

# Hierarchical Bayesian Model을 이용한 GCMs 의 최적 Multi-Model Ensemble 모형 구축

## Optimal Multi-Model Ensemble Model Development Using Hierarchical Bayesian Model Based

권 현 한\*·민 영 미\*\*·Saji N. Hameed\*\*\*  
Hyun-Han Kwon·Young-Mi Min·Saji N. Hameed

### 요 지

In this study, we address the problem of producing probability forecasts of summer seasonal rainfall, on the basis of Hindcast experiments from a ensemble of GCMs(cwb, gcps, gdaps, metri, msc\_gem, msc\_gm2, msc\_gm3, msc\_sef and ncep). An advanced Hierarchical Bayesian weighting scheme is developed and used to combine nine GCMs seasonal hindcast ensembles. Hindcast period is 23 years from 1981 to 2003. The simplest approach for combining GCM forecasts is to weight each model equally, and this approach is referred to as pooled ensemble. This study proposes a more complex approach which weights the models spatially and seasonally based on past model performance for rainfall. The Bayesian approach to multi-model combination of GCMs determines the relative weights of each GCM with climatology as the prior. The weights are chosen to maximize the likelihood score of the posterior probabilities. The individual GCM ensembles, simple poolings of three and six models, and the optimally combined multimodel ensemble are compared.

**핵심용어 : Multi-Model Ensemble GCMs, Bayesian Weighting, Summer Rainfall,**

## 1. 서론

수자원분야에서 기상자료를 활용한 홍수 및 가뭄 예측은 중요한 연구테마로서 지난 10년 동안 국외를 중심으로 많은 연구가 진행되고 있다. 특히 여름 강수량은 우리나라의 수자원 관리에서 50%이상을 차지한다고 해도 과언이 아니다. 이러한 점에서 여름 계절 강수량과 기상인자와의 상관성을 검토함과 동시에 이들 기상인자를 토대로 유량을 예측하는 연구는 브라질, 미국 등 수력발전이 발전한 나라들에서 실질적인 연구가 진행되고 있다.

지금까지의 연구를 살펴보면 주로 기상인자와 상관성을 토대로 수문사상을 평가한 연구들로서 Lall과 Mann(1995)은 SSA(Singular Spectrum Analysis)와 MTM(Multi-taper method)를 이용하여 Great Salt Lake의 월별 용적변화에 대한 시계열, 월 강우량, 온도 그리고 대기순환의 관계를 해석하였다. Moon과 Lall(1996)은 기후변동성과 수문학적 시스템에 잠재된 동역학적 관계를 이해하기 위하여

\* 정회원·한국건설기술연구원 선임연구원 · E-mail : hkwon@kict.re.kr

\*\* APEC Climate Center· Research Scientist

\*\*\* APEC Climate Center·Senior Research Scientist

미국의 Great Salt Lake와 남방진동지수(Southern Oscillation Index, SOI)를 포함한 여러 대기 순환지수를 이용한 저빈도 관계를 연구한 바 있다. Kwon 등(2009)은 중국 양쯔강의 여름평균유량과 최대유량을 3개월 이전에 예측하기 위해서 해수면온도, Snowpack 등을 예측인자로 하여 모형을 구축한 사례가 있다.

본 연구에서는 1개월 이전에 제공되는 GCMs Ensemble로부터 우리나라에 가장 적합한 Multimodel Ensemble의 대한 확률적 예측모형을 구축하고 각 모형에 대한 적합성을 평가하고자 한다. APEC Climate Center로부터 총 8개의 GCMs(cwb, gcps, gdaps, metri, msc\_gem, msc\_gm2, msc\_gm3 and ncep) Hindcast Experiment를 제공받았으며 이를 활용하여 연구를 진행하였다. 자료는 1981년부터 2003년까지 총 23년의 JJA(June-July-August)계절을 대상으로 하였다.

최적의 Multimodel Ensemble(MME)를 구축하기 위해서 본 연구에서는 Hierarchical Bayesian 모형을 구축하였으며 이를 통해 모형이 가지는 전체적인 불확실성 및 최적의 가중치를 추정하였다.

## 2. 연구 방법

최적의 MME를 구성하기 위한 식은 다음과 같이 간단히 나타낼 수 있다. 여기서  $w_i$ 는 각 GCM 모형에 주어지는 가중치를 나타내며  $f(x_i)$ 는 각 GCM이 가지는 Ensemble로부터 추정되는 확률밀도함수를 나타낸다.  $Y$ 는 실측강수량을 나타낸다.

$$Y = \sum_{i=1}^M w_i f(x_i) \quad (1)$$

여기서  $w_i$ 는 정규분포로 가정하였으며 실측치  $Y$ 와  $f(x_i)$ 는 모두 Lognormal 분포를 따른다고 가정하였다.

$$Y_t \sim LN(\mu_t, \sigma) \quad (2)$$

$$w_i \sim N(\mu_{wi}, \sigma_{wi}) \quad (3)$$

$$f(x_i) \sim LN(\mu_{fi}, \sigma_{fi}) \quad (4)$$

$$\mu_t = w_i \mu_{fi} \quad (5)$$

식 1에서 5까지 모형의 계층적 구조를 간단히 나타내었다. 이외에도 각 매개변수간의 관계등 더 복잡한 형태로 모형이 구성되게 된다. 이러한 계층적 구조를 가지는 특정 Bayesian 모형을 Hierarchical Bayesian 모형이라고 한다. 계층적 모형에 대해서 자세히 살펴보면 다음과 같다.

계층적(hierarchical) Bayesian 모형은 2개의 단계를 가지고 있기 때문에 “Hierarchical” 이라는 이름으로 불리어진다. 우선 상위단계에서, 각각의 모형 매개변수들은 다변량 정규분포와 같은 다변량 확률분포로 나타내어질 수 있다고 가정하며 이러한 확률분포들은 또한 평균과 공분산을 이용하여 분포형의 특성이 결정되어진다. 하위단계에서는 우리가 모의하고자 하는 변량의 확률을 상위 단계에서 매개변수들을 이용한 선형 및 비선형 모형으로 추정하게 된다. 여기서 각 변량은 특정 확률분포를 가지게 되며 Bayesian 방법을 통해 사후분포를 추정하게 되는 것이다. 계층적 Bayesian 모형의 가장 큰 장점은 모형을 통해 각 매개변수들이 갖는 불확실성이 파악된다는 것이다. 이를 통해 각 매개변수들의 유의성을 쉽게 판단할 수 있다(권현한과 문영일, 2007).

본 연구에서, 계층적 Bayesian 모델을 구성하기 위해서 Markov Chain Monte Carlo(MCMC)을 이용하였다. MCMC기법은 주어진 다변량 확률분포가 복잡하여 이를 따르는 iid 난수를 얻을 수 없는 경우에 사용가능한 기법으로서 iid 난수 대신 Markov Chain 난수를 추출하여 사용한다. Markov Chain을 통해 난수를 발생시킨다고 해서 정확하게 관심이 되는 확률분포를 따르지 않지만 이를 일정 시간동안 반복 후에 얻어지는 난수들은 추출을 원하는 분포에 수렴하게 된다. 따라서 MCMC기법은 복잡한 다변량 확률분포 및 매개변수의 추정을 요하는 문제에서 주로 사용되며 또한 Bayesian 통계 기법에서 사후분포의 추론의 이용될 수 있다(권현한과 문영일, 2007, Kwon et. al, 2008; 2009).

### 3. 연구결과 및 결론

그림 1은 CWB GCM 모형으로부터 추정된 10개의 Ensemble Member의 여름강수량과 단순히 산술 평균된 Ensemble Mean 결과를 그림 1에 나타내었다. 그림처럼 각 모형 간에 매우 큰 변동성을 나타내고 있어 여름강수량 예측이 그만큼 쉽지 않음을 알 수 있다.

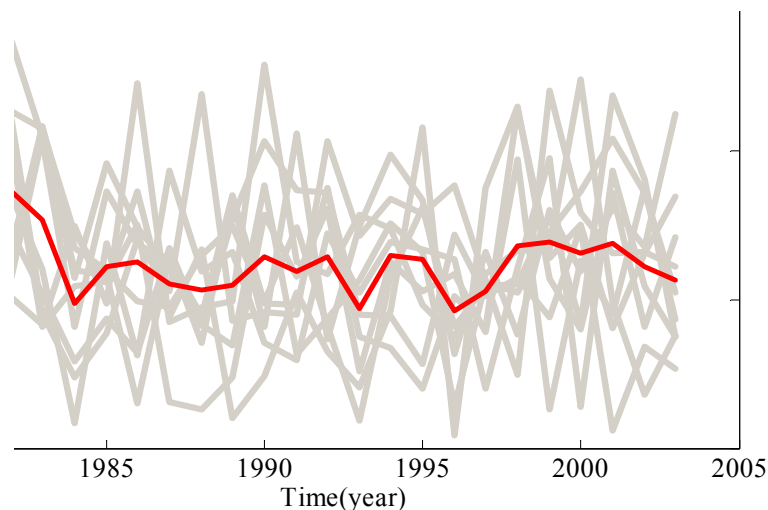


그림 1. CWB GCM 모형으로부터 추정된 여름강수량과 Ensemble Mean 결과

그림 2는 식(1)에서  $w_i$  해당하는 값에 대한 확률분포를 나타낸다. 그림에서 보는 것처럼 Bayesian 방법을 적용함으로써 각 GCM들에 대한 가중치에 대한 확률분포를 추정할 수 있다. 앞서 가정했듯이 가중치  $w_i$ 는 정규분포를 가정하였으므로 사후분포 또한 정규분포로서 추정되어야 한다. 가중치가 확률분포로서 결정되므로 가중치에 대한 적합성을 용이하게 평가할 수 있다. 즉 그림 2에 GCPS GCM의 경우 0을 기준으로 대부분의 밀도가 양의 값을 가지고 있으므로 그 영향 또한 일관성을 가지고 MME 결과치에 영향을 미치고 있다는 것을 의미한다. 반대로 NCEP GCM의 경우 0을 중심으로 거의 대칭형태를 나타내고 있다. 이는 MME 결과치에 거의 영향을 미치지 못하고 통계적으로 Noise로서 작용함을 의미한다. 이들 가중치에 대한 보다 자세한 값은 표 1에 나타내었다. 표에서 Bold체로 나타낸 값은 타 GCM에 비해 영향정도가 큼을 의미한다.

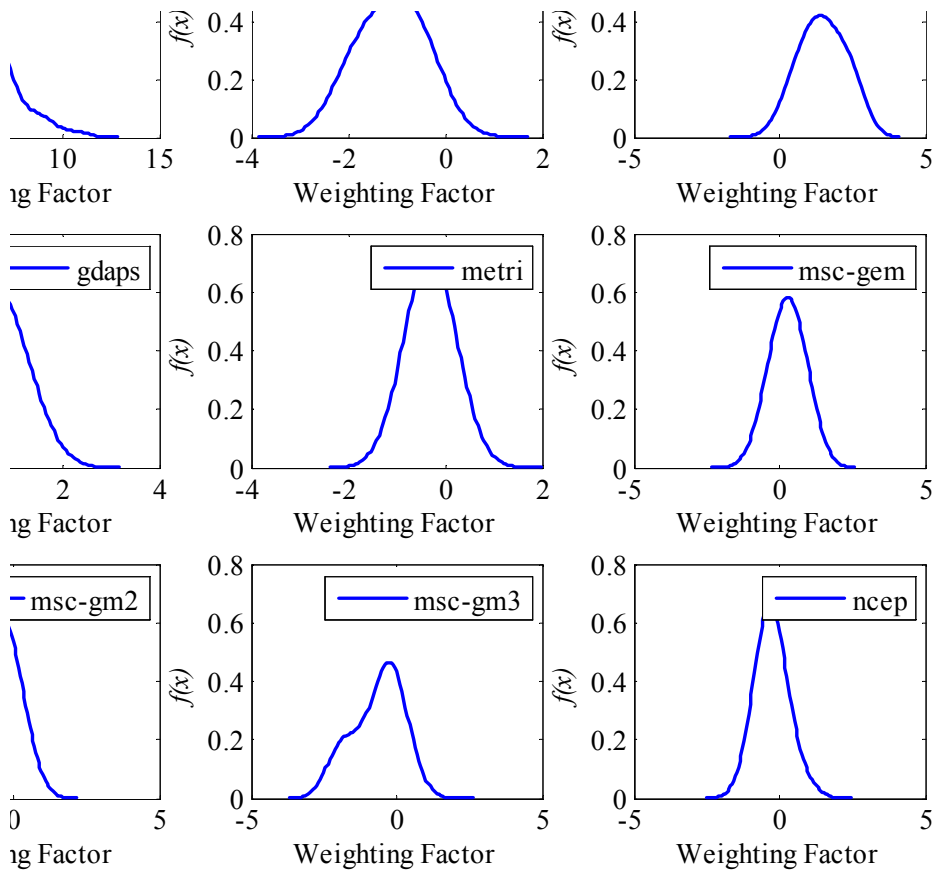


그림 2. MME 추정에서 각 GCM의 가중치와 확률분포

표 1. MME 추정에서 각 GCM의 가중치와 확률분포

GCM s	Mean	St. Dev.	2.50%	Median	97.50%
0 ffset	6.809	1.353	4.676	6.72	10.17
<b>cwb</b>	<b>-1.115</b>	<b>0.577</b>	<b>-2.094</b>	<b>-1.037</b>	<b>-0.1037</b>
<b>gps</b>	<b>1.492</b>	<b>0.7216</b>	<b>0.1338</b>	<b>1.43</b>	<b>2.596</b>
<b>gdaps</b>	<b>0.6718</b>	<b>0.393</b>	<b>-0.03142</b>	<b>0.6109</b>	<b>1.47</b>
<b>metri</b>	<b>-0.3071</b>	<b>0.2159</b>	<b>-0.6736</b>	<b>-0.3111</b>	<b>0.1134</b>
m_sc_gem	0.2737	0.444	-0.5772	0.2834	1.009
m_sc_gm_2	-0.3195	0.4011	-0.9956	-0.4169	0.4684
<b>m_sc_gm_3</b>	<b>-0.6603</b>	<b>0.7732</b>	<b>-2.029</b>	<b>-0.3997</b>	<b>0.4789</b>
ncep	-0.2692	0.3712	-0.9076	-0.3299	0.7549

그림 3은 Bayesian 모형을 통해 Fitting한 결과를 나타낸다. 여기서 결과는 Fitting과정을 통해 각 GCM이 가지는 영향정도를 평가하기 위함이지 실측값과의 상관성 비교는 의미가 없다. 즉 그림 3에서 보이는 결과는 과대적합(overfitting)된 결과이다. 그림에서 보듯이 상당한 불확실성이 존재함을 알 수 있다. 3개의 자료를 제거한 후 Blind Forecasting을 실시한 결과를 보면 2001년에는 비교적 정확한 예측이 수행되고 있으나 2002년의 경우 Fitting의 경우처럼 200mm 가까운 차이를 보이고 있다. 그러나 불확실성 구간 안에 실측값이 포함되어 있다. 마지막으로 2003년의 경우 Fitting의 경우보다 더욱 왜곡된 결과를 보여주고 있으며 불확실성 구간 밖에서 예측이 이루어지고 있다.

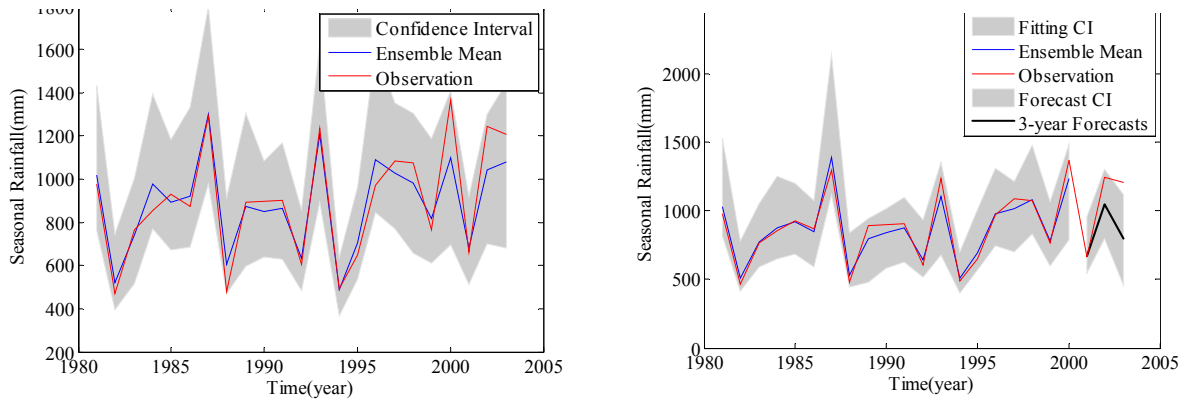


그림 3. Hierarchical Bayesian 모델을 통한 Fitting 결과 및 3-Year Blind Forecasts

본 연구에서는 기존 MME방법을 개선하기 위해서 Hierarchical Bayesian방법을 도입하여 23년간의 여름강수량에 대해서 적용하였다. 적용결과 각 GCM이 가지는 예측력에 정도를 확률밀도함수로 정량화 할 수 있었다. 또한 이들 결과를 활용하여 MME를 구축하였으며 전체적인 불확실성 정량화뿐만 아니라 비교적 만족한 개선된 결과를 확인할 수 있었다. 그러나 보다 정확한 지역적 예측을 위해서는 SST와 같은 기상상태변량이 예측인자로 포함되어야 할 것으로 판단된다.

### 참고문헌

1. Kwon, H.-H., U Lall, (2008), Simulation of Daily Rainfall Scenarios with Interannual and Multidecadal Climate Cycles for South Florida, *Stochastic Environmental Research and Risk Assessment*, DOI 10.1007/s00477-008-0270-2.
2. Kwon, H.-H., C. Brown, K.-Q Xu and U. Lall (2009), Seasonal and Annual Maximum Streamflow Forecasting using Climate Information: Application to the Three Gorges Dam in the Yangtze River Basin, *Hydrological Sciences Journal*, in press.
3. 권현한, 문영일, 김병식, 윤석영, Bayesian Markov Chain Monte Carlo 기법을 통한 NWS-PC 강우-유출 모형 매개변수의 최적화 및 불확실성 분석, 대한토목학회 논문집, 대한토목학회, 제 28호, 4B호, 2008.
4. 권현한, 문영일, 기상정보 및 태풍특성을 고려한 계절 강수량의 확률론적 모형 구축, 대한토목학회 논문집, 대한토목학회, 제 27호, 1B호, 2007.



2009년 한국수자원학회

# 학술발표회

녹색의 땅, 생명의 물

- 기간 : 2009년 5월 21일(목)~22일(금)
- 장소 : 용평리조트
- 주최 : 한국수자원학회
- 후원 : 한국과학기술단체총연합회, 한국학술진흥재단



모시는글 ▶

목차 ▶

Student Competition ▶

구두발표 ▶

포스터발표 ▶

검색 ▶

도움말 ▶

나가기 ▶