, by the Korea Meteorological Administration

2.3 Reflecting climate information: Croley-Wilks approach

Consider a set of historical climate series $\{v_i\}$ for $i = 1, 2, \dots, N$, e.g., v_i can be monthly precipitation. To each is assigned a value $x_i = g(v_i)$ which reflects the more detailed and complete series (Stedinger and Kim 2010). Here x_i can be the monthly

streamflow in this study. Croley II (2000, 2003) and Wilks (2002) expressed that the conditional distribution function $D[g(v)|H]$ is summarized by the probability of the selected set of climate variables that $g(v)$ fall into the intervals below-normal, normal, and above-normal. The hydrologic information employed to determine the conditional distribution of streamflow is symbolized by H . Their algorithm adjusts the

probabilities assigned to the different series so as to achieve the target probabilities using the values of the selected variables, $g(v_i)$ but only to the extent that they determine whether a given climate series, v_i , is in the below-normal, normal, or above-normal range for $g(v_i)$ (Stedinger and Kim 2010). After the non-parametric approach—developed by Croley II 2003 and Wilks (2002)—that assigns the same probability to climate scenarios included in each category, Stedinger and Kim (2010) proposed a simple and general approach, PDF-ratio approach, that makes use of the entire $D[g(v)|H]$ distribution, as well as the individual values of the selected variables $g(v_i)$ associated with each v_i to generate series-probability pairs $\{(v_i, q_i)\}$. These pairs provide a better approximation of the entire $D[g(v)|H]$ distribution. Nonetheless, there was no significant difference in accuracy of updated streamflow forecasts between Croley-Wilks approach and PDF-ratio approach (Kim et al. 2016). Hence, Croley-Wilks approach was adopted for its straightforward implementation.

Given a N number of historical climate scenarios, x_i , for $i = 1, 2, \dots, N$, let x_i be the monthly precipitation variable that has quantiles x_b and x_a such that

$$0.333 = F(x_b) = F(x_a) - F(x_b) = 1 - F(x_a), \tag{1}$$

where $F(x)$ is a cumulative distribution function and x_a and x_b are the upper and lower terciles, respectively, defining above-normal and below-normal ranges (Stedinger and Kim 2010). As mentioned in the previous section, if no forecast information is provided, the standard practice is to assign an equal weight, $1/N$, to each climate scenario.

For the terciles x_a and x_b , let us assume that the probabilistic precipitation forecast (PPF) is given as

$$\begin{aligned} F(x_b) &= p_b \\ F(x_a) - F(x_b) &= p_n \\ 1 - F(x_a) &= p_a, \end{aligned} \tag{2}$$

where $p_b + p_n + p_a = 1$. The three interval probabilities, p_b , p_n , and p_a , are defined as the below-normal, normal, and above-normal probabilities, respectively.

To reflect climate forecast information so that the prior probabilities, $1/N$, on the x_i can be updated to new probabilities, the Croley-Wilks approach (Croley II 2000, 2003; Wilks 2002) has been used in this study. This probability adjustment technique assigns the same probability to the climate scenarios in each selected interval (Croley II 2000, 2003; Wilks 2002).

Given that the climate forecast information is provided, prior probabilities for each historical climate series, $1/N$, are simply updated to p_b/N_b , p_n/N_n , and p_a/N_a for below-normal, normal, above-normal, respectively. N_b , N_n , and N_a are the number of historical climate scenarios that belong to below-normal, normal, and above-normal range, respectively (i.e., $N_b + N_n + N_a = N$). These Croley-Wilks probabilities, p_b/N_b ,

p_n/N_n , and p_a/N_a , provide a simple solution that matches the required probabilities for below-normal, normal, and above-normal events (Stedinger and Kim 2010). Figure 2 presents an illustration for a better understanding of the weights adjustment. Based on the adjusted probabilities of the climate scenarios in each interval, the ensemble mean of streamflow forecast, μ_x , is calculated as

$$\mu_x = \mu_x^b + \mu_x^n + \mu_x^a \tag{3}$$

where $\mu_x^b = \sum(p_b/N_b)x_i$ (only when $x_i \leq x_b$), $\mu_x^n = \sum(p_n/N_n)x_i$ (only when $x_b < x_i \leq x_a$), and $\mu_x^a = \sum(p_a/N_a)x_i$ (only when $x_a < x_i$). Note that without climate forecast information, μ_x is calculated as a value of the simple mean of x_i , i.e., $\mu_x = \frac{1}{N} \sum x_i$.

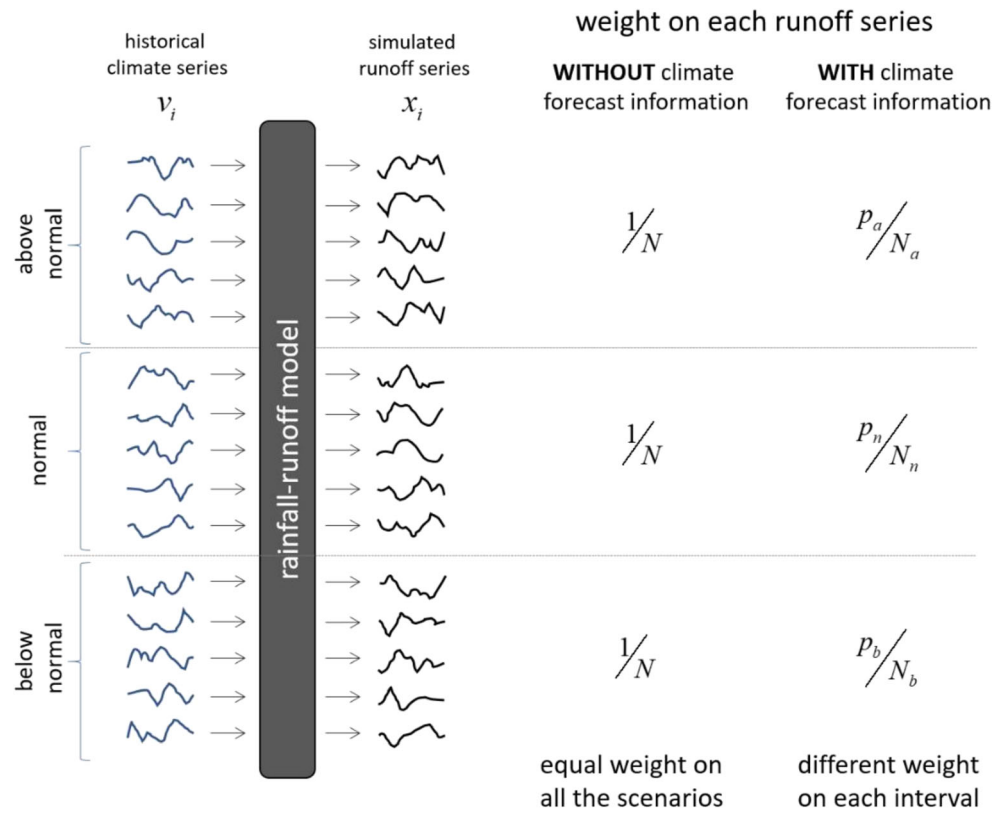
2.4 Forecast verification metrics

Various verification metrics are used for the evaluation of forecast quality. First, the Nash-Sutcliffe Efficiency (NSE) is used to evaluate the quantitative error between observed and predicted values. NSE is a non-dimensional coefficient that can range from $-\infty$ to 1 (Nash and Sutcliffe 1970). It is computed by standardizing the mean squared error between forecasts and observations. Essentially, the closer the NSE is to 1, the more accurate the forecast is. The normalized-root mean squared error (N-RMSE) is also used. RMSE is frequently used to measure the differences between values predicted by a model and the observed values. N-RMSE is used because normalizing the RMSE facilitates comparison between data sets with different scales (e.g., different mean values of monthly streamflow across all the season).

Additionally, probability of detection (POD) is used to evaluate the accuracy of categorical forecasts. The POD is simply a ratio that indicates the number of occasions an event occurred on the date it was forecasted to occur (Wilks 2011). In this study, a 3×3 contingency table for the categorical forecast verification situation is used, as shown in Fig. 3. The categories are divided into below-normal, normal, and above-normal, with x_a^{obs} and x_b^{obs} as the upper and lower terciles of observations. Here, the totals for each of the nine possible forecast and event pair outcomes are denoted by the letter a through i . The POD is given by the proportion of correct forecasts (denoted as “hit” in the contingency table). That is, in the 3×3 contingency table represented in the Fig. 3, the POD would be $(a + e + i)/(a + b + c + d + e + f + g + h + i)$.

For verification of categorical probabilistic forecast, this study uses the ranked probability skill score (RPSS), the skill score for a collection of rank probabilistic score (RPS) values relative to the RPS computed from the climatological probabilities. RPS is the most commonly used measure capable of increasingly penalizing forecasts, as more probability is

Fig. 2 An illustration of the weight adjustment for each interval with climate forecast information. N_b , N_n , and N_a are the number of historical climate scenarios that belong to below-normal, normal, and above-normal intervals, respectively. N is the total number of historical climate scenarios. p_b , p_n , and p_a are the precipitation probabilities for the below-normal, normal, and above-normal intervals, respectively



assigned to event categories further removed from the actual outcome (Wilks 2011). Discrete metric such as NSE and N-RMSE simply quantify the accuracy of deterministic forecast which is ensemble mean value. On the other hand, POD and RPSS evaluate the accuracy of categorical probabilistic forecast. These two measures provide a threshold value which is driven by reference forecast (historical mean termed as climatology) so that it can evaluate comparative superiority over climatology.

2.5 Realization of synthetic probabilistic forecasts

To evaluate the role of climate forecasts on the ESP accuracy, multiple sets of PPF were synthetically generated. The synthetic probabilistic forecast sets were sampled from historical forecasts. Using the target value of POD, which ranges from 0 to 1 increasing by 0.1 intervals, PPF sets were generated for each province (11 in total). The two steps used for generating the synthetic PPF are as follows:

Fig. 3 Contingency table for the 3×3 categorical forecast verification

		observation		
		below normal x_a^{obs}	normal x_b^{obs}	above normal
forecast	below normal x_a^{obs}	a (hit)	b	c
	normal x_b^{obs}	d	e (hit)	f
	above normal x_b^{obs}	g	h	i (hit)

Table 1 Historical forecast samples issued to date

Below-normal			Normal			Above-normal		
p_b	p_n	p_a	p_b	p_n	p_a	p_b	p_n	p_a
<i>0.50</i>	0.40	0.10	0.25	0.30	0.45	0.10	0.40	<i>0.50</i>
<i>0.50</i>	0.35	0.15	0.30	0.30	0.40	0.15	0.35	<i>0.50</i>
<i>0.50</i>	0.30	0.20	0.40	0.30	0.30	0.20	0.30	<i>0.50</i>
<i>0.55</i>	0.35	0.10	0.45	0.30	0.25	0.10	0.35	<i>0.55</i>
<i>0.55</i>	0.30	0.15	:	:	:	0.15	0.30	<i>0.55</i>
<i>0.60</i>	0.30	0.10	0.25	0.60	0.15	0.10	0.30	<i>0.60</i>
<i>0.60</i>	0.25	0.15	0.30	0.60	0.10	0.15	0.25	<i>0.60</i>
<i>0.65</i>	0.25	0.10	0.20	0.65	0.15	0.10	0.25	<i>0.65</i>
<i>0.70</i>	0.20	0.10	0.15	0.65	0.20	0.10	0.20	<i>0.70</i>

A forecast sample in which below-normal probability is equal or greater than 0.5 becomes a member of the below-normal forecast. On the other hand, a forecast sample in which above-normal probability is equal or greater than 0.5 becomes a member of the above-normal forecast. If a forecast does not belong to either below or above-normal, it becomes a member of normal forecast. The number of samples for both below and above-normal is 9, whereas the number of samples for normal is 40. Italicized values represent the events which probability is greater than 50%

- 1) Establish a group of historic forecasts: obtain historic forecasts that were issued up to the present, and categorize them into three intervals (below-normal, normal, and above-normal) as shown in Table 1.
- 2) Conduct random sampling: extract samples randomly from the group of historic PPF based on the POD target values. For example, if observed precipitation belongs to the below-normal interval and the given value of target POD is 0.5, half of the forecasts are randomly sampled from the historical forecasts that belong to the below-normal category, while the other half of the forecasts are

randomly sampled among the historical forecasts that do not belong to the below-normal category. Figure 4 shows an illustrative example of the synthetic PPF generation. First, a single forecast for each category is randomly selected among all the historic samples. Then, a final forecast is selected among these three candidates based on given weights (i.e., extraction probability) that vary corresponding to the value of the target POD. This is repeated until the end of the time step and separately generated for each province. Fig. 4 presents an example of a selected synthetic forecast.

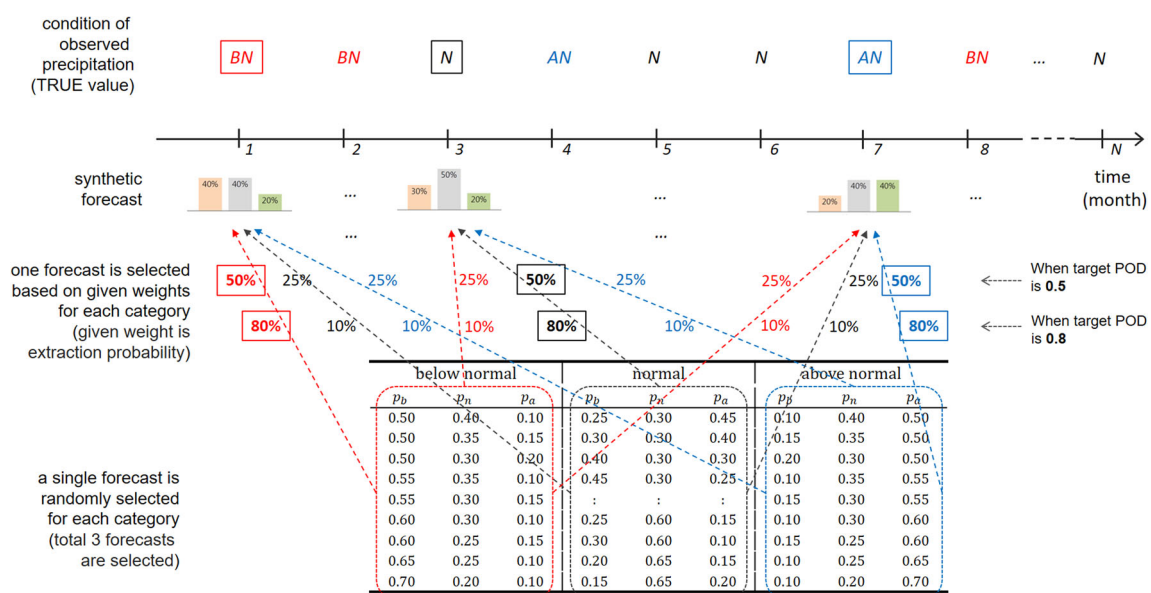


Fig. 4 An illustration of the synthetic PPF generation

Table 2 A list of the 35 watershed used in this study

No.	Name	Drainage areas (km ²)	Reservoir capacity (million cubic meters)
1	Soyanggang	2703	2900
2	Chungju	6648	2750
3	Heongseong	209	87
4	Gwangdong	125	13
5	Dalbang	29	9
6	Andong	1584	1248
7	Imha	1361	595
8	Seongdeok	41	28
9	Yeongju	500	181
10	Gunwi	88	49
11	Kimcheonbuhang	82	54
12	Bohyunsan	33	22
13	Hapcheon	925	790
14	Namgang	2285	309
15	Miryang	95	74
16	Yeongcheon	235	103
17	Angye	7	18
18	Gampo	4	3
19	Woonmoon	301	160
20	Daegok	58	36
21	Sayeon	67	30
22	Daeam	77	13
23	Seonam	1	2
24	Yeoncho	12	5
25	Gucheon	13	10
26	Yongdam	930	815
27	Daechong	3204	1490
28	Seomjin	763	466
29	Juam	1010	457
30	Juam regulator	135	250
31	Buan	59	50
32	Boryeong	164	117
33	Jangheung	193	191
34	Sueo	49	31
35	Pyeongnim	20	10

3 Application

3.1 Study area and data sets

This study tested 35 watersheds in which dams are operated below the watershed outlet. Table 2 presents a list of the 35 watersheds used in this study, while Fig. 5 shows the locations of the 35 watersheds across South Korea. Observed meteorological data sets from 1966 to 2016—daily precipitation, maximum and minimum temperature, and average wind speed series—

were collected from 60 automated synoptic observing system (ASOS) locations of the KMA. Daily potential evapotranspiration series are estimated by the FAO Penman-Monteith equation No. 56 method (Allen et al. 1998). These data sets were converted into the mean areal values for each test watershed through the Thiessen Polygon method (Brassel and Reif 1979).

3.2 Rainfall-runoff model

The tank model with soil moisture structure, a modified conceptual rainfall-runoff model, was used in this study as a hydrologic model for runoff simulation. With four tanks and a soil moisture structure, the tank model measures the net stream discharge as

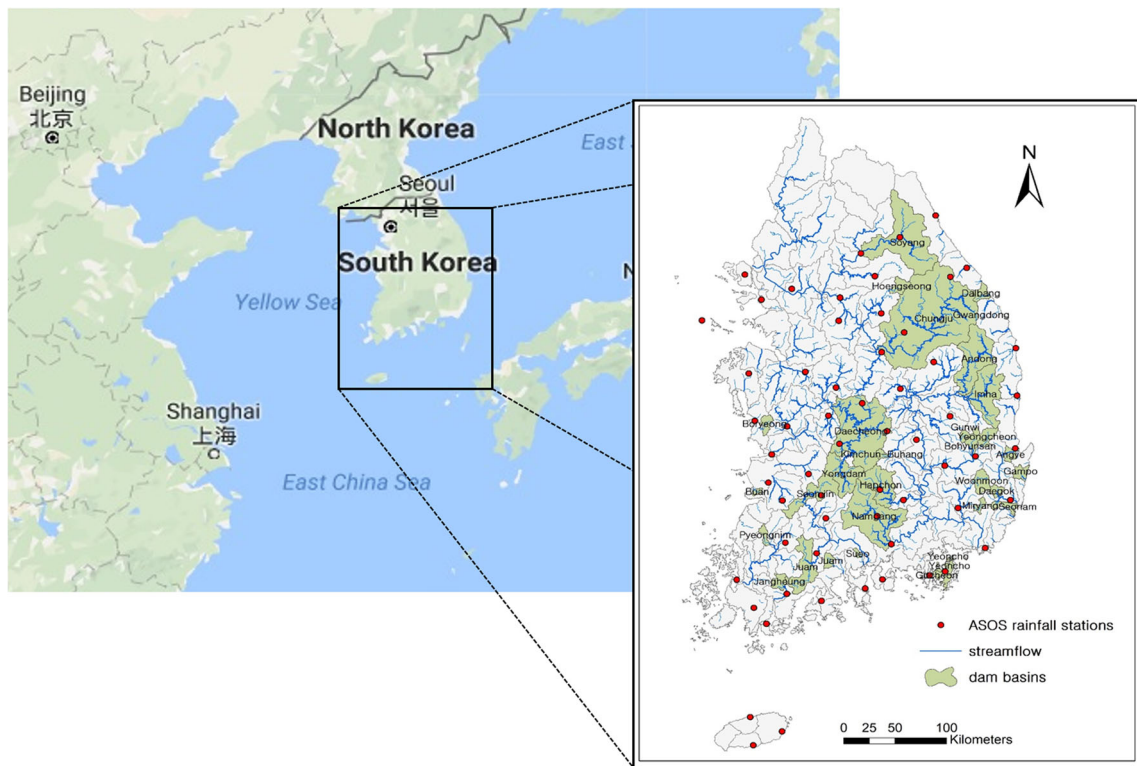


Fig. 5 Application sites: 35 watersheds (dam basins) across South Korea

the sum of the discharges from the side outlets of the tanks (Sugawara 1995). To consider the snow accumulation-melting module, the modified tank model developed by McCabe and Markstrom (2007) was used. The daily time series of the average basin precipitation, temperature, and potential evapotranspiration were used as input data. The parameters of the model were estimated using the shuffled complex evolution algorithm, one of the population-evolution-based global optimization methods

(Duan et al. 1992). Seo and Kim (2018) provide an informative schematic diagram of the tank model.

3.3 Modeling framework

Figure 6 exhibits the modeling framework used for this study that analyzes the role of synthetic forecasts on the accuracy of ESP. First, 1-month ahead streamflow forecasts were obtained

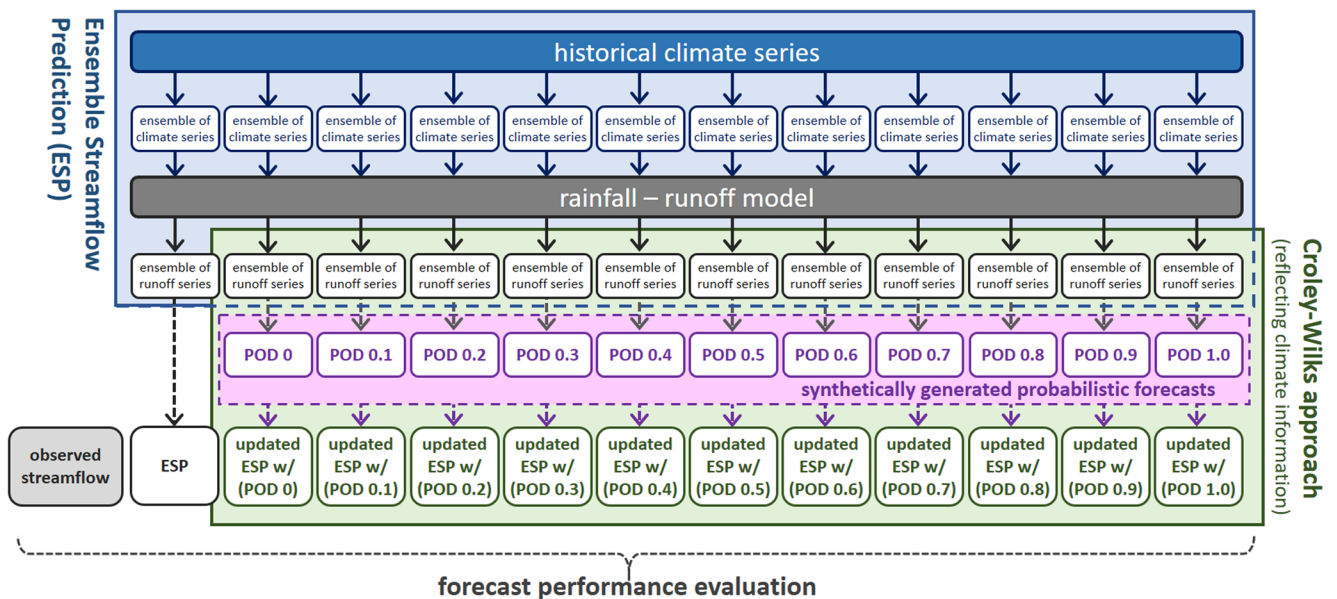


Fig. 6 A diagram of this study’s modeling framework

Table 3 N-RMSE results of 1-month ahead streamflow forecast: ESP and ESP-CW (ESP updated using the Croley-Wilks approach with PPF)

N-RMSE Season	All the forecasts			Forecasts under below-normal conditions		
	ESP ①	ESP-CW ②	②-①	ESP①	ESP-CW ②	②-①
JFM	0.61	0.60	-0.01	0.65	0.64	-0.01
AMJ	1.35	1.14	-0.21	8.44	6.49	-1.95
JAS	1.24	1.28	0.04	7.19	7.53	0.34
OND	0.87	0.86	-0.01	1.23	1.11	-0.12

The values in the table are the averaged values across 35 watersheds. A negative value of ②-① indicates that ESP-CW is more accurate than ESP

by ESP. Then, the results of ESP (ensemble mean of simulated runoff scenarios) were updated with a set of synthetic PPF using the Croley-Wilks approach. Finally, the updated streamflow forecasts were evaluated by comparing them to the observed streamflow series. The application period of this study was from January 1971 to December 2013. The generation of the synthetic PPF was repeated 10 times for each POD case, resulting in total 110 sets (10 sets × 11 POD cases).

4 Results

4.1 Streamflow forecast performance under actual climate forecasts

Before analyzing the role of climate forecasts on streamflow forecast using synthetic PPF, we evaluated the streamflow forecast performance with actual PPF data. Hence, synthetically generated probabilistic forecasts (purple color in Fig. 6) are replaced with the actual probabilistic forecasts for this test. Since KMA has been reporting PPF from June 2014, a total 43 months of streamflow forecast results (June 2014–December 2017) were analyzed.

4.1.1 Deterministic forecast verification

To evaluate deterministic streamflow forecast skill, N-RMSE was calculated for both the ESP and the updated ESP based on Croley-Wilks approach (hereinafter, ESP-CW). Table 3

presents the N-RMSE results for four different seasons (JFM, Jan-Feb-Mar; AMJ, Apr-May-Jun; JAS, Jul-Aug-Sep; OND, Oct-Nov-Dec). Though ESP-CW reduces errors when compared with ESP (except during JAS), the differences were not significant during JFM and OND. The results were also similar under the below-normal condition (i.e., when observed precipitation belongs to the below-normal interval). Thus, this suggests that the impact of PPF on streamflow forecast accuracy is insufficient. Nonetheless, this result does not confirm that ESP-CW is not useful. If more accurate probabilistic forecasts are provided, it would be expected that ESP-CW can improve streamflow forecast accuracy up to a certain level. This is tested using synthetic climate forecasts in the following “Section 4.2.”

4.1.2 Categorical forecast verification

To evaluate categorical streamflow forecast skill, the POD was calculated for both ESP and ESP-CW. Table 4 presents the POD results for the four seasons. Similar to the results of N-RMSE, ESP-CW slightly increases the forecast accuracy when compared with ESP. During AMJ season, ESP-CW is able to improve forecast skill in terms of both N-RMSE and POD metrics. On the other hand, during the other seasons, it appears that ESP-CW is not able to sufficiently increase forecast accuracy compared with ESP. It infers that the PPF is not accurate enough to enhance ESP performance. Thus, we analyzed the forecasting skill of PPF itself. We calculated the RPSS of PPF with climatology (i.e., the values of p_b , p_n , and

Table 4 POD results of 1-month ahead streamflow forecast: ESP and ESP-CW (ESP updated by Croley-Wilks approach using PPF)

POD Season	All the forecasts			Forecasts under below-normal conditions		
	ESP ①	ESP-CW ②	②-①	ESP①	ESP-CW ②	②-①
JFM	0.56	0.56	0.00	0.37	0.41	0.04
AMJ	0.26	0.33	0.07	0.12	0.24	0.12
JAS	0.29	0.32	0.03	0.15	0.20	0.05
OND	0.56	0.56	0.00	0.42	0.43	0.01

The values in the table are the averaged values across 35 watersheds. A positive value of ②-① means that ESP-CW is more accurate than ESP

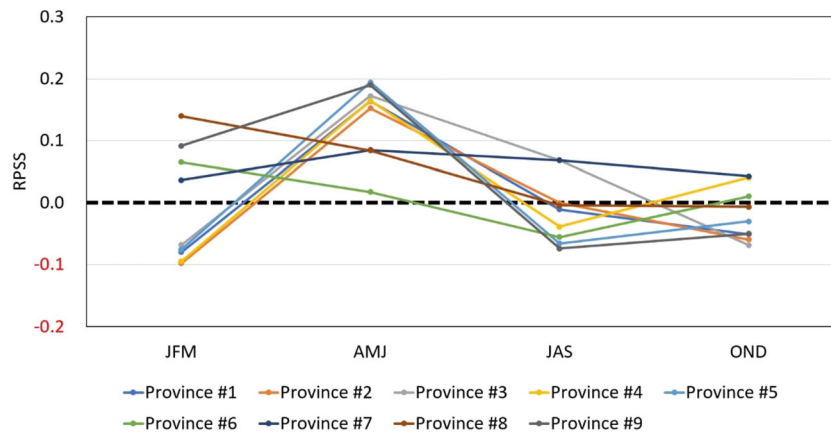


Fig. 7 Ranked probability skill score (RPSS) of the PPF. Province #1 Seoul-Incheon-Gyeonggi-do; #2 Gangwondo Youngseo; #3 Gangwondo Youngdong; #4 Daejeon-Sejong-Chungcheongnam-do; #5

Chungcheongbuk-do; #6 Gwangju-Jeollanam-do; #7 Jeollabuk-do; #8 Busan-Ulsan-Gyeongsangnam-do; #9 Daegu-Gyeongsangbuk-do

p_a are 0.333) as a reference forecast. RPSS values were calculated for 9 provinces in South Korea. Figure 7 shows the RPSS results for the four seasons. Except for AMJ season, RPSS values were below zero in more than 5 provinces, which means the PPF does not provide better predictions than climatology. Though the length of available forecast data sets is not long enough to be statistically reliable, this result shows that the current skill of PPF is not conducive (at least for long-range prediction) to significantly improving streamflow forecasting accuracy. Hence, under the assumption that climate forecasting skill would be improved with rapidly evolving technology, the role of climate forecasting on streamflow forecast accuracy is quantified subsequently.

4.2 Streamflow forecast performance with synthetic climate forecasts

Streamflow forecast performance was evaluated with synthetically generated PPF. Forty-three years of monthly streamflow forecasts (January 1971–December 2013), with 1-month ahead forecasts, were analyzed.

4.2.1 Deterministic forecast verification

To evaluate deterministic forecast skill, N-RMSE and NSE were calculated for both ESP and ESP-CW. Figure 8 presents both the N-RMSE and NSE results for the four different seasons. In terms of ESP skill, the N-RMSE of ESP were 0.91, 0.90, 0.82, and 1.02 for JFM, AMJ, JAS, and OND, respectively. The NSE of ESP were 0.25, 0.05, 0.03, and 0.21 for JFM, AMJ, JAS, and OND, respectively. Note that, in this study, an individual result of each watershed is not plotted in order to avoid messy and complex plots. To analyze overall performance, the results from 35 watersheds were averaged to be plotted. Overall, ranges of the performance index values across 35 watersheds were rather negligible.

For ESP-CW, it was found that N-RMSE decreases as the POD of PPF increases, while NSE increases with the POD of PPF—a result that could be easily anticipated. Interestingly, we found that the mean values of N-RMSE and NSE under ESP-CW, possessing 50% of the POD of PPF, were close to the values of N-RMSE and NSE under ESP. In other words, it implies that POD of PPF should be greater than 50% to be adequately used for updating ESP skill. Since the current PPF cannot provide reliable information compared with climatology, utilizing it to update ESP would be not acceptable.

Our study then asked—what if the PPF skill improve? When the POD of PPF becomes 100%, the mean of the NSE values of ESP-CW becomes 0.41, 0.37, 0.37, and 0.40 for JFM, AMJ, JAS, and OND, respectively. With perfect categorical forecast information, the NSE values of ESP-CW converged to around 0.4, regardless of the season. This is compelling because the NSE values of ESP varied across seasons. Without PPF information, the NSE values of ESP were much lower in AMJ and JAS, which streamflow variability are greater than the other two seasons. However, NSE values of ESP-CW in AMJ and JAS seasons rapidly increase as PPF performance increases. Eventually, with 100% of the POD of PPF, NSE values of ESP-CW reach a certain value across all the seasons. Though 0.4 of NSE may not be an ideal value for 1-month ahead streamflow forecast, it would be difficult to expect NSE values greater than 0.4 with categorical probabilistic forecast information given the significant uncertainty in long-range PPF. It is a given that it is impossible to predict the exact amount of next month's precipitation. Thus, it is difficult to expect forecasting that can assign 100% probability to any given interval with the current level of technology in the field.

4.2.2 Categorical forecast verification

In this section, categorical streamflow forecast skill was evaluated for both ESP and ESP-CW with synthetically generated

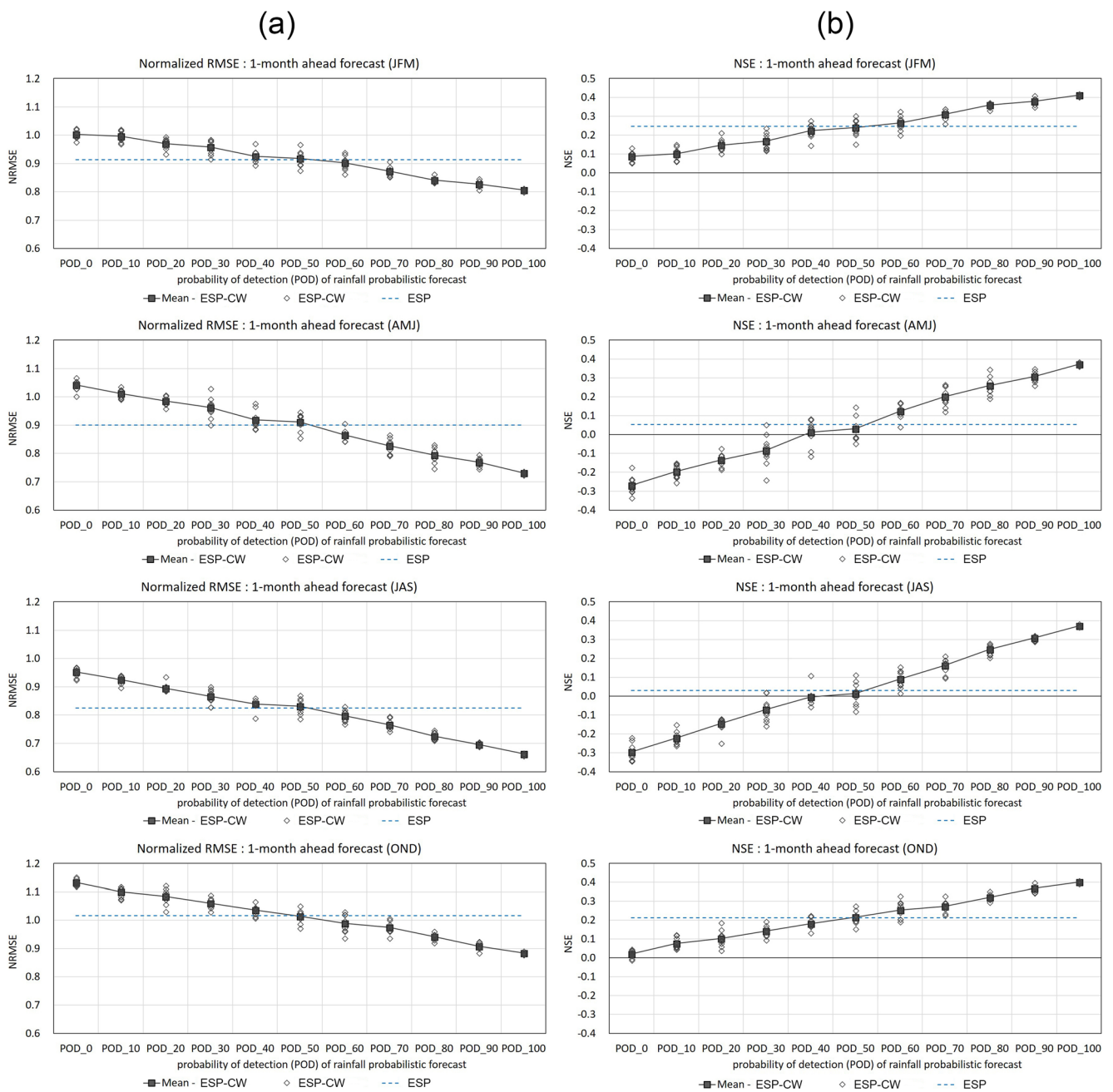


Fig. 8 Deterministic forecast evaluation results of 1-month ahead streamflow forecast for ESP and ESP-CW: **a** N-RMSE and **b** NSE. The values in the figure are the averaged values across 35 watersheds. Ten

different sets for synthetic PPF were tested for each POD of PPF case. In total, 11 cases for the POD of PPF were tested

PPF. Figure 9 presents the POD of streamflow forecasts for the four seasons. The POD of ESP was 0.52, 0.38, 0.36, and 0.47 for JFM, AMJ, JAS, and OND, respectively. Similar to the results of NSE values, the POD of ESP-CW reached a certain value across all seasons with 100% of the POD of PPF. The POD of ESP was also much lower in AMJ and JAS. However, when accurate categorical PPF are provided, it is expected that the categorical streamflow forecast would have approximately 0.55 of POD (of ESP-CW) across all seasons.

Figure 10 presents the POD of streamflow forecasts across all 35 watersheds for four seasons. It shows that the POD of ESP-CW in AMJ and JAS rapidly increase as PPF performance increases, whereas the POD of ESP-CW in JFM and OND does not vary significantly. Thus, there is much potential for increases in streamflow forecast skill in seasons where the streamflow variability is greater. With 100% of the POD of PPF, the POD of ESP-CW eventually converges to a similar value across all seasons.

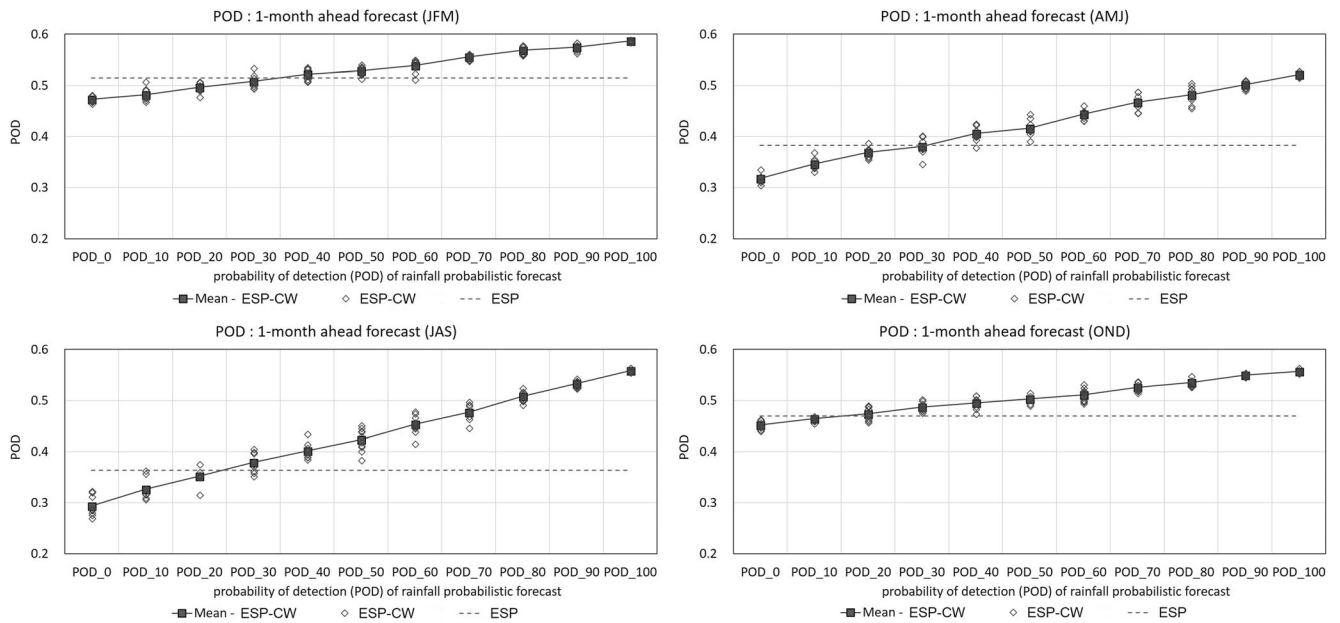


Fig. 9 Categorical forecast evaluation results (POD) of 1-month ahead streamflow forecast for ESP and ESP-CW. The values in the figure are the averaged values across 35 watersheds. Ten different sets for synthetic

precipitation probability series were tested for each POD case. A total of 11 cases for the POD of PPF were tested

Figure 11 presents the POD of streamflow forecasts for the four seasons under below-normal conditions only. The POD of ESP was 0.35, 0.19, 0.17, and 0.39 for JFM, AMJ, JAS, and OND, respectively. The POD of ESP was also much lower in AMJ and JAS. However, under below-normal conditions, the mean POD of ESP-CW reached different values across the four seasons—0.44, 0.31, 0.32, and 0.45 for JFM, AMJ, JAS, and OND, respectively, with 100% of the POD of PPF. These results clearly show that there is little potential for an increase in streamflow forecast skill under below-normal conditions. As highlighted in “Section 4.1,” ESP itself does not have good prediction skill under below-normal conditions. Thus, especially under below-normal conditions, it would be difficult to improve the streamflow forecast accuracy even

with the perfect categorical PPF, unless the forecasting performance of ESP is considerably enhanced.

Lastly, Fig. 12 presents boxplots that show the range of RPSS values across 11 different cases of POD of PPF for the four seasons. Each boxplot represents 350 values of RPSS that are 10 different sets for each watershed (10 sets × 35 watersheds). The median of the RPSS values of ESP (across 35 watersheds) was 0.08, −0.01, −0.01, and 0.04 for JFM, AMJ, JAS, and OND, respectively. The negative median value of RPSS of ESP in AMJ and JAS infers that ESP cannot provide better forecasts than climatology in more than half of 35 watersheds. It also illustrates that ESP-CW has potential to increase forecast quality if accurate categorical PPF are provided. When the value of POD of PPF is greater

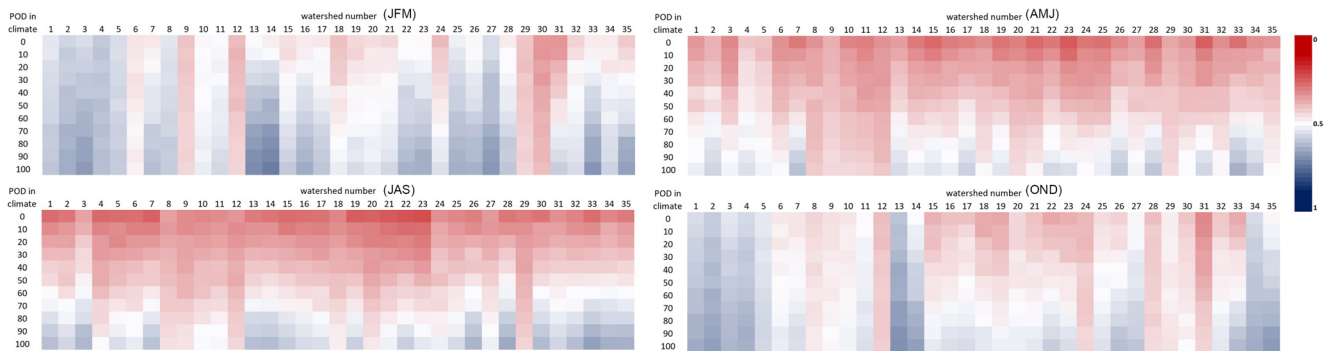


Fig. 10 Categorical forecast evaluation results (POD) of 1-month ahead streamflow forecast for ESP-CW. The values in the figure are the

averaged values across 10 different sets for the synthetic precipitation probability series. A total of 11 cases for POD of PPF were tested

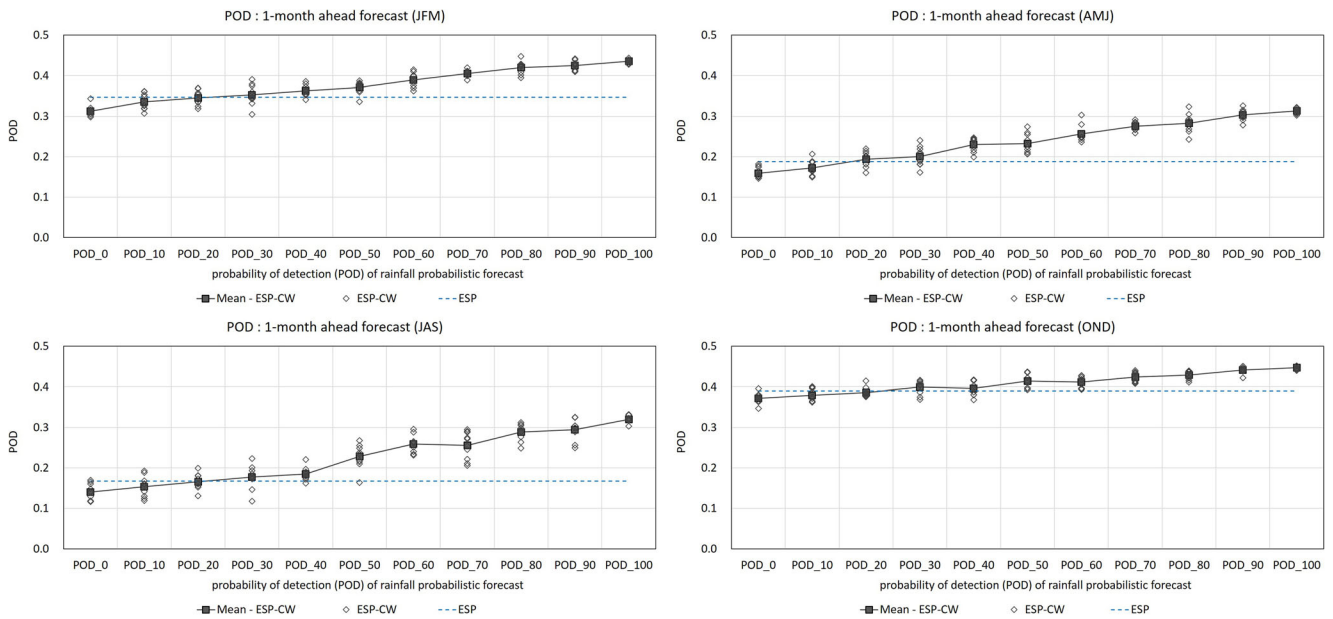


Fig. 11 Categorical forecast evaluation results (POD) of 1-month ahead streamflow forecast for ESP and ESP-CW under below-normal conditions. The values in the figure are the averaged values across 35

watersheds. Ten different sets for synthetic precipitation probability series were tested for each POD case. A total of 11 cases for POD of PPF were tested

than 50%, the values of RPSS of ESP-CW become greater than those of ESP. On the other hand, there are wide ranges in box and whiskers which means the values of RPSS vary a lot across 35 different watersheds. There are several potential reasons that the performance of the streamflow forecasts

varies a lot across different watersheds. One might be different calibration performances of hydrologic model parameters, while the other might be different climate regimes across the target watersheds. Overall analysis on this issue would be beyond of the scope of this study.

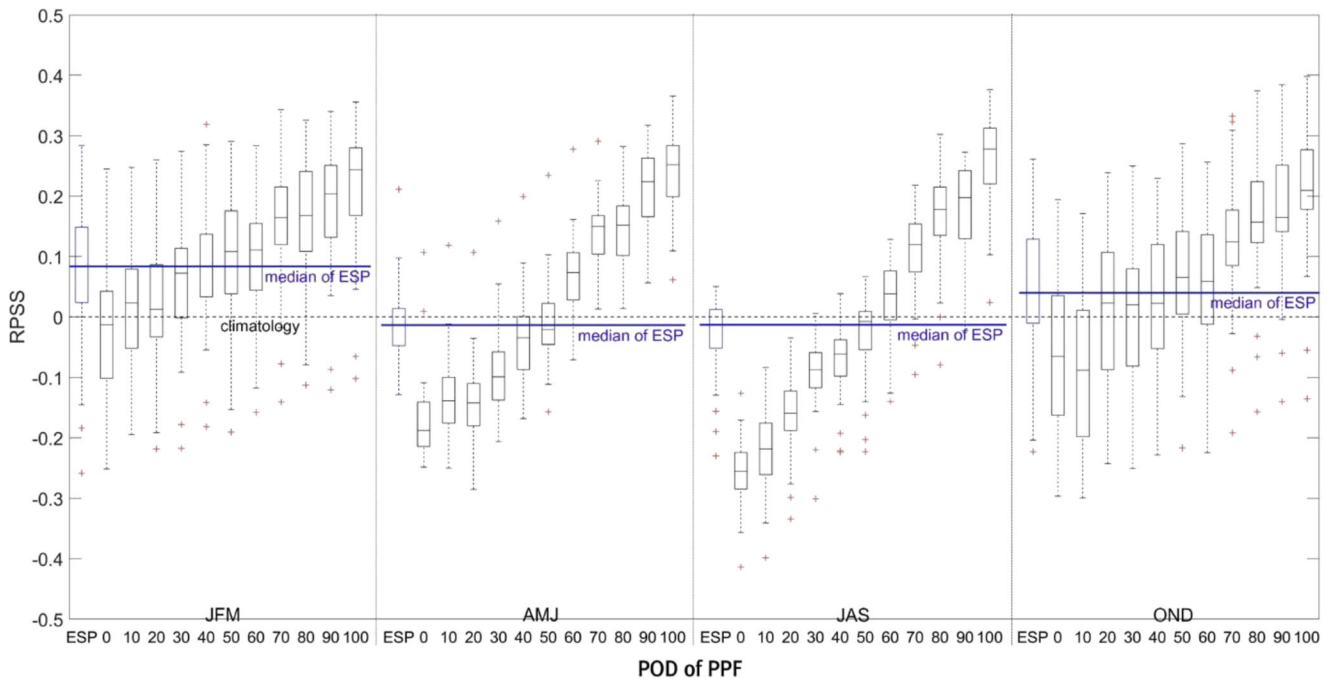


Fig. 12 Ranked probability skill score (RPSS) of 1-month ahead streamflow forecast for ESP and ESP-CW. Box plots display the range of each RPSS value across ten different sets for synthetic precipitation

probability series for each 35 watersheds (i.e., 350 values). A total of 11 cases for POD of PPF were tested for each season—JFM, AMJ, JAS, and OND

5 Conclusions

In the current study, a modeling framework was developed to estimate the role of climate information (PPF) on the accuracy of the streamflow forecast (ESP). We initially found that the actual PPF issuing in South Korea is still not accurate enough to enable streamflow forecast accuracy to be significantly increased. Using a set of synthetic PPF, changes in ESP skill corresponding to the varying accuracy of PPF were evaluated on 35 target watersheds in South Korea. With perfect categorical climate forecast information (i.e., 100% of the POD of PPF), NSE of ESP-CW reaches a similar value (approximately 0.4) across all the seasons. We found that there is much potential for the enhancement of streamflow forecast skill in seasons with greater streamflow variability. With 100% of the POD of PPF, the POD of ESP-CW eventually converges to a certain value (approximately 55%) across all the seasons. Nonetheless, we also found that there is less potential for an increase in streamflow forecast skill under below-normal conditions.

Aside from the findings of the current study, a significant amount of research has been conducted on improving streamflow forecast skill. Regardless of forecasting methods, Mendoza et al. (2017) discussed that there is a general consensus in the research community on the main opportunities to improve streamflow prediction skill (Maurer et al. 2004; Wood and Lettenmaier 2008; Yossef et al. 2013). These include obtaining enhanced knowledge of initial hydrologic conditions (IHC) and weather and climate information during the forecasting period. In terms of harnessing climate information, Mendoza et al. (2017) demonstrated that ensemble forecast errors can be reduced by introducing climate information, despite some discrepancies between techniques. However, it seems that the current weather forecasting technology (at least in Korea) is not yet satisfactory. Besides, even if highly accurate climate information (i.e., 100% POD on PPF) is incorporated into a streamflow forecast technique like the ESP, it would not be able to guarantee streamflow forecast results as accurate as the climate information.

There is still much room for improvement in streamflow forecast skill. Past and ongoing studies have aimed to enhance data assimilation techniques for better IHCs (DeChant and Moradkhani 2011; Huang et al. 2017) and pursue pre- and post-processing methods for the dynamic prediction of streamflow (Wood and Schaake 2008; Yang et al. 2018). Besides, the techniques mentioned above can be incorporated into a variety of climate information such as the climate index (Silva et al. 2017) and probabilistic and deterministic climate forecasts (Bennett et al. 2016). It should be noted that the key aim of the current study was not a development of a state-of-the-art streamflow forecast technique but a discussion about the potential impacts of the accurate climate forecasts on the hydrologic predictions.

The proposed modeling framework can quantify the magnitude of potential improvements in hydrological predictability under the assumption that much more accurate climate information will be available in the future. Although it was beyond the scope of this study, we expect that this potential improvement in hydrological predictability might result from advanced streamflow forecast techniques. We anticipate that this modeling framework can be applied to any other regions across a wide range of climate regimes.

Acknowledgments All authors contributed to the study conception and design. Material preparation, data collection, and first draft writing were performed by Seung Beom Seo, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding information This research was supported by a grant (NRF-2017R1A6A3A11031800) through the Young Researchers program funded by the National Research Foundation of Korea. The authors also thank for University of Seoul for their support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M. (1998) FAO irrigation and drainage paper no. 56. Rome: FAO UN 56(97):e156
- Beckers JV, Weerts AH, Tjeldeman E, Welles E (2016) ENSO-conditioned weather resampling method for seasonal ensemble streamflow prediction. *Hydrol Earth Syst Sci* 20(8):3277–3287. <https://doi.org/10.5194/hess-20-3277-2016>
- Bennett JC, Wang QJ, Li M, Robertson DE, Schepen A (2016) Reliable long-range ensemble streamflow forecasts: combining calibrated climate forecasts with a conceptual runoff model and a staged error model. *Water Resour Res* 52(10):8238–8259. <https://doi.org/10.1002/2016WR019193>
- Bradley AA, Habib M, Schwartz SS (2015) Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach. *Water Resour Res* 51(9):7382–7400. <https://doi.org/10.1002/2014WR016811>
- Brassel KE, Reif D (1979) A procedure to generate Thiessen polygons. *Geogr Anal* 11(3):289–303. <https://doi.org/10.1111/j.1538-4632.1979.tb00695.x>
- Clark MP, Serreze MC, McCabe GJ (2001) Historical effects of El Niño and La Niña events on the seasonal evolution of the montane snowpack in the Columbia and Colorado river basins. *Water Resour Res* 37:741–757. <https://doi.org/10.1029/2000WR900305>
- Croley TE II (2000) Using meteorology probability forecasts in operational hydrology. ASCE Press, Reston
- Croley TE II (2003) Weighted-climate parametric hydrologic forecasting. *J Hydrol Eng* 8(4):171–180. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2003\)8:4\(171\)](https://doi.org/10.1061/(ASCE)1084-0699(2003)8:4(171))
- Day GN (1985) Extended streamflow forecasting using NWS-RFS. *J Water Resour Plann Man* 111(2):157–170. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1985\)111:2\(157\)](https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157))
- DeChant CM, Moradkhani H (2011) Improving the characterization of initial condition for ensemble streamflow prediction using data

- assimilation. *Hydrol Earth Syst Sci* 15:3399–3410. <https://doi.org/10.5194/hess-15-3399-2011>
- Demargne J, Wu L, Regonda SK, Brown JD, Lee H, He M, Schaake J (2014) The science of NOAA's operational hydrologic ensemble forecast service. *Bull Am Meteorol Soc* 95(1):79–98. <https://doi.org/10.1175/BAMS-D-12-00081.1>
- Dettinger MD, Diaz HF, Meko DM (1998) North-south precipitation patterns in western North America on interannual-to-decadal time-scales. *J Clim* 11:3095–4111. [https://doi.org/10.1175/1520-0442\(1998\)011<3095:NSPPIW>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3095:NSPPIW>2.0.CO;2)
- Duan Q, Sorooshian S, Gupta V (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour Res* 28(4):1015–1031. <https://doi.org/10.1029/91WR02985>
- Faber BA, Stedinger JR (2001) Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts. *J Hydrol* 249(1–4):113–133. [https://doi.org/10.1016/S0022-1694\(01\)00419-X](https://doi.org/10.1016/S0022-1694(01)00419-X)
- Franz KJ, Hartmann HC, Sorooshian S, Bales R (2003) Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin. *J Hydrometeorol* 4(6):1105–1118. [https://doi.org/10.1175/1525-7541\(2003\)004<1105:VONWSE>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1105:VONWSE>2.0.CO;2)
- Gershunov A (1998) ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: implications for long-range predictability. *J Clim* 11:3192–3203. [https://doi.org/10.1175/1520-0442\(1998\)011<3192:EIOIER>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3192:EIOIER>2.0.CO;2)
- Gobena AK, Gan TY (2010) Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *J Hydrol* 385(1):336–352. <https://doi.org/10.1016/j.jhydrol.2010.03.002>
- Grantz K, Rajagopalan B, Clark M, Zagona E (2005) A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resour Res* 41(10). <https://doi.org/10.1029/2004WR003467>
- Herr HD, Krzysztofowicz R (2010) Bayesian ensemble forecast of river stages and ensemble size requirements. *J Hydrol* 387:151–164. <https://doi.org/10.1016/j.jhydrol.2010.02.024>
- Huang C, Newman AJ, Clark MP, Wood AW, Zheng X (2017) Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States. *Hydrol Earth Syst Sci* 21(1):635–650. <https://doi.org/10.5194/hess-21-635-2017>
- Hwang Y, Clark MP, Rajagopalan B (2011) Use of daily precipitation uncertainties in streamflow simulation and forecast. *Stoch Env Res Risk A* 25(7):957–972
- Jeong DI, Kim Y-O (2005) Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrol Process* 19:3819–3835. <https://doi.org/10.1002/hyp.5983>
- Kalra A, Ahmad S, Nayak A (2013) Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns. *Adv Water Resour* 53:150–162. <https://doi.org/10.1016/j.advwatres.2012.11.003>
- Kelman J, Stedinger JR, Cooper LA, Hsu E, Yuan S (1990) Sampling stochastic dynamic programming applied to reservoir operation. *Water Resour Res* 26(3):447–454. <https://doi.org/10.1029/WR026i003p00447>
- Kim Y-O, Jeong DI, Kim HS (2001) Improving water supply outlook in Korea with ensemble streamflow prediction. *Water Int* 26(4):563–568. <https://doi.org/10.1080/02508060108686957>
- Kim Y-O, Jeong DI, Ko IH (2006) Combining rainfall-runoff model outputs for improving ensemble streamflow prediction. *J Hydrol Eng* 11(6):578–588. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:6\(578\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(578))
- Kim, H.S., Jeon, K.I., Kang, S.-U., Nam, W.S. (2016) The study on the weighting method of ESP based on probabilistic long-term forecast. The proceeding of KSCE 2016 convention, ICC Jeju, Oct. 2016, Korean Society of Civil Engineers
- Krzysztofowicz R (2001) Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. *J Hydrol* 249(1–4):69–85. [https://doi.org/10.1016/S0022-1694\(01\)00413-9](https://doi.org/10.1016/S0022-1694(01)00413-9)
- Labadie JW (2004) Optimal operation of multireservoir systems: state-of-the-art review. *J Water Resour Plan Manag* 130(2):93–111. [https://doi.org/10.1061/\(ASCE\)0733-9496\(2004\)130:2\(93\)](https://doi.org/10.1061/(ASCE)0733-9496(2004)130:2(93))
- Li Q, Chen J (2014) Teleconnection between ENSO and climate in South China. *Stoch Env Res Risk A* 28(4):927–941
- Liu P, Lin K, Wei X (2015) A two-stage method of quantitative flood risk analysis for reservoir real-time operation using ensemble-based hydrologic forecasts. *Stoch Env Res Risk A* 29(3):803–813
- Maurer EP, Lettenmaier DP, Mantua NJ (2004) Variability and potential sources of predictability of North American runoff. *Water Resour Res* 40:W09306. <https://doi.org/10.1029/2003WR002789>
- McCabe, G.J., Markstrom, S.L. (2007) a monthly water-balance model driven by a graphical user interface (no. 2007-1088). Geological Survey (US)
- Mendoza PA, Wood AW, Clark E, Rothwell E, Clark MP, Nijssen B, Brekke LD, Arnold JR (2017) An intercomparison of approaches for improving operational seasonal streamflow forecasts. *Hydrol Earth Syst Sci* 21(7):3915–3935. <https://doi.org/10.5194/hess-21-3915-2017>
- Najafi MR, Moradkhani H, Piechota TC (2012) Ensemble streamflow prediction: climate signal weighting methods vs. climate forecast system reanalysis. *J Hydrol* 442:105–116. <https://doi.org/10.1016/j.jhydrol.2012.04.003>
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models. Part I: a discussion of principles. *J Hydrol* 10:282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Olsson J, Lindströma G (2010) Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. *J Hydrol* 350(1–2):14–24. <https://doi.org/10.1016/j.jhydrol.2007.11.010>
- Renner M, Werner MGF, Rademacher S, Sprockereef E (2009) Verification of ensemble flow forecasts for the river Rhine. *J Hydrol* 376(3–4):463–475. <https://doi.org/10.1016/j.jhydrol.2009.07.059>
- Seo SB, Kim Y-O (2018) Impact of spatial aggregation level of climate indicators on a national-level selection for representative climate change scenarios. *Sustainability* 10(8). <https://doi.org/10.3390/su10072409>
- Seo SB, Kim Y-O, Kang S-U, Chun GI (2019) Improvement in long-range streamflow forecasting accuracy using the Bayesian method. *Hydrol Res* 50:616–632. <https://doi.org/10.2166/nh.2019.098>
- Silva CB, Silva MES, Ambrizzi T (2017) Climatic variability of river outflow in the Pantanal region and the influence of sea surface temperature. *Theor Appl Climatol* 129(1–2):97–109. <https://doi.org/10.1007/s00704-016-1760-7>
- Stedinger JR, Kim YO (2010) Probabilities for ensemble forecasts reflecting climate information. *J Hydrol* 391(1–2):9–23. <https://doi.org/10.1016/j.jhydrol.2010.06.038>
- Sugawara, M. (1995) Tank model. Computer models of watershed hydrology. Singh, V. P. (Ed.). Highlands ranch, CO: water resources publications
- Wilks DS (2002) Realizations of daily weather in forecast seasonal climate. *J Hydrometeorol* 3(2):195–207. [https://doi.org/10.1175/1525-7541\(2002\)003<0195:RODWIF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0195:RODWIF>2.0.CO;2)
- Wilks DS (2011) Statistical methods in the atmospheric sciences (Vol. 100). Academic press, San Diego
- Wood AW, Lettenmaier DP (2008) An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys Res Lett* 35:L14401. <https://doi.org/10.1029/2008GL034648>

- Wood AW, Schaake JC (2008) Correcting errors in streamflow forecast ensemble mean and spread. *J Hydrometeorol* 9:132–148. <https://doi.org/10.1175/2007JHM862.1>
- Yang T, Tao Y, Li J, Zhu Q, Su L, He X, Zhang X (2018) Multi-criterion model ensemble of CMIP5 surface air temperature over China. *Theor Appl Climatol* 132(3–4):1057–1072. <https://doi.org/10.1007/s00704-017-2143-4>
- Yossef NC, Winsemius H, Weerts A, Van Beek R, Bierkens MFP (2013) Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. *Water Resour Res* 49:4687–4699. <https://doi.org/10.1002/wrcr.20350>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.