

# 수문기상정보를 이용한 여름 유량의 Ensemble 예측

## Ensemble Forecasting of Summer Seasonal Streamflow Using Hydroclimatic Information

권현한\*, 문영일\*\*

Kwon, Hyun-Han / Moon, Young-Il

### 요 지

우리나라 수자원 관리에서 여름 유량은 이수 및 치수 측면에서 매우 중요한 역할을 한다. 이러한 점에서 여름유량의 예측 가능성을 검토하는 것은 수자원 관리에 유연성을 주는 동시에 상대적으로 위험도를 저감시킬 수 있는 역할을 할 수 있다. 따라서 본 연구의 목적은 여름 계절 유량을 대상으로 기상인자와의 상관성 분석을 통해 유량 예측을 위한 수문기상정보(hydroclimatics)를 전 지구적으로 검토하고 최종적으로 불확실성을 고려할 수 있는 Ensemble예측을 실시하고자 한다. Ensemble예측은 설정 가능한 입력 자료를 통하여 다수의 출력자료를 얻는 방법론으로서 불확실성이 큰 기상 및 수문기상자료 분석에 주로 이용되고 있다. 본 연구에서는 해수면온도(sea surface temperature), 해수면기압(sea level pressure)과 방출장파복사에너지(outgoing longwave radiation)를 주요 기상인자로 고려하였으며 예측모형으로는 Cross Ensemble(out of bagging)방법에 근거한 Support Vector Machine 모형을 이용하였다. 분석결과 주요 기상인자와 50%이상의 상관관계를 보이고 있으며 다소 합리적인 예측 결과를 제시하여 주고 있어 수자원 관리를 위한 보조수단으로 이용이 가능할 것으로 사료된다.

**핵심용어 : Ensemble 예측, 수문기상정보, 여름유량, Support Vector Machine**

### 1. 서 론

우리나라 수자원 관리에서 여름 유량은 이수 및 치수 측면에서 매우 중요한 역할을 한다. 이러한 점에서 여름유량의 예측 가능성을 검토하는 것은 수자원 관리에 유연성을 주는 동시에 상대적으로 위험도를 저감시킬 수 있는 역할을 할 수 있다. 우리나라를 포함한 극동아시아 지역은 여름 Monsoon의 영향을 많이 받으며 기상학적인 관점에서 매우 중요한 위치를 차지하고 있으며 이를 물리적, 수치해석적 방법을 통해 GCM(Global Circulation Model)모형을 구성하여 예측 및 모의를 실시하고 있다. 그러나 GCM 모형은 전 지구적인 관점에서는 어느 정도 신뢰성을 확보한 예측이 가능하나 수문학적에서 이를 활용하기에는 낮은 해상도의 문제점을 가지고 있다. 그럼에도 불구하고 이들 정보를 수문학적 변량들의 예측에 활용하고자 하는 연구는 점점 증가하고 있는 추세에 있다.

이러한 관점에서 본 연구의 목적은 여름 계절 유량을 대상으로 기상인자와의 상관성 분석을 통해 유량 예측을 위한 수문기상정보(hydroclimatics)를 전 지구적으로 검토하고 최종적으로 불확실성을 고려할 수 있는 Ensemble예측을 실시하고자 한다. Ensemble예측은 설정 가능한 입력 자료를 통하여 다수의 출력자료를 얻는 방법론으로서 불확실성이 큰 기상 및 수문기상자료 분석에 주로 이용되고 있다. 본 연구에서는 실제 관측자료와 GCM(Global Circulation Model) 모형을 통해 추정된 강수량을 예측자(predictor)로 고려하였으며 예측을 위해서 모든 기상학적 변량들은 예측 시점이전에 자료를 이용하였다. 예측모형으로

\* 정회원 · Columbia University Researcher · E-mail : hk2273@columbia.edu

\*\* 정회원 · 서울시립대학교 토목공학과 부교수 · E-mail : ymoon@uos.ac.kr

서는 Cross Ensemble(out of bagging)방법에 근거한 Support Vector Machine 모형(권현한과 문영일, 2006)을 이용하였다.

## 2. 본 론

### 2.1 계절 유량과 기상자료

소양강 댐에 월 단위 유량에서 6월부터 8월까지의 유입량을 평균하여 여름 유량을 추출하였다. 현재 범지구적인 기상현상을 연구하기 위한 기상자료로 이용되어지고 있는 자료들은 NOAA(National Oceanic and Atmospheric Administration)를 중심으로 만들어지고 있으며, 이들 자료들은 SST(Sea Surface Temperature), OLR(Outgoing Longwave Radiation), SLP(Sea Level Pressure), PNA (Pacific North American Index), CNP(Central North Pacific Index) NAO (North Atlantic Oscillation), Nino1.2, Nino3.4, Nino3, Nino4, WP (West Pacific Index), QBO (Quasi-Biennial Oscillation), PDO (Pacific Decadal Oscillation), NP (North Pacific Index), AO (Arctic Oscillation) 등이 있으며 이들 중 본 연구에서는 비교적 중요도가 높은 SST, SLP와 OLR을 해석대상으로 하였다. 또한, 유럽기상예측국(European Centre for Medium-Range Weather Forecasts)에 Ensemble 예측모형인 DEMETER(Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction) GCM모형으로부터 한반도에 해당하는 강수량을 추출하였다.

### 2.2 계절 유량과 기상인자와의 상관성분석

소양강의 여름 계절 유량과 3가지 주요 기상인자 즉, SST, OLR, SLP, DEMETER GCM 예측 여름강수량과의 상관성을 평가하였으며 그림 1-4에 나타내었다. 상관관계 평가는 실제유량의 예측가능성을 평가하기 위해서 3개월 이전의 Global 기상자료를 대상으로 하였으며 3가지 인자 모두 소양강 여름유량과 원격상관관계를 확인할 수 있었다. 그림 1은 SST와의 상관관계를 나타내며 태평양상의 경도 150°W-160°W와 위도 20°S-30°S에서 60%이상의 높은 상관성을 나타내고 있다. 그림 2는 OLR과의 의미하며 인도양의 100°E-120°E와 위도 20°S-30°S과 태평양상의 경도 150°W-160°W와 위도 30°S-35°S에서 50%이상의 상관성을 나타내고 있다. 또한 SLP와의 상관관계를 그림 3에 나타내었으며 인도양상의 경도 120°E-140°E와 위도 15°N-20°N과 태평양상의 경도 170°E-170°W와 위도 45°N-60°N에서 50%이상의 상관성을 나타내고 있다. 마지막으로 DEMETER GCM 모형의 예측 여름강수량과의 상관관계를 그림 4에 나타내었으며 비교적 높은 예측결과를 나타내 주고 있다.

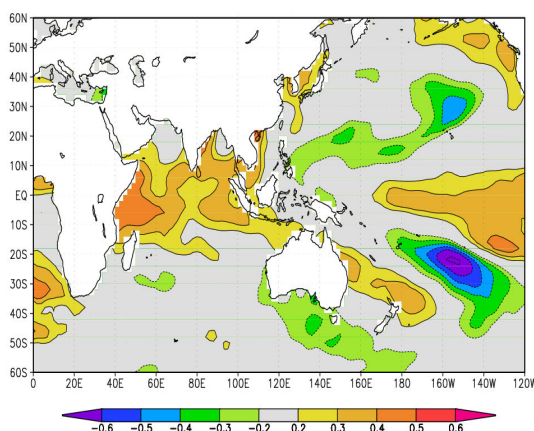


그림 1. 소양강 댐의 여름유입량과 3개월 이전 SST(3월-5월)와의 상관관계

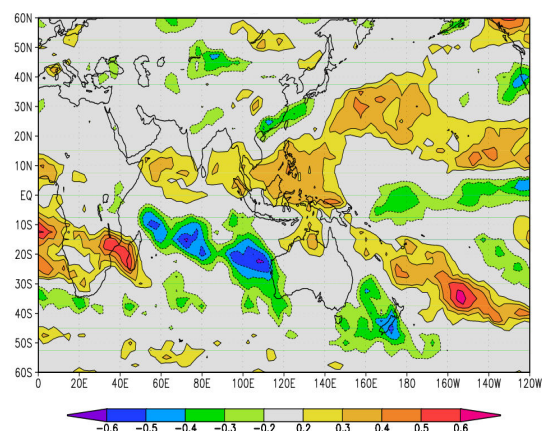


그림 2. 소양강 댐의 여름유입량과 3개월 이전 OLR(3월-5월)과의 상관관계

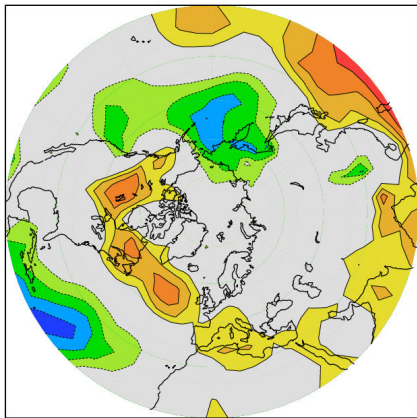


그림 3. 소양강 댐의 여름유입량과 3개월 이전 SLP(3월-5월)와의 상관관계

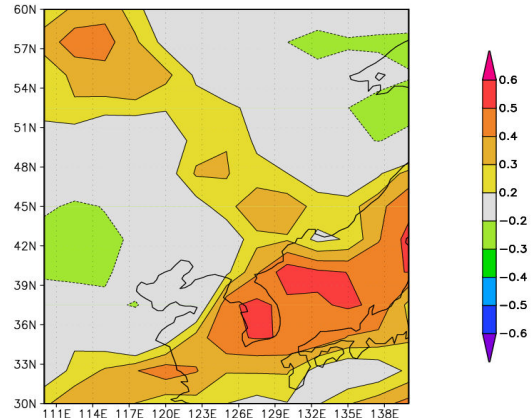


그림 4. 소양강 댐의 여름유입량과 DEMETER GCM 여름 예측강수량(6월-7월)과의 상관관계

표 1. 추출된 기상인자와 소양강댐 유입량과의 상관관계

	OLR Box-1	OLR Box-2	SLP Box-1	SLP Box-2	SST	GCM(Rainfall)
상관계수	-50%	60%	43%	-42%	-61%	56%

### 2.3 Support Vector Machine Ensemble 예측 모형

소양강댐 여름유입량 예측모형을 구성하는데 있어서 본 연구에서는 Vapnik(1995)이 제안한 비선형 학습(learning) 예측 모형인 Support Vector Machine(SVM)을 도입하였다. SVM의 장점은 무엇보다도 훈련을 위한 자료 Vector에 내재한 정보를 인식하는 능력이 뛰어나다는 점과 상대적으로 낮은 공간의 결정 평면 집단을 사용한다는 것이다(권현환과 문영일, 2006). 또한 Ensemble 유량 예측 모형을 구성하기 위해서 Breiman (1996)이 제시한 "Bootstrap ensembles" 또는 "Out of Bagging"이라 불리는 Cross Ensemble 모형을 구성하였다. 최근에 많은 연구에서 예측을 위한 모형을 구성하는데 있어서 단일 모형만을 이용한 예측은 그릇된 예측결과를 나타낼 수 있다고 지적하고 있다(Breiman, 1996).

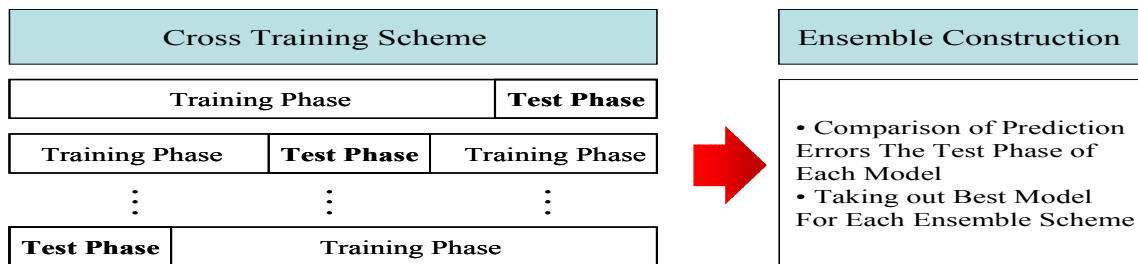


그림 5. Ensemble 모형 구성을 위한 Cross Ensemble 모형

### 2.4 예측결과 분석

분석에 앞서 구성된 모형의 적합성을 평가하기 위해서 다음 표 2와 같은 지표들을 이용하였다. 여기서  $y_t$ 는 관측치를  $\hat{y}_t$ 은 예측치를 나타내며  $E(y)$ 는 평균을  $\theta$ 는 나타낸다. 평균제곱오차(Root Mean Square Error, RMSE)는 모형의 수행결과가 평균적으로 어느 정도 오차가 발생하는지를 나타내는 지표로 일종의 평균치라고 할 수 있다. 모형효율성계수(Coefficient of Efficiency)는 통계적 기준으로서 편의를 줄일 수 있는 무차원계수이다. 예측치와 관측치가 일치할수록 1에 가까워지는 성질이 있다. 이 기준은 무차원양으로서 자료의 개수에 관계없이 절대적 평가기준이 될 수 있다. 평균오차(Mean Absolute Error, MAE)는 예측값과 실제값의 절대 오차를 평균한 값으로 관측치와 예측치의 차이가 적으면 적을 수록 좋은 예측값

을 나타낸다. 마지막으로 모형효율성계수 E는 이상치에 민감한 단점이 있으며 이를 보완할 수 있는 일치 계수(Index of Agreement, IOA)를 이용하였으며 상대적으로 극치값에 민감하지 않은 통계치이다.

표 2. Different efficiency measures used in this study

Statistics	Formula
RMSE	$Q_1 = RMSE = \sqrt{n^{-1} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$
Mean absolute error	$Q_2 = MAE = n^{-1} \sum_{t=1}^n  y_t - \hat{y}_t $
Coefficient of efficiency	$Q_3 = E = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (\hat{y}_t - E(\hat{y}_t))^2}$
Index of agreement	$Q_4 = d = 1 - \frac{\sum_{t=1}^n  y_t - \hat{y}_t }{\sum_{t=1}^n  y_t - E(\hat{y}_t)  +  \hat{y}_t - E(\hat{y}_t) }$

표 3. 소양강 여름 유입량 예측 결과의 통계지표

Statistics	Training	Predict
RMSE	4.64	9.09
MAE	3.66	6.76
Index of agreement	0.90	0.74
Coefficient of efficiency	0.74	0.50
Correlation Coefficient	0.92	0.91

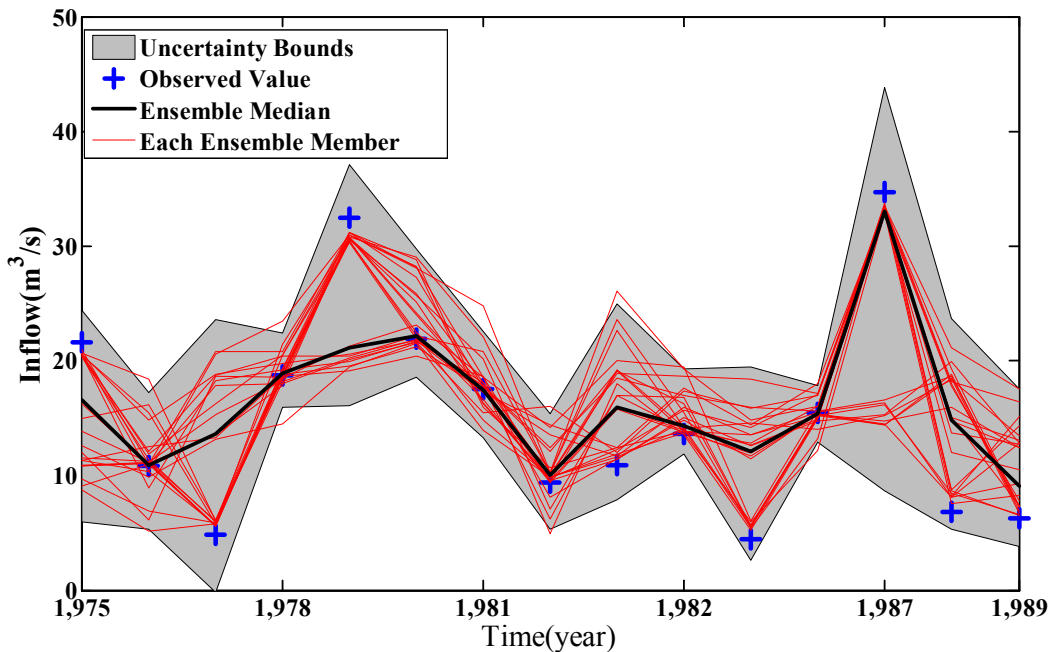


그림 6. Training 구간의 Ensemble 예측 결과

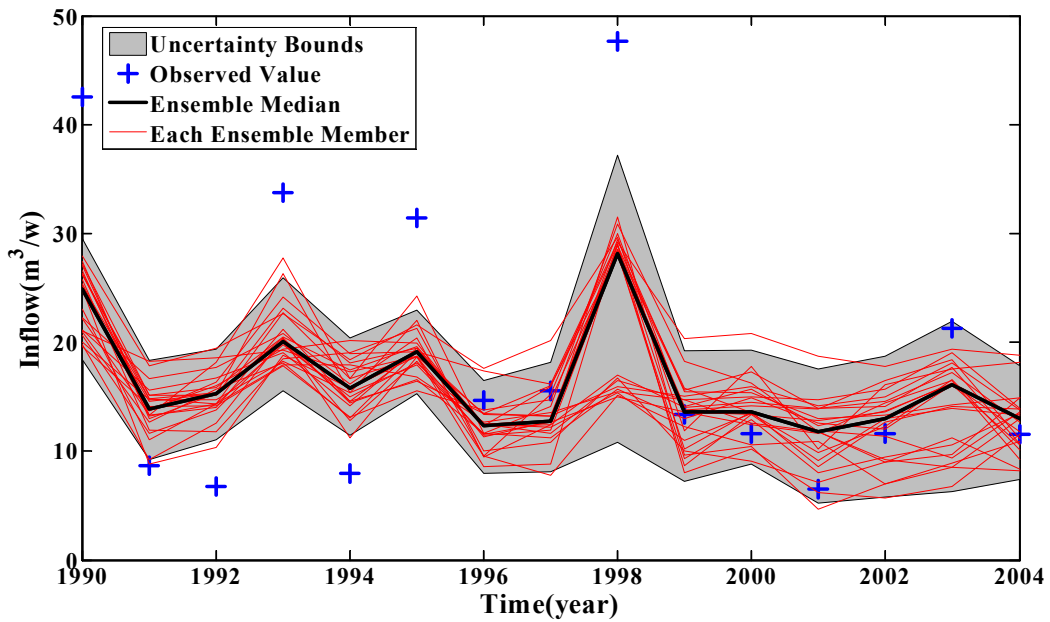


그림 7. Testing 구간의 Ensemble 예측 결과

그림 6과 7은 Training과 Testing 구간의 Ensemble 예측결과와 이에 따른 5%와 95%의 불확실구간을 나타내고 있다. 예측을 실시하는데 있어서 전반부 15년은 Training을 위한 자료로 후반부 15년은 모형을 검증하기 위한 기간을 활용하였으며 Testing 구간에서는 함수의 반복추정 없이 전반부자료를 토대로 추정된 모형을 통해 예측을 실시하였다. 그림 6에서 보듯이 실측자료의 거동을 매우 잘 묘사하고 있으며 불확실성 구간 역시 합리적인 결과를 나타내고 있다. 그림 7은 실제 예측결과를 나타내며 극치사상에 대해서 약간 과소 추정되는 문제점을 보여주고 있으나 표 3에 통계치를 고려해보면 기상인자를 고려한 예측이 합리성을 가지고 있는 것으로 사료된다.

### 3. 결 론

본 연구에서는 우리나라의 대표적인 다목적댐인 소양강댐의 30년간의 계절유량과 Global 격자를 갖는 기상인자 및 GCM 강수량 예측자료와의 상관관계를 평가하고 SVM Ensemble 모형을 통해 예측가능성을 평가하였다. 분석결과 3개월 이전에 SST, SLP와 OLR 및 동일기간의 GCM 강수량 예측결과와 50%이상의 상관관계를 보이고 있으며 비선형 Ensemble 예측 모형의 입력 자료로 활용하였다. 예측결과 Training 및 Testing 예측결과에서 0.9이상의 예측성을 보여주고 있으며 수자원관리에 유연성을 더해줄 수 있는 불확실성 또한 평가가 가능하였다. 본 논문에서는 15년 자료만을 가지고 모형을 구성한 후 15년을 예측한 결과로서 만약, 모형을 계속 갱신(updating)하면서 예측을 실시한다면 보다 정확한 예측 또한 가능하리라 사료된다.

#### 감사의 글

본 연구의 일부는 건설교통부 한국건설교통기술평가원의 이상기후대비시설기준강화 연구단에 의해 수행되는 2005 건설기술기반구축사업(05-기반구축-D03-01)에 의해 지원되었습니다.

#### 참 고 문 헌

- 권현한, 문영일 (2006) Support Vector Machine과 상태공간모형을 이용한 단변량 수문 시계열의 동역학적 비선형 예측모형, 대한토목학회논문집, 대한토목학회, 제 26호, 3B.
- Vapnik, V. (1995) The nature of statistical learning theory. Springer, New York.
- Breiman, L (1996) Bagging predictors., Machine Learning, vol. 24, no. 2, pp. 123-140.