Improvement in long-range streamflow forecasting accuracy using the Bayes’ theorem
Seung Beom Seo, Young-Oh Kim, Shin-Uk Kang and Gun Il Chun

ABSTRACT
This study has developed a hydrologic forecasting system for correcting the systematic bias inherent in hydrologic simulations based on the Bayes’ theorem. The observed climatology was used as prior information, and results of a linear regression model that describes the relationship between ‘the observed streamflow’ and ‘the mean of the Ensemble Streamflow Prediction (ESP) forecasts’ was used to form a likelihood function. The Bayes’ theorem was then applied to produce posterior information for the streamflow forecast. Thirty-five watersheds, in which a dam is operated, were tested in this study, and the forecast accuracy was evaluated. It was found that the developed Bayesian ESP (B-ESP) model is capable of improving the forecast accuracy of the ESP. It was found that the forecasting accuracy was improved for all the different lengths of lead-times with the B-ESP model. Nonetheless, the B-ESP model obtained lower RPSS values than the ESP, while its deterministic forecasting accuracy was better than the ESP. This is due to the intrinsic attribute of the Bayesian inference.

Key words | Bayes’ theorem, Ensemble Streamflow Prediction, hydrologic forecast, likelihood, posterior information

INTRODUCTION
Accurate hydrologic forecasts help water resources agencies formulate appropriate plans for available water resources. Forecasts at multiple timescales enable management decisions across various time horizons (Labadie 2004). Seasonal streamflow forecasts benefit a range of water resources management activities (Zhao et al. 2016), such as drought mitigation (Steinemann 2006), flood preparation (Ding et al. 2015; Pappenberger et al. 2015) and reservoir operation (Georgakakos et al. 2012). Besides, many forecast users are interested in sub-seasonal (e.g., monthly) streamflow prediction, which can help water managers achieve efficient short-term decision-making (Alemu et al. 2011; Zhao & Zhao 2014).
Ensemble Streamflow Prediction (ESP) technique (Day 1985) has been popular in the past (Franz et al. 2003; Jeong & Kim 2005; Kim et al. 2006; Gobena & Gan 2010; Najafi et al. 2012). The ESP uses a hydrological model with weather scenarios sampled from the historical record. The soil moisture conditions are initialized at the time of the forecast, and then the hydrological model runs with sampled alternate weather inputs to generate an ensemble of simulated streamflow forecasts.

Meanwhile, Coelho et al. (2004) introduced the Bayes’ theorem to improve the forecast accuracy of El Niño-Southern Oscillation (ENSO). They enhanced the prediction ability of the sea surface temperature in December by combining an empirical regression model with a raw coupled model ensemble that gives forecasts based on the Bayes’ theorem. Since then, the Bayes’ theorem has been applied to improve hydrological forecast accuracy. Luo
et al. (2007) used this Bayes’ theorem to merge ensemble seasonal climate forecasts, which were generated by multiple climate models, for better forecasting. Fang et al. (2009) improved their summer rainfall forecasting ability over the Yangtze River valley using the Bayes’ theorem. Yoon et al. (2012) applied the Bayes’ theorem for seasonal precipitation forecasts over the contiguous United States and compared its results with other methods. More recently, Bradley et al. (2015) utilized the Bayes’ theorem to the climate index weighting of ensemble forecasts. Moreover, there are many other studies that have applied the Bayesian approach for probabilistic hydrologic forecast. Krzysztofowicz (1999) proposed a Bayesian forecasting system (BFS) that can be used for probabilistic forecasting using the deterministic hydrologic model of any complexity. The BFS provides an ideal theoretic framework for uncertainty quantification, which can be broken down into precipitation and hydrologic uncertainty (Han & Coulibaly 2017). Precipitation and hydrologic uncertainty are then integrated to produce a probabilistic forecast (Krzysztofowicz 1999, 2002; Krzysztofowicz & Kelly 2000; Krzysztofowicz & Herr 2001). Since then, other various approaches using Bayesian methods have been proposed to estimate the probability distribution of the predictand that are conditioned based on the available information, e.g., Model Conditional Processor (MCP) (Todini 2008), Bayesian Ensemble Forecast (BEF) (Reggiani & Weerts 2008; Reggiani et al. 2009; Herr & Krzysztofowicz 2010, 2015), Bayesian Joint Probability (BJP) (Wang et al. 2009; Wang & Robertson 2011; Zhao et al. 2015) and Integrated Bayesian Uncertainty Estimator (IBUNE) (Ajami et al. 2007). These studies have been developed to assess the predictive uncertainty in flood forecasting systems. Their key objective is to provide more reliable flood forecasting that can consider all sources of uncertainties.

Nonetheless, there remains room for improvement in hydrologic forecasts. We rather focused on correcting the systematic bias in hydrologic simulations driven by a deterministic model and utilized the overall uncertainty in probabilistic forecasts to assign weight to the correcting factor. For short-term forecasting such as flood forecasting, deterministic forecasting is extremely challenging due to the large natural variability. On the other hand, for mid- or long-term forecasting, more accurate deterministic forecasting can be expected, as natural variability diminishes due to being temporally aggregated. Therefore, we used the simple Bayes’ theorem to correct the systematic bias in hydrological forecasting. Although there are some studies that utilized this simple Bayes’ theorem for hydrologic forecasting after Coelho et al. (2004), there is a limited number of studies that use the Bayes’ theorem for updating the probabilistic streamflow forecast. Bae et al. (2017) developed a hydrological drought forecasting system based on the Bayes’ theorem. However, they only considered the predictability of the streamflow variable and overlooked the resolution of ensemble forecasts. In this study, we consider two different sources of uncertainties in the derivation of the likelihood function, which are the predictability of the variable and the resolution of the ensemble forecasts.

Herr & Krzysztofowicz (2015) developed a posterior density function of the forecast conditioned on antecedent observations (river stages) up to the forecast time. BJP model (Wang et al. 2009) also uses antecedent streamflow as the initial condition predictor for ensemble streamflow forecasting (e.g., Zhao et al. 2016; Bennett et al. 2017). Further, the Ensemble Kalman Filter (Evensen 2003) and Particle Filter methods have been widely used for the data assimilation of observations to improve hydrologic prediction (e.g., Xie & Zhang 2010; Chen et al. 2011; Fan et al. 2017). The fundamental function of this type of data assimilation is to quantify errors in both the hydrological model and observations along with updating hydrological model states in a manner that optimally combines model simulations with observations (Clark et al. 2008). Thus, the role of the observed antecedent conditions in reducing errors in hydrologic forecasting has been demonstrated by a number of studies. In terms of ESP, since the soil’s initial moisture conditions at the time of the forecast are driven by the observed climate data sets, it is inferred that ensemble forecasts driven by ESP are already affected by the observed antecedent conditions.

This study developed a hydrologic forecasting model based on Bayes’ theorem, which includes a prior information and likelihood function. This model is then evaluated in multiple dam watersheds across South Korea. For model fitting, the climatological forecasts are used as prior information, and a regression model that describes the relationship between ‘the streamflow at the time of the forecast’ and ‘the mean ensemble forecast driven by ESP’ is
applied to form a likelihood function. The Bayes’ theorem is then applied to produce posterior information for the streamflow forecast, and the forecast accuracy is evaluated by comparing the results to those of the climatology and the traditional ESP.

The rest of this paper is structured as follows. The ‘Methods’ section describes the theoretical background of the key methodologies. Then, information about a case study is presented in the subsequent section. Following this, the results of the case study are illustrated in terms of calibration and performance validation of the developed Bayesian ESP forecasting model. Finally, the paper ends with the ‘Discussion’ and ‘Concluding remarks’ sections.

METHODS

Rainfall–runoff model: tank model

A modified conceptual rainfall–runoff model (tank model with soil moisture structure) was used as a deterministic hydrologic model for the runoff simulation in this study. Having four tanks and a soil moisture structure, the tank model simulates the net stream discharge as the sum of the discharges from the side outlets of the tanks (Sugawara 2013). In order to consider the snow accumulation–melting module, the modified tank model developed by McCabe & Markstrom (2007) was used. When the temperature is below the value of the specified threshold (\(T_{\text{snow}}\)), all precipitation is considered to be snow. If the temperature is greater than the value of the additional threshold (\(T_{\text{rain}}\)), all precipitation is considered to be rain. Within the range defined by \(T_{\text{snow}}\) and \(T_{\text{rain}}\), the amount of precipitation that becomes snow decreases linearly from 100 to 0% of the total precipitation. This relation is expressed as Equation (1):

\[
P_{\text{snow}} = P \times \left[ \frac{T_{\text{rain}} - T}{T_{\text{rain}} - T_{\text{snow}}} \right] \tag{1}
\]

\(P_{\text{rain}}\) is then computed as \(P_{\text{rain}} = P - P_{\text{snow}}\). For snow melt module, the fraction of snow storage (\(\text{snostor}\)) that melts in a month (snow melt fraction, SMF) is computed from mean monthly temperature (T) and a maximum melt rate (\(\text{meltmax}\)); \(\text{meltmax}\) is often set to 0.5 (McCabe & Wolock 1999). The fraction of snow storage that melts in a month is computed as Equation (2):

\[
\text{SMF} = \frac{T - T_{\text{snow}}}{T_{\text{rain}} - T_{\text{snow}}} \times \text{meltmax} \tag{2}
\]

If the computed SMF is greater than \(\text{meltmax}\), then SMF is set to \(\text{meltmax}\). The amount of snow that is melted in a month (SM) is computed as \(SM = \text{snostor} \times \text{SMF}\).

The daily time series of the basin average precipitation, temperature and potential evapotranspiration were used as input data. The parameters of the model are estimated using the shuffled complex evolution algorithm, one of the population-evolution-based global optimization methods (Duan et al. 1992). Figure 1 illustrates the schematic diagram of the tank model. The tank model parameters for the target watersheds were calibrated and validated by a previous study (Seo & Kim 2018). NSE values for the 35 watersheds ranged from 0.68 to 0.91, and percent BIAS values ranged from −1.57 to 8.93.

Ensemble Streamflow Prediction (ESP)

ESP has been a well-known probabilistic forecasting technique in operational hydrology since it was first introduced in Korea in 2001 (Kim et al. 2001). The ESP runs a rainfall–runoff model with observed meteorological inputs to generate an ensemble of possible streamflow (or runoff) hydrographs (Kim et al. 2001; Gobena & Gan 2010; Najafi et al. 2012). In the ESP technique, all the meteorological scenarios are input into the rainfall–runoff model under an assumption that they are likely to occur in the future. The ESP produces an ensemble of streamflow forecasts by inputting different meteorological traces given the same initial condition at the time of the forecast. Since the initial conditions of the rainfall–runoff model vary depending on the time of the forecast, the ensemble of streamflow forecasts varies as per the initial conditions even for the same meteorological inputs. The initial condition for each point of forecast time is set up by executing the tank model with the observed forcing data sets for a spin-up period (for the previous 5 years). Therefore, the generated...
streamflow ensemble is also a function of the current hydrological states driven by the rainfall–runoff model. Thus, the technique is sometimes called a conditional Monte Carlo simulation approach (Day 1985).

Bayes’ theorem

Bayes’ theorem defines the manner in which prior distribution can be updated when new information is provided (as a form of the likelihood function). The Bayesian inference derives the posterior probability as a consequence of two antecedents, namely, the prior probability and the likelihood function derived from the statistical model for the observed data as shown in Equation (3):

\[ p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \] (3)

where \( \theta \) is a random variable of interest (i.e., dam inflow in this study), \( y \) is new information for the random variable of interest (i.e., ESP in this study), \( p(y|\theta) \) is the likelihood function and \( p(y) \) is the marginal probability of \( y \). In the context to be used, the prior distribution \( p(\theta) \) describes the probability forecast and the posterior distribution \( p(\theta|y) \) describes the updated forecast given \( y \).

All the probability density functions (pdf) are assumed to be a normal distribution function for the direct evaluation of Equation (3).

Prior distribution

The prior distribution can be simply derived from the climatological distribution of dam inflow from historical observations. A normal distribution is fit to the data as the prior distribution \( p(\theta) \sim N(\theta_o, \sigma_o^2) \), where \( \theta_o \) is the mean of the climatological distribution and \( \sigma_o \) is the standard deviation of the climatological distribution. It is noted that other distributions are also possible for the prior distribution. For example, Coelho et al. (2004) used the prediction from an empirical model for this purpose.

Likelihood function

The likelihood function, \( p(y|\theta) \), expresses the probability of the forecast \( y \) given the observed dam inflow \( \theta \). The likelihood function can be estimated from the historical performance of the forecast driven by a model based on the hindcasts (Luo et al. 2007). Since there are an ensemble of dam inflow forecasts (obtained by ESP) for a single observation, conditional distributions are employed to estimate
the likelihood function using the conditional distribution of the ensemble mean, given the observation $p(y|\theta)$, and the conditional distribution of the ensembles, given their mean value $p(y|\bar{y})$, as shown in Equation (4):

$$p(y|\theta) = p(y|\bar{y})p(\bar{y}|\theta)$$

(4)

The relationship between the ensemble mean and the observation is expressed as a linear regression model, shown as follows (Coelho et al. 2004):

$$\bar{y} = \alpha + \beta \theta + \epsilon$$

(5)

where $\alpha$ and $\beta$ are the intercept and slope parameters of the regression model, respectively.

The variable, $\epsilon$, represents the zero-mean residuals of the regression model. It is assumed to be normally distributed, and its variance, $\sigma^2$, reflects the efficiency of the regression. With the linear regression model, $\bar{y}$ follows a normal distribution, $p(\bar{y}|\theta) \sim N(\alpha + \beta \theta, \sigma^2)$. Assuming the ensembles of the current forecast are normally distributed around the mean with the variance, $\sigma^2$, the conditional distribution of $y$ given the ensemble mean can be expressed as $p(y|\bar{y}) \sim N(y, \sigma^2)$.

With an assumption that the current forecast variance, $\sigma^2$, is independent of the long-term linear regression error variance, $\sigma^2$ (Luo et al. 2007), it results in the likelihood function given below:

$$p(y|\theta) \sim N(\alpha + \beta \theta, \sigma^2 + \sigma^2)$$

(6)

As shown in Equation (6), the variance in the likelihood function is composed of two sources – $\sigma^2$, which represents the efficiency of the linear regression that relates the ensemble mean forecast to the matching observation and $\sigma^2$, which represents the spread of the ensemble members around the mean. The larger $\sigma^2$ is the less efficient linear regression in terms of explaining the relationship between $\bar{y}$ and $\theta$ (Luo et al. 2007). This suggests a less skilful forecast model, resulting in a lower weight when merged with the prior. On the other hand, $\sigma^2$, which is the variance of the ensemble members of the current forecast, represents the uncertainties in the forecast system. The greater $\sigma^2$ represents the larger uncertainties in the forecast system and, therefore, the less contribution this forecast system provides.

### Posterior distribution

Based on the Bayes’ theorem, when both the prior distribution and likelihood function follow a normal distribution, the posterior distribution $p(\theta|y)$ can be analytically derived and is also found to follow a normal distribution (Lee 1997; Coelho et al. 2004). The posterior distribution is then given by the following equation:

$$p(\theta|y) = p(\theta|\bar{y}, \sigma^2) \sim N(\theta_p, \sigma_p^2)$$

(7)

with mean $\theta_p$ and variance $\sigma_p^2$ calculated as follows:

$$\frac{1}{\sigma_p^2} = \frac{1}{\sigma^2} + \frac{\beta^2}{\sigma_y^2}$$

(8)

$$\frac{\theta_p}{\sigma_p^2} = \frac{\theta_o}{\sigma^2} + \frac{\beta^2}{\sigma_y^2} \left( \frac{\bar{y} - \alpha}{\beta} \right)$$

(9)

It should be noted that the posterior distribution is conditioned on the entire distribution of $y$, not just on the mean $\bar{y}$.

### Forecast performance verification metrics

Some verification metrics are used for evaluation of forecast quality. First, Nash-Sutcliffe efficiency (NSE) is used to evaluate the quantitative error between observed and predicted values. NSE is a non-dimensional coefficient which can range from $-\infty$ to 1 (Nash & Sutcliffe 1970). NSE is computed through the standardization of the mean squared error between forecasts and observations. Essentially, the closer the NSE is to 1, the more accurate the forecast is.

In addition, probability of detection (POD) is used to evaluate the accuracy of categorical forecasts. The POD is simply the fraction of those occasions when the forecast event occurred on which it was also forecast (Wilks 2011). In this study, a $3 \times 3$ contingency table for categorical forecast verification situation is used, as shown in Figure 2. The categories are divided into ‘below normal’, ‘normal’
and ‘above normal,’ given $x_{ob}^a$ and $x_{ob}^b$ are upper and lower terciles of observations. The values of the thresholds, $x_{ob}^a$ and $x_{ob}^b$, are obtained as 66.7 and 33.3 percentile value of the observations, respectively. Here, the counts for each of the nine possible forecast/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 $3 \times 3$ contingency table represented in Figure 2, the POD would be $(a + e + i)/(a + b + c + d + e + f + g + h + i)$.

For verification of probabilistic forecast in this study, ranked probability skill score (RPSS), the skill score for a collection of ranked probabilistic score (RPS) values relative to the RPS computed from the climatological probabilities, is used. RPS is the most commonly used measure that is capable of penalizing forecasts increasingly, as more probability is assigned to event categories further removed from the actual outcome (Wilks 2011). Readers are referred to Wilks (2011) for details of RPS. Once the posterior distribution is determined, the three interval probabilities, $p_b$, $p_n$, and $p_a$ (called the ‘below normal’, ‘normal’ and ‘above normal’ probability) are calculated by inputting observed terciles ($x_{ob}^a$ and $x_{ob}^b$) to posterior cumulative distribution function, as seen below:

\[
F_{post}(x_{ob}^b) = p_b \\
F_{post}(x_{ob}^a) - F_{post}(x_{ob}^b) = p_n \\
1 - F_{post}(x_{ob}^a) = p_a
\]  

(10)

where $F_{post}$ is a posterior cumulative distribution function.

Note that these three interval probabilities are categorical probabilistic forecasts of the B-ESP model. If we use prior cumulative distribution function for Equation (10), we then obtain categorical probabilistic forecast of the ESP.

**CASE STUDY**

**Application sites and data sets**

Thirty-five watersheds, in which a dam is operated, were tested in this study. Figure 3 shows the locations of the 35 watersheds across South Korea along with 60 ASOS (Automated Synoptic Observing System) rainfall stations used in this study. Table 1 presents a list of the 35 dams applied in this study. Historical observed meteorological data sets from 1966 to 2016 – daily precipitation, maximum and minimum temperature, and average wind speed series – were collected from 60 ASOS locations of the Korea Meteorological Administration (KMA). Daily potential evapotranspiration series were 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 by the Thiessen polygon method (Brassel & Reif 1979).

Daily observed streamflow (dam inflow) series from 1966 to 2016 at 35 dam sites were collected from the K-water Institute. Parameters of the tank model were estimated using the observed data sets from 1971 to 2000, and the model performance was validated comparing simulated streamflow to the observed streamflow from 2001 to 2015. NSE values for the 35 watersheds ranged from 0.68 to 0.91, and percent BIAS values ranged from –1.57 to 8.93.

**Modelling framework**

**Bayesian ESP fitting**

One advantage of the developed B-ESP model is that it can be easily set up for any scale of lead-time. In this study, four different lead-times were modelled using B-ESP: (i) 1-week ahead weekly forecast, (ii) 1-month ahead monthly forecast, (iii) 2-month ahead monthly forecast and (iv) 3-month ahead monthly forecast. As presented in the
’Bayesian approach’ section, $\alpha$, $\beta$, $\sigma^2$ can be obtained by the regression model fitted using ‘the ensemble mean forecasts’ and ‘matching observations’ during the hindcast period, which is from 1971 to 2010 in this study. These three parameters of B-ESP were estimated for each time step of a year, i.e., in the case of the monthly forecast model, we
obtained 12 different values for each parameter, while we obtained 53 different values for the weekly forecast model. The setting-up stage is defined as ‘B-ESP fitting’. B-ESP fitting was implemented for four different lead-times and 35 test watersheds each.

### Bayesian ESP forecast

With a group consisting of the three B-ESP parameters, we can obtain the updated forecast (the posterior mean and variance from the B-ESP) using the mean and variance of the ESP at the time of the forecast. This stage is defined as ‘B-ESP forecast’. Figure 4 illustrates the modelling framework of this study, including the ESP, B-ESP fitting and B-ESP forecast.

### RESULTS

#### Determination of the posterior distribution

The B-ESP models were fitted using the ESP forecasts from 1971 to 2010. Figure 5 illustrates two examples of the determination of the posterior distribution. The blue-coloured histogram represents the ESP forecasts, while the short-dashed line, long-dashed line and solid line represent prior distribution, likelihood distribution and posterior distribution respectively.

When the variances of the prior distribution and likelihood distribution are similar, the posterior mean is located approximately at the middle of the prior and likelihood mean (as shown in Figure 5(a)). On the other hand, when the variances of the prior distribution and likelihood distribution are extremely different (one is much larger than the other), the posterior mean is located close to either the prior or likelihood mean in which the variance is much smaller (as shown in Figure 5(b)). The posterior variance cannot be larger than the smaller variance among the prior and likelihood variance, and it becomes half of the prior variance when both the prior and likelihood variance are the same.

### Bayesian ESP fitting

Figure 6 presents the performance of the two 1-month ahead forecast models (ESP and B-ESP) for the B-ESP fitting period. As shown in Figure 6(a), the B-ESP model has

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Drainage areas (km²)</th>
<th>Reservoir capacity (million cubic meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Soyanggang</td>
<td>2,703</td>
<td>2,900</td>
</tr>
<tr>
<td>2</td>
<td>Chungju</td>
<td>6,648</td>
<td>2,750</td>
</tr>
<tr>
<td>3</td>
<td>Heongseong</td>
<td>209</td>
<td>87</td>
</tr>
<tr>
<td>4</td>
<td>Gwangdong</td>
<td>125</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>Dalbang</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Andong</td>
<td>1,584</td>
<td>1,248</td>
</tr>
<tr>
<td>7</td>
<td>Imha</td>
<td>1,361</td>
<td>595</td>
</tr>
<tr>
<td>8</td>
<td>Seongdeok</td>
<td>41</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>Yeongju</td>
<td>500</td>
<td>181</td>
</tr>
<tr>
<td>10</td>
<td>Gunwi</td>
<td>88</td>
<td>49</td>
</tr>
<tr>
<td>11</td>
<td>Kimcheonbuhang</td>
<td>82</td>
<td>54</td>
</tr>
<tr>
<td>12</td>
<td>Bohyunsan</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>13</td>
<td>Hapcheon</td>
<td>925</td>
<td>790</td>
</tr>
<tr>
<td>14</td>
<td>Namgang</td>
<td>2,285</td>
<td>309</td>
</tr>
<tr>
<td>15</td>
<td>Miryang</td>
<td>95</td>
<td>74</td>
</tr>
<tr>
<td>16</td>
<td>Yeongcheon</td>
<td>235</td>
<td>103</td>
</tr>
<tr>
<td>17</td>
<td>Anye</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>18</td>
<td>Gampo</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>Woommoon</td>
<td>301</td>
<td>160</td>
</tr>
<tr>
<td>20</td>
<td>Daegok</td>
<td>58</td>
<td>36</td>
</tr>
<tr>
<td>21</td>
<td>Seyeon</td>
<td>67</td>
<td>30</td>
</tr>
<tr>
<td>22</td>
<td>Daeam</td>
<td>77</td>
<td>13</td>
</tr>
<tr>
<td>23</td>
<td>Seonam</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>Yeoncho</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>25</td>
<td>Gucheon</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>Yongdam</td>
<td>930</td>
<td>815</td>
</tr>
<tr>
<td>27</td>
<td>Daecheong</td>
<td>3,204</td>
<td>1,490</td>
</tr>
<tr>
<td>28</td>
<td>Seomjin</td>
<td>763</td>
<td>466</td>
</tr>
<tr>
<td>29</td>
<td>Juam</td>
<td>1,010</td>
<td>457</td>
</tr>
<tr>
<td>30</td>
<td>Juam regulator</td>
<td>135</td>
<td>250</td>
</tr>
<tr>
<td>31</td>
<td>Buan</td>
<td>59</td>
<td>50</td>
</tr>
<tr>
<td>32</td>
<td>Boryeong</td>
<td>164</td>
<td>117</td>
</tr>
<tr>
<td>33</td>
<td>Jangheung</td>
<td>193</td>
<td>191</td>
</tr>
<tr>
<td>34</td>
<td>Sueo</td>
<td>49</td>
<td>31</td>
</tr>
<tr>
<td>35</td>
<td>Pyeongnim</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 4 | Modelling framework for Bayesian ESP.

Figure 5 | Examples of determination of the posterior distribution: (a) 1-month ahead forecast for 1989 December at Soyang Dam watershed and (b) 1-month ahead forecast for 2001 March at Namgang Dam watershed. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2019.098.
larger NSE values than the ESP across all the 35 watersheds. Although NSE values of the B-ESP model are still low in summer (June–August), it is found that the B-ESP model is capable of reducing forecasting errors in the ESP across all the seasons. In terms of POD, the B-ESP model also outperformed the ESP. Nonetheless, it is not easy to discern the differences in the POD values in Figure 6(b). Hence, averaged NSE and POD values for the four seasons (JFM, AMJ, JAS and OND) across all the 35 watersheds are presented in Figures 7 and 8, respectively.

As shown in Figure 7(a)–7(d), the B-ESP model yielded larger NSE values than the ESP regardless of the different lead-time. This proves that the B-ESP, which corrects the systematic bias in ensemble mean forecasts obtained from the ESP, can efficiently reduce forecasting errors. Furthermore, the longer the forecast lead-time is, the less efficient both the ESP and B-ESP are. In case of the 2- and 3-month ahead forecasts, surprisingly, climatology forecast shows better forecast efficiency than the ESP. This is because the impact of soil moisture initialization vanishes as the forecast lead-time gets longer. Besides, it can also be caused by the less efficient derivation of the regression model (e.g., OND season of 2- and 3-month ahead forecasts). Figure 8(a)–8(d) also illustrate that the B-ESP model
has larger POD values than the ESP across most seasons. Although the forecast accuracy in summer (high-flow season in South Korea) was still poor, it is found that the B-ESP model has a greater capacity of increasing forecasting accuracy as compared to the ESP. Further, regardless of the forecast model, the forecast performance diminished as the lead-time extended, as was expected.

However, in terms of RPSS, the B-ESP model received lower scores than the ESP. Figure 9 presents the RPSS of the two 1-month ahead forecast models (the ESP and the
B-ESP). As shown, most of the RPSS values of the B-ESP model were below zero (‘–’ icon), except some watersheds in winter and spring. This entails that the B-ESP model was even less accurate than the climatology forecast. Since posterior precision is given by the sum of the prior precision and the data (likelihood) precision, the posterior variance of the B-ESP model is always less than the ESP. Hence, when forecast accuracy is low for both the B-ESP and ESP models, the B-ESP models will suffer a greater penalty regarding the RPS values than the ESP model. Since variance of the posterior distribution always decreases when Bayes theorem is employed, the B-ESP model can lead to quite a narrow resolution of the ensemble spreads. Therefore, although the B-ESP model was capable of reducing errors in forecasting accuracy, it may not outperform the ESP model in terms of probability forecast verification in some cases.

Figure 10 presents the averaged RPSS across all the 35 watersheds for the 1 month-ahead forecast. In AMJ and JAS seasons, when streamflow variabilities are higher...
than the others (JFM and OND), both the forecast models (the ESP and the B-ESP) were not as accurate as climatology. It appears to be extremely challenging to obtain accurate long-term forecasts, especially for the high-flow seasons in which the flow variability is large. As a result of the relative comparison to the ESP model, the B-ESP model was able to produce better forecast performance in terms of the deterministic forecast accuracy, while the B-ESP model was not as reliable as the ESP in terms of probabilistic forecast verification metrics.

**Bayesian ESP forecast verification**

The forecast performance of the two forecast models (the ESP and the B-ESP) was verified from their forecast results from 2011 to 2016. With the estimated B-ESP parameters using the B-ESP fitting period (1971–2010), forecasts were produced for the following 6 years. In Figure 11, the averaged POD values for the 6 years across all 35 watersheds are presented. Similar to the B-ESP fitting results, the B-ESP model is capable of further increasing POD values as compared to the ESP model, except during the AMJ and JAS seasons of the 1 week-ahead forecast model. Nonetheless, the B-ESP model's forecasting performance is still extremely low for the JAS season, which is the flood season in South Korea. Although there were multi-year droughts during the verification period in South Korea, it appears that the long-term forecasting ability with regard to the summer season needs to be improved.

Figure 12 shows the RPSS of the 1-month ahead forecast models (the ESP and the B-ESP) for the verification period. Similar to Figure 9, most of the RPSS values of the B-ESP model were below zero, which entails that the B-ESP model was even less accurate than the climatology forecast. Because the B-ESP model tends to assign less probability to both the tails in its distribution, the narrow spread of the posterior distribution tends to have a larger penalty in RPS scores. Figure 13 presents the averaged RPSS across all 35 watersheds for the 1 month-ahead forecast during the verification period. Similar to the fitting period, both the forecast models (the ESP and the B-ESP) were not as accurate as climatology during AMJ and JAS seasons. Moreover, the B-ESP model obtained lower RPSS values than the ESP across all the seasons, although its deterministic forecasting accuracy was better than the ESP.

**Figure 11** | Forecast performance verification – the averaged POD across all 35 watersheds (B-ESP model forecasts from 2011 to 2016): (a) 1 week-ahead forecast, (b) 1 month-ahead forecast, (c) 2 month-ahead forecast and (d) 3 month-ahead forecast.
DISCUSSION

Data transformation

The proposed B-ESP model follows a basic assumption that the hydrological model forecasts are normally distributed. However, the hydrological model prediction variables are often highly skewed, and their errors are typically abnormal (Schoups & Vrugt 2010). Besides, the normal model allows the flow to be negative, although it rarely occurs. Hence, many hydrologic forecast models use transformation methods for data normalization and convert them into the original space after obtaining the forecasts. From a simple logarithm transformation, Box–Cox transformation (Box & Cox 1964) and Yeo–Johnson power distribution (Yeo & Johnson 2000) have been widely applied (Bates & Campbell 2001; Thyer et al. 2002; Yang et al. 2007; Engeland et al. 2010; Wang & Robertson 2011). Recently, Wang et al. (2012) proposed a log-sinh transformation, which can compensate for the limitations of the Box–Cox transformation, and successfully applied it to the BJP modelling approach for ensemble forecasts (Zhao et al. 2016).

Nonetheless, we have used original meteorological and hydrological data sets without data transformation despite...
the observed hydrologic data sets being skewed for many of the dry months. First, we applied simple log-transformation and Box–Cox transformation schemes before applying the B-ESP. However, after transforming back to the original space, their forecast performance with regard to the posterior distribution was reduced, i.e., both errors in quantitative and probabilistic forecasts were increased as compared to the B-ESP forecasts using the original data sets. The reason behind obtaining these unexpected results should be investigated further, but it is deemed to be beyond the scope of this study. A comprehensive analysis of the error propagation caused by data transformation would be implemented by a future study. In this study, to resolve an issue pertaining to the potential negative forecasts, the likelihood mean is bound to zero to prevent the posterior mean from ever becoming negative. Further, it is assumed that the negative space (below zero threshold) of the posterior distribution is lumped to the zero flows.

Probabilistic forecasts

As mentioned in the section ‘Bayesian ESP fitting’, the posterior precision is always higher as compared to the prior (e.g., climatology) precision. Thus, the posterior variance is less than the prior variance. Due to this intrinsic attribute of the Bayesian inference, the B-ESP model tends to assign less probability to both tails in its distribution (Figure 5(a) is an example of this attribute). Consequently, the B-ESP model obtained lower RPSS values than the ESP despite its deterministic forecasting accuracy being enhanced. This is due to RPSS being a measure that is sensitive to distance. It penalizes forecasts increasingly as more probability is assigned to the event categories that are further removed from the actual outcome (Wilks 2011). This limitation of the B-ESP model in terms of probability forecasts should be overcome through future studies.

CONCLUDING REMARKS

This study developed a hydrologic forecasting model based on the Bayes’ theorem using climatology as prior information. Through a comprehensive application of the 35 dam watersheds across South Korea, it was found that the developed B-ESP model is capable of improving the forecast accuracy of the ESP model.

For deterministic forecast verification, the proposed B-ESP model produces the posterior mean, as it updated forecasting and successfully improved the forecasting accuracy compared to the ESP model. Another merit of this B-ESP model is that it can easily be applied to any lead-time for the forecast. In this study, four different lengths of lead-times (i.e., 1 week, 1 month, 2 month and 3 month) were tested, and the forecasting accuracy was improved for all the cases. On the other hand, in terms of probabilistic forecast verification, it was found that the B-ESP model lessens the performance despite its superiority with regard to deterministic forecast accuracy. This is because of the attribute of the Bayesian inference that increases precision by combining prior information with the likelihood derived from the new information. We believe that this limitation can be resolved by future work, which is being undertaken on the categorized fitting of the B-ESP model depending on the initial condition on the point of the forecast.

Long-term hydrologic forecasting has been challenging due to its low accuracy and high uncertainty. It is expected that the B-ESP model, a simple but capable of correcting errors in the traditional forecasting models, sheds light on how the current hydrological forecasting system in Korea can be improved. Further, the B-ESP model also can be applied to real-time hydrological forecasting, as it is flexible enough to be applied to any lead-time. Corresponding future efforts would be undertaken for the evaluation of the potential improvement in real-time forecasting.

ACKNOWLEDGEMENTS

This research was supported by a research project (201700970) in advanced, probability-based hydrological forecasting technology funded by K-water. This research was also supported by a grant (NRF-2017R1A6A3A11031800) through the Young Researchers program, which is funded by the National Research Foundation of Korea. The authors thank National Drought Information Analysis Centre of K-water for providing the data sets. The authors declare that they have no conflict of interest.
REFERENCES


