

## 현 기후예측시스템에서의 기온과 강수 계절 확률 예측 신뢰도 평가

현유경<sup>1)\*</sup> · 박진경<sup>2)</sup> · 이조한<sup>1)</sup> · 임소민<sup>1)</sup> · 허솔잎<sup>1)</sup> · 함현준<sup>3)</sup> ·  
이상민<sup>1)</sup> · 지희숙<sup>1)</sup> · 김윤재<sup>1)</sup>

<sup>1)</sup>국립기상과학원 현업운영개발부 기후모델개발팀, <sup>2)</sup>서울대학교 지구환경과학부,  
<sup>3)</sup>기상청 기후과학국 수문기상팀

(접수일: 2020년 3월 13일, 수정일: 2020년 5월 20일, 게재확정일: 2020년 6월 8일)

### Reliability Assessment of Temperature and Precipitation Seasonal Probability in Current Climate Prediction Systems

Yu-Kyung Hyun<sup>1)\*</sup>, Jinkyung Park<sup>2)</sup>, Johan Lee<sup>1)</sup>, Somin Lim<sup>1)</sup>, Sol-Ip Heo<sup>1)</sup>, Hyunjun Ham<sup>3)</sup>,  
Sang-Min Lee<sup>1)</sup>, Hee-Sook Ji<sup>1)</sup>, and Yoonjae Kim<sup>1)</sup>

<sup>1)</sup>Climate Model Development Team, Operational Systems Development Department,  
National Institute of Meteorological Sciences, Jeju, Korea

<sup>2)</sup>School of Earth and Environmental Sciences, Seoul National University, Seoul, Korea

<sup>3)</sup>Hydrometeorological and Meteorological Drought Team, Climate Science Bureau,  
Korea Meteorological Administration, Seoul, Korea

(Manuscript received 13 March 2020; revised 20 May 2020; accepted 8 June 2020)

**Abstract** Seasonal forecast is growing in demand, as it provides valuable information for decision making and potential to reduce impact on weather events. This study examines how operational climate prediction systems can be reliable, producing the probability forecast in seasonal scale. A reliability diagram was used, which is a tool for the reliability by comparing probabilities with the corresponding observed frequency. It is proposed for a method grading scales of 1-5 based on the reliability diagram to quantify the reliability. Probabilities are derived from ensemble members using hindcast data. The analysis is focused on skill for 2 m temperature and precipitation from climate prediction systems in KMA, UKMO, and ECMWF, NCEP and JMA. Five categorizations are found depending on variables, seasons and regions. The probability forecast for 2 m temperature can be relied on while that for precipitation is reliable only in few regions. The probabilistic skill in KMA and UKMO is comparable with ECMWF, and the reliabilities tend to increase as the ensemble size and hindcast period increasing.

**Key words:** Climate prediction systems, seasonal forecast, probabilistic skill, reliability diagram

## 1. 서 론

최근 전 세계적으로 발생하는 기상 이변과 홍수, 가뭄, 태풍, 지구온난화 등의 기상재해는 예측하기 어렵

고, 갈수록 피해 규모가 증가하고 있다. 이러한 사회 경제적 수요에 대응하고 가치있는 정보를 제공하기 위해 신뢰할 수 있는 장기예보 자료의 생산은 더욱 중요해지고 있다. 기상청의 장기예보는 이러한 기상 현상을 예측하여 사전에 대비 할 수 있게 하는 한편, 다양한 분야의 의사결정 과정에서 중요한 역할을 담당하고 있다.

한국(KMA)과 미국(NOAA)이 운영 중인 WMO의 장기예보 다중모델 앙상블 선도센터(Lead Center for

\*Corresponding Author: Yu-Kyung Hyun, Climate Model Development Team, Operational Systems Development Department, National Institute of Meteorological Sciences, 33, Seohobuk-ro, Seogwipo-si, Jeju-do 63568, Korea.

Phone: +82-64-780-6573, Fax: +82-64-738-6514  
E-mail: ykhyun@korea.kr

Long-Range Forecast Multi-Model Ensemble, LC-LRFMME)는 전 세계 장기예보 센터들, GPC (Global Producing Center)로 부터 예측 자료를 수집하여 예측을 위한 표준화 된 자료를 제공하고, 다중 모델의 앙상블 예측기술 개발 등의 공식 활동을 이어오고 있다. 또한 최근의 장기예보에 대한 세계적인 수요를 반영하여, 2013년부터 WMO는 계절예측 정보를 계절내(2주~2개월) 시간 규모로까지 확장하여 단기와 장기의 예측 공백을 메우기 위한 계절내~계절예측(Subseasonal-to-Seasonal, S2S) 프로젝트를 조직하고 운영하고 있다. 한국 기상청은 한영 기상청 간 과학협약을 맺고 기후예측시스템, GloSea5 (Global Seasonal forecast system version 5) (MacLachlan et al., 2015)를 공동 운영 중이며, GPC로서 장기예보 선도센터와 S2S 프로젝트의 데이터 베이스에 예측 결과를 제공하고 있다.

GloSea5의 예측 성능과 예측 자료의 활용성을 평가하기 위한 여러 연구가 수행되어 왔다. GloSea5의 주별 예측자료를 이용하여 500 hPa 지위고도와 기온, 강수의 예측성이 확인된 바 있다(Ham et al., 2017, 2019; Heo et al., 2019). 가을 및 겨울철의 ENSO 지수, 가을철 인도양 쌍극자 지수(Indian Ocean Dipole, IOD)에 대한 6개월 이상의 예측성이 검증되었으며(Jung et al., 2015; Lim et al., 2018), Lee and Kwon (2015)는 인도양과 북서태평양 지역을 예측인자로 하여 동아시아 여름철 강수 편차의 예측 성능을 진단하였다. 그리고 계절내 규모에서의 해빙, 지위고도, 성층권 돌연승온, 표층 해류에 대한 예측성 평가가 이루어졌다(Jung et al., 2016; Kim et al., 2018b; Kim et al., 2018c; Lee et al., 2018; Park et al., 2018; Song et al., 2018). 또한 Kim et al. (2018a)는 계절내예측에서의 주요한 변동 모드인 MJO와 GloSea5에서의 원격상관에 대한 진단을 수행하였고, Lee et al. (2016)과 Park et al. (2017)에서는 현업 단기예보에 활용하는 UM 대기모델과 GloSea5의 예측 성능을 비교하여 대기와 해양의 상호작용에 의해 발생하는 MJO와 태풍의 예측 성능에 있어서 결합모델인 GloSea5가 더 우수함을 확인하였다. 이외에도 가뭄 전망을 위해 가뭄지수의 예측성 평가(Son et al., 2015)나, 용담댐 유역을 대상으로 GloSea5의 시, 공간적인 재현성을 검증한 연구(Moon et al., 2016) 등이 있었다.

이와 같이 다양한 시간 규모에서 예측 선행 시간에 따른 모델의 오차와 예측 성능이 결정론적으로 검증되었으나, 확률론적 관점에서의 정량적 검증 연구는 국내에서는 아직 수행되지 않았다. 이 연구에서는 Weisheimer and Palmer (2014)의 신뢰도 곡선(Reliability Diagram, RD) 방법을 이용하여 한국(KMA)과 영국기상청(UKMO) GloSea5의 기온과 강수의 확률 예측 신뢰도를 정량적으로 평가해보고자 한다. 그리고 유럽

중기예보센터(ECMWF), 미국국립환경예측센터(NCEP), 일본기상청(JMA)의 분석 가능한 자료를 확보하여 비교해 봄으로써 현 기후예측시스템에서 생산되는 계절 예측의 신뢰 수준에 대한 현황을 제시하고자 하였다.

기후예측시스템의 예측 수준을 높이기 위해 근본적으로 모델 자체의 역학이나 물리과정을 개선하는 방법 외에도, 예측시스템의 운영 설계, 즉, 초기화 요소와 기법, 앙상블 생성 방법과 앙상블의 수, 후처리 단계에서의 보정 등의 개선을 통하여 예측성을 높이는 방법들이 고민되고 있다. 이 연구에서 수집된 각 현업 센터의 자료는 별개의 예측시스템에서 여러 초기화 기법 혹은 여러 앙상블 생성법을 사용하고 있어, 이의 영향을 직접적으로 비교하고 판단하기는 어렵다. 그러나 이 연구에서는 앙상블 초기 시간, 앙상블 수 그리고 기후모의 기간 등과 같이 예측시스템 간 비교 자료의 선택을 달리하여 확률 예측 신뢰도를 산출하고 역학/물리 과정의 직접적인 개선 이외에 기후예측 시스템 운영 방법 측면에서 어떤 요소들이 예측성 향상에 민감하게 작용할 수 있는지에 대한 정보를 정량 확률예측 관점에서 파악해보고자 하였다.

현재 단중기 예보(short- and medium-range forecast) 이후와 기후변화 전망(climate change projection) 이내 사이를 지칭하는 용어로 ‘장기예측(long-range prediction)’, ‘기후예측(climate prediction)’, ‘계절예측(seasonal prediction)’ 등이 혼용되어 사용되고 있으며, ‘예보(forecast)’와 ‘예측(prediction)’ 또한 장기예보/장기예측 혹은 계절예보/계절예측 등과 같이 혼용되어 사용되고 있다. ‘예보’가 미리 일어날 일을 예측하여 알린다는 사전적 의미로 본다면 ‘예보’와 ‘예측’의 의미가 크게 다르지 않을 것이나, 예보는 과거 기상 상태의 영향이 지속될 것이라고 볼 수 있는 2주 이내의 시간에서 좀 더 객관적인 근거를 가지고 앞서 ‘fore-’ 알린다 ‘-cast’는 의미로 더 많이 사용되고 있다면, 예측은 이 보다는 긴 규모에서 주관적인 판단이 들어 갈 수 밖에 없는 변화를 짐작한다는 의미를 가지고 있다. 따라서 이 논문에서는 용어를 통일하여 사용하고자 단중기 예보와 기후변화 전망 사이를 ‘기후예측’이라고 통일하여 지칭하고, ‘기후예측’의 규모에서 2주~2개월 사이를 ‘계절내예측’, 3개월 이상에서 1년 이내의 규모를 ‘계절예측’이라고 구별지어 사용하였다. 우리나라 기후업무규정에서는 ‘장기예보’를 11일~3개월 이내, ‘기후전망’을 3개월 이후의 미래 상황을 예상하는 것으로 정의하고 있으며, ‘기후예측’은 ‘장기예보’와 ‘기후전망’을 포함한 기간으로 정의하고 있다. 따라서 기상청의 공식 예보를 이야기하는 경우는 규정 그대로 ‘장기예보’라고 하였으며, 한국기상청 뿐만 아니라 각 센터들의 공식예보에 대해 언급하는 경우에도 이에 대해서는 ‘장기예보’라는 용어를 통일하여 사용하였다.

**Table 1.** Summary of data from five operational centers.

		KMA	UKMO	ECMWF	NCEP	JMA
System		GloSea5		System4	CFS2	JMA/MRI-CPS2
CONFIG	Atmos	UM8.6	UM10.3	IFS Cycle 36r4	GFS4	JMA-GSM
	Ocean	NEMO3.4		NEMO3.0	GFDL MOM4	MRI.COM
RES	Atmos	N216L85		T <sub>1</sub> 255L91	T126L64	T <sub>1</sub> 519L60
	Ocean	ORCA025L75		ORCA1L42	1/3°-1°L50	1°lon, 0.3-0.5°lat L52
HCST Init	Atmos	ERA-Interim		ERA-interim	CFSR	JRA55
	Ocean	NEMOVAR		ORA-S4	CFSR	MOVE/MRI.COM-G2
Reference		MacLachlan et al. (2015)		Molteni et al. (2011)	Saha et al. (2014)	Takaya et al. (2018)
HCST Period		1991~2010 (20 yr)	1993~2015 (23 yr)	1981~2010 (30 yr)	1983~2010 (28 yr)	1979~2010 (36 yr)
ENS for each time		3	3 / 7	15 / 51	4	5
Init times/month		4	4	1	5	2
ENS size/month		240	276 / 644	450 / 1530	560	360
Init times/total		7	7	1	5	2
ENS size/total		420	483	450	560	360

## 2. 분석 자료

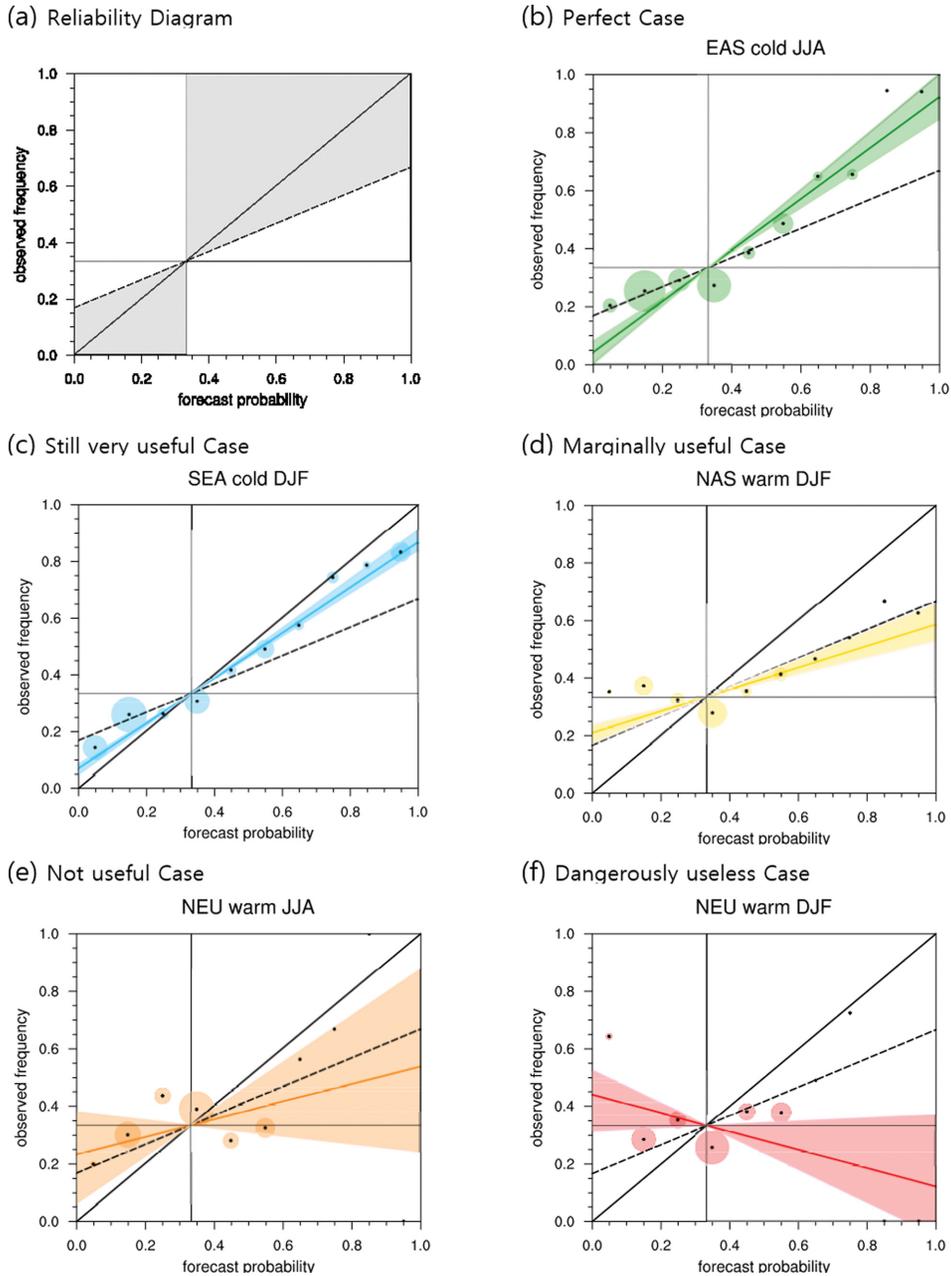
분석에 사용된 자료는 한국기상청(KMA), 영국기상청(UKMO)과 유럽 중기예보센터(ECMWF), 미국국립환경예측센터(NCEP), 일본기상청(JMA) 5개 기관의 예측 결과이며, 기온과 강수의 과거기후모의(Hindcast) 자료를 분석하였다. 각 기관의 시스템 구성은 Table 1에 표기하였다. 한국의 예측시스템 GloSea5는 Hadley Centre Global Environmental Model version 3 (HadGEM3)를 기반으로 하는 전 지구 대기-해양/해빙 결합 시스템이다. Hindcast의 경우 한 달에 4번(1, 9, 17, 25일) 초기장을 가지고, 한 초기장 당 3개의 통계적 기법(Stochastic)을 이용한 앙상블을 생산하고 있으므로 한달에 총 12개 20년치 240개의 앙상블을 생산한다. 영국 기상청의 경우는 한국과 같은 시스템이지만 Hindcast 기간이 23년으로 총 276개의 앙상블을 생산하는데, 최근 통계적 앙상블 생성을 3개에서 7개로 확장하여 총 644개의 앙상블이 생산되고 있다. ECMWF의 경우 한 번의 초기시간에 51개의 앙상블을 30년치 1,530개 생산하고 있는데, 입수하여 분석한 자료는 15개의 앙상블 결과로 총 450개 앙상블이다. 따라서 450개 앙상블 결과를 분석하였으며, 이후 1,530개의 앙상블을 사용한 Weisheimer and Palmer (2014) 논문 결과와 비교하였다. 미국과 일본의 앙상블은 각각 560개와 360개

이다(Table 1; ENS size/month).

이와 같이 현재 분석에 이용된 5개 기관의 현업 운영 설계, 즉, 초기 시간, 통계적 앙상블 생성 수, Hindcast 기간 등이 모두 다른 것을 알 수 있는데, 모델의 성능을 동일선 상에 두고 분석하기 위해서는 같은 초기시간에 생산된 같은 앙상블 수를 가지고 분석하는 것이 가장 정확한 비교라고 생각할 수 있을 것이다. 그러나 독립적으로 개발운영되는 시스템을 동일하게 맞추어 분석하는 것은 불가능하고, 각 기관 고유의 앙상블 생성 전략 또한, 모델 자체가 아닌 운영 시스템의 성능이라고 볼 수 있기 때문에 기관별 현업 운영 설계를 그대로 가용한 앙상블을 사용하여 분석에 이용하였다.

이 연구에서는 Weisheimer and Palmer (2014)와 같은 방법으로 확률 예측성을 검증하여 비교해 보고자 5월, 11월 생산된 각 기관의 앙상블을 사용하여 여름철 JJA와 겨울철 DJF의 신뢰도를 산출하였다. 다만 앙상블 수에 대한 의존성을 우선 제거하고 예측성을 먼저 평가하고자 KMA와 UKMO의 경우 5월과 11월 이전 2회, 이후 1회의 앙상블을 추가하여 7회의 초기장 총 420개, 483개 앙상블을 확보하여 분석하였다(Table 1; ENS size/total). 그리고 이 후 앙상블의 수와 초기 시간, Hindcast 기간을 달리하여 예측성에 미치는 영향을 평가하였다.

신뢰도를 검증하기 위한 관측 자료로 ERA-Interim



**Fig. 1.** Examples of reliability categorization. (a) is a schematic of a reliability diagram without any data points. The thick diagonal line indicates perfect reliability. The horizontal and vertical lines are the climatological frequency and the dotted line is no skill line. The grey area marks a region where the forecast data has a positive value in Brier skill score. From (b) to (f) are examples of each case. (b) is ‘perfect’, (c) is ‘still very useful’, (d) is ‘marginally useful’, (e) and (f) are ‘not useful’ and ‘dangerously useless’ cases.

기온(2m temperature)과 Global Precipitation Climatology Project (GPCP) 강수 자료를 이용하였으며, 해상도를 통일하기 위해 위경도 2.5°로 선형 내삽하여 분석하였다.

### 3. 신뢰도 등급 산출 방법

모델의 확률 예측 신뢰 수준은 앙상블 예측에서 나

**Table 2.** Five categorizations of reliability.

Category	Reliability	Description
1	Perfect	The uncertainty range of the reliability slope includes the perfect slope and falls into the skillful BSS area
2	Still very useful	The uncertainty range of the reliability line at minimum slope $\geq 0.5$
3	Marginally useful	The slope of the reliability line $\geq 0$ and not includes the perfect line
4	Not useful	The slope of the reliability line cannot be distinguished within its uncertainties from 0
5	Dangerously useless	The reliability line $< 0$

타난 3분위 이벤트(Below Normal, Normal, Above Normal)가 관측에서 발생하는 빈도와 비교하여 얼마나 일치하였는지를 본다. 확률 예측의 검증을 위해 대표적으로 쓰이는 방법으로, 평균 제곱 오차(Mean Squared Error) 와 비슷한 개념으로 예측과 관측 빈도 간 차를 제곱 평균한 BS (Brier Score) 가 있는데, 모든 예측이 100% 다 맞았을 경우는 0이 되고, 반대로 100%로 틀렸을 경우 1의 값을 가진다. BS에 기반하여 BSS (Brier Skill Score)는 BS의 기후값을 reference로 비교하여, 모든 예측이 적중했을 때 1이 되도록 한 검증법이다. 또한 관측과 예측에서 모두 적중한 경우(Hit)와 관측에 없었던 경우를 잘못 예측한 경우(False Alarm), 관측에서 나타났으나 예측하지 못한 경우(Miss)를 조합하여, 얼마나 많은 비율로 예측을 적중했는지 혹은 틀렸는지의 비를 보는 Probability of Detection (POD), False Alarm Rate (FAR), Threat Score (TS) 등의 검증 방법이 있다.

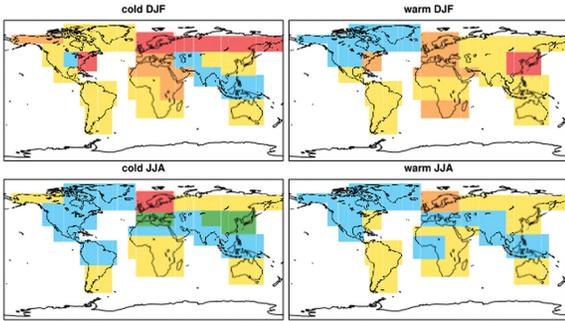
이러한 확률 검증 방법 중 RD (Reliability Diagram)은 앙상블 예측 확률(x 축) 대비 실제 관측에서 나타난 빈도(y 축)를 표시하는 방법으로, Fig. 1a에서 보이는 기울기가 1인 대각선과 정확하게 일치하면 완벽한 예측(Perfect Forecast)이 되고, 대각선 위에 존재하면 과대 예측, 아래에 위치하면 과소 예측한 것으로 해석할 수 있다. Figure 1a에서 가로 실선과 세로 실선은 기후적 발생 확률(33%)이며, 회색 음영은 BSS가 양의 값을 가지는 구간으로 점선은 완벽한 예보를 나타내는 대각선과 기후적 발생 확률의 중앙 값이며 기후적 발생 확률 대비 유용한 예측인지의 여부를 판단하는 기준선(No skill line)이다. 즉, 확률 예측 결과가 회색 음영구간(BSS > 0)에 들어가는 경우를 유용한 예측으로 판단한다.

RD는 이렇게 그래프를 통해 가시적으로 예측 결과를 분석할 수 있는 널리 쓰이는 검증 방법 중에 하나이다. 다만 이 방법은 그래프를 통해 분석결과를 쉽게 알 수 있는 장점이 있지만 정량적으로 예측이 잘 맞았는지에 대한 기준을 제시해 주지 않는다. 이

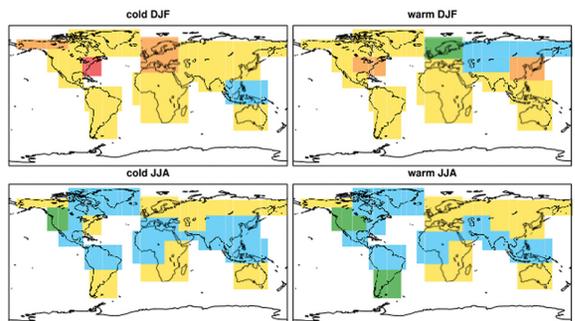
러한 점을 보완하고자 Weisheimer and Palmer (2014)에서는 RD를 통한 검증 결과를 신뢰도 표를 이용하여 정량화하고 5개 카테고리 나눈 구간에 따른 예측성을 쉽게 파악 할 수 있는 방법을 제시하였다. 먼저 기후값을 기준으로 하여 여름철과 겨울철 기온, 강수의 계절 예측을 3분위(평년 이상, 평년 비슷, 평년 이하)로 구분하고, 평년 비슷 범위는 연도별 자료를 순차적으로 배열한 후 표본의 수가 동일하게 되도록 3등분하여 중앙에 속하는 표본 값의 범위를 택하는 순위(Ranking) 방법을 이용하였다. 3분위의 각 분위 별 예측 확률을 구하고, Bootstrap 기법을 이용하여 무작위로 표본을 선택하여 선형회귀선을 계산하는 과정을 1000번 반복하였다. 이 과정을 통해 적정 선형회귀선과 오차 범위(75% confidence)를 구하고 앙상블 멤버의 스프레드를 고려하여 5개의 신뢰도 등급으로 나누었다(Table 2). 그리고 이를 ECMWF 예측 결과에 적용하여 전지구의 지역별 신뢰도 맵을 제시하였다.

Figures 1b-f는 신뢰도 1에서 5등급까지 결정된 경우에 대한 예시를 보였다. 점으로 찍힌 값이 예측확률 대비 관측에서 나타난 빈도 값이며, 점을 중심으로 한 동그라미의 크기는 해당 예측 확률 구간의 예측 횟수와 비례한다. 음영으로 색칠된 구간은 선형회귀선의 오차 범위이다. Figure 1b는 EAS (East Asia) 지역의 추운 여름을 예측했던 경우로 신뢰도 1등급(Perfect)은 선형회귀선의 오차 범위가 일직선의 대각선(Perfect line)을 포함하고 BSS가 양의 값인 구간에 있어 완벽한 예보로 판단될 때이다. Figure 1c는 SEA (Southeast Asia)의 추운 겨울을 예측했던 2등급(Still very useful)의 경우로 선형회귀선의 오차 범위가 최소 0.5 이상의 선형 관계를 가지므로 매우 유용한 수준의 확률 예측임을 뜻한다. Figure 1d는 NAS (North Asia)의 따뜻한 겨울을 예측했던 3등급(Marginally useful) 사례로 선형회귀선의 오차 범위가 최소 0 이상, 즉 climatology를 포함하지 않아 기후적 발생 확률보다 유용한 예측이나 perfect line을 포함하지 않을

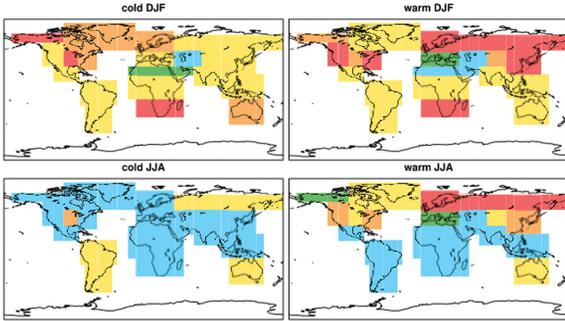
(a) KMA



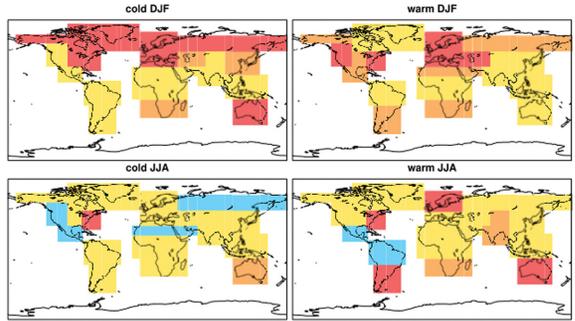
(b) UKMO



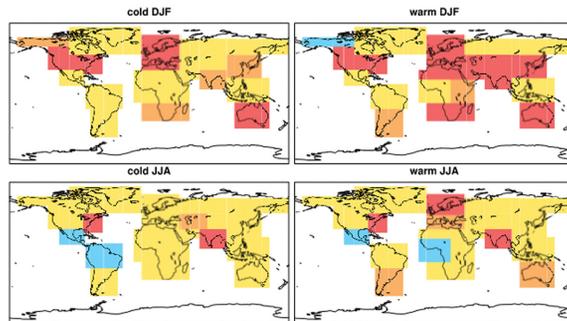
(c) ECMWF



(d) NCEP



(e) JMA



① perfect ② still useful ③ marginally useful ④ not useful ⑤ dangerous

**Fig. 2.** The reliability map of 2 m temperature from (a) KMA, (b) UKMO, (c) ECMWF, (d) NCEP and (e) JMA on below DJF (upper left), above DJF (upper right), below JJA (lower left) and above JJA (lower right). Green, blue, yellow, orange and red-colors are shaded based on reliability categorization from perfect to dangerous forecast.

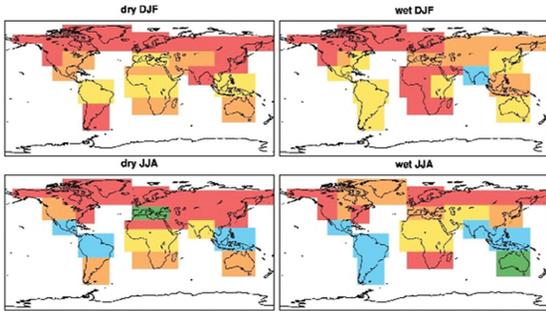
때를 나타내며 확률 예측을 신뢰할 수 있는 경계 등급이다. Figure 1e는 NEU (Northern Europe)의 따뜻한 여름을 예측한 4등급 (Not useful)으로 예측 확률과 관측 빈도의 선형 관계가 없을 때, Fig. 1f는 같은 NEU의 따뜻한 겨울을 예측한 5등급(Dangerously useless)으로 두 변수가 음의 선형관계를 가지고 있어 신뢰할 수 없는 수준의 확률 예측임을 뜻 한다. 좀 더 자세한 방법은 Weisheimer and Palmer (2014)를 참고할 수 있다.

## 4. 결 과

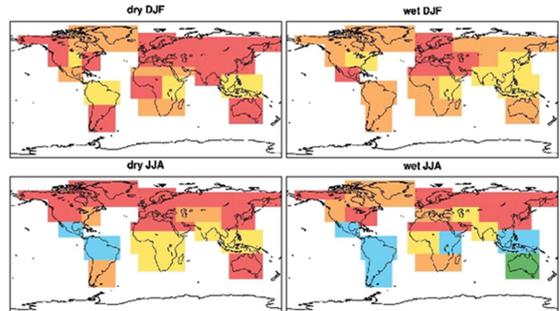
### 4.1 기온 및 강수 계절 확률예측 신뢰도

Figure 2와 Figs. 3a-f는 각 기관별 기온 예측과 강수에 대한 신뢰도 맵이다. 관측 기온이 3분위 확률 분포에서 평년보다 추울 것/건조할 것으로 예측한 겨울 (cold/dry DJF)과 더울 것/습할 것으로 예측한 겨울 (warm/wet DJF), 그리고 평년보다 기온이 낮을 것/건조할 것으로 예측한 여름(cold/dry JJA)와 더울 것/습

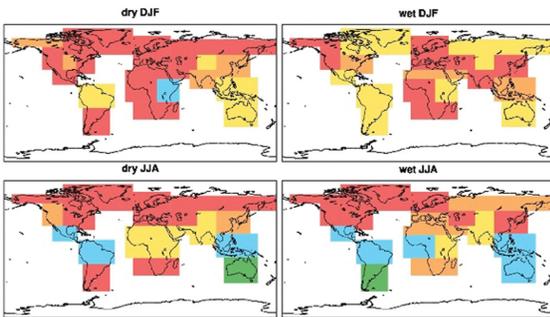
(a) KMA



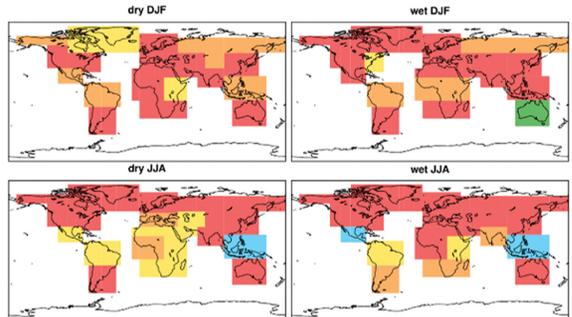
(b) UKMO



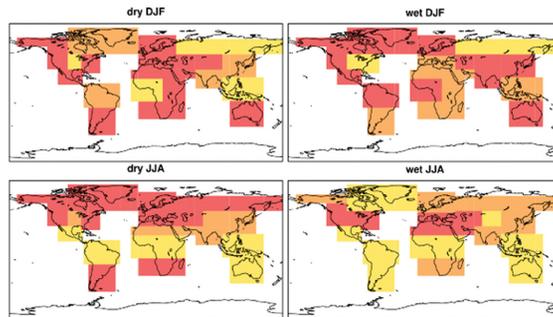
(c) ECMWF



(d) NCEP



(e) JMA



1 perfect 2 still useful 3 marginally useful 4 not useful 5 dangerous

Fig. 3. Same as Fig. 2 except for precipitation.

할 것으로 예측한 여름(warm/wet JJA)에 대하여, 전 지구를 21개 지역(Giorgi and Francisco, 2000)으로 나누어 지역별 신뢰도를 산출하였다. 그리고 가장 좋은 1등급 ‘Perfect Case’부터 마지막 5등급인 ‘Dangerously Useless Case’까지 순서대로 초록, 파랑, 노랑, 주황 그리고 빨간색으로 표시하였다.

Figure 2에서 가장 큰 특징은 모든 현업 기관의 결과에서 여름철 기온의 계절예측 신뢰도가 겨울에 비해 높은 것을 알 수 있다. 이 결과는 GloSea5 결과에 대해 지표기온이 통계적으로 유의한 수준에서 여름에 시작된 예측의 예측성이 높게 나타난 선행 연구(Jung

et al., 2015)와 850 ha 기온의 RMSE가 여름철에 작고 겨울철에 크게 나타난 선행 연구에서도 확인할 수 있다(Ham et al., 2017). 그러나 평년보다 기온이 낮을 것(cold)이라고 예측한 결과와 높을 것이라고(warm) 예측한 결과에 대해서는, 지역별로 기관별로 차이는 있겠으나 대체로 비슷한 신뢰도를 보였다. 또 하나의 특징은 대부분 예측 스کیل이 없는 것으로 판단되는 붉은색 지역이 중고위도 지역에 많이 위치하고 있어, 선행 연구(Phelps et al., 2004; Peng et al., 2011; Kim et al., 2012)에서 알려진 바와 같이, 열대지역의 신뢰도가 중고위도에 비해 높은 것을 확인할 수 있었다.

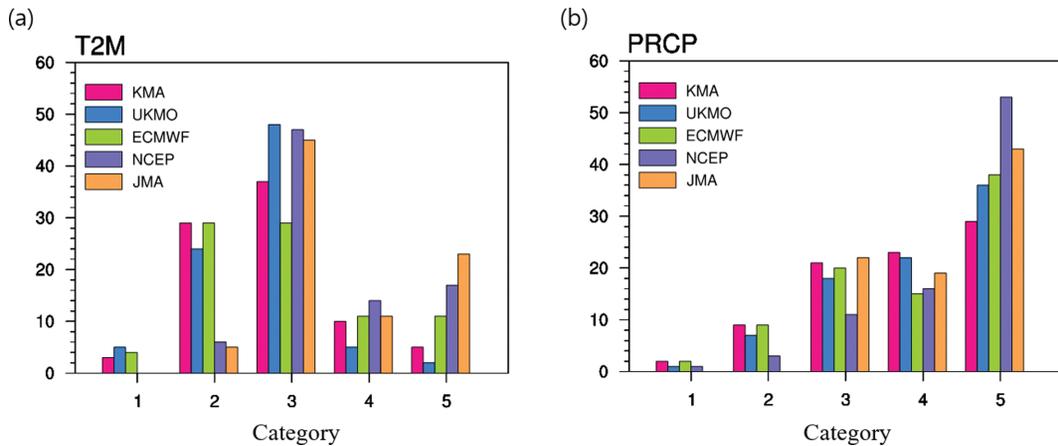


Fig. 4. Histogram of the numbers of regions in each reliability category for different centers. (a) is for 2 m temperature and (b) is for precipitation, and red color is for KMA, blue is UKMO, green is ECMWF, purple is NCEP and orange is JMA.

Figures 2a, b에서 KMA와 UKMO의 GloSea5는 대기 모델 버전과 Hindcast 기간의 차이가 있었으나 거의 동일한 시스템이다. 따라서 KMA의 추운 겨울(cold DJF) 유라시아 북쪽의 신뢰도가 UKMO에 비해 좋지 않은 5등급으로 평가되었으나 따뜻한 겨울 북미의 예측성은 KMA가 더 좋게 나타나는 등 조금씩 차이가 나타나지만 전반적인 신뢰도는 유사하게 나타났다. 전체 신뢰도 등급별 수는 뒤에 Fig. 4에서 보일 것이다. Figure 2c의 ECMWF System4와 GloSea5 결과를 비교해 보았을 때 유사한 신뢰도를 보이고 있었으며, ECMWF에서 따뜻한 겨울과 여름(warm DJF, warm JJA) 유라시아 대륙의 북쪽 예측성과 겨울철 아프리카 남쪽의 예측성이 유독 좋지 않은 특징이 있었다. Figure 2d NCEP과 Fig. 2e JMA의 경우 1등급의 신뢰도를 나타내는 지역이 없었으며 상대적으로 신뢰도 등급이 낮게 평가되었다. 이는 NCEP과 JMA 시스템 해상도가 수평 100 km 연직 60층 정도로 GloSea5와 System4의 60 km/연직 85층, 80 km/연직 91층에 비해 상대적으로 저해상도인 이유도 있을 수 있을 것이고, JMA의 앙상블 수가 분석 기관들 중 가장 작은 360개인 것도 불리하게 작용할 수 있었을 것으로 생각된다. 하지만 Fig. 5에서 240개의 앙상블 만으로 GloSea5 신뢰도가 상대적으로 높았던 것과 낮은 신뢰도를 앙상블 수만으로는 설명할 수 없을 것으로 생각된다.

Figure 3에서 강수의 경우, 기온에 비하여 신뢰도 등급이 많이 낮아 5개 기관 모두에서 4, 5등급의 신뢰도가 대다수를 차지 하였으며, 기온과 동일하게 여름철 신뢰도가 겨울철보다 상대적으로 높은 경향을 보였다. Figures 3a-c의 GloSea5와 System4에서 평년보다 강수가 많을 때(wet)의 예측 신뢰도가 강수가 적을 때(dry)보다 높은 경향이 있었으며 남반구의 신뢰

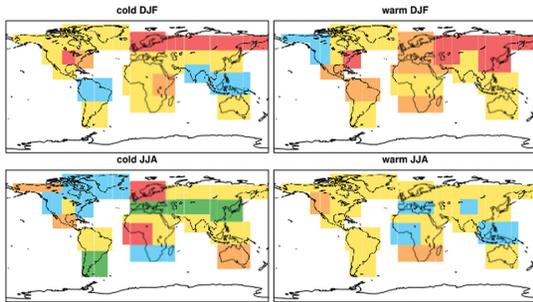
도가 북반구에 비해 높은 경향이 있었다. 특이한 점은 특히 평년보다 강수가 많을 때의 여름철(wet JJA) 호주 지역은 KMA, UKMO, ECMWF에서 거의 완벽한 수준의 1, 2등급 신뢰도를 나타냈다.

기온과 강수의 신뢰도 등급에 따라 구분된 전체 지역 수를 히스토그램으로 Fig. 4에 나타냈다. 왼쪽 Fig. 4a와 오른쪽 Fig. 4b이 각각 기온과 강수에 대한 각 기관별 신뢰도 등급의 지역 수이다. 기온의 계절 확률 예측 신뢰도는 절반에 가까운 49%가 3등급에 위치하고 있었으며, 71%가 2, 3등급에 위치하고 있는 것을 볼 수 있었다. 그러나 강수의 경우는 48%가 5등급의 신뢰도를 나타내고 있어 현재 강수 계절예측의 어려움과 한계를 확인 할 수 있었다. 기관에 따라 차이가 존재하였으나, 평균적으로 기온의 경우 높은 신뢰도(1, 2등급)를 보이는 예측과 낮은 신뢰도(4, 5등급)를 보이는 경우가 반반인 반면, 강수의 신뢰도는 1, 2등급에 해당하는 경우가 전체의 8% 밖에 해당되지 않았다.

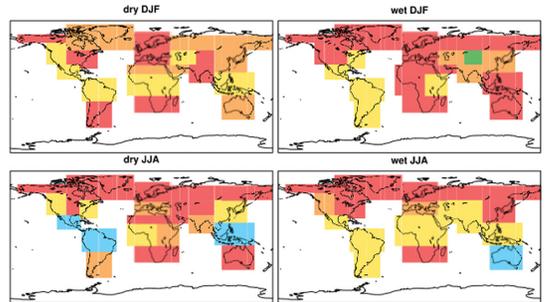
#### 4.2 앙상블 초기 시간에 따른 신뢰도 비교

강수의 2절에서 언급한대로, KMA와 UKMO GloSea5의 경우 Hindcast 앙상블 수를 다른 기관과 유사하게 확보하기 위해, 5월과 11월 각 4번의 초기장에 이전 2회, 이후 1회의 초기 시간을 추가하여 분석에 사용하였다. 초기 시간을 확장하여 앙상블 수를 420개로 증가시킨 경우, 앙상블이 기존 240개인 경우와 비교하여 어느 정도 차이가 있는지 살펴보고자 하였다. 그리고 앙상블 240개라 하더라도 초기 시간이 예측 계절과 먼 4월 말/10월 말이 포함된 이른 초기장(4/17, 4/25, 5/1, 5/9 | 10/17, 10/25, 11/1, 11/9)의 경우와 6월 1일과 12월 1일이 포함된 늦은 초기장(5/9, 5/17,

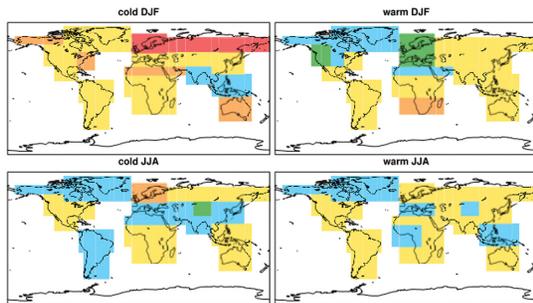
(a) T2M Early Init; pre 17th, 25th and 1st, 9th of the mon



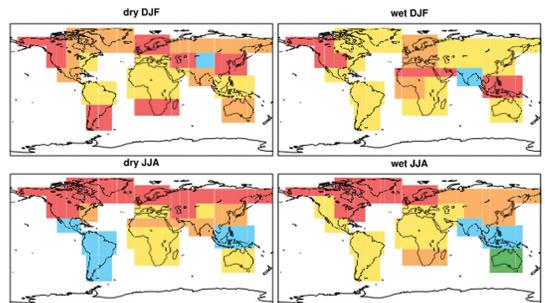
(b) PRCP Early Init; pre 17th, 25th and 1st, 9th of the mon



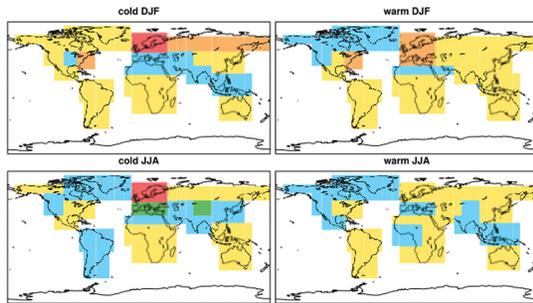
(c) T2M Org Init; 1st, 9th, 17th, 25th of the mon



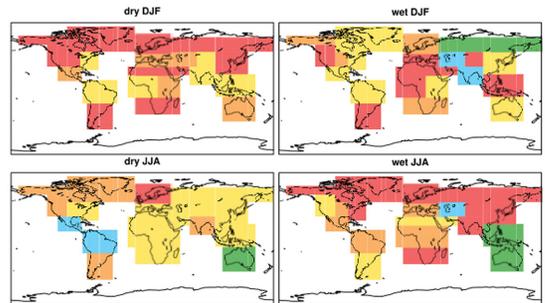
(d) PRCP Org Init; 1st, 9th, 17th, 25th of the mon



(e) T2M Late Init; 9th, 17th, 25th and next 1st of the mon



(f) PRCP Late Init; 9th, 17th, 25th and next 1st of the mon



**Fig. 5.** The reliability map of 2 m temperature (left panel) and precipitation (right panel) from KMA's 240 ensembles. Upper, middle and bottom panels are for different initial times (early, original and late initial times for each).

5/25, 6/1 | 11/9, 11/17, 11/25, 12/1)을 사용한 경우를 비교하여 목표 계절에 대한 선행 시간에 따라 예측성이 어떠한 차이를 보이는지도 살펴보았다(Fig. 5, Table 3).

Figure 5는 KMA GloSea5 기온(왼쪽 패널)과 강수(오른쪽 패널)의 위 줄부터 이른 초기장, 기존 초기장, 늦은 초기장을 사용한 경우에 대한 신뢰도 맵이다. 그리고 Table 3은 KMA와 UKMO의 결과를 표로 나타낸 것이다. 기온의 경우 늦은 초기장 즉, 초기화 한 날짜가 예측 시간에 가까운 값을 쓴다고 하더라도 무조건 신뢰도가 증가하는 것은 아니었고 지역과 사례에 따라 다르게 나타났다. 유럽의 경우는 따뜻한 겨울(warm DJF)을 예측한 경우 기존 한달 초기장을 쓴

경우가 초록색 1등급으로 신뢰도가 가장 높게 나타났으며, 추운 여름(cold JJA)을 예측한 동아시아, 중국, 남유럽 지역의 예측성은 이른 초기장을 썼을 때 1등급으로 상향되었다. 이는 기후예측시스템 예측성을 평가해 보면 한 주는 예측성이 전체적으로 상향하고 그 다음주는 전반적으로 떨어지고 다시 상향하는 특성이 나타나는 경우가 발생하는데, 특정 대기 상태일 때 모델 예측에 더 유리한 혹은 예측성이 더 지속되는 의존성이 존재함을 알 수 있다. 그러나 Table 3에서 볼 수 있듯 늦은 초기장을 사용했을 때 전체적으로 4, 5등급 신뢰도 지역이 줄어드는 경향을 확인할 수 있었다. Figure 2a의 420개 앙상블을 사용한 경우와 원

**Table 3.** The number of regions that is initialized from different initial dates of KMA, UKMO on five reliability categories summed over all four events for 2 m temperature (T2M) and precipitation (PRCP).

Center	Initial date	Category of reliability									
		T2M					PRCP				
		1	2	3	4	5	1	2	3	4	5
KMA	Early	5	13	44	13	9	1	4	26	12	41
	Original	3	29	37	10	5	2	9	21	22	30
	Late	2	27	48	5	2	4	5	28	19	28
UKMO	Early	0	19	49	11	5	2	3	31	14	34
	Original	5	24	48	5	2	1	7	18	22	36
	Late	2	24	55	2	1	0	6	45	15	17

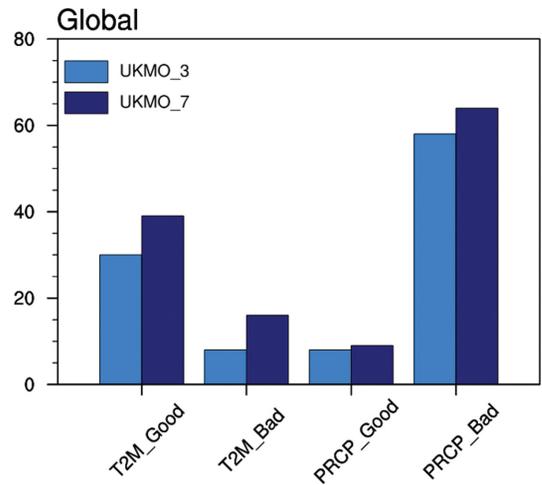
래의 한 달 240개 초기장을 쓴 Fig. 5c을 쓴 경우 비슷한 결과를 보였으며, 가장 늦은 초기장을 쓴 Fig. 5e와 비교했을 때에는, 5등급 지역이 없어지는 결과로 볼 때, 앙상블을 전후로 확장한 경우보다는 최근 초기장을 쓰는 것이 계절예측에 더 유용한 것으로 판단되었다. 강수의 경우도 초기 시간에 대한 효과가 지역마다 다른 경향으로 나타났으나, 특히 wet DJF와 dry JJA에 늦은 초기장을 쓸수록 예측성이 증가하는 것을 확인할 수 있었으며, 표에서 볼 수 있듯 5등급 신뢰도 지역이 감소하는 경향이 늦은 초기장을 사용하였을 때 뚜렷이 나타났다.

선행 시간이 짧은 앙상블이 모든 경우 유리하게 작용하는 것이 아니라는 결과는, 예측 선행 시간이 계절 규모 이상으로 점점 증가함에 따라 선행 시간에 따른 영향보다는 앙상블 크기의 효과가 점점 더 크게 발휘되고 있기 때문인 것으로 판단되었다. 어느 정도의 예측 규모까지 선행시간의 효과가 큰 지, 아니면 앙상블 수의 효과가 큰지를 구별해 보고자 계절규모가 아닌 계절내 선행시간에 대한 신뢰도 맵을 산출해 보고자 하였으나, 이 경우 선행 분석시간이 주 단위로 짧아짐에 따라 분석에 사용하는 앙상블 수도 줄어들 수밖에 없고, 표본이 작은 경우 결과가 중간 등급인 3등급으로 집중되는 경향이 나타나 이 연구의 신뢰도 맵 방법은 계절규모 이내의 예측성 평가에는 적절하지 않은 것으로 판단되었다.

**4.3 앙상블 확대에 따른 신뢰도 비교**

앞서 Table 1에서 설명한 바와 같이 최근 영국에서는 Hindcast의 통계적 앙상블 수를 한 초기시간 당 3개(총276개)에서 7개(총644개)로 증가시켜 생산하기 시작하였다. 따라서 이 두 경우를 비교하여 초기시간이 아닌 통계적 앙상블 확대 생산에 따른 영향을 비교해 보았다.

Figure 6는 UKMO GloSea5의 기온과 강수의 등급별 신뢰도의 지역 수를 1, 2등급의 높은 신뢰도 지역



**Fig. 6.** Histogram of the numbers of regions in ‘Good’ (category 1 and 2) and ‘Bad’ (category 4 and 5) reliability cases from UKMO. Left two bars are for 2 m temperature and right two are precipitation. Blue color means 3-stochastic ensemble case per initial time, and dark blue is 7-stochastic ensemble.

(Good) 과 4, 5등급의 낮은 신뢰도 지역(Bad)을 나누어 표시한 히스토그램이다. 기온의 경우 앙상블을 확대한 경우 확대하지 않은 경우에 비해 3등급에서 1, 2등급으로 신뢰도가 증가하여 앙상블 증가에 대한 효과를 볼 수 있으나 이전보다 4, 5등급의 지역도 증가하였다. 강수의 경우 오히려 4, 5등급 지역이 증가하는 결과가 나타나 강수의 경우 앙상블을 증가한 효과가 예측성의 향상으로 나타나지 못했다. 그러나 여름과 겨울 계절로 따로 나누어 분석해 보았을 때, 앙상블 증가로 3등급의 지역이 2등급으로 옮겨지는 효과가 여름철 기온에서 두드러지게 나타났다(not shown).

앙상블의 수가 많은 ECMWF의 기온에 대해 앙상블의 수와 기후 기간 등을 달리하여 예측 신뢰도를

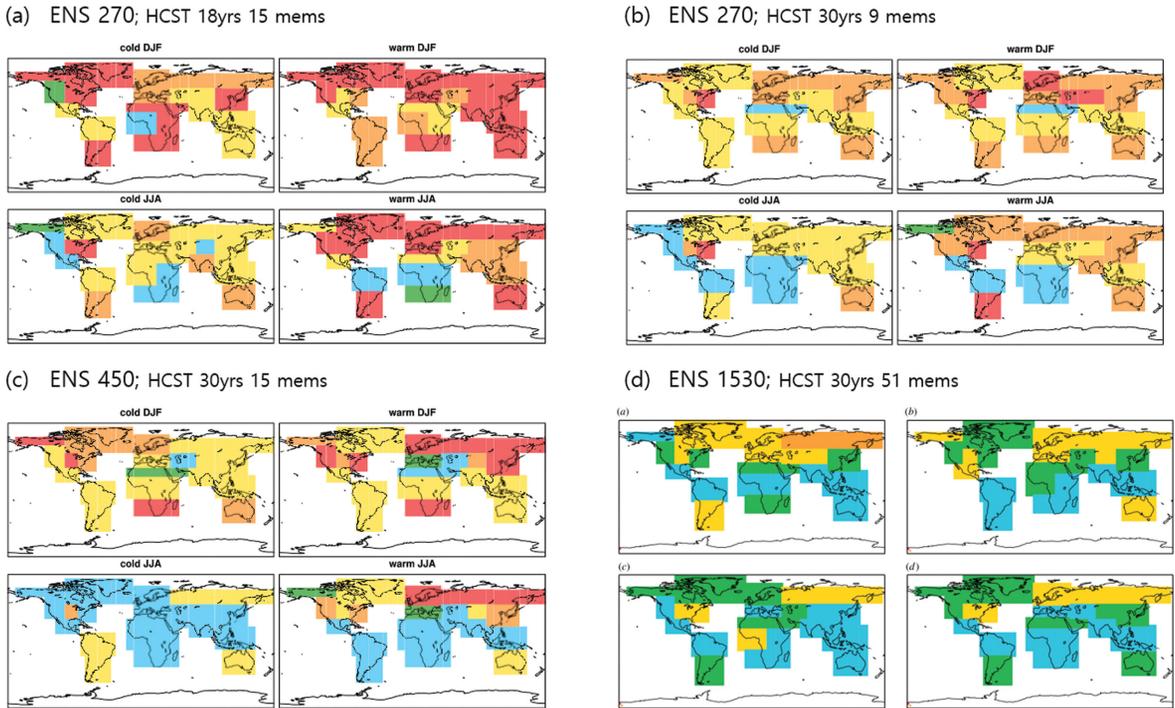


Fig. 7. The reliability map of 2 m temperature from ECMWF for different ensemble sizes and hindcast period. (d) is 'Fig. 4' in Weisheimer and Palmer (2014).

Table 4. Same as in Table 3, but for that is based on different ensembles of UKMO and ECMWF.

Center	Total Runs	Category of reliability									
		T2M					PRCP				
		1	2	3	4	5	1	2	3	4	5
KMA	<b>Total 240</b>   20yrs/StcE03/Iinit04	3	29	37	10	5	2	9	21	22	30
	<b>Total 276</b>   23yrs/StcE03/Iinit04	5	24	48	5	2	1	7	18	22	36
UKMO	<b>Total 644</b>   23yrs/StcE07/Iinit04	9	30	29	10	6	1	8	11	16	48
	<b>Total 270</b>   18yrs/StcE15/Iinit01	3	9	24	14	34	3	6	15	19	41
ECMWF	<b>Total 270</b>   30yrs/StcE09/Iinit01	1	15	33	27	8	0	13	31	28	12
	<b>Total 480</b>   30yrs/StcE15/Iinit01	4	29	29	11	11	2	9	20	14	39

평가해보았다(Fig. 7). 먼저 Fig. 7c는 Fig. 2c와 동일한 그림으로, ECMWF의 30년 기후모의기간(1981~2010년)에 대한 15개 앙상블 총 450개 앙상블 멤버를 이용하여 분석한 결과이다. Figure 7a는 기후모의기간을 18년(1993~2010년)으로 축소한 270개 앙상블의 결과이며, Fig. 7b는 기후기간은 30년으로 그대로 두고 9개의 앙상블을 선택하여 같은 270개 앙상블로 분석한 신뢰도 결과이다. 기후모의 기간을 1993~2010년의 18년으로 설정한 것은 각 센터들의 공통된 과거 기후기간을 추출하여 분석하였기 때문이며, 이 경우 270개

앙상블 수가KMA/UKMO의 GloSea5 한 달 240개/276개 앙상블 수와도 비슷하여 일관성있게 비교분석 할 수 있기 때문이다. 마지막으로 Fig. 7d는 Weisheimer and Palmer (2014) 논문의 Fig. 4로 30년 기후기간에 51개 앙상블을 써서 총 1,530개의 앙상블 멤버로 산출한 신뢰도 결과이다.

확률 예측의 신뢰도는 앙상블 멤버 수가 증가하면서 함께 증가하는 경향을 살펴 볼 수 있었다. 그러나 Fig. 7a와 Fig. 7b의 결과를 비교해 보았을 때에 같은 270개의 앙상블 수라고 할지라도 기후 기간을 증가시

키는 것이 예측성 향상에 더 도움이 되는 것을 알 수 있었다. 앙상블이 1,530개로 증가함에 따라 예측성의 향상이 크게 눈에 띄었다.

이 결과들을 강수도 함께 Table 4에 표시하였다. KMA, UKMO, ECMWF 앙상블이 240, 276, 270개로 유사한 경우 이들 세 기관의 확률예측 신뢰도는 기온의 경우 3등급, 2등급 순으로 많이 나타나고 있었다. 그러나 비슷한 앙상블 수라도 ECMWF의 Hindcast 기간이 18년으로 짧은 경우는 동일 ECMWF의 30년 Hindcast 경우 뿐만 아니라 KMA와 UKMO의 경우에 비해서도 5등급의 신뢰도 지역이 많이 나타났다. 한편, 강수의 경우 ECMWF의 신뢰도가 상대적으로 1, 2등급의 신뢰도 지역이 많이 나타나 높은 예측성을 보였다.

따라서 앙상블 수가 400개 전후(Fig. 2)나 200개 전후로 유사한 경우 KMA나 UKMO, ECMWF 세 기관의 예측성은 유사하게 평가되었다. 또한, 같은 앙상블 수를 사용한다면 개별 앙상블의 수를 늘리는 것보다 Hindcast 기간을 늘리는 것이 계절 예측성을 증가시키는 데 더 유리하게 작용하는 것을 알 수 있었다.

## 5. 결론 및 토의

이 연구에서는 KMA/UKMO의 GloSea5와 각 세계 현업 센터들의 Hindcast 자료를 이용하여 기온과 강수의 계절 확률 예측 신뢰도를 평가하였다.

현재 기후 예측 현업 센터들에서 생산하고 있는 기온의 확률 예측은 계절 규모에서 유용하게 활용할 수 있는 수준이었으며, 강수의 경우는 일부 지역에 한해 신뢰할 수 있는 수준으로 나타났다. ECMWF의 신뢰도는 KMA/UKMO GloSea5와 유사한 수준이었고, 선행 연구(Weisheimer and Palmer, 2014)에 비해 신뢰도 등급이 낮게 산출된 것은 3배가 넘는 앙상블 멤버 수의 차이 때문인 것으로 보인다.

이 연구에서는 2장에서 언급한 바와 같이 운영시스템 간의 초기 시간 차이를 같게 맞추고 분석하기 불가능한 사항을 감안하여 분석에 이용하였다. 그리고 초기 시간에 따른 민감도 분석한 결과 계절 규모에서 앙상블 멤버를 확보하기 위한 초기 시간의 사용은 앙상블 수 보다는 민감하지 않는 것으로 판단되었다. 즉, 계절내예측의 경우 앙상블의 예측 초기 시간에 더 민감할 수 있을 것이나, 계절 규모 예측의 경우 초기시간보다는 앙상블 수의 확보가 더 중요한 것으로 생각된다. 그러나 초기 시간에 따른 예측 성능의 차이를 보였고, 선행 시간이 짧은 자료를 사용할 수록 5등급의 신뢰도 지역이 줄어드는 경향을 확인할 수 있었다. 그러나 항상 초기시간이 가깝거나 앙상블의 수가 많은 경우 예측성에 유리한 것만은 아니었는데 이는 특

정 대기 상태를 초기장으로 모델을 수행했을 때 예측에 더 유리한 혹은 예측성이 더 길게 지속되는 의존성이 존재하기 때문인 것으로 판단된다. 이러한 예측성의 평가는 평가 방법에 따라서 의존적일 수 있다.

이 연구에서 사용한 검증 방법은 Hindcast 기간과 초기장 수, 앙상블 멤버로 구성된 표본의 수에 따라 민감하게 반응하였다. 표본이 클수록 신뢰도 등급이 높거나 낮은 방향으로 뚜렷하게 구별되지만, 표본이 적어질수록 중간 등급인 3등급 수준의 신뢰도 등급으로 편중되는 경향이 있었다. 어느 정도의 표본이 충분한 수 인지에 대해서는 추가적인 실험이 필요할 것으로 생각되었다. 또한 신뢰도가 결정되는데 있어 신뢰도 표 x 축의 높은 예측 확률 구간, 즉 모든 앙상블 멤버가 동일한 확률로 삼분위수를 예측했을 때, 그때의 관측 발생 여부가 매우 중요하게 작용하였다. 바꾸어 말해 신뢰도가 하락하는 대부분의 경우는 다른 구간에서 예측 확률이 관측빈도와 유사해도, 예측 확률 값이 높은 구간에서 모델이 관측 발생빈도에 비해 과대하게 모의하는 경우였다. 따라서 이 논문에서 사용한 신뢰도 표 방법은 높은 예측 확률 구간에서의 표본이 많이 확보될수록 확률 예측을 정량적으로 검증할 수 있는 효과적인 방법이었으며, 따라서 현재로써는 많은 앙상블을 확보할 수 있는 계절 예측 평가에 더 적합한 방법이었다. 이 방법은 현업 기후예측시스템의 버전 업그레이드, 역학/물리과정 등의 변경에 따른 예측 성능의 검증을 위해 활용될 수 있을 것이다. 이 논문에서 사용한 평가 방법을 계절내 예측성에 활용하기 위해서는 초기 시간에 더욱 민감할 수 있을 것이며, 기관 간 공통 초기 시간을 맞추기 위해서는 한 초기 시간 당 좀 더 많은 앙상블이 확보되어야 하거나, 아니면 추가적인 실험을 해야 가능할 것으로 판단되었다.

향후 GloSea5에서 ECMWF 만큼의 앙상블과 Hindcast 기간을 확보한 후 선행 연구만큼의 예측성이 나타나지는지를 확인해보는 연구도 필요할 것으로 생각되며, 앙상블이 확보된다면, 초기 시간, 앙상블 수와 Hindcast 기간 간의 민감도와 계절 예측성이 최대에 이르는 앙상블 수에 대한 정량적인 판단도 가능할 것이라고 생각된다.

## 감사의 글

이 연구는 기상청 국립기상과학원 「장기예측시스템 개발」(KMA2018-00322)의 지원으로 수행되었습니다.

## REFERENCES

Giorgi, F., and R. Francisco, 2000: Uncertainties in regional

- climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dyn.*, **16**, 169-182.
- Ham, H., D. Won, and Y.-S. Lee, 2017: Performance assessment of weekly ensemble prediction data at seasonal forecast system with high resolution. *Atmosphere*, **27**, 261-276, doi:10.14191/Atmos.2017.27.3.261 (in Korean with English abstract).
- \_\_\_\_\_, S.-M. Lee, Y.-K. Hyun, and Y. Kim, 2019: Performance assessment of monthly ensemble prediction data based on improvement of climate prediction system at KMA. *Atmosphere*, **29**, 149-164, doi:10.14191/Atmos.2019.29.2.149 (in Korean with English abstract).
- Heo, S.-I., Y.-K. Hyun, Y. Ryu, H.-S. Kang, Y.-J. Lim, and Y. Kim, 2019: An assessment of applicability of heat waves using extreme forecast index in KMA climate prediction system (GloSea5). *Atmosphere*, **29**, 257-267, doi:10.14191/Atmos.2019.29.3.257 (in Korean with English abstract).
- Jung, M.-I., S.-W. Son, and H.-S. Kang, 2015: Assessment of 6-month lead prediction skill of the GloSea5 hindcast experiment. *Atmosphere*, **25**, 323-337, doi:10.14191/Atmos.2015.25.2.323 (in Korean with English abstract).
- \_\_\_\_\_, \_\_\_\_\_, Y. Lim, K. Song, D. Won, and H.-S. Kang, 2016: Assessment of stratospheric prediction skill of the GloSea5 hindcast experiment. *Atmosphere*, **26**, 203-214, doi:10.14191/Atmos.2016.26.1.203 (in Korean with English abstract).
- Kim, H., M.-I. Lee, D. Kim, H.-S. Kang, and Y.-K. Hyun, 2018a: Representation of boreal winter MJO and its teleconnection in a dynamical ensemble seasonal prediction system. *J. Climate*, **31**, 8803-8818, doi:10.1175/JCLI-D-18-0039.1.
- Kim, H., S.-W. Son, K. Song, S.-W. Kim, H.-S. Kang, and Y.-K. Hyun, 2018b: Prediction skill of GloSea5 model for stratospheric polar vortex intensification events. *J. Korean Earth Sci. Soc.*, **39**, 211-227, doi:10.5467/JKES.2018.39.3.211 (in Korean with English abstract).
- Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF system 4 and NCEP CFSv2 retrospective forecast for the northern hemisphere winter. *Climate Dyn.*, **23**, 2957-2973, doi:10.1007/s00382-012-1364-6.
- Kim, S.-W., H. Kim, K. Song, S.-W. Son, Y. Lim, H.-S. Kang, and Y.-K. Hyun, 2018c: Subseasonal-to-seasonal (s2s) prediction skills of GloSea5 model: Part 1. Geopotential height in the northern hemisphere extratropics. *Atmosphere*, **28**, 233-245, doi:10.14191/Atmos.2018.28.3.233 (in Korean with English abstract).
- Lee, H., P.-H. Chang, K. Kang, H.-S. Kang, and Y. Kim, 2018: Assessment of ocean surface current forecasts from high resolution Global Seasonal Forecast System version 5. *Ocean and Polar Research*, **40**, 99-114, doi:10.4217/OPR.2018.40.3.099 (in Korean with English abstract).
- Lee, K.-J., and M. Kwon, 2015: A prediction of northeast Asian summer precipitation using teleconnection. *Atmosphere*, **25**, 179-183, doi:10.14191/Atmos.2015.25.1.179 (in Korean with English abstract).
- Lee, S.-M., H.-S. Kang, Y.-H. Kim, Y.-B. Byun, and C. Cho, 2016: Verification and comparison of forecast skill between global seasonal forecasting system version 5 and unified model during 2014. *Atmosphere*, **26**, 59-72, doi:10.14191/Atmos.2016.26.1.059 (in Korean with English abstract).
- Lim, S.-M., Y.-K. Hyun, H.-S. Kang, and S.-W. Yeh, 2018: Prediction skill of East Asian precipitation and temperature associated with El Niño in GloSea5 hindcast data. *Atmosphere*, **28**, 37-51, doi:10.14191/Atmos.2018.28.1.037 (in Korean with English abstract).
- MacLachlan, C., and Coauthors, 2015: Global Seasonal Forecast System version 5 (GloSea5): a high-resolution seasonal forecast system. *Q. J. Roy. Meteor. Soc.*, **141**, 1072-1084, doi:10.1002/qj.2396.
- Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (System4). ECMWF Tech. Memo. No. 656, 49 pp, doi:10.21957/4nery093i.
- Moon, S., S. Han, K. Choi, and J. Song, 2016: Data processing system and spatial-temporal reproducibility assessment of GloSea5 model. *J. Korea Water Resour. Assoc.*, **49**, 761-771, doi:10.3741/JKWRA.2016.49.9.761 (in Korean with English abstract).
- Park, J., H.-S. Kang, Y.-K. Hyun, and T. Nakazawa, 2018: Predictability of the arctic sea ice extent from s2s multi model ensemble. *Atmosphere*, **28**, 15-24, doi:10.14191/Atmos.2018.28.1.015 (in Korean with English abstract).
- Park, S., D.-J. Kim, S.-W. Lee, K.-W. Lee, J. Kim, E.-J. Song, and K.-H. Seo, 2017: Comparison of extended medium-range forecast skill between KMA ensemble, ocean coupled ensemble, and GloSea5. *Asia-Pac. J. Atmos. Sci.*, **53**, 393-401, doi:10.1007/s13143-017-0035-2.
- Peng, P., A. Kumar, and W. Wang, 2011: An analysis of seasonal predictability in coupled model forecasts. *Climate Dyn.*, **36**, 637-648, doi:10.1007/s00382-009-0711-8.

- Phelps, M. W., A. Kumar, and J. J. O'Brien, 2004: Potential predictability in the NCEP CPC dynamical seasonal forecast system. *J. Climate*, **17**, 3775-3785, doi:10.1175/1520-0442(2004)017o3775: PPITNC42.0.CO;2.
- Saha, S., and Coauthors, 2014: The NCEP climate forecast system version 2. *J. Climate*, **27**, 2185-2208, doi:10.1175/JCLI-D-12-00823.1.
- Son, K.-H., D.-H. Bae, and H.-S. Cheong, 2015: Construction and evaluation of GloSea5-based hydrological drought outlook system. *Atmosphere*, **25**, 271-281, doi:10.14191/Atmos.2015.25.2.271 (in Korean with English abstract).
- Song, K., H. Kim, S.-W. Son, S.-W. Kim, H.-S. Kang, and Y.-K. Hyun, 2018: Subseasonal-to-Seasonal (S2S) prediction of glosea5 model: Part 2. stratospheric sudden warming. *Atmosphere*, **28**, 123-139, doi:10.14191/Atmos.2018.28.2.123 (in Korean with English abstract).
- Takaya, Y., and Coauthors, 2018: Japan Meteorological Agency/Meteorological Research Institute-Coupled Prediction System version 2 (JMA/MRI-CPS2): Atmosphere-land-ocean-sea ice coupled prediction system for operational seasonal forecasting. *Climate Dyn.*, **50**, 751-765, doi:10.1007/s00382-017-3638-5.
- Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. R. Soc. Interface*, **11**, 20131162, doi:10.1098/rsif.2013.1162.