

다중 기후모델 앙상블에서 성능과 독립성 모수를 결정하는 새로운 통계적 방법

홍주영¹, 신용관¹, 신이레¹, 박정수¹

전남대학교 수학/통계학과¹

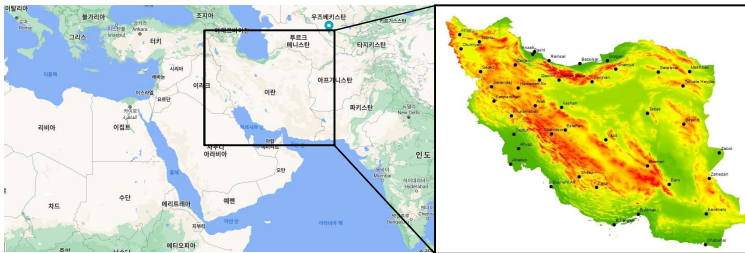
Email : hjy_stat@naver.com

서론

연구자들은 대기대순환모델과 같은 기후모델자료들을 사용해서 미래의 기후변화를 전망한다. 이때, 미래에 대한 예측 불확실성이 매우 크기 때문에 개별 기후모델들을 사용하여 다중모델 앙상블을 구성하고, 구성된 앙상블 모델을 이용하여 미래를 예측하는 방법으로 예측결과와 불확실성을 줄이며 신뢰도를 높일 수 있다. Knutti et al.(2017)은 다른 모델과 구별되는 특성을 가진 독립적인 기후모델들이 점차 줄어들 것이라고 설명하며, 다중모델 앙상블을 구성할 때 성능과 독립성을 모두 고려하는 PI(Performance & Independence)-가중법이 필요하다고 주장했다. PI-가중법을 적용하기 위해서는 각 모델에 주어지는 가중치의 정도를 조절하는 2가지의 모양모수를 결정해야 한다. 이것은 단일 잔류 교차검증(leave-one-out crossvalidation)으로 구할 수 있지만, 이는 너무 많은 계산비용을 필요로 한다. 따라서 엔트로피와 카이제곱 통계량의 유의확률을 이용한 **2가지 모양모수를 결정하는 더욱 간단한 방법을 제안한다.**

분석 자료

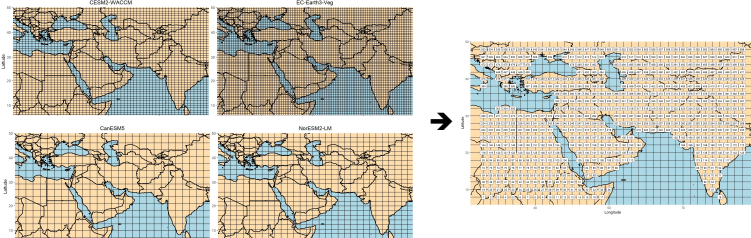
- 변수 : **연간 일일 최대 강수량 (AMP1)**, 연간 5일 최대 강수량 (AMP5), 연간 총 강수량 (ATP), 연간 최대 연속 강수량일수 (AMCWD), 연간 최대 연속 무강수량일수 (AMCDD)
- 분석 지역 : 중동 (이란) (위도 : 8° ~ 48°, 경도 : 16° ~ 86°)
- 관측 자료 : 이란 42개 지점 (1971 ~ 2014, Prof. Khadijeh Javan)



- 기후 시나리오 모델 : CMIP6 24종 (ESGF)
- 시나리오 : Historical (1850 ~ 2014), SSP2-4.5, SSP3-7.0, SSP5-8.5 (2015 ~ 2100)

No.	Model	Institution	Resolution (Lon x Lat Level)
1	ACCESS-CM2	CSIRO, ARCCSS (Australian Res Council Centre of Excellence for Clim System Sci)	192x144 L85
2	ACCESS-ESM1-5	Commonwealth Scientific & Industrial Res Organ (CSIRO)	192x145 L38
3	BCC-CSM2-MR	Beijing Clim Center	320x160 L46(T106)
4	CanESM5	Canadian Centre Clim Model & Analysis, Enviro & Clim Change (CCCma)	128x64 L49(T63)
5	CESM2-WACCM	Nat Center for Atmos Res, Clim & Global Dynamics Lab (NCAR)	288x192 L70
6	CESM2	Nat Center for Atmos Res, Clim & Global Dynamics Lab (NCAR)	288x192 L32
7	CMCC-CM2-SR5	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici(CMCC)	288x192 L30
8	CMCC-ESM2	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici(CMCC)	288x192 L30
9	EC-Earth3-Veg-LR	EC-Earth consortium, Swedish Météo & Hydro Inst/SMHI, Sweden	320x160 L62(TL159)
...
24	TaiESM1	Research Center for Environmental Changes, Academia Sinica, Nankang (AS-RCEC)	288x192 L30

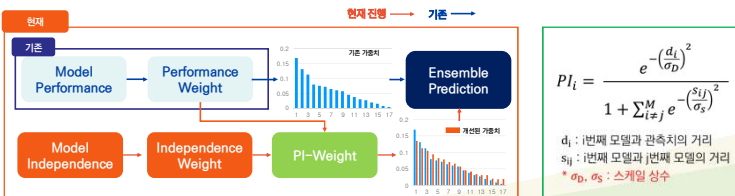
- 24종 모델의 해상도 통일 (2° x 2°) 및 격자 추출 (568개 격자점)



분석 방법 및 결과

Performance-Independence (PI) Weight, Knutti (2017)

- 모델 성능 뿐만 아니라 모델 유사성을 고려하여 **좀 더 신뢰 할 수 있는 앙상블** 예측을 수행



σ_D 와 σ_S 를 결정하는 방법 - Perfect model test, Ruth Lorenz (2018)

하나의 모델이 사실인 것으로 가정하고, 다른 모든 모델을 활용해서 그것을 예측 (일종의 Leave-One-Out Cross Validation)

- 한 모델을 사실로 가정할 때마다, 41x41의 σ 의 조합으로 다중 모델 앙상블을 구성함
- 사실로 가정된 모델의 10~90 백분위수의 사이에 다중 모델 앙상블의 값 80% 이상이 포함되게 하는 σ_D, σ_S 중 가장 작은 값을 선택 (aggressive weighting을 방지) (일종의 Grid Search)

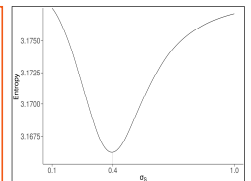
σ_D 와 σ_S 를 결정하는 방법 - 새롭게 제안하는 방법

- σ_S 의 선택

$$I_i(\sigma_S) = \frac{s_i(\sigma_S)}{\sum_{j=1}^M s_j(\sigma_S)} \quad s_i(\sigma_S) = \frac{1}{1 + \sum_{j \neq i} \exp\left(-\frac{S_{ij}}{\sigma_S}\right)}$$

$$E(\sigma_S) = -\sum_{i=1}^M I_i(\sigma_S) \log I_i(\sigma_S)$$

→ l-weight의 Entropy정의 (불확실성의 척도, σ_S 의 함수)



- σ_S 를 0에서 1까지 0.01씩 커지도록 변화하며 엔트로피를 계산
- 엔트로피(불확실성)를 최소로 하는 σ_S 를 선택

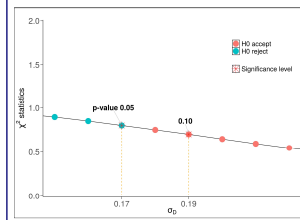
σ_S 가 0에 가까워지면, 모든 l-weight는 등가중치가 되므로 **최소의 불확실성으로 l-weight를 다르게 줄 수 있는 σ_S 를 선택**

- σ_D 의 선택

$$\chi^2_0(\sigma_D) = \sum_{i=1}^M \frac{1/M - P_i(\sigma_D)^2}{1/M}$$

H_0 : 모든 가중치는 동일하다 ↔ $P_i = \frac{1}{M}$ for all i
 H_1 : 어떤 가중치는 동일하지 않다 ↔ $P_i \neq \frac{1}{M}$ for some i

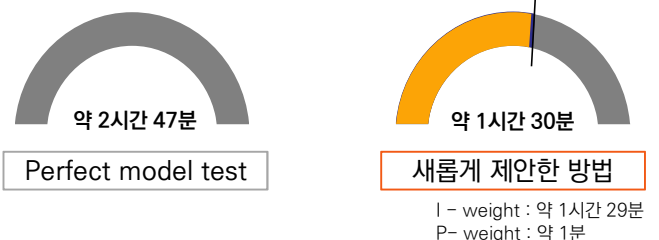
등가중치가 아닌 가중치를 선택하고 싶으므로,
귀무가설을 기각하는 σ_D 를 선택해야함
 → $\sigma_D^* = \max\{\sigma_D : p\text{-value}(\sigma_D) < \alpha\}$
 $p\text{-value}(\sigma_D) = \Pr[\chi^2 > \chi^2_0(\sigma_D) | H_0]$



p-value는 Dirichlet 분포로부터 난수를 생성하여 **Monte-Carlo Simulation** 방법으로 계산함

- Dirichlet 분포로부터 random weights $P_i^{(k)}$ 를 발생, $i = 1, \dots, M$ (under H_0)
- $\chi^2_{(k)} = \sum_{i=1}^M \frac{1/M - P_i^{(k)}(\sigma_D)^2}{1/M}$ 을 계산
- 위의 2개 단계를 K번 (ex) 1000번) 반복
- $p\text{-value}(\sigma_D) = \sum_{k=1}^K \frac{1[\chi^2_{(k)} > \chi^2_0(\sigma_D)]}{K}$ 를 계산, $1[\cdot]$ 는 지시함수(A의 조건에 따라 0 또는 1값을 가짐)

- 소요 시간 비교



* 테스트 환경
 OS : Debian GNU/Linux 10 (buster)
 CPU : Intel(R) Xeon(R) Platinum 8260 CPU @ 2.30GHz
 사용 언어 : R (v.3.6.3)

참고 문헌

- Lukas Brunner et al (2019) Quantifying uncertainty in European climate projections using combined performance independence weighting, Environ. Res. Lett. 14
- Ruth Lorenz et al (2018) Prospects and caveats of weighting climate models for summer maximum temperature projections over North America. Journal of Geophysical Research: Atmospheres, 123, 4509–4526
- Knutti, R. et al(2017), A climate model projection weighting scheme accounting for performance and interdependence, Geophys. Res. Lett., 44, 1909–1918
- Sheldon M. Ross (1997), Simulation Second Edition, 190–193
- Shin et al (2021), Future Projections and Uncertainty Assessment of Precipitation Extremes in the Korean Peninsula from the CMIP6 Ensemble with a Statistical Framework, Atmosphere, 12, 97