Article

Forecasting Seasonal Inflow to Reservoirs Combining Copula-Based Bayesian Network Method with Drought Forecasting

Kwanghoon Kim [1], Sangho Lee [1,*] and Youngkyu Jin [1]

[1] Department of Civil Engineering, Pukyong National University, Busan 48513, Korea, Rep. of; gnsdlekdzizi@gmail.com (K.K.); accvn75@gmail.com (Y.J.)

* Correspondence: peterlee@pknu.ac.kr; Tel.: +82-51-629-6076

Abstract: Especially for drought periods, the higher the accuracy of reservoir inflow forecasting, the more reliable the water supply from a dam. The article focuses on probabilistic forecasting of seasonal inflow to reservoirs and determines estimates from the probabilistic seasonal inflow according to drought forecast results. The probabilistic seasonal inflow was forecasted by a copula-based Bayesian network employing a Gaussian copula function. Drought forecasting was performed by calculation of the standardized streamflow index value. The calendar year is divided into four seasons; the total inflow volume of water to a reservoir for a season is referred to as the seasonal inflow. Seasonal inflow forecasting curves conforming to drought stages produce estimates of probabilistic seasonal inflow according to the drought forecast results. The forecasted estimates of seasonal inflow were calculated by using the inflow records of Soyanggang and Andong dams in the Republic of Korea. Under the threshold probability of drought occurrence ranging from 50 to 55 %, the forecasted seasonal inflows reasonably matched critical drought records. Combining the drought forecasting with the seasonal inflow forecasting may produce reasonable estimates of drought inflow from the probabilistic forecasting of seasonal inflow to a reservoir.

Keywords: drought; copula; Bayesian network; inflow; reservoir

---

1. Introduction

The higher the accuracy of reservoir inflow forecasting, the more reliable the water supply from a dam, especially for drought periods. Ensemble streamflow prediction (ESP) has been a method to forecast streamflow [1-6]. ESP is a typical method of probabilistic forecasting which combines a river basin's initial conditions and past weather conditions, which can reappear in the future, to forecast basin runoff. The forecast results of ESP, however, are greatly influenced by the accuracy of the watershed runoff model as well as the method used to assign weights to the flow scenarios [7]. ESP also assumes that the rainfall scenarios which occurred in the past will also occur in the future. Because of this assumption, ESP can have limitations in properly forecasting severe droughts that have not occurred in the past if the results are not adequately analyzed or combined with other techniques.

Bayesian networks are a probabilistic forecasting method for forecasting future streamflow. Recently in the field of hydrology, research is actively being performed on using Bayesian networks for inflow forecasting and drought forecasting. Madadgar and Moradkhani [8] used a copula-based Bayesian network to forecast flows and droughts in the Gunnison River in the United Sates and the flow forecast results were compared with flow forecast results that used the ESP technique. Madadgar and Moradkhani [8] divided the calendar year into four seasons and forecast the flow for a season based on the flow of the previous season. The flow forecast results that used a copula-based Bayesian network showed a tendency to forecast an increase in the streamflow of a season if the previous season's streamflow increased. However, the streamflow forecast results, which used the ESP technique, showed no tendency with regards to the flow status of the previous season. In

Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233

Madadgar and Moradkhani's [9] later research, they used a copula-based Bayesian network to forecast droughts in the Gunnison River Basin temporally and spatially. Kim et al. [10] developed BAYES-ESP which applies Bayesian theory to the ESP technique to forecast the monthly inflow of multipurpose dams in the Republic of Korea. The monthly inflow forecast results for the Chungju Dam using BAYES-ESP were better than the ESP forecast results that had been too small compared to observed values, but the predicted inflow was far larger than the observed inflow in the period of severe drought in 2015 and the error was very large. Shin et al. [11] used a Bayesian network to perform meteorological drought forecasts and evaluate the accuracy of drought forecasts based on the Republic of Korea's weather observatories. Shin et al. [11] concluded that drought forecasts that use Bayesian networks can be extended for use for not only meteorological droughts but also hydrological droughts.

The goal of this study is to forecast the probabilistic seasonal inflows of dams while taking drought forecast results into account so that dams can reliably provide water. The probabilistic seasonal inflows are forecast results using a copula-based Bayesian network. The drought forecasts for considering future droughts are performed by calculating the drought index. Probabilistic seasonal inflow forecasting combining the drought forecast results has the advantage of being able to forecast reservoir inflow while taking into account the droughts that could occur in the future.

2. Status of the dams for the research

To perform probabilistic forecasting of seasonal inflow and evaluate the results, we must have data that has been recorded over a long period of time. The Korean dams that are suitable for performing our research are the Soyanggang Dam and the Andong Dam, which are relatively important among all the multipurpose dams in the Republic of Korea and have long periods of recordkeeping. The two dams are located on the Han River basin and the Nakdong River basin, respectively (Figure 1). Data on the Soyanggang Dam's inflow has been provided since January 1974, and data on the Andong Dam's inflow has been provided since January 1977. The basin area of the Soyanggang Dam is 2,703 km^2. The Soyanggang Dam's total storage volume is 2.90 billion m^3, making it the largest multipurpose dam in the Han River basin. The Soyanggang Dam's yearly average inflow is 1.75 billion m^3, and it provides 1.213 billion m^3 of water for use every year. The Andong Dam is located on the Nakdong River, and its basin area is 1,584 km^2. The Andong Dam's total storage volume is 1.24 billion m^3 and it is the largest multipurpose dam in the Nakdong River basin. The Andong Dam's yearly average inflow is 0.940 billion m^3 and it provides 0.926 billion m^3 of water for use every year.

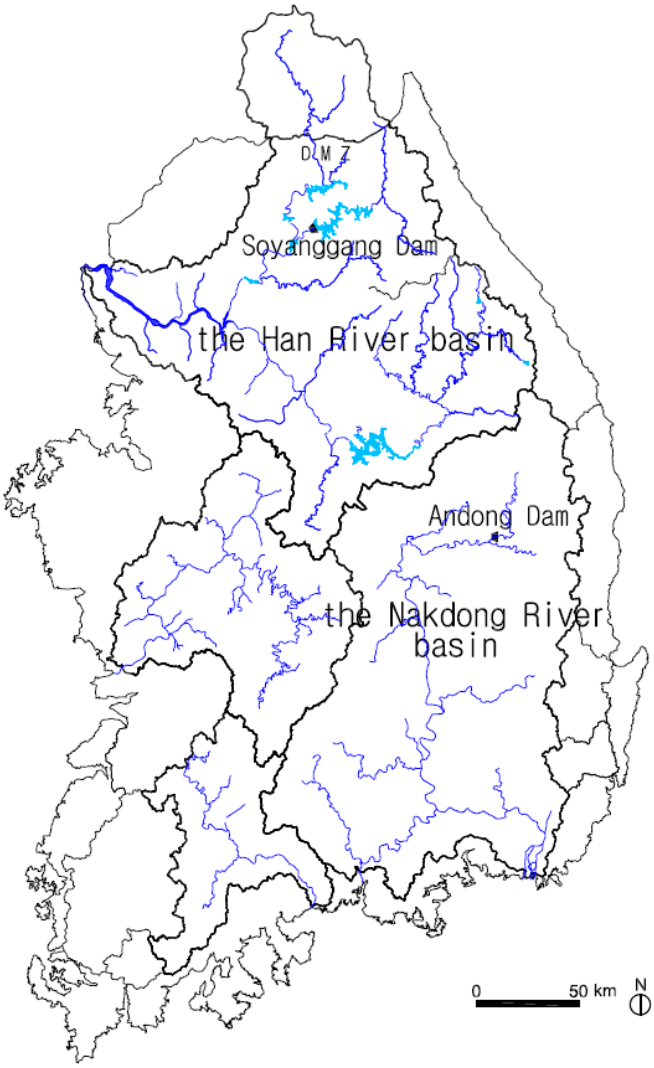
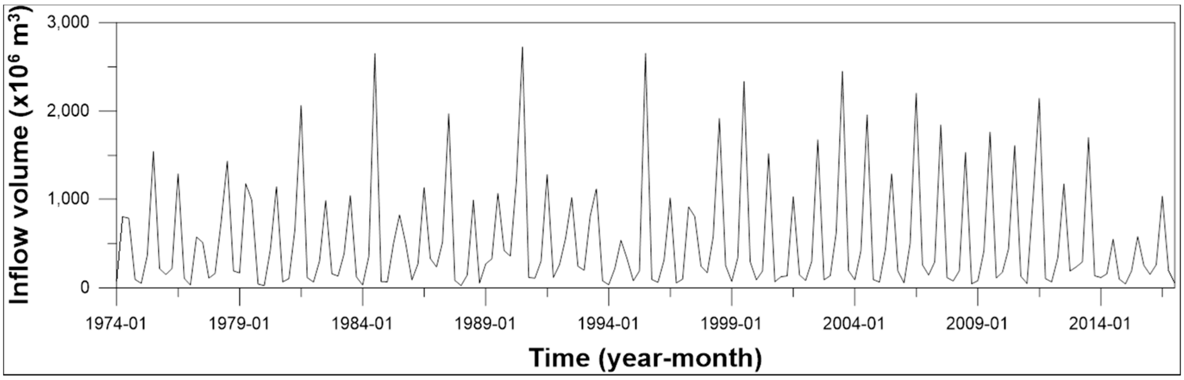


Figure 1. Locations of the Soyanggang Dam and Andong Dam.

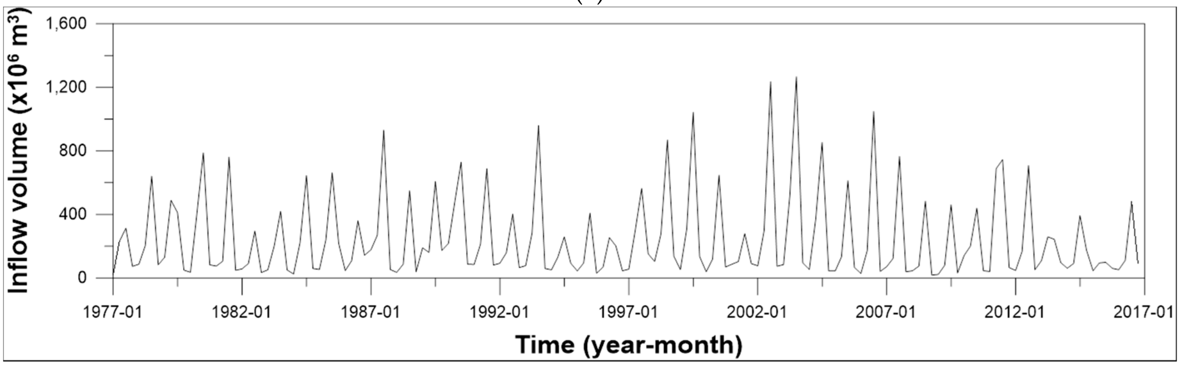
The Andong Dam's inflow data were collected from January 1977 to December 2016 and the Soyanggang Dam's inflow data were collected from January 1974 to December 2016. To forecast seasonal inflow, the calendar year was divided into four seasons: spring is from April to June; summer is from July to September; fall is from October to December; winter is from January to March. Total inflow volume of water to a reservoir for a season is referred to as the seasonal inflow. Fig. 2 shows the Soyanggang Dam and Andong Dam's seasonal inflow time series. The Soyanggang Dam's average summer inflow is 1.439 billion m^3. The year in which the summer inflow was the smallest was 2014. At that time, the summer inflow was 0.513 billion m^3, which is around 35 % of the average inflow. Also, in 2015 the Soyanggang Dam's summer inflow was 0.578 billion m^3, which is 40 % of the average inflow. The Andong Dam's average summer inflow is 607 million m^3. In 2015 when the Andong Dam's summer inflow was the smallest, it was 98 million m^3, which is 16 % of the average inflow. Unlike the Soyanggang Dam, the Andong Dam's 2013 summer inflow was smaller than in 2014. The 2013 summer inflow was 243 million m^3, which is 40 % of the average inflow. As can be realized from the two dam's inflow states, the drought that occurred in the Republic of Korea from 2013 to 2015 was fairly severe. We used the seasonal inflow data from the first year of observation to 2010 to train the copula-based Bayesian network model. The seasonal inflow data from 2011 and 2016 were used to verify the forecast results from the trained copula-based Bayesian network module.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233



(**a**)



(**b**)

Figure 2. Seasonal inflow of dams. (a) Soyanggang Dam; (b) Andong Dam.

3 Methodology

3.1 Procedures of seasonal inflow forecasting

The following shows the overall procedures used to forecast a seasonal inflow to a reservoir combining the copula-based Bayesian network method with drought forecasting: 1) a preparation for a current seasonal inflow and its standardized streamflow index (SSI): 2) drought forecasting using the SSI from a Bayesian network (BN): 3) taking a seasonal inflow forecasting curve developed from a BN conforming to the forecasted drought stage: and 4) deciding a forecasted seasonal inflow on the selected seasonal inflow forecasting curve. The following sections describe the details of the above procedures.

3.2 Copula-based Bayesian network

Bayesian networks are probabilistic models that describe the conditional dependencies of a set of random variables via directed acyclic graphs (DAG). A DAG represents the sequence of events in a direct ordering with no direct circuits. The random variables (x) that evolve over time (e.g., streamflow or drought states) can be shown in a DAG and their probabilistic queries can be represented within a Bayesian network. The joint probability density function of the set of random variables in vector **x** $(x_{t_1}, x_{t_2} \cdots, x_{t_n})$ forming a Bayesian network can be written as the product of individual density functions conditional on their parent variables [8,12]. If the dependency ordering of random variables exactly follows the temporal sequence and the parent variables of x_{t_i} is the set of all prior variables $(x_{t_{i-1}}, x_{t_{i-2}} \cdots, x_{t_1})$, the joint probability density function of the **x** can be written as Equation (1).

$$f(\mathbf{x}) = f(x_{t_1}, \cdots, x_{t_n}) = \prod_{\forall t \in T} f(x_{t_i} | x_{t_{i-1}}, \cdots, x_{t_1}) \tag{1}$$

where Π is the product operator, x_{t_i} represents the random variable at time t_i, and T is the length of the time period over which the random variables evolve. That is the chain rule in the probability

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at *Water* **2018**, *10*, 233; doi:10.3390/w10020233

theory, and the conditional probabilities from Equation (1) can be shown simply by introducing copulas as in Equation (2) [8].

$$f\left(x_{t_n} \middle| x_{t_{n-1}}, \cdots, x_{t_1}\right) = \frac{c(u_{t_n}, \cdots, u_{t_1}) \prod_{i=t_1}^{t_n} f_{x_i}(x_i)}{c(u_{t_{n-1}}, \cdots, u_{t_1}) \prod_{i=t_1}^{t_{n-1}} f_{x_i}(x_i)} \tag{2}$$

where c is the copula density function, u_{t_n} means $F_{x_n}(x_n)$, and $f(\cdot)$ is the density function of a marginal distribution. A detailed description of copula-based Bayesian network is found in Madadgar and Moradkhani [8].

In this study, Equation (3) was used to calculate the conditional probability of two successive seasons' inflows.

$$f\left(x_{t_2} \middle| x_{t_1}\right) = \frac{c(u_2, u_{t_1}) f_{x_{t2}}(x_{t_2}) f_{x_{t1}}(x_{t_1})}{f_{x_{t1}}(x_{t_1})} = c(u_2, u_{t_1}) f_{x_{t2}}(x_{t_2}) \tag{3}$$

here, x_{t_2} is the seasonal inflow to be forecast, and x_{t_1} is the previous season's inflow. In the hydrological field, the Archimedean and elliptical families have been the main copula families. Madadgar and Moradkhani [8] showed that the Gaussian copula is the most suitable for forecasting the flow of the Gunnison river in the United States. Also, Yoo et al. [13] used a copula function on Korean weather observatories and performed bivariate frequency analysis for droughts. They presented results showing that the Gaussian copula was the most suitable for the data from 32 out of the 57 target weather observatories. In this study, we selected the Gaussian copula as the copula function to calculate Equation (3). The Gaussian copula is in the elliptical family, and it is easy to calculate compared to other copula functions. This is because the parameters do not need to be estimated and it can be calculated using only the correlation coefficient between two variables. The Gaussian copula equation is as follows [14]:

$$c(u_1, u_2, \cdots, u_n; R) = \frac{1}{|R|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} u'(R^{-1} - I)u\right] \tag{4}$$

here, $u = (u_1, u_2, \cdots, u_n)'$ and $u_i = \Phi^{-1}\left(F_i(X_i)\right)$. The $\Phi^{-1}\left(F_i(X_i)\right)$ is the inverse function of the cumulative normal distribution function, and R is the correlation coefficient matrix made from ρ, the correlation coefficient between two variables.

In this study, we used the probability distribution of the next season's inflow calculated via Equation (3) from the previous season's inflow to make inflow forecasts. Figure 3 shows the probability distribution of the Andong Dam's summer inflow when the spring flow is 100 million m^3, calculated using Equation (3). According to the probability distribution, when the spring inflow is 100 million m^3, the summer inflow with the greatest probability density function value is around 480 million m^3. The next-season inflow with the greatest probability density function value is used as the inflow forecast value when a drought is not forecast in an inflow prediction that incorporates drought forecasting, which will be discussed later.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1
Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233

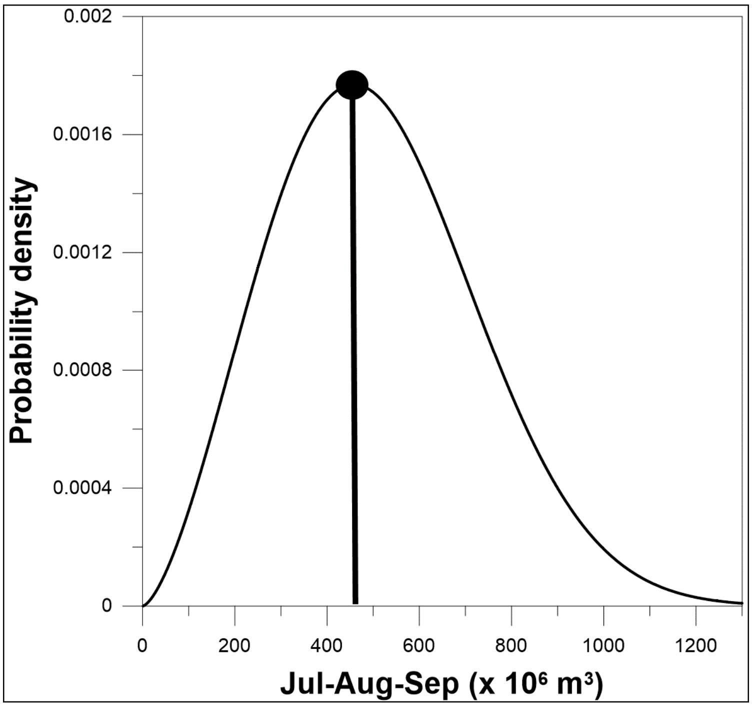


Figure 3. Probability density function of summer inflow conditioned on a spring inflow of 100 million m^3.

3.3 Drought Forecasting Using Drought Index

Determining the forecast inflow to be the inflow that has the largest probability density function value (as in Figure 3) has the limitation of not taking the probability of a drought into account. Because of this, in this study we also performed a drought forecast to create an inflow forecast that has a focus on droughts. The drought forecast is performed using the standardized streamflow index (SSI), which is one of the drought indexes. The SSI is similar in concept to the standardized precipitation index developed by Mckee et al. [15], but it is a drought index that is derived from calculating the lack of streamflow rather than the lack of precipitation. Table 1 shows the categories of drought stages using SSI [16]. A negative SSI is categorized as a drought and a larger negative value means a more severe drought.

Table 1. Classification of drought severity by the range of SSI.

SSI range	Drought category
$2.00 \leq Z$	Extreme wet
$1.50 \leq Z < 2.00$	Very wet
$1.00 \leq Z < 1.50$	Moderately wet
$0.00 \leq Z < 1.00$	Near normal
$-1.00 \leq Z < 0.00$	Mild drought (D1)
$-1.50 \leq Z < -1.00$	Moderate drought (D2)
$-2.00 \leq Z < -1.50$	Sever drought (D3)
$Z < -2.00$	Extreme drought (D4)

To calculate the SSI for each season, we must first determine the probability distribution of the seasonal inflow. A set of distributions is tested to find the best one fitted to the seasonal inflow. The following five distributions are considered in this study: lognormal, gamma, Gumbel, Weibull, and Gaussian distributions. The method of maximum likelihood estimation (MLE) is used to estimate the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233

parameters of each distribution. To find the best distribution fitted to the seasonal inflow, the Kolmogorov-Smirnov (K-S) test [17] is applied. K-S test returns the D_n value. D_n is the maximum value among the differences between the theoretical cumulative probability distributions derived from estimated parameters and the probability distributions derived from observed data. The value of D_n is limited by the number of data values and the significance level of α. A probability distribution in which D_n is calculated to be beyond these values is considered to be inappropriate for the data. In this study, we compared the D_n of several probability distributions and chose the probability distributions with the smallest calculated D_n as the seasonal inflow probability distribution. If an optimal probability distribution is determined for each seasonal inflow, calculating the cumulative probability for a certain inflow becomes possible, and this can be used to calculate the SSI. In this study, we used the method proposed by Abromowitz and Stegun [18] to calculate the SSI, which uses cumulative probability, the equation for which is as follows.

$$t = \sqrt{ln\left(\frac{1}{\left(H(x)\right)^2}\right)}, \quad 0.0 < H(x) \leq 0.5 \tag{5}$$

$$t = \sqrt{ln\left(\frac{1}{1-\left(H(x)\right)^2}\right)}, \quad 0.5 < H(x) \leq 1.0 \tag{6}$$

$$Z = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), \quad 0.0 < H(x) \leq 0.5 \tag{7}$$

$$Z = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right), \quad 0.5 < H(x) \leq 1.0 \tag{8}$$

here, $H(x)$ is the cumulative probability calculated from the probability distribution function; c_0 is 2.515517; c_1 is 0.802853; c_2 is 0.010328; d_1 is 1.432788; d_2 is 0.189267; d_3 is 0.001308.

If the seasonal SSI is calculated through the process above, Equation (3) is used to calculate the probability distribution of the next season's SSI from the previous season's SSI. The non-exceedance probability of the next-season SSI probability distribution in which the SSI is less than or equal to zero is the probability that a drought will occur in the next season. Figure 4 shows the probability distribution of the Andong Dam's summer SSI when the spring SSI is -1. Here, the area of the shaded part is the probability that the SSI is less than zero in summer. Drought forecasting is performed by comparing the probability that the SSI is less than zero and the threshold probability of drought occurrence. If the probability of a drought occurring next season exceeds the threshold probability of drought occurrence, then a drought is predicted for the next season, otherwise a drought is not predicted.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233

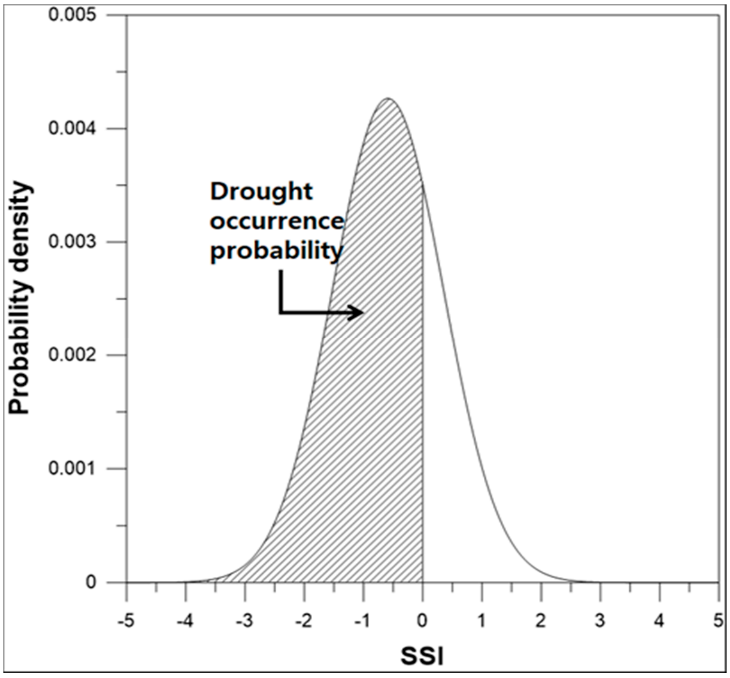


Figure 4. Probability density function of summer SSI conditioned on a spring SSI of -1.

3.4 Inflow Forecasting Combined with Drought Forecasting

If we want to forecast the inflow using the drought forecast results, we must estimate the inflow, which represents each drought stage. To do this we estimated the seasonal inflow forecasting curves conforming to drought stages. The seasonal inflow forecasting curves conforming to drought stages are estimated through the following method:

(1) Use the probability distribution of the next season inflow (like that in Figure 5) to calculate the cumulative probability that corresponds to a specific inflow of the next season.

(2) Use the cumulative probability to calculate the SSI.

(3) When the calculated SSI is the same as the lower bound value of each drought stage, set the corresponding inflow as the inflow that represents that drought stage. (For D4, the inflow where the SSI is -2.5)

(4) Repeat steps (1) to (3) for all seasons and inflows.

If seasonal inflow forecasting curves conforming to drought stages are made, we must determine which curve will be selected to forecast the inflow when a drought has been forecast. If a dam is being operated, especially for the purpose of municipal and irrigation water supplies, a drought over two successive seasons could be fatal. Furthermore, continuous droughts in spring and summer can be very fatal due to the climate characteristics of Korea where rain is concentrated in the summer season. In this study, if a drought occurred in the previous season and a drought is forecast for the following season, the following season's forecast inflow is determined through the inflow forecasting curve of the drought stage that is raised one stage above the drought stage of the previous season. On the other hand, if a drought is not forecast, the inflow with the largest probability density function value is used as the following season's inflow forecast value, as explained in Section 3.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at *Water* **2018**, *10*, 233; doi:10.3390/w10020233

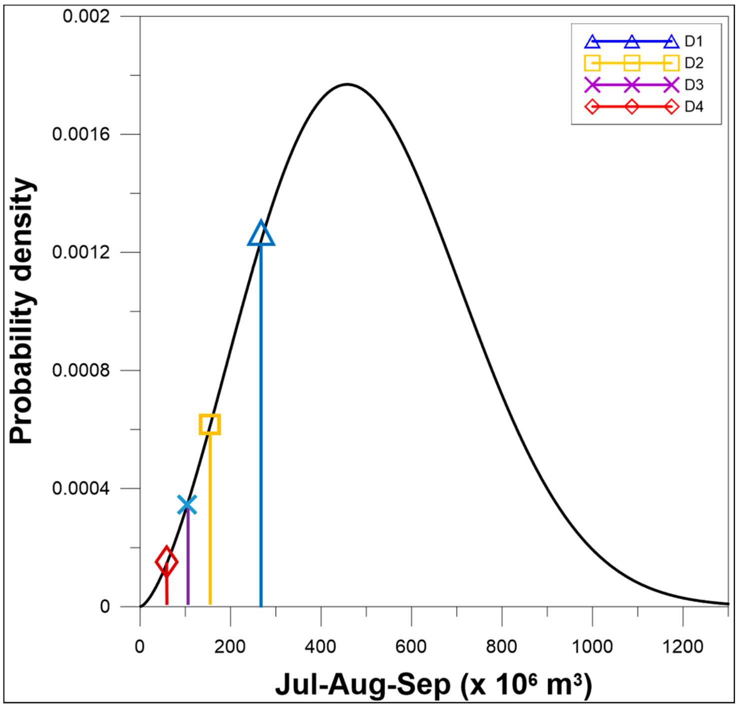


Figure 5. Example of estimating representive inflows corresponding to lower bounds of SSI ranges of drought stages.

4. Research Results

4.1 Seasonal Inflow Forecasting Curves Conforming to Drought Stages

To create a copula-based Bayesian network, we must first determine the best distribution of the seasonal inflow. Figure 6 and Figure 7 show histograms and five kinds of probability distribution curves calculated from the parameters estimated by the MLE from the seasonal inflow data for the Soyanggang Dam and the Andong Dam, respectively. Table 2 shows the results of the K-S test with a significance level α set at 0.05. In Table 2, the values shown in bold are the smallest values of D_n that do not exceed the threshold values and lognormal distribution is found to be the best fit to the seasonal inflow volumes except summer inflow volumes of Andong Dam.

Next, Table 3 shows the results of calculating the correlation coefficient of the inflows between the two successive seasons, which is a variable used in the Gaussian copula function. Looking at Table 3, the correlation coefficients of the inflows between two successive seasons at the Soyanggang Dam are not large with the exception of winter and spring. The small correlation coefficient between spring and summer inflows means that the linear dependence between spring and summer inflows is weak. On the other hand, in the case of the Andong Dam, the inflow correlation coefficients between winter and spring and between spring and summer were found to be larger than those from other seasons. These results mean that there is strong linear dependence between the inflows in spring and summer and the inflows in the previous seasons.

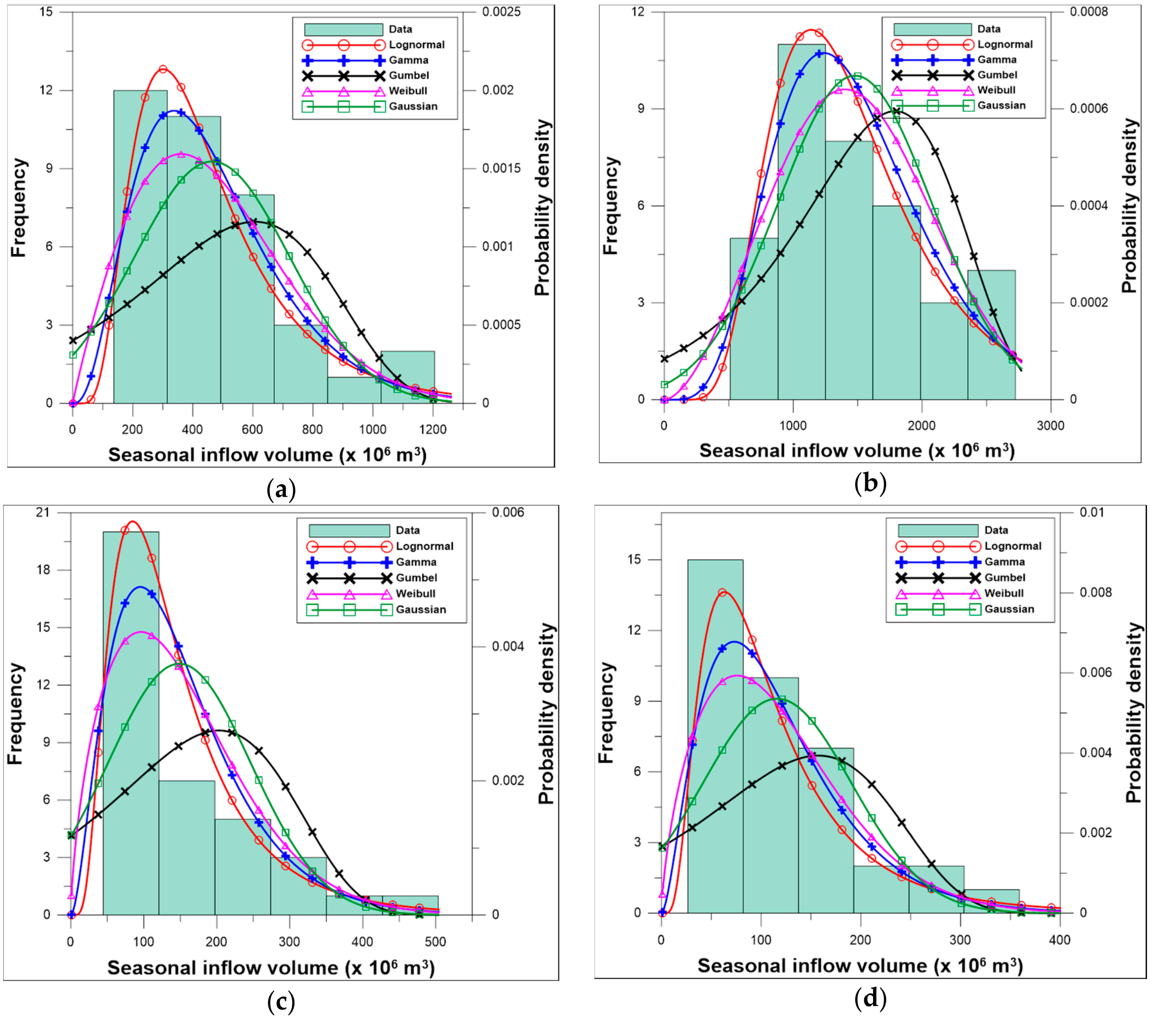
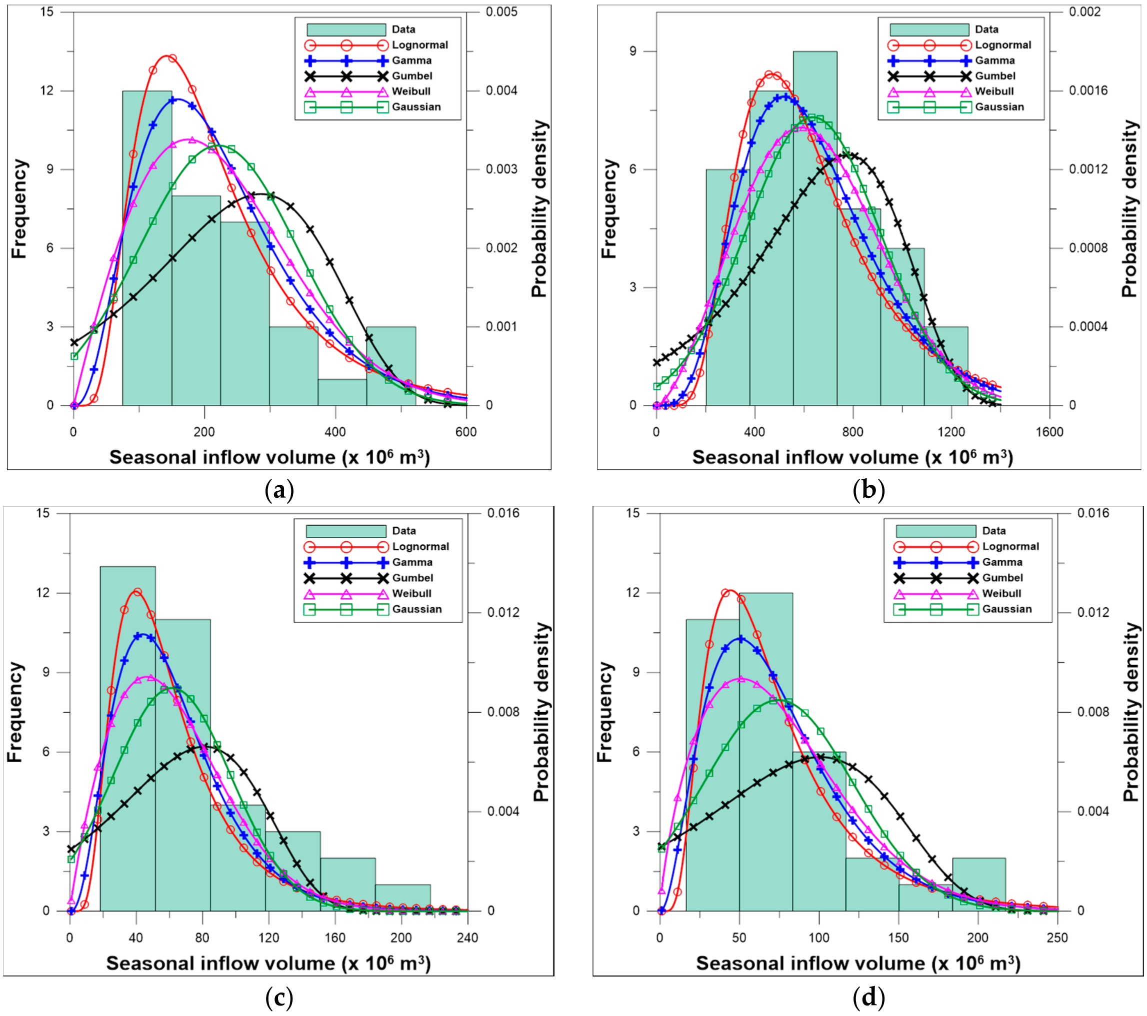


Figure 6. Histogram against the five distributions of the Soyanggang Dam seasonal inflow volumes during the training period of 1974-2010. (a) Spring; (b) Summer; (c) Fall; (d) Winter.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233



245 **Figure 7.** Histogram against the five distributions of the Andong Dam seasonal inflow volumes
246 during the training period of 1977-2010. (a) Spring; (b) Summer; (c) Fall; (d) Winter.

247

Table 2. The D_n value of the K-S test.

Dam	Distribution	Season				Threshold value
		Spring	Summer	Fall	Winter	
Soyanggang	Lognormal	**0.0636**	**0.0898**	**0.1242**	**0.0724**	0.219
	Gamma	0.0898	0.1110	0.1583	0.1062	
	Gumbel	0.2029	0.1534	0.2334	0.2189	
	Weibull	0.1127	0.1261	0.1601	0.1169	
	Gaussian	0.1570	0.1438	0.2082	0.1691	
Andong	Lognormal	**0.1001**	0.1057	**0.0813**	**0.0915**	0.232
	Gamma	0.1123	0.0771	0.1072	0.1267	
	Gumbel	0.1930	0.1288	0.2396	0.2457	
	Weibull	0.1038	**0.0736**	0.1380	0.1394	
	Gaussian	0.1205	0.0932	0.1713	0.1871	

Table 3. The correlation coefficient of each season’s inflow to that of prior season.

Dam	Spring - Summer	Summer - Fall	Fall - Winter	Winter - Spring
Soyanggang	0.0543	-0.1770	0.0739	0.4781
Andong	0.4828	0.1281	0.1831	0.3767

The previously determined probability distribution of seasonal inflow and the Gaussian copula function were used to create the two dams' copula-based Bayesian networks. Figure 8 and Figure 9 show the seasonal inflow forecasting curves conforming to drought stages that were determined using the networks. In the figures, the horizontal axis denotes the previous season's inflow and the vertical axis denotes the next season's inflow. The black circles in the image show the observed seasonal inflows from 2011 to 2016. In Figure 8(a) the spring inflow forecasting curve conforming to drought stages shows a trend due to the effect of the correlation coefficient of inflow between winter and spring. On the other hand, there are few such trends for other seasons. Looking at the Soyanggang Dam's summer inflow forecasting curve conforming to drought stages, there are two observed inflows that are much less than the other observed inflows. The two values, which represent the observed inflows in summer 2014 and 2015 when a severe droughts occurred at the Soyanggang Dam, are located close to the curve that corresponds to drought stage D3. If a drought had occurred in the spring of 2014 and 2015 and a drought was forecast for summer, it would have been possible to forecast the inflows corresponding to these two observed values.

At the Andong Dam, the correlation coefficients of inflows for winter-spring and spring-summer were relatively large. As a result, the slopes of the Andong Dam's spring inflow forecasting curves and summer inflow forecasting curves conforming to drought stages were larger than the slopes of the other two seasons' curves. At the Andong Dam as well, there are two observed inflows located in the vicinity of D2 and D3 curves on the summer inflow forecast curves. These points are the observed summer inflows for 2013 and 2015. As for the inflows corresponding to these two points, if droughts had occurred in the spring of 2013 and 2015 and droughts had been forecast for the summers, they could have been adequately forecast.

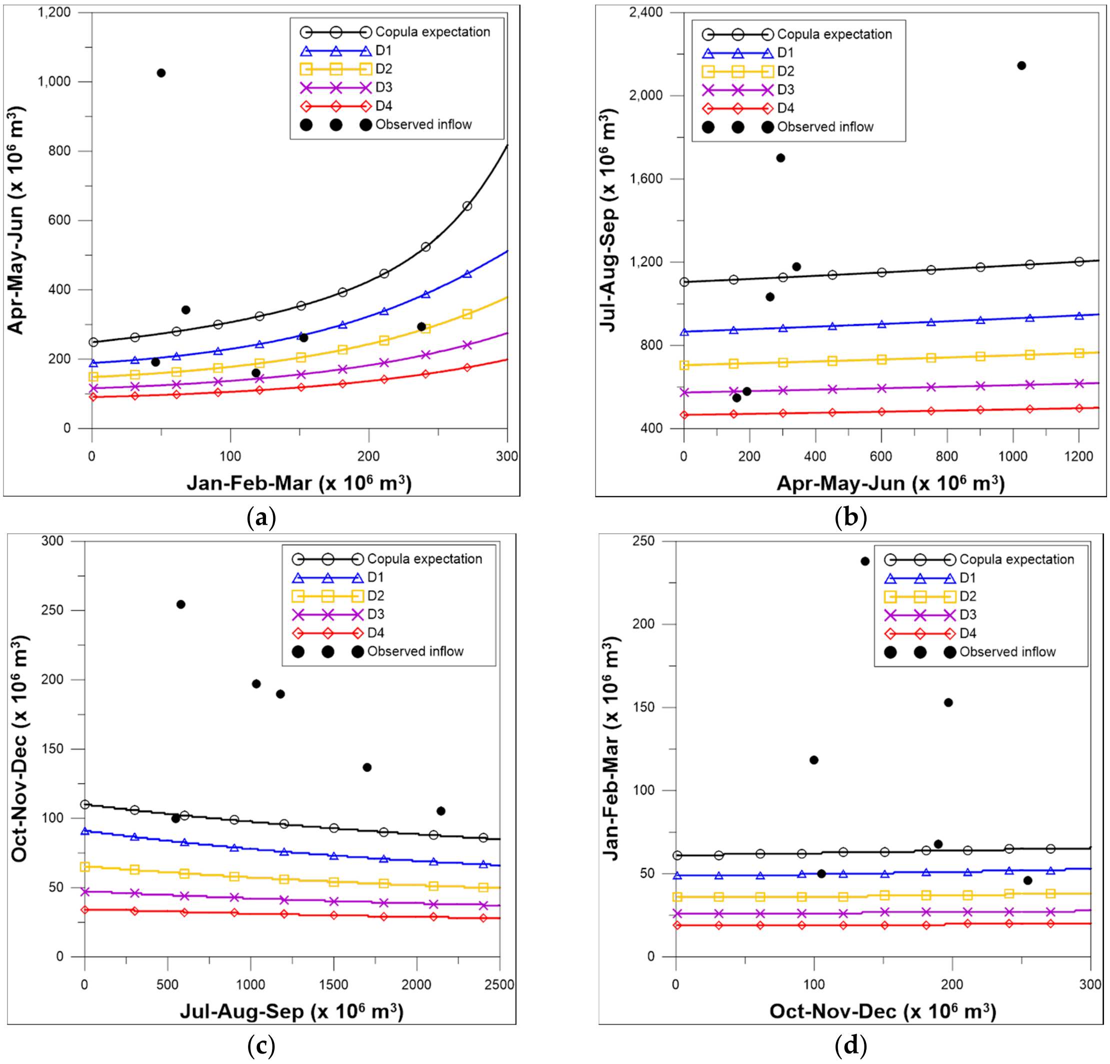


Figure 8. Seasonal inflow forecasting curves conforming to drought stages for Soyanggang Dam. (a) Spring; (b) Summer; (c) Fall; (d) Winter.

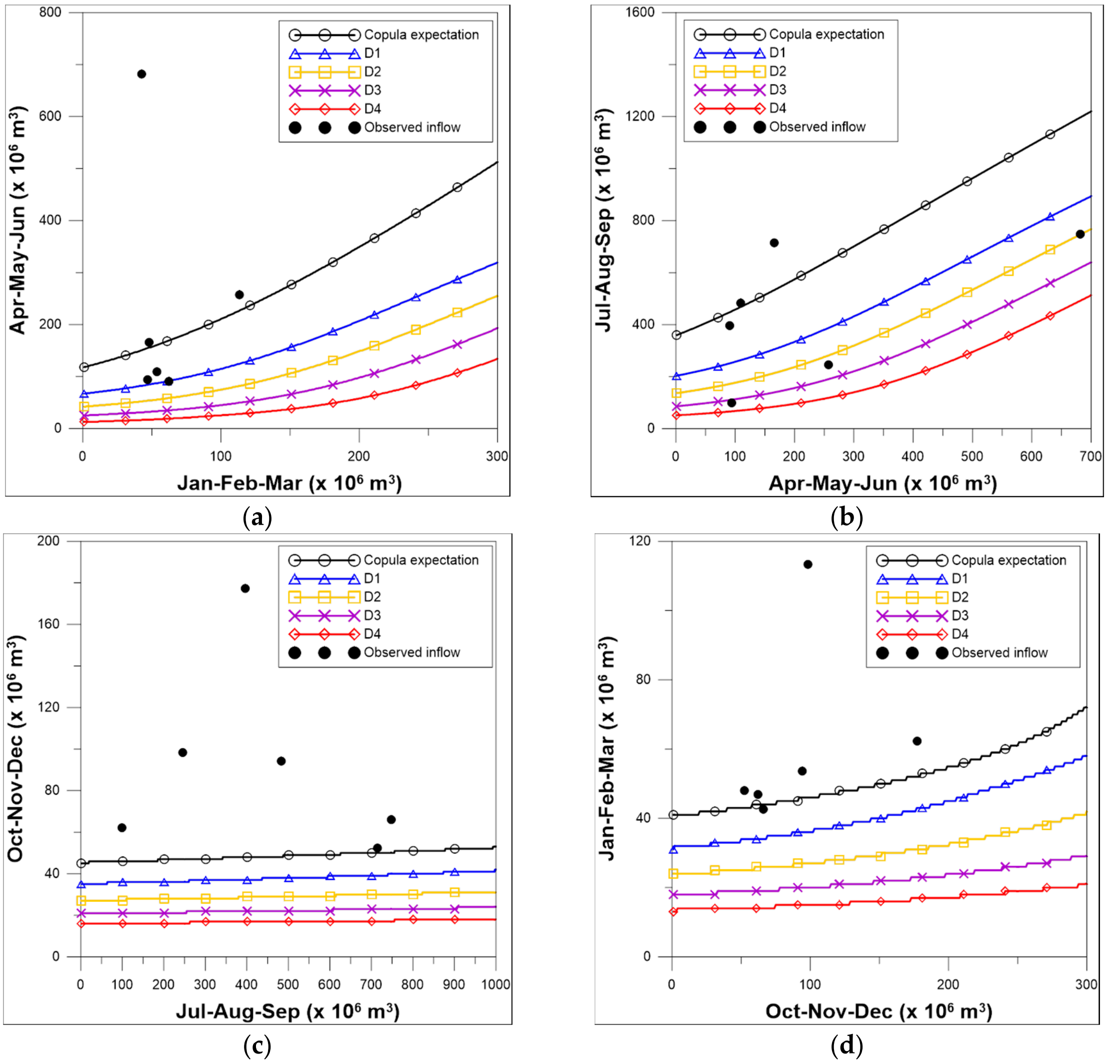


Figure 9. Seasonal inflow forecasting curves conforming to drought stages for Andong Dam. (a) Spring; (b) Summer; (c) Fall; (d) Winter.

4.2 Seasonal Drought Forecast Results

Figure 10 shows the results of calculating the seasonal SSI of the two dams to forecast the two dams' seasonal droughts. The black colored portions in Figure 10 are the areas where the SSI is below zero, which can be distinguished as droughts. Looking at the two dams' seasonal SSI, it can certainly be recognized that severe and moderate droughts occurred from 2013 to 2015.

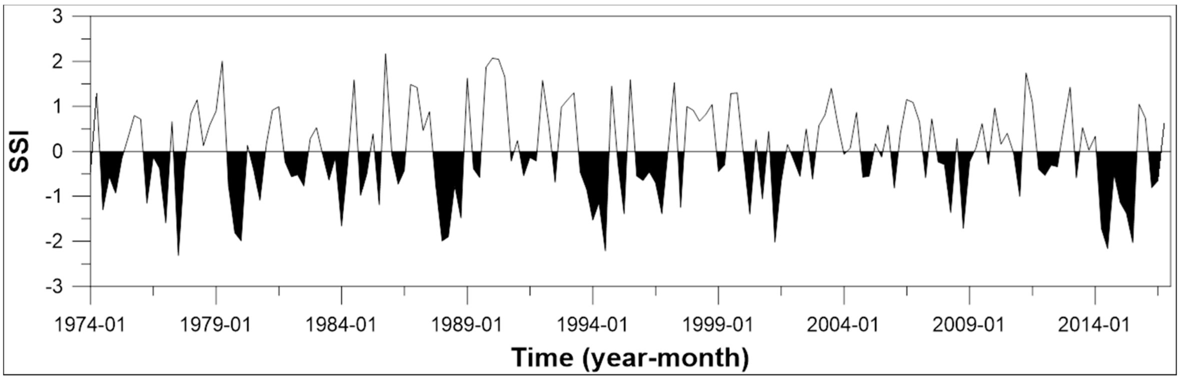
First, we evaluated the drought forecasting accuracy during the training period of the Bayesian network. The results showed that at the Soyanggang Dam, drought forecasting based on the previous period had a 75 % probability of success, and at the Andong Dam, drought forecasting based on the previous period had a 65 % probability of success, meaning the probability of drought forecasting success at both dams was high.

Figure 11 shows the results of calculating the probability of a drought occurring at the two dams by season from 2011 to 2016. There was an over 50 % probability of a drought occurring at the Soyanggang Dam in the summer of 2014 and the summer of 2015, when extreme droughts with SSI below -2 occurred. These results mean that drought forecasts were reliable for the two seasons if the threshold probability of drought occurrence was set at 50 %. However, at the Andong dam, in the summer of 2015 when an extreme drought occurred, the probability of a drought occurring was fairly high, but in the summer of 2013 when a moderate drought with an SSI of -1.2 occurred, the probability of a drought occurring was calculated to be below 50 %. It is believed that these results occurred due to the large correlation coefficient between spring and summer inflows. Specifically, the 2013 spring

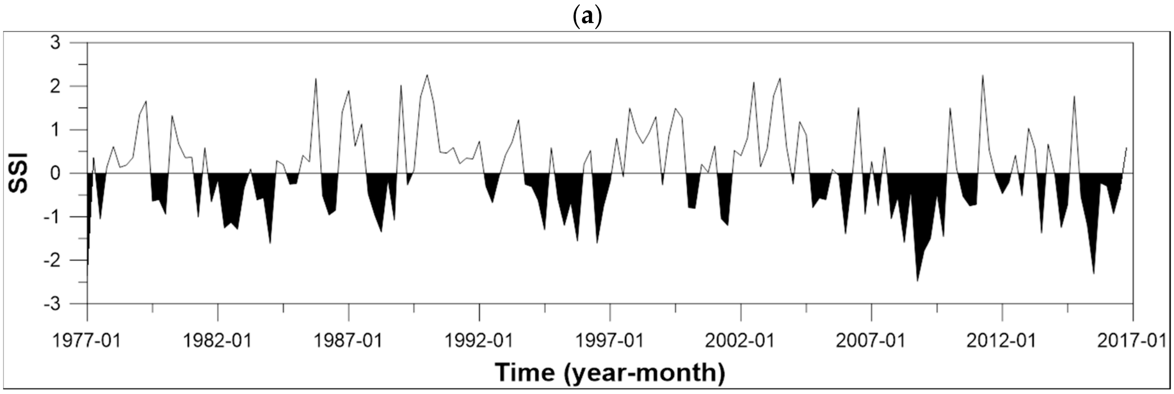
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 January 2018 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233

inflow at the Andong Dam was greater than the average spring inflow. Because of this, the summer inflow, which has a large correlation coefficient with the spring inflow, was predicted to also be large, so the probability of a drought occurring in the summer of 2013 was calculated to be very low. Unlike 2013, the probability of a drought occurring in the summer of 2014 was calculated to be high because the spring inflow was small. This means that the drought forecasting method has limitations in forecasting moderate droughts that randomly occur without relation to the prior season's inflow at dams with a large correlation coefficient for the inflows of successive seasons. On the other hand, it was successful at adequately forecasting extreme droughts that occur after some continued droughts.

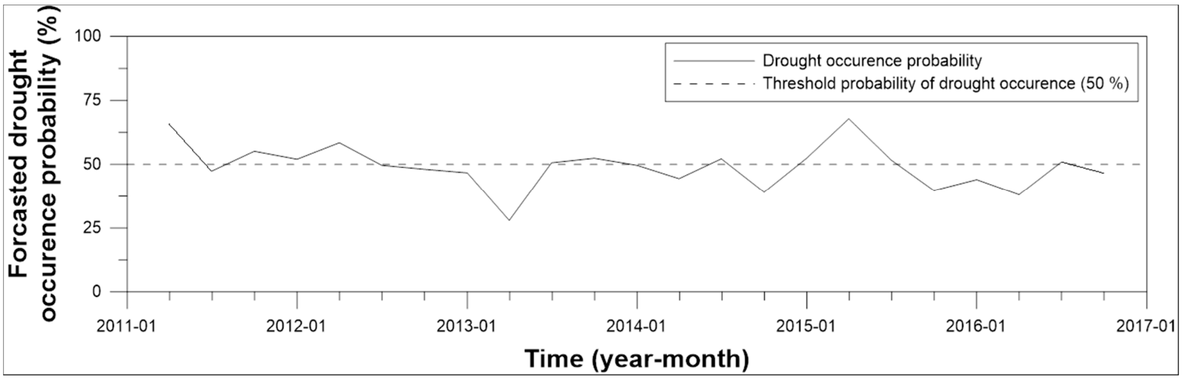


(**a**)

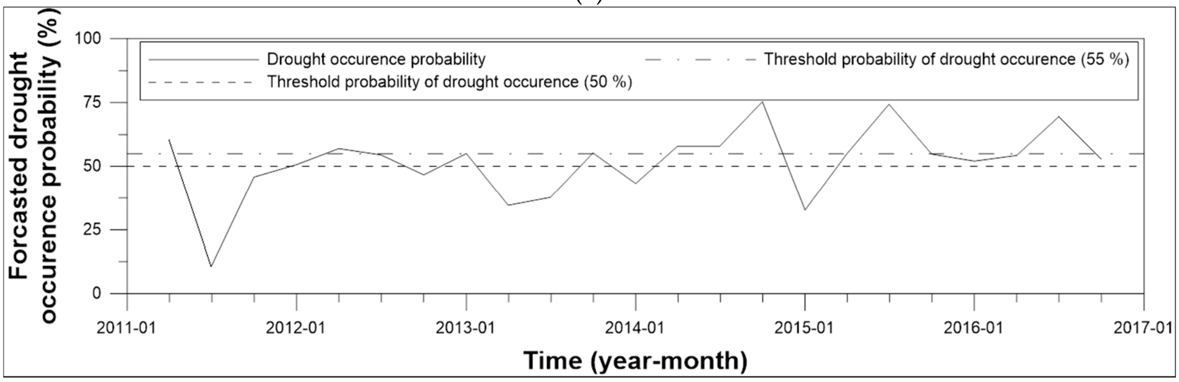


(**b**)

Figure 10. Seasonal SSI. (a) Soyanggang Dam; (b) Andong Dam.



(a)



(b)

Figure 11. The results of seasonal drought forecast. (a) Soyanggang Dam; (b) Andong Dam.

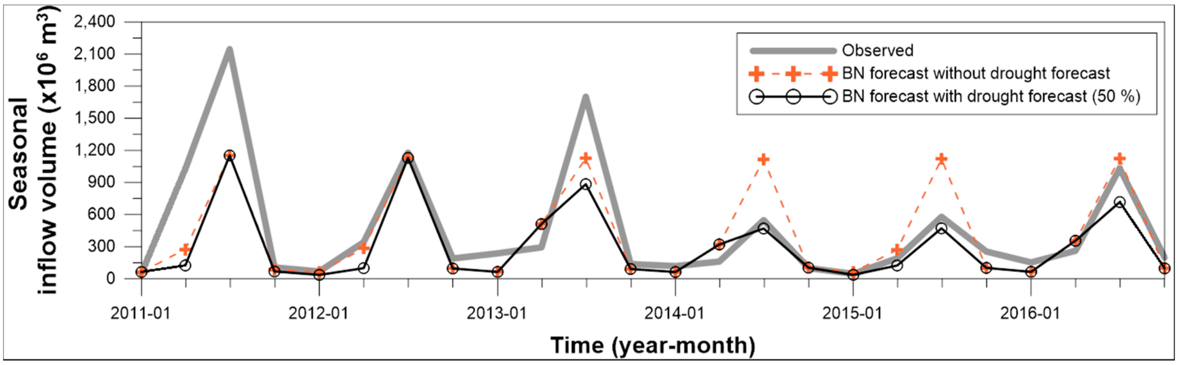
4.3. Seasonal Inflow Forecast Results

The results of seasonal inflow forecasting without or with drought forecasting under a threshold probability of drought occurrence of 50 % are shown in Figure 12 and Tables 4 and 5. In the results without drought forecasting, there was a failure to forecast the inflow for the summer seasons when both dams experienced drought. The results of forecasting inflow at the Soyanggang Dam with the threshold probability of drought occurrence set at 50 % had inflow forecast absolute errors of 14.2 % and 18.6 % for the summers of 2014 and 2015, respectively, during which extreme drought occurred. Those results reasonably match the extreme drought records. The reason for this is that D3 droughts occurred in the spring of both years, so summer droughts were predicted and a D4 inflow forecasting curve was used. In the results of Andong Dam inflow forecasting combined with drought forecasting at a threshold probability of drought occurrence of 50 %, the inflow forecast for summer 2015, which had an extreme drought, was 32.1 %, which reasonably matches the extreme drought record. However, the inflow forecast for summer 2013, which had a moderate drought, failed due to the limitations of drought forecasting previously mentioned. Also, it showed a tendency to forecast little inflows for most seasons. We believe that it showed this tendency because the threshold probability of drought was set at 50 % and the correlation coefficient between spring and summer inflows were high. At the Andong Dam, the correlation coefficient between spring and summer inflows was high so the summer forecast inflow changed linearly according to the spring inflow. This characteristic had an effect on the correlation coefficient of the SSI as well, and even though the spring inflow corresponded to a very weak drought stage, the probability of a drought occurrence in summer increased. That is, in the case of the Andong Dam, setting the threshold probability of drought at 50 % means that frequent drought forecasts can occur. In actuality, during the 40 years for which Andong Dam inflow data exists, there have been 67 seasons where the actual drought index was below zero out of a total of 160 seasons. On the other hand, when the threshold probability of drought was set at 50 %, the number of times a drought was forecast for all seasons was 82, which is 15 more than the actual number of times droughts occurred.

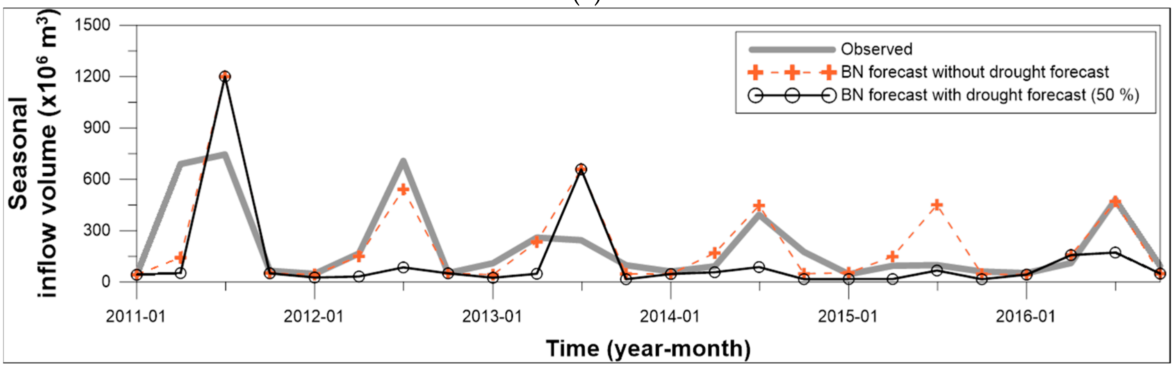
 doi:10.20944/preprints201801.0029.v1

Peer-reviewed version available at *Water* **2018**, *10*, 233; doi:10.3390/w10020233

Figure 13 shows the results when the threshold probability of drought occurrence for the Andong Dam is increased to 55 % to improve on this trend and the inflow is forecast. Looking at Figure 13, this improved the tendency to forecast small inflow during many seasons including the summer of 2012. Also, the absolute error of the inflow forecast for the summer of 2015, when an extreme drought occurred, was further improved to 9.4 %.



(**a**)



(**b**)

Figure 12. The results of seasonal inflow forecasting without or with drought forecast under the threshold probability of drought occurrence to 50 %. (a) Soyanggang Dam; (b) Andong Dam.

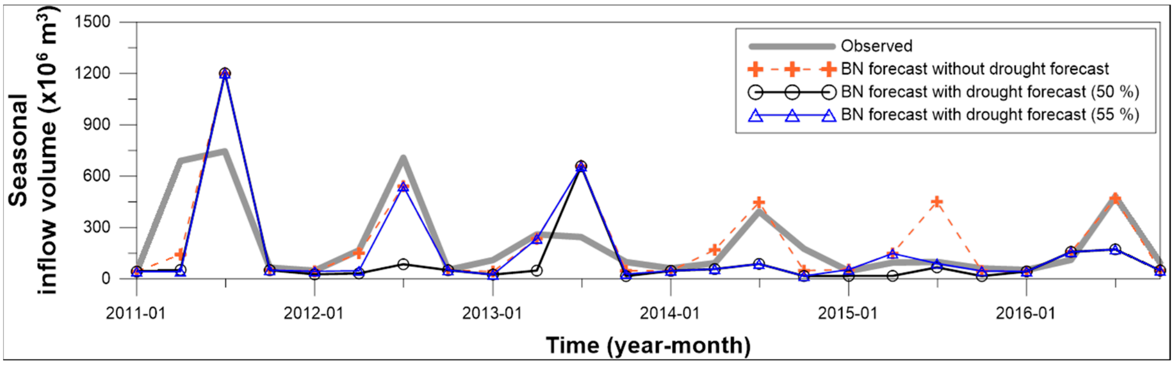


Figure 13. The results of seasonal inflow forecasting without or with drought forecast under the threshold probability of drought occurrence ranging 50 to 55 % for Andong Dam.

Peer-reviewed version available at Water 2018, 10, 233; doi:10.3390/w10020233

Table 4. Seasonal inflow forecasting errors for Soyanggang Dam.

Cases	Absolute error for summer 2014 (%)	Absolute error for summer 2015 (%)	Range of absolute error for summer (%)	Range of absolute error for all seasons (%)
BN forecast without drought forecast	103.6	93.3	4.1 ~ 103.6	3.2 ~ 103.6
BN forecast with drought forecast by 50 % criteria	14.2	18.6	4.1 ~ 48.0	3.2 ~ 100.1

Table 5. Seasonal inflow forecasting errors for Andong Dam.

Cases	Absolute error for summer 2013 (%)	Absolute error for summer 2015 (%)	Range of absolute error for summer (%)	Range of absolute error for all seasons (%)
BN forecast without drought forecast	169.7	355.6	2.2 ~ 355.6	1.2 ~ 355.6
BN forecast with drought forecast by 50 % criteria	169.7	32.1	32.1 ~ 169.7	1.2 ~ 169.7
BN forecast with drought forecast by 50 % criteria	169.7	9.4	9.4 ~ 169.7	1.2 ~ 169.7

5. Conclusion

In this study, we used a copula-based Bayesian network combined with drought forecasting to forecast the seasonal inflows of multipurpose dams. The study's target dams were the Soyanggang Dam and the Andong Dam and 2011 to 2016 seasonal inflow data was used to evaluate the accuracy of the forecast results. The calendar year was divided into four seasons: spring is from April to June; summer is from July to September; fall is from October to December; winter is from January to March. From the results of determining the probability distribution of seasonal inflow, lognormal distribution was found to be the best fit to the seasonal inflow volumes except summer inflow volumes of the Andong Dam. At the Soyanggang Dam, the correlation coefficient of inflows between two successive seasons was very small with the exception of winter and spring. In the case of the Andong Dam, the correlation coefficients of inflows for winter-spring and spring-summer were large, but the correlation coefficients of inflows for the other two seasons were calculated to be small. We used the copula-based Bayesian network to determine the seasonal inflow forecasting curves conforming to drought stages, and then we used the standardized streamflow index (SSI) to forecast seasonal droughts. In the drought forecasts, we calculated the probability distribution of the SSI for a season based on the previous season's SSI, found the non-exceedance probability of the drought index being less than or equal to zero, and compared it to the threshold probability of drought occurrence.

We drew the following conclusions from the study results. The results of seasonal inflow forecasting by a Bayesian network without drought forecasting are very unsuitable as future inflow data for operating a dam. If drought forecasting is not used, then the inflow of the next season is simply the value with the largest probability density function out of the next-season probability inflows that are based on the previous season. Because of this characteristic, the inflow forecast results were very poor when droughts occurred at the two dams.

When drought forecasting is used, there is a need to vary the threshold probability of drought occurrence according to the hydrologic characteristics of the dam. We believe that the reason the proper threshold probability of drought occurrence is different for each dam is that the correlation coefficients of inflows between successive seasons are different for the two dams. According to the results of this study, the threshold probability of drought occurrence must be slightly larger than 50 % for dams with a large correlation coefficient of inflows between successive seasons, and the threshold probability of drought occurrence does not need to be made large for dams with a small correlation coefficient of inflows between successive seasons.

Under the above limitations, reasonable estimates of drought inflow to a reservoir may be forecast by combining the drought forecasting with the probabilistic forecasting of seasonal inflow using a copula-based Bayesian network.

In the future, one can apply the techniques used in this study to various other dams, which might screen proper threshold probabilities of drought occurrence according to the dams' hydrological characteristics. We can also consider methods for finding a threshold probability of drought occurrence that minimizes the absolute error using an optimization technique. Finally, if weather forecasts are reliable, they can be used in drought forecasting.

Acknowledgments: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2017R1A2B2003715). This work was also supported by the Korea Agency of Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 17AWMP-B083066-04).

Author Contributions: Youngkyu Jin and Sangho Lee conceived the study methods; Kwanghoon Kim performed the numerical calculations of the study contents; Kwanghoon Kim and Sangho Lee analyzed the study results and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jeong, D.I.; Kim, Y.O. Forecasting monthly inflow to Chungju dam using ensemble streamflow prediction. *KSCE J. Civ. Eng.* **2002**, *22*, 321-331.
2. Croley, T.E. *Using meteorology probability forecasts in operational hydrology*; ASCE Press: Reston, VA, USA, 2000, ISBN 978-0-7844-0459-1.
3. Croley, T.E. Climate-biased storm-frequency estimation. *J. Hydrol. Eng.* **2001**, *6*, 275-283, doi: 10.1061/(ASCE)1084-0699(2001)6:4(275).
4. Stedinger, J.R.; Kim, Y.O. Updating ensemble probabilities based on climate forecasts. In Proceedings of Conference on Water Resources Planning and Management and Symposium on Managing the Extremes Floods and Drought, Roanoke, VA, USA, May 2002.
5. Jeong, D.I.; Kim, Y.O.; Cho, S.Z.; Shin, H.J. A study on rainfall-runoff models for improving ensemble streamflow prediction – I. Rainfall-runoff models using artificial neural networks. *KSCE J. Civ. Eng.* **2002**, *23*, 521-530.
6. Jee, Y.G.; Kim, S.J.; Kim, P.S. Forecasting monthly inflow for the storage management of small dams. In Proceedings of the Korea Water Resources Association, Iksan, Republic of Korea, May 2005; pp. 85-89.
7. Jin, Y. Reservoir operations applying a discrete hedging rule with ensemble streamflow prediction to cope with droughts. M. thesis, Pukyong National University, Busan, Republic of Korea, Feb 2016.
8. Madadgar, S.; Moradkhani, H. A Bayesian framework for probabilistic seasonal drought forecasting. *J. Hydrometeorol.* **2013**, *14*, 1685-1705. doi:10.1175/JHM-D-13-010.1.

9. Madadgar, S.; Moradkhani, H. Spatio-temporal drought forecasting within Bayesian networks. *J. Hydrol.* **2014**, *512*, 134-146. doi:10.1016/j.jhydrol.2014.02.039.
10. Kim, S.H.; So, J.M.; Kang, S.U.; Bae, D.H. Development and evaluation of dam inflow prediction method based on Bayesian method. *J. Korea Water Resour. Assoc.* **2017**, *50*, 489-502. doi:10.3741/JKWRA.2017.50.7.489.
11. Shin, J.Y.; Ajmal, M.; Yoo, J.Y.; Kim, T.W. A Bayesian network-based probabilistic framework for drought forecasting and outlook. *Adv. Meteorol.* **2016**, *2016*, doi:10.1155/2016/9472605.
12. Russel, S.J.; Peter, N. *Artificial intelligence: A modern approach 3rd ed*: Prentice Hall: Upper Saddle River, NJ, USA, 2009, ISBN 978-0136042594.
13. Yoo, J.Y.; Shin, J.Y.; Kim, D.; Kim, T.W. Drought risk analysis using stochastic rainfall generation model and copula functions. *J. Korea Water Resour. Assoc.* **2013**, *46*, 425-437. doi:10.3741/JKWRA.2013.46.4.425.
14. Zezula, I. On multivariate Gaussian copula. *J. Stat. Plan.* **2009**, *139*, 3942-3946. doi:10.1016/j.jspi.2009.05.039.
15. Mckee, T.B.; Doesken, N.J.; Kleist, J. The relationship of drought frequency and duration to time scales. American Meteorological Society 8th Conference on Applied Climatology, Anaheim, CA, USA, Jan 1993.
16. Son, K.H.; Bae, D.H.; Chung, J.S. Drought analysis and assessment by using land surface model on South Korea. *J. Korea Water Resour. Assoc.* **2011**, *44*, 667-681. doi:10.3741/JKWRA.2011.44.8.667.
17. Massey, F.J. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68-78. doi:10.2307/2280095.
18. Abramowitz, M.; Stegun, I.A. *Handbook of mathematical function with formulas, graphs, and mathematical tables;* U.S. Department of Commerce: Washington D.C., WA, USA, 1964, ISBN 9781591242178.