

## 앙상블 예측을 이용한 충주댐의 월 유입량 예측

### Forecasting Monthly Inflow to Chungju Dam using Ensemble Streamflow Prediction

정대일\* · 김영오\*\*

Jeong, Dae-II · Kim, Young-Oh

#### Abstract

Probabilistic approaches are useful for long-term hydrologic forecasting because input uncertainties are relatively large. Probabilistic forecasting specifies a probability distribution of a forecast variable, and thus makes it possible to consider risks in operating water resources systems. This study suggests a probabilistic forecasting method for the monthly inflow to Chungju Dam using Ensemble Streamflow Prediction (ESP), employing current hydrological states, and historical precipitation scenarios. To improve the forecast accuracy, two weighting schemes, called the Croley, and the PDF-Ratio methods, assigned a unique weighting to each ESP scenario. This study proved that ESP has an ability to forecast the inflow to the Chungju Dam inflow forecasting. The Croley and PDF-Ratio methods, using the probabilistic forecasts of rainfall, performed better than the uniform weighting method, but neither of the first two methods was not superior. This study attempted to disaggregate the uncertainty associated with the ESP forecasts into the hydrologic, and the meteorologic uncertainties. This uncertainty analysis showed the hydrologic uncertainty dominated the monthly inflow forecast during winter, while the meteorologic uncertainty did so during the other seasons.

**Keywords :** ensemble streamflow prediction (ESP), probabilistic forecasting, Chungju dam, weighting

#### 요 지

중장기 유량예측은 입력변수의 불확실성이 크므로 확률론적 예측이 유리하다. 확률론적 예측은 예측변수의 분포를 제시할 수 있으므로 위험도를 고려한 수자원 관리가 가능하다. 본 연구는 현재 유역의 상태와 과거에 재현되었던 강우를 결합하는 확률론적 예측기법인 앙상블 예측(Ensemble Streamflow Prediction)을 이용하여 충주댐의 월 유입량에 대한 확률론적 예측방안을 제시하고 그 효용성을 검증하였다. 또한, 기상 입력변수의 불확실성으로 인해 야기되는 예측오차를 기상 확률예보자료를 이용하여 극복할 수 있는 방법으로 Croley방법과 PDF-Ratio방법을 소개하고 적용성을 검토하였다. 검토결과 앙상블 예측기법의 효용성을 증명하였으며, 기상 확률예보를 이용한 가중값의 부여로 인한 확률예측의 정확성의 향상을 확인하였다. 그러나 Croley방법과 PDF-Ratio방법의 정확성 개선능력은 거의 비슷하였다. 또한, 기초적인 불확실성 정량화를 실시하여 강우량이 적은 겨울철 예측의 경우 수문학적 불확실성이 지배적이며, 상대적으로 강우량이 큰 봄, 여름, 가을의 경우는 기상학적 불확실성이 지배적임을 증명하였다.

**핵심용어 :** 앙상블 예측, 확률론적 예측, 충주댐, 가중값

\*서울대학교 지구환경시스템공학부, 박사과정(E-mail: jung922@snu.ac.kr)

\*\*정회원 · 서울대학교 지구환경시스템공학부, 조교수(E-mail: yokim05@snu.ac.kr)

# 1. 서 론

수문학적 예측은 단기와 중장기 예측으로 구분할 수 있다. 단기 유량예측에는 확정론적 강우-유출모형을 많이 이용하며 주로 홍수조절 목적으로 사용된다. 반면, 중장기 유량예측은 입력변수 등의 불확실성이 크므로 확률론적 모형을 사용하는 것이 바람직하다. 우리나라의 경우 치수 등을 위한 단기 실시간 유량예측은 그 긴박한 필요성으로 인하여 연구와 실무에서 관심의 대상이 되고 있으나, 한 달 이상의 중장기 유량예측에 대한 관심과 연구는 상대적으로 미비한 실정이다. 그러나 월 단위 이상의 장기 용수공급과 저수지 관리를 위해서는 중장기 유량 예측정보가 필수적이므로 이에 대한 활발한 연구가 요구되고 있다.

확률론적 예측정보의 필요성은 Missouri와 Mississippi강 유역에서 1993년 발생한 대홍수(Great Flood)를 계기로 확대되었다. NOAA(National Oceanic and Atmospheric Administration)의 재난조사 연구팀은 1994년 사건조사를 통해, 한정된 값을 포함하는 확정론적 예측의 한계성에 대해 언급하고, 확률론적 예측의 필요성을 강조하였으며(Krzysztofowicz, 2001), 이후 이러한 요구가 현재에 이르기까지 계속되고 있다.

미국 기상청인 NWS(National Weather Service)에서는 확률론적 예측기법의 하나인 Ensemble Streamflow Prediction(ESP; 이하 앙상블 예측)을 수문학적 예측시스템의 현대화 작업의 핵심사항으로 선정하여 그 연구에 박차를 가하고 있다(Fread, 1998). 앙상블 예측은 현재 유역의 상태와 과거에 재현되었던 강우를 결합하는 확률론적 예측기법이다. 또한 강우-유출모형의 입력변수가 되는 기상변수의 예측자료를 이용하여 입력변수가 가지는 불확실성을 감소시킴으로서 정확한 유량을 예측할 수 있으므로 우리나라의 장기 유량예측을 위해 적용할 경우 효율적인 수자원의 계획과 관리에 큰 도움을 줄 수 있을 것이다.

앙상블 예측의 국내적용 사례로 Kim et al.(2001)은 건설교통부 하천계획과와 한국건설기술연구원에서 매 월초 발간하고 있는 '물공급전망'의 개선을 위해 앙상블 예측을 제시하였다. 즉, 물공급전망의 핵심이 되는 5대강 유역의 유량예측 중 금강유역에 앙상블 예측기법을 적용하여 그 효용성을 입증하였고, 이를 통한 중장기 예측방안을 제시하였다.

수문예측은 기상정보의 적절한 이용을 통해 그 불

확실성이 감소할 수 있으며 이러한 노력이 학계에서 부단히 계속되고 있다. 수문예측에 이용되는 대표적인 기상정보로는 NOAA와 Environment Canada(EC)에서 실시하고 있는 기상 확률예보를 들 수 있으며, 이를 수문예측 과정에 포함시키는 연구는 Croley II (1996, 1997, 2001)와, Stedinger and Kim(2002) 등이 있다.

본 연구에서는 앙상블 예측을 이용하여 충주댐 월 유입량 시나리오를 생성하고, 여기에 기상 확률예보를 이용하기 위해 Croley방법과 Stedinger and Kim이 제시한 PDF-Ratio방법을 적용하여 그 실용성을 검증하고자 한다.

# 2. 앙상블 예측

## 2.1 기본개념

앙상블 예측은 강우-유출모형에 미래에 일어날 가능성이 있는 모든 강우 시나리오(즉, 강우 앙상블)를 입력하여 그 결과물로 다수의 유량 시나리오(즉, 유량 앙상블)를 얻는 기법이다(그림 1). 예를 들어, 1951년부터 2000년까지 50년 동안의 10월 강우자료가 존재한다면 이를 강우-유출모형에 입력하여 2001년 10월 유량 예측으로 50개의 유량 시나리오를 생성해 내는 기법이다. 강우-유출모형의 초기조건은 예측하는 시점(즉, 위의 예에서는 10월 초에 따라 변하므로, 같은 강우 시나리오를 사용한다 하여도 초기조건에 따라 유량 시나리오가 다르게 생성된다.

앙상블 예측에서 가장 핵심이 되는 사항은 생성된 유량 시나리오에 기중값을 부과하는 과정이며, 이를 통해 입력자료의 불확실성을 극복할 수 있다. 기중값을 부여하는 가장 간단한 방법으로는 모든 시나리오의 중요도가 동일하다고 가정하는 방법, 즉 예를 들어 50개의 시나리오에 각각 1/50의 확률을 부과하는 방법이다. 조금 발전된 방법을 고안해 보면, 예측대상

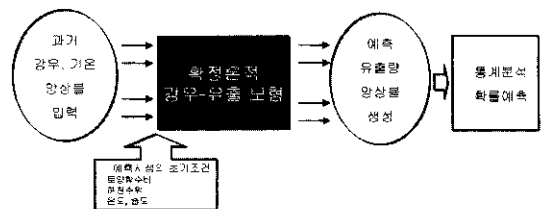


그림 1 앙상블 예측과정

의 수문사상과 비슷한 과거의 수문사상을 가진 연도에 해당하는 시나리오에는 보다 높은 가중값(또는 확률)을 부여하는 방법이다. 수문학적 상황을 판단하는 근거에는 강우의 예측정보, 엘니뇨/라니냐의 ENSO 정보 등이 있을 수 있다.

## 2.2 기상 확률예보를 이용한 가중값

### 2.2.1 기상 확률예보

본 연구에서는 기상 확률예보를 이용하여 유량 양상비에 가중값을 부여하는 연구를 수행하였다. 기상 확률예보 중 대표적인 사례로 NOAA에서 실시하고 있는 예보형식을 소개하면 식 (1a), (1b)와 같다.

$$\Pr[R_{sep,98} \leq r_{sep,0.333}] = 0.153 \quad (1a)$$

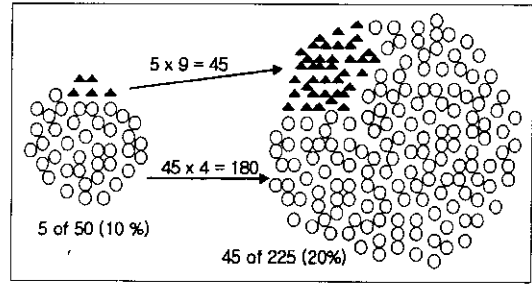
$$\Pr[R_{sep,98} > r_{sep,0.667}] = 0.273 \quad (1b)$$

식 (1a)와 식 (1b)가 1998년 9월의 강우 확률예측의 결과라고 하면,  $R_{sep,98}$ 은 예측월(1998년 9월)의 강우량을 나타내며  $r_{sep,0.333}$ 은 과거(예: 1961~1990)의 9월 강우 중 누가확률 33.3%에 해당하는 강우량 값을 나타낸다. 즉 식 (1a)는 1998년 9월에 내릴 강우는 과거강우의 누가확률 33.3%에 해당하는 강우 보다 작거나 같을 확률이 15.3%이며, 식 (1b)는 누가확률 66.7%에 해당하는 강우보다 클 확률이 27.3%란 뜻이다. 이 경우 98년 9월의 강우가 과거강우의 누가확률 33.3%~66.7% 사이에 위치할 확률은 100%에서 15.3%와 27.3%를 뺀, 57.4%가 된다.

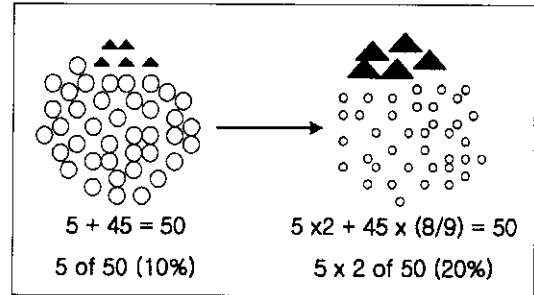
### 2.2.2 Croley방법에 의한 예측 시나리오의 가중값

Croley는 그의 책(2000)과 연속된 연구(1996, 1997, 2001)에서 기상 확률예보를 고려하여 표본 시나리오에 가중값을 부여하는 구체적인 기법을 제안하였으며 이를 정리하면 아래와 같다.

2001년 9월의 강우량( $R_{sep,01}$ )이 7 mm 보다 작을 확률이 20%라는 기상 확률예보가 있다고 하자(즉,  $\Pr(R_{sep,01} < 7 \text{ mm}) = 0.2$ ). 그림 2(a)와 2(b)의 왼쪽이 과거 9월의 강우( $r_{sep}$ )의 시나리오라면,  $r_{sep}$ 이 7 mm 보다 작을(삼각형) 확률은 5/50으로 10%이고, 클(원) 확률은 90%란 것을 알 수 있다. 과거 시나리오를 이용하여 예측월 2001년 9월 강우량 예보( $\Pr(R_{sep,01} < 7 \text{ mm}) = 0.2$ )에 맞는 시나리오를 재구성하려면 그림 2(a)의 오른쪽과 같이 중복추출을 하거나 그림 2(b)의



(a) 중복추출로 구성된 표본시나리오



(b) 가중값을 이용한 표본 시나리오

그림 2 기상 확률예보를 고려한 표본 시나리오의 재구성

방법처럼  $r_{sep} > 7 \text{ mm}$ 인 삼각형에는 2의 가중값을,  $r_{sep} < 7 \text{ mm}$ 인 원에는 8/9의 가중값을 부여한다. 이와 같이 표본 시나리오를 재구성한다는 것은 동일한 확률(1/50)을 가진 시나리오중 일정한 조건(즉, 강우 확률예보)에 해당하는 시나리오에 더 큰 확률을 부여한다는 의미가 된다.

기상 확률예보가 하나 이상 존재하는 경우 위에 설명한 원리에 근거하여 Croley방법으로 가중값을 구하는 방법을 설명하면 아래와 같다.

$n$ 개의 시나리오의 가중값을 구하기 위해  $n$ 개의 기상 확률예보가 존재한다면 단순히 연립방정식을 풀어 가중값을 구할 수 있다. 그러나 대부분의 경우 기상 확률예보자료가 적기 때문에 Croley(2000)은 가중값( $w_i$ )과 1의 차의 제곱을 최소로 하는 목적함수와 기상 확률예보가 제약조건인 최적화문제를 풀어 가중값을 구하는 방법을 제안하였다. 즉 제약조건으로 이용되는 기상 확률예보의 기본형식인 식 (2)와 같은 예보가  $m$ 개 존재한다면 이를 이용해 식 (3)의 최적화 문제를 해결해 가중값을 구할 수 있다.

$$\Pr[X \leq c] = \beta \quad (2)$$

여기서  $c$ 는  $X$ 의 과거 시나리오 중 예측구간이 되는

누가확률(예: 33.3%)값이며,  $\beta$ 는 예측된 발생확률을 의미한다.

$$\min \sum_{i=1}^n (w_i - 1)^2 \quad (3)$$

$$\text{subject to } \sum_i a_{k,i} w_i = e_k \quad k = 1, \dots, m$$

식 (3)에서  $e$ 는 각 기상 확률예보에서의 시나리오 개수( $n$ )와 발생확률( $b$ )의 곱이며,  $a$ 는 식 (4a), (4b)와 같이 기상 확률예보의 조건에 따라 시나리오  $x_i$ 가  $c$ 보다 작거나 같으면 1을 그렇지 않으면 0의 값을 갖는다.

$$a_i = 1 \quad \text{if } x_i \leq c \quad (4a)$$

$$a_i = 0 \quad \text{if } x_i > c \quad (4b)$$

과거 기상자료와 기상 확률예보로 구한 가중값은, 과거 기상자료에 대응하는 예측유량 시나리오의 확률로 적용할 수 있다.

Croley방법의 단점은 가중값을 구하기 위해 다수의 기상 확률예보가 존재해야 한다는 점이다. 또한, 기상변수의 분포를 무시하고 과거자료를 이용해 분할한 예측구간 내에 속하는 시나리오는 모두 같은 확률을 갖는다.

2.2.3 PDF-Ratio방법에 의한 예측 시나리오의 가중값  
Stedinger and Kim(2002)은 Croley방법과는 달리 확률밀도함수를 이용하여 가중값을 구하는 PDF-Ratio 방법을 제안하였으며 이를 정리하면 아래와 같다.

기온, 강우, 유량 등  $n$ 개의 과거 수문기상 시나리오를  $X(= x_1, \dots, x_n)$ 라고 하고, 현재 유역의 상태, 지구 기상시스템 등과 같은 모든 수문학적 정보를  $H$ , 과거의 수문기상 시나리오의 분포를  $D(X)$ 라고 한다면  $X$ 의 수문학적 조건부확률은  $D(X|H)$ 로서 표현할 수 있다. PDF-Ratio방법은  $H$ 를 정확히 알 수는 없으나, 분포를 알 수 있는 예측변수  $X$ 의 조건부 확률예측  $D(X|H)$ 를 이용하여 기상 시나리오의 가중값을 수정하는 방법이다.

Croley방법에서는 식 (1a), (1b)와 같이 확률예보가 3가지 구간으로 주어질 경우 예측변수  $X$ 도 3가지 구간으로 나누어 각 구간의 과거 시나리오에 모두 같은

확률을 부여하였다. 즉 기상 확률예보의 조건을 만족시키기 위해 각 구간에 해당하는 시나리오의 확률을 일률적으로 높이거나 낮추는 불연속 비모수방법이다. 그러나 PDF-Ratio방법은 예측변수의 분포를 고려하여 예측변수에 대하여 연속인 가중값을 구하는 방법이다. 또한 가중값을 구하기 위해 많은 기상 확률예보자료를 요구하는 Croley방법과는 달리 하나의 기상 확률예보만으로도 가중값을 구할 수 있다.

기상변수  $X$ 에 대한 확률밀도함수를  $f_0(x)$ 라하고 수문학적 정보에 따른 조건부 확률밀도함수를  $f_1(x) = f(X|H)$ 라 하면 구하려는 가중값, 즉 시나리오  $x_i$ 에 대응하는 가중값  $w_i$ 는 식 (5)를 이용하여 구할 수 있다.

$$w_i = (f_1(x_i)/f_0(x_i)) \quad (5)$$

### 3. 앙상블 예측기법의 적용

#### 3.1 적용대상 및 연구방법

앙상블 예측의 국내 적용 가능성을 검토하기 위해 충주댐 유역을 연구대상으로 선정하였다. 우리나라와 같이 강우자료가 유량자료보다 일반적으로 풍부한 경우 앙상블 예측을 이용하면 강우자료만큼의 유량자료를 생성할 수 있다는 장점을 살릴 수 있다. 충주댐은 1978년 6월에 착공하여 1985년 10월에 완공되었으며 1986년부터 유입량이 실측되었다. 충주댐 유역의 강우자료는 1965년부터 존재하므로 앙상블 예측을 적용하기에 적당한 대상이라고 할 수 있다.

본 연구에서는 탱크모형을 이용하여 1996~1999년까지 4년간의 예측을 실시하여 예측유입량 앙상블을 생성하고 이를 실측유입량과 비교, 검토하였다. 1966년부터 1995년까지의 30개의 강우자료를 탱크모형에 입력하여 예측년도(1996~1999)의 매월마다 30개의 유입량 앙상블을 생성하였다.

탱크모형을 초기화하기 위한 방법을 설명하면, 탱크모형에 예측하고자 하는 월 이전의 과거 강우자료를 입력하여 예측시점의 토양함수 등의 초기조건이 자동으로 결정되도록 하였다. 예를 들어 1996년 5월의 유입 앙상블을 생성하는 경우는, 1995년 5월부터 1996년 4월까지의 단일한 과거 강우자료를 이용하여 탱크모형을 초기화시킨 후, 1966년부터 1995년까지의 30개의 강우 앙상블을 각각 탱크모형에 입력함으로써 96년 5월의 30개 유입량 앙상블을 생성하였다.

다음으로 Croley방법과 PDF-Ratio방법을 이용하여 각 유입량 시나리오에 해당하는 가중값을 각각 구한 후, 유량 확률예보의 정확성이 개선되었는지 여부를 검토하였다.

### 3.2 강우-유출모형의 검증

본 연구에서 사용한 탱크모형은 장기 일 유출량자료 구축을 위해 한국건설기술연구원(강경석 등, 2001)에서 사용한 모형을 이에 필요한 매개변수와 함께 그대로 사용하였다. 이 모형은 4개의 탱크로 이루어진 4단 탱크모형으로서, 각 탱크의 유출공(5개), 침투공(3개), 유출공 높이(4개), 초기저류고(4개) 등 16개의 매개변수로 구성되어 있다.

강우자료는 한국건설기술연구원에서 장기 일 유출량 자료를 구축을 위해 만든 1966~1998년까지의 유역의 평균 강우자료를 사용하였으며, 1999년의 유역평균강우는 각 소유역에 해당하는 강우관측소의 강우자료를 수자원관리종합정보시스템(<http://wamis.kowaco.or.kr/>)에서 구한 후 산술평균하여 탱크모형의 입력자료로 사용하였다.

탱크모형을 본격적으로 사용하기에 앞서 모형이 유입량을 정확히 모의 할 수 있는가에 대한 검증이 필요하였다. 이 검증 작업을 위해 충주댐 유역의 1996~1999년 실측 강우를 탱크모형을 이용하여 모의한 후 충주댐의 실측유입량과 비교하였다.

그림 3는 예측기간인 4년간(1996~1999년)의 충주댐 실측유입량과 탱크모형을 이용하여 모의한 유입량의 결과를 그림으로 나타낸 것이다. 대부분의 경우 실측 유입량과 탱크모형에 의해 모의된 유입량이 잘 일치하고 있으나, 유입량이 다른 계절에 비해 상대적으로 적은 겨울철의 경우는 두 유입량 간에 상당한 차이가

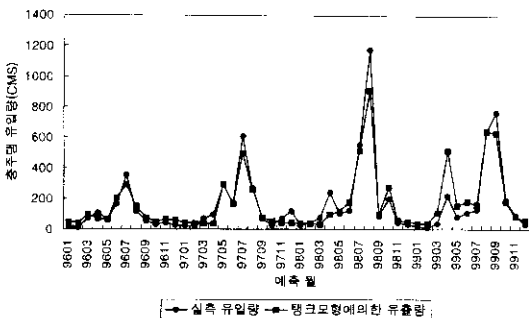


그림 3 탱크모형 유입량과 실측유입량의 비교

표 1 탱크모형 유입량과 실측유입량의 비교

	연간	겨울	봄	여름	가을
BIAS (CMS)	2.65	12.60	15.64	-20.77	3.11
RMSE (CMS)	74.46	31.70	103.31	90.86	47.40
R. B.*	0.02	0.42	0.13	-0.06	0.02

\*R. B.: Relative Bias = Bias/평균

있음을 알 수 있었다.

표 1은 1996~1999년까지의 연간, 겨울(12~2월), 봄(3~5월), 여름(6~8월), 가을(9~11월)에 대한 Bias, RMSE(Root Mean Square Error), Relative Bias를 구한 것이다. 겨울철의 Bias를 평균값으로 나눈 값인 Relative Bias를 보면 연간이나, 다른 계절에 비해 매우 큰 값을 알 수 있다. 결국 본 연구에서 사용할 탱크모형은 봄, 여름, 가을의 경우에는 강우의 변화에 따라 유입량을 잘 모의하고 있으나, 겨울철에는 그렇지 못하다는 결론에 이르게 되었다.

### 3.3 예측유입량 앙상블의 생성

앙상블 예측의 효용성을 검증하기 위하여 탱크모형으로 1996~1999년까지 4년간의 예측유입량 앙상블을 생성하였다. 그림 4는 4월의 유입량 앙상블을 그래프로 도시해본 결과이다. 1996년 4월, 1997년 4월, 1998년 4월, 1999년 4월 모두 1966년~1995년 동안의 4월 강우 앙상블을 이용하여 예측유입량 앙상블을 생성하였으므로, 예측시점의 초기조건이 비슷한 경우에는 유입량 앙상블도 거의 비슷하였으나(예: 그림 4의 1997년 4월(9704)과 1998년 4월(9804)의 유입량 앙상블), 그렇지 않은 경우는 각 예측월의 초기조건에

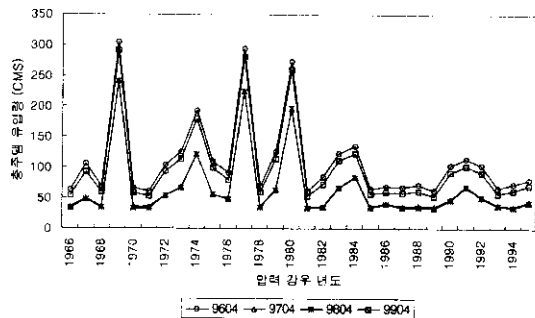


그림 4 탱크모형으로 생성된 예측유입량 시나리오의 예(4월)

따라 서로 다른 유입량 앙상블이 생성되었다.

### 3.4 기상 확률예보의 생성과 가중값의 계산

충주댐 유입량의 확률예측에 기상예보를 이용하기 위해서는 2.2.1절에서 설명한 기상 확률예보자료가 존재해야 한다. 우리나라의 경우 기상청에서 발행하는 ‘월간산업기상정보’에 월별과 순별(상순, 중순, 하순)로 기온과 강우에 대한 예보를 실시하고 있다. 그러나 이는 확률예보가 아닌 과거 30년간의 기상자료의 평균값을 기준으로 ‘평년보다 조금 적겠음’, ‘평년과 비슷하겠음’ 등의 예보형식이며, 조금 적음과 많음의 기준도 명확하지 않아 현재 자료의 형태로 볼 때 예측 유량 시나리오에 적용할 가중값을 구하기 위한 기상 확률예보자료로 정량화하여 이용할 수 없었다. 그러나 기상 확률예보, 특히 강우 확률예보는 유입량 앙상블의 가중값을 구하기 위한 결정적인 자료이며, 본 연구의 목적은 앙상블 예측의 적용성에 대한 검토이므로 충주댐 유역의 실측 강우자료를 이용하여 기상 확률예보자료를 다음과 같이 인위적으로 생성하였다.

충주댐 유역의 강우자료(1966~1999)를 이용하여 예측기간인 1966~1995년까지의 자료를 가지고, 각각 한달, 상순, 중순, 하순에 대한 누가확률 33.3%, 66.7%에 해당하는 강우를 구하여 이를 기준으로 세구간을 설정하였다. 다음으로 1996~1999년까지 4년동안 매월 실측 강우가 위치하는 구간의 확률을 33.3%가 아닌 50%로 강우가 발생한 부분의 확률을 높여주었다. 그리고 나머지 두 구간은 동일하게 25%로 부과하였다. 즉 확률예보가 실측 강우가 일어날 구간을 맞출 확률은 50%이고 그렇지 못할 경우도 50%확률을 갖도록 하였다.

Croley방법에 의한 가중값을 구하기 위해 본 연구에서는 Great Lakes Environmental Research Laboratory(GLERL)에서 개발한 소프트웨어인 Derivative Outlook Weight를 사용하였다. 이 소프트웨어와 메뉴얼은 웹 사이트 <http://www.glerl.noaa.gov/wr/OutlookWeight.html>에서 다운로드가 가능하다. Croley방법으로 가중값을 구하기 위해서는 2.2.2절에 소개한 식 (3)의 최적화 문제를 해결해야 하며, 이를 위해 위 소프트웨어는 라그랑지안 방법을 이용하고 있다. 본 연구에서는 앞서 설명한 과정으로 생성한 한달, 상순, 중순, 하순의 강우 확률예보를 식 (3)의 제약조건으로 사용하였다.

PDF-Ratio방법을 이용하기 위해서는 우선 과거 강우자료에 대한 분포의 검정을 실시해야 한다. 충주댐 월 강우자료의 경우 정규분포나 감마분포로 검정되는 월도 있었으나 일관성을 기하기 위하여 모든 월을 대수정규분포로 가정하여 가중값을 구하였다. 검정된 각 월의 과거 강우자료의 분포  $f_0(x)$ 와 강우 확률예보의 분포  $f_1(x)$  또한 대수정규분포로 동일하고, 다만 평균과 분산만이 변한다고 가정하면, 식 (2)를 이용하여 강우량 시계열의 가중값  $w_i$ 를 구할 수 있다.

표 2(a)는  $f_0(x)$ 의 월별 평균( $\mu_0$ )과 표준편차( $\sigma_0$ )를 표 2(b)는 강우 확률예보에 따른  $f_1(x)$ 의 평균( $\mu_1$ )과 표준편차( $\sigma_1$ )를 구한 결과이다. 대수정규분포로 가정된 확률밀도함수  $f_0(x)$ 와  $f_1(x)$ 를 식 (5)에 대입하여 가중값을 구하였다. Croley방법에서는 한달, 상순, 중순, 하순 등 네가지 강우 확률예보를 사용하였으나 PDF-Ratio방법에서는 한달만의 강우 확률예보를 사용하였다.

그림 5는 Croley방법과 PDF-Ratio방법으로 구한 1996년 4월의 가중값을 비교한 그림이다. 각 가중값의 합은 시나리오 개수와 같은 30이어야 하며 가중값을 30으로 나눈 값이 각 연도에 해당하는 유량시나리오의 확률이다. 두 방법의 가중값이 일치하지는 않으나 크고 작음의 경향이 많이 유사함을 확인할 수 있다.

### 3.5 앙상블 예측의 정확성 평가

#### 3.5.1 확률예보의 정확성 평가방안

예측유량 앙상블을 아무런 처리 없이 예보할 경우 이를 이용하여 수자원 관리를 해야하는 실무자들이나, 일반인들은 예측의 결과를 이해하기 힘들 것이다. 앙상블 예측의 결과로 생성된 예측유량 앙상블을 통계 분석을 통해 확률로 예보할 수 있는 방안을 설명하면 아래와 같다.

유량의 확률예보는 통상 유량을 몇 개의 구간으로 나누어 이에 해당하는 발생확률을 예보하는 과정이다. 즉 발생 가능한 유량을 3개의 구간으로 분할한다면, 이는 저수(low flow), 평수(normal flow), 풍수(high flow)를 대표한다고 할 수 있다. 그러나, 이러한 구간을 나누는 기준이 국내에서는 제시되어 있지 못한 실정이므로, 확률예보는 먼저 이 구간을 과거자료에 근거하여 결정하는 작업에서 시작하여야 한다. 즉, 예측 지점의 과거 유량자료의 적합분포를 선정하고 이 분포를 이용하여 동일한 발생확률을 갖도록 유량의 범위를 분할한다. 만약 예측구간을 그림 6(a)에서와 같

표 2  $f_0(x)$  및  $f_1(x)$ 의 월별 평균과 표준편차

(a) $f_0(x)$ 의 월별 평균( $\mu_0$ )과 표준편차( $\sigma_0$ )												
	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
$\mu_0$	25.27	26.08	42.08	65.23	73.88	122.12	261.52	239.73	126.36	43.31	34.87	19.44
$\sigma_0$	22.11	20.81	30.02	45.92	31.70	74.30	113.80	134.95	89.86	40.24	23.27	11.63

(b) $f_1(x)$ 의 월별 평균( $\mu_1$ )과 표준편차( $\sigma_1$ )													
1996	$\mu_1$	46.97	23.04	70.21	57.18	65.50	173.86	247.62	208.93	129.77	88.91	56.40	38.21
	$\sigma_1$	70.89	19.90	83.35	27.35	40.93	145.86	77.19	150.91	235.75	156.42	62.60	55.06
1997	$\mu_1$	22.60	32.21	37.41	57.43	96.30	111.43	247.62	208.93	112.64	40.07	30.90	17.56
	$\sigma_1$	34.11	92.80	44.41	64.97	60.18	42.06	77.19	150.91	74.72	70.49	14.57	10.00
1998	$\mu_1$	22.60	23.04	37.41	105.63	70.19	173.86	247.62	322.79	112.64	88.91	30.91	18.75
	$\sigma_1$	34.11	19.90	44.41	119.50	20.57	145.86	77.19	233.15	74.72	156.42	34.30	27.03
1999	$\mu_1$	22.60	32.21	70.21	105.63	96.30	173.86	231.96	322.79	291.95	88.91	30.90	18.75
	$\sigma_1$	34.11	92.80	83.35	119.50	60.18	145.86	155.45	233.15	530.37	156.42	14.57	27.03

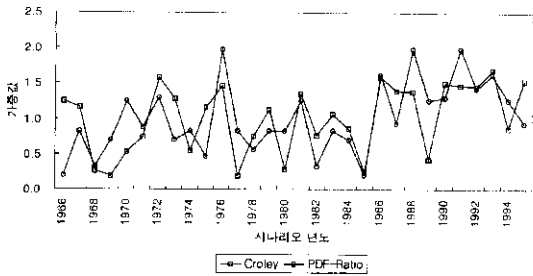
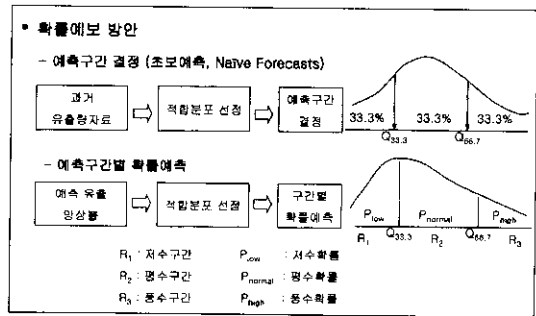


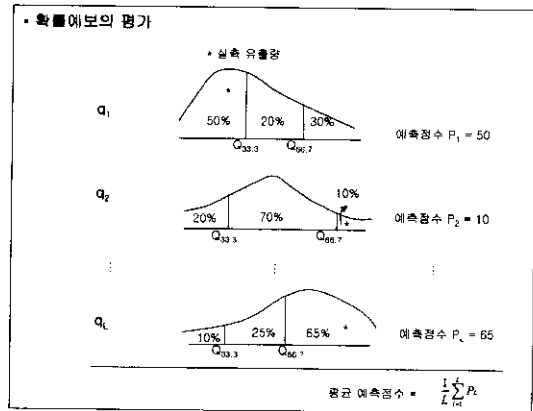
그림 5 Croley방법과 PDF-Ratio방법의 가중값 비교 (1996년 4월)

이 3개 구간으로 나눈다고 가정한다면, 누가확률 33.3%와 66.7%에 해당하는 유량(각각  $Q_{33.3}$ 과  $Q_{66.7}$ )을 기준으로,  $\sim Q_{33.3}$ 을 저수,  $Q_{33.3} \sim Q_{66.7}$ 을 평수,  $Q_{66.7} \sim$ 를 풍수구간으로 나눌 수 있다. 만일 별다른 예측기법이 존재하지 않는다면 미래 유량은 이러한 과거자료에 의존하여 각 구간에 대하여 동일하게 33.3%의 확률로 예보하여야 할 것이며, 이를 “초보예측(Naive Forecasts)”이라 한다(김영오 등, 2001).

만일 앙상블 예측과 같은 예측기법이 존재한다면, 이제 이 예측정보에 근거하여 유량( $q$ ) 확률분포를 갱신하여야 할 것이다. 즉 앙상블 예측의 결과로 얻어진 예측유량 앙상블에 대해 적합분포(이하 앙상블 예측분포)를 선정하고, 과거분포에서 결정한  $Q_{33.3}$ 과  $Q_{66.7}$ 에 해당하는 누가확률을 앙상블 예측분포로부터 읽어 저수, 평수, 풍수의 발생확률(각각  $P_{low}$ ,  $P_{normal}$ ,



(a) 확률예보 방안



(b) 확률예보의 평가

그림 6 확률예보 방안과 평가

$P_{high}$ )을 식 (6)과 같이 계산한다.

$$P_{low} = \Pr[q \leq Q_{33.3}] \quad (6a)$$

표 3 과거자료(1966-1995)를 이용한 유입량 구간의 설정

(단위: CMS)

	3월	4월	5월	6월	7월	8월	9월	10월	11월
Q33.3	43.23	80.11	67.36	61.26	269.10	239.18	163.08	53.54	42.54
Q66.7	76.23	153.89	119.72	118.73	490.35	400.03	285.69	87.13	62.49

표 4 균일가중값 부여방법의 예측점수

(단위: %)

	3월	4월	5월	6월	7월	8월	9월	10월	11월	평균
1996	28.1	40.7	12.1	21.5	39.7	33.0	48.3	73.1	0.0	32.9
1997	14.6	9.3	12.5	44.8	33.7	25.6	69.9	38.1	9.3	28.6
1998	12.7	7.3	50.7	21.2	47.6	37.8	53.9	100.0	0.0	36.8
1999	82.2	12.5	55.1	65.1	29.4	39.1	25.9	90.5	100.0	55.5
평균	34.4	17.4	32.6	38.2	37.6	33.9	49.5	75.4	27.3	38.5*

\*는 평균 예측점수

$$P_{normal} = \Pr[q < Q_{66.7}] - \Pr[q < Q_{33.3}] \quad (6b)$$

$$P_{high} = 1 - \Pr[q < Q_{66.7}] \quad (6c)$$

필요에 따라 구간별 발생확률을 제시함과 더불어 앙상블 예측분포의 중앙값 등 여러 가지 통계값을 실무자의 목적에 따라 이용할 수도 있다.

확률예측의 정확도는 예측점수로 평가할 수 있다. 예측점수는 간단히 정의하자면 실측유량이 발생한 구간의 예측확률을 의미한다. 유량( $q$ )에 대하여  $i$ 번의 확률예보를 실시하였다면, 그림 6(b)에서 알 수 있듯이, 각 예측에서 실측유량이 발생한 구간의 예측점수( $P_i, i = 1, \dots, l$ )를 구할 수 있다.  $q_1$ 의 경우처럼 저수(low flow)의 예측확률값이 다른 구간보다 큰 50%이고 실측유량이 이 구간에서 발생하였다면 확률예측은 어느 정도 정확했다고 말할 수 있다. 그러나  $q_2$ 의 경우처럼 풍수(high flow)가 갖는 확률이 10%로 다른 구간의 확률보다 작음에도 불구하고 실측유량이 풍수구간에 위치했다고 한다면 확률예보가 정확했다고 말할 수는 없을 것이다. 이와 같이 예측을 수행한 후 일정한 검증기간  $n$ 에 대하여 예측점수를 평균한 값을 평균 예측점수라고 하며, 만약 평균 예측점수가 33.3%보다 크다면 전체적인 앙상블 예측은 초보예측보다 정확하여 예측기법으로 유효하다고 할 수 있다.

### 3.5.2 결과분석

먼저 예측점수를 이용하여 앙상블 예측에 대해 평가를 실시하였다. 충주댐의 경우 실측유입량은 1974년 이후부터 존재하므로 자료의 양이 적어 적합한 분포

를 찾기에는 부족하였다. 이를 보충하기 위해 탱크모형과 과거 강우자료(1966~1973)를 이용하여 8년간의 과거유입량 시나리오를 추가하고, 총 30개의 유입량 시나리오로 이용하여 각 월의 적합분포를 선정하고  $Q_{33.3}$ 과  $Q_{66.7}$ 을 구하였다(표 3).

본 연구에 사용한 탱크 모형은 12월~2월 사이의 비교적 강우량이 적은 겨울철의 경우에는 유입량을 잘 모의하지 못하는 경향이 뚜렷하였으므로 예측점수를 구하는 대상 월은 3월~11월로 하였다. 표 4는 가중값을 부여하지 않은 경우, 다시 말해서 각 예측유입량 시나리오에 모두 균일한 가중값(1/30)을 부여한 경우(이하 균일가중값 부여방법의 결과이다. 가중값을 부여하지 않은 경우에도 평균 예측점수가 초보예측의 33.3%보다 큰 38.5%로서 앙상블 예측이 효용성을 가짐을 알 수 있었다.

다음으로 Croley방법과 PDF-Ratio방법에 의한 가중값을 예측유입량 시나리오에 적용하여 예측점수를 계산하였으며 가중값 계산방법에 따른 평균 예측점수는 표 5와 같다.

Croley방법의 평균 예측점수는 46.6%로서 균일가중값 부여방법의 평균 예측점수보다 8.1%가 증가하였으며 PDF-Ratio방법의 평균 예측점수는 45.9%로서 7.4%가 증가하였다. 즉 강우 확률예보를 이용하여 유

표 5 가중값 부여방법별 평균 예측점수(단위 : %)

	균일가중값 부여방법	Croley 방법	PDF-Ratio 방법
평균 예측점수	38.5	46.6	45.9



표 6 방법별 예측분포의 중앙값과 실측유입량 비교

	균일가중값 부여방법	Croley 방법	PDF-Ratio 방법
Bias	-42.92	-47.52	-47.11
RMSE	173.80	162.50	163.11
$\sqrt{\text{Variability}}$	170.20	157.04	157.81

입량 앙상블에 가중값을 부여할 경우 확률예측의 정확성이 증가함을 증명하였다. 또한 Croley방법의 평균 예측점수가 PDF-Ratio방법보다 0.7%가 크게 나타났으나, Croley방법이 우수하다고 할만큼 큰 차이를 보이지는 않았다.

마지막으로 앙상블 예측분포의 중앙값과 실측유입량을 비교하였다. 표 6은 각 예측유입량 앙상블의 중앙값과 실측유입량을 비교하여 그 오차를 Bias, RMSE,  $\sqrt{\text{Variability}}$ 로 나타낸 결과이다. 여기서 Variability는 오차의 분산을 의미한다. 균일가중값 부여방법보다 Croley방법과 PDF-Ratio방법의 모든 오차지표가 작음을 확인하였다. 이는 균일가중값 부여방법보다 가중값을 부여한 경우 중앙값과 실측유입량과의 오차가 작다는 것을 의미하며, 예측의 정확도가 증가하였다는 의미로 해석할 수 있다. 그러나 역시 Croley방법과 PDF-Ratio방법간의 우열을 판가름할 정도의 큰 차이는 보이지 않았다.

### 3.6 불확실성의 정량화

수문예측에서 발생하는 불확실성은 크게 수문모형이 불완전하기 때문에 발생하는 수문학적 불확실성 (Hydrological Uncertainty)과 수문모형의 입력변수로 사용되는 미래의 발생 가능한 강우량과 기온 등의 기상변수에 의해 발생하는 기상학적 불확실성(Meteorological Uncertainty)으로 나눌 수 있다. 그러나 수문학적 불확실성과 기상학적 불확실성은 비선형으로 복합되어 발생하므로, 발생원인에 의한 불확실성의 크기를 정량화 하는 것은 어려운 일이다.

탱크모형이 완벽하다면 각 예측월에 대한 실측유입량( $Q_o$ )과 실측 강우량을 이용하여 탱크모형으로 모의한 유입량( $Q_s$ )은 같아야할 것이다. 그러나 실제 문제에서 이러한 예는 드물다.  $Q_o$ 와  $Q_s$ 의 차이를 본 연구에서는 탱크모형의 불완전에서 발생된 오차( $E_h$ ), 즉 수문학적 불확실성에 의한 오차라고 정의하였다. 각 예측유입량 앙상블( $q_i, i = 1, 2, \dots, n$ )에  $E_h$ 를 더한

후, 실측유입량과 예측유입량 앙상블 사이에서 발생한 오차는 유입량 앙상블을 만들기 위해 실측 강우가 아닌 과거의 강우자료들을 이용했기 때문에 생긴 오차( $E_m$ ), 즉 기상학적 불확실성에 의한 오차라고 할 수 있을 것이다. 식 (7a), (7b), (7c)는  $E_h, Q_{f,med}, E_m$ 의 정의를 식으로 표현한 것이다. 여기서  $Q_{f,med}$ 는 예측유입량 앙상블에서  $E_h$ 만큼을 더한 후 구한 중앙값이다.

$$E_h = Q_o - Q_s \quad (7a)$$

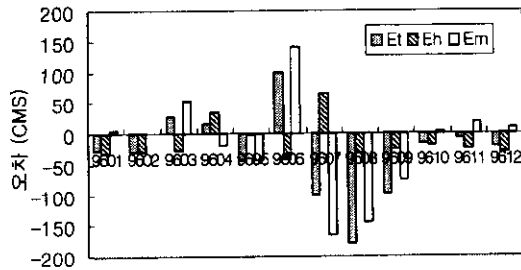
$$Q_{f,med} = (q_i + E_h)_{med} \quad (i = 1, 2, \dots, n) \quad (7b)$$

$$E_m = Q_o - Q_{f,med} \quad (7c)$$

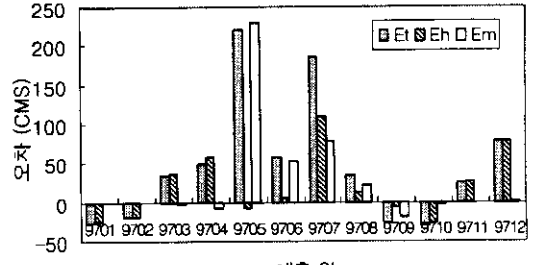
그림 7은 각 예측월의 총 오차 중에서 탱크모형에 의해 발생한 오차( $E_h$ )와 예측유입량 앙상블을 구하기 위해 사용한 과거 강우량에 의해 발생한 오차( $E_m$ )를 각 예측월별로 막대그래프를 이용하여 그린 것이다. 겨울철인 12월, 1월, 2월의 경우는 총 오차 중에서 탱크모형으로 인한 오차가 강우량에 의한 오차보다 훨씬 더 큼을 보여 주고 있으며 이는 앞서 표 1에 나타난 결과와 일치한다. 이 결과는 예측의 정확도를 증가시키기 위해서 탱크모형의 개선이 보다 시급함을 다시 한번 시사한다. 또한, 겨울철을 제외한 대부분의 경우에는  $E_m$ 이  $E_h$ 보다 더 큼을 알 수 있다. 즉 입력변수인 강우량의 불확실성으로 인한 오차가 탱크모형의 오차보다 더 큼을 알 수 있다. 따라서, 이 기간에는 주 입력변수인 강우량으로부터 기인하는 불확실성을 줄이기 위해 노력해야 하며, 강우예보 등의 정보를 이용하여 이러한 입력변수의 불확실성을 줄이는데 주력해야 함을 시사한다.

## 4. 결론 및 향후 연구

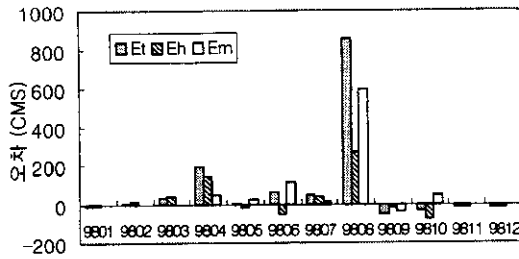
앙상블 예측을 통해 1996~1999년 동안 실시한 충주댐 월 유입량의 확률예보는 가중값을 부여하지 않아도 평균 예측점수가 38.5%로서 초보예측 보다는 우수함이 입증되었다. 또한 강우 확률예보를 이용하여 예측유입량 시나리오에 가중값을 부여할 경우, 평균 예측점수는 Croley방법에서 8.1%, PDF-Ratio방법에서 7.4%의 개선효과를 보였다. 두 방법의 평균 예측점수 차이는 0.7%로 Croley방법의 평균 예측점수가 약간 높았으나 Croley방법이 더 우수하다고 할만큼 큰 차이는 아니었다. 또한 Croley방법은 예측기



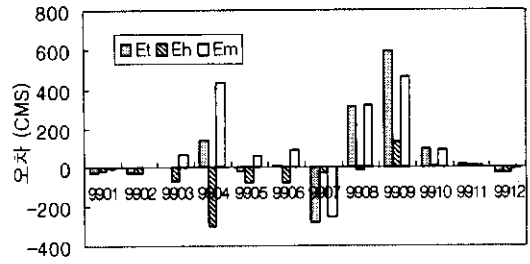
예측 월  
(a) 1996년



예측 월  
(b) 1997년



예측 월  
(c) 1998년



예측 월  
(d) 1999년

그림 7 불확실성 분석결과

간이 한달, 상순, 중순, 하순인 4가지의 강우 확률예보를 이용한 반면 PDF-Ratio방법은 한달 강우 확률예보만을 이용하여도 기중값을 구할 수 있음을 확인하였다.

예측분포의 중앙값과 실측유입량을 Bias 등의 오차지표로 비교한 결과 균일기중값 부여방법의 오차지표가 기중값 부여방법의 오차지표보다 크게 나타나, 기중값을 부여할 경우 예측분포의 중앙값과 실측유입량의 오차가 작아짐을 확인하였다. 두 기중값 부여방법의 오차지표 역시 거의 비슷하여 두 방법의 정확성 개선능력이 거의 비슷함을 확인하였다.

예측분포의 중앙값, 탱크모형의 모의유입량, 그리고 실측유입량을 이용하여 실시한 불확실성에 관한 분석에서, 강우가 적은 겨울철의 경우 탱크모형의 불확실성에서 발생하는 오차가 큰 반면, 상대적으로 강우가 많은 봄, 여름, 가을철의 경우에는 탱크모형에 의한 오차보다는 입력변수의 불확실성에서 발생하는 오차가 매우 크게 나타났다. 따라서, 수문모형의 불확실성에서 발생하는 오차가 큰 겨울철의 예측오차를 감소시키기 위해서는 강우-유출모형의 개선이 필요하며, 기상변수의 불확실성에서 발생하는 오차가 큰 봄, 여름, 가을

의 경우는 정확한 기상 확률예보를 이용하여 예측오차를 감소시켜야 한다는 결론에 이르게 되었다.

본 연구에서 드러난 문제점과 향후연구에 대해 언급하면 첫째, 강우-유출모형의 불완전성과 개선의 필요성을 들 수 있다. 본 연구에서 사용한 탱크모형은 겨울철에는 유량을 상대적으로 정확히 모의하지 못한다는 결점을 나타내었으므로 보다 신뢰성 있는 결론을 도출하기 위해서는 탱크모형의 매개변수의 개선이나 보다 정확한 강우-유출모형으로의 교체 적용이 필요하다.

둘째, 기상 확률예보자료의 확보이다. 본 연구에서 사용한 강우량 확률예보는 인위적으로 생성한 것이므로 실제 충주댐 유입량 확률예보에 적용하기 위해서는 충주댐 유역의 강우 확률예보가 필수적이다. 그러나 현재 우리 나라의 기상청 등을 비롯한 기상 예보 기관에서는 강우량 확률예보를 실시하고 있지 않는 실정이다. 따라서, 강우량 확률예보의 실시와 확보가 기중값을 부여하기 위해 시급히 해결되어야 할 문제이며 현재 우리나라의 기상 예측정보의 실정에 맞는 기중값 부여방안에 관한 연구도 함께 진행되어야 한다.

## 참고문헌

- 강경석(2001) 탱크모형을 이용한 장기 일 유출량 자료 구축. 2001년 학술발표회 논문집(I), 한국수자원학회, pp. 245-250.
- 김영오, 정대일, 김형섭, 이길성(2001) 앙상블 예측을 통한 물공급전망 개선방안. 2001년 학술발표회 논문집(1), 한국수자원학회, pp. 304-308.
- Croley II, T.E. (1996) Using NOAA's new climate outlooks in operational hydrology. *Journal of Hydrologic Engineering*, ASCE, 1(3), pp. 93-102.
- Croley II, T.E. (1997) Mixing probabilistic meteorology outlooks in operational hydrology. *Journal of Hydrologic Engineering*, ASCE, 2(4), pp. 161-168.
- Croley II, T.E. (2000) *Using Meteorology Probability Forecasts In Operational Hydrology*, ASCE Press.
- Croley II, T.E. (2001) Climate-biased storm-frequency estimation. *Journal of Hydrologic Engineering*, ASCE, 6(4), 275-283.
- Fread, D.L. (1998) A perspective on hydrologic prediction trends. *Special Symposium on Hydrology, American Meteorological Society*, Phoenix, Arizona, pp. J1 - J6.
- Kim, Y.O., Jeong, D.I., and Kim, H.S. (2001) Improving Water Supply Outlooks in Korea with ensemble streamflow prediction, *Water International*, 26(4), pp. 563-568.
- Krzysztofowicz, R. (2001) The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, 249(1-4), pp. 2-9.
- Stedinger, J.R. and Kim, Y.-O. (2002) Probabilities for ensemble forecast reflecting climate information. *Journal of Hydrologic Engineering*, submitted.
- (접수일:2002.1.14/심사일:2002.2.25/심사완료일:2002.2.25)