

기상청 기후예측시스템(GloSea6) 과거기후 예측장의 앙상블 확대와 초기시간 변화에 따른 예측 특성 분석

김지영* · 박연희 · 지희숙 · 현유경 · 이조한

국립기상과학원 기후연구부

(접수일: 2022년 10월 19일, 수정일: 2022년 11월 27일, 게재확정일: 2022년 12월 13일)

Assessment of the Prediction Derived from Larger Ensemble Size and Different Initial Dates in GloSea6 Hindcast

Ji-Yeong Kim*, Yeon-Hee Park, Heesook Ji, Yu-Kyung Hyun, and Johan Lee

Climate Research Department, National Institute of Meteorological Sciences, Jeju, Korea

(Manuscript received 19 October 2022; revised 27 November 2022; accepted 13 December 2022)

Abstract In this paper, the evaluation of the performance of Korea Meteorological Administration (KMA) Global Seasonal forecasting system version 6 (GloSea6) is presented by assessing the effects of larger ensemble size and carrying out the test using different initial conditions for hindcast in sub-seasonal to seasonal scales. The number of ensemble members increases from 3 to 7. The Ratio of Predictable Components (RPC) approaches the appropriate signal magnitude with increase of ensemble size. The improvement of annual variability is shown for all basic variables mainly in mid-high latitude. Over the East Asia region, there are enhancements especially in 500 hPa geopotential height and 850 hPa wind fields. It reveals possibility to improve the performance of East Asian monsoon. Also, the reliability tends to become better as the ensemble size increases in summer than winter. To assess the effects of using different initial conditions, the area-mean values of normalized bias and correlation coefficients are compared for each basic variable for hindcast according to the four initial dates. The results have better performance when the initial date closest to the forecasting time is used in summer. On the seasonal scale, it is better to use four initial dates, where the maximum size of the ensemble increases to 672, mainly in winter. As the use of larger ensemble size, therefore, it is most efficient to use two initial dates for 60-days prediction and four initial dates for 6-months prediction, similar to the current Time-Lagged ensemble method.

Keywords: Climate prediction system, GloSea6, Seasonal forecasting, Ensemble size

1. 서 론

단일 예보에서 앙상블 예보로의 패러다임 전환은 1992년에 여러 현업운영 센터의 전지구 앙상블 예측

시스템 구현과 함께 시작되었다(Steven Tracton and Kalnay, 1993; Buizza and Palmer, 1995). 단일 예보과 달리 앙상블 예측은 미래 대기 상태의 가능한 변동 범위를 나타내며(Vitart and Takaya, 2021), 앙상블 예측시스템은 수치예보 분야에서 가장 선진화된 기술로 효율적인 앙상블 예측을 통해 단일 예측에 비해 향상된 성능으로 예보 정확도를 개선할 수 있다. 앙상블 예측의 시작으로 정확도 높은 예보와 더불어 활용도 높은 예보 생산에 대한 요구가 지속적으로 증가되고 있다. 앙상블 예측

*Corresponding Author: Ji-Yeong Kim, Climate Research Department, National Institute of Meteorological Sciences, 33, Seohobuk-ro, Seogwipo-si, Jeju-do, 63568, Korea.
Phone: +82-64-780-6636, Fax: +82-64-738-9071
E-mail: jykim0810@korea.kr

시스템은 앙상블 생성 기법, 앙상블 크기, 예보 생산 일자 등에 따라 예측능력이 좌우될 수 있으며, 특히 앙상블 크기의 증가는 부족한 스프레드를 늘리고, 극단치를 감소시키며, 해상도 증가보다도 예측성 향상에 효과적이라고 알려져 있다(Buizza and Palmer, 1998; Mullen and Buizza, 2002).

이에 앙상블 확대에 따라 예측시스템 성능을 살펴본 연구들이 많이 진행되어 왔다. Kumar and Hoerling (2000)은 앙상블 모델을 이용해 계절예측 성능을 분석하여 앙상블 크기가 클수록 패턴상관계수가 증가함을 보였고, Palmer et al. (2004)에서는 열대지역 여름철 강수에 대한 예측능력을 앙상블 9멤버부터 54멤버까지 증가시켜 살펴보았을 때, 앙상블이 증가할수록 시스템의 신뢰성과 BSS (Brier Skill Score)가 개선된 바 있다. Scaife et al. (2014)은 1993년부터 2012년 겨울철을 최대 24개 멤버로 예측한 결과 앙상블 크기가 증가할수록 North Atlantic Oscillation 상관계수가 증가함을 보였다.

국내에서도 최근 앙상블 크기에 대한 예측능력을 분석한 연구가 지속적으로 이뤄지고 있는데, 특히 Hyun et al. (2020)에서는 기상청 기후예측시스템, GloSea5 (Global Seasonal Forecasting System version5)를 포함한 여러 기후예측모델 기후장 결과를 이용해 계절 확률예측 신뢰도를 평가하여, 계절내예측의 경우 초기시간 설정에 민감도가 높고 계절예측의 경우 앙상블의 크기가 예측성 향상에 더 중요한 요소임을 확인했다.

그러나 앙상블 멤버수의 확대는 자원과 비용에 대한 문제 또한 수반되고 있다. 현업 시스템에서 적분 가능한 앙상블의 크기는 전산 자원의 한계로 인해 대부분 작은 편이다. 따라서 전산 자원과 예측능력을 모두 고려해 적합한 앙상블 멤버 수를 찾는 것이 중요하다. Buizza and Palmer (1998)는 최소 2개에서 최대 32개의 앙상블 멤버로 ECMWF (European Centre for Medium Range Weather Forecasts)의 500 hPa 지위고도 예측성능을 분석했을 때 적어도 8멤버 이상에서 평균 제곱근 오차 (Root Mean Square Error, RMSE)의 개선효과가 컸음을 밝혔다. Kumar et al. (2001)은 0.4~0.5의 Signal-to-Noise Ratio일 때 10~20개 멤버의 앙상블 크기가 계절예측에 적합하며 더 큰 앙상블 개수일 때 도출될 수 있는 성능과 비슷한 수준임을 보인 바 있다. Park et al. (2021)에서는 GloSea5 앙상블을 56개까지 확대한 실험을 통해 현업 앙상블 12개 결과와 비교하였을 때, 멤버 수가 증가함에 따라 주요 대기변수의 결정론적 검증 결과가 개선되었고, 특히 평소에 변동성이 큰 영역에서 효과가 더 크게 나타났다. 그러나 초기시간의 개수와 조합에 따라 변수별, 검증방법별로 앙상블 결과의 예측성능이 다름을 보여 무조건적인 앙상블 확대보다는 적절한 개수를 찾아내는 것이 중요함을 시사했다. 또한, 계절내

예측에서 앙상블 수가 비슷할 때 시간지연 기법의 앙상블 생성이 다른 방법보다 이점이 있다는 연구 결과(Vitar and Takaya, 2021)를 통해 시간지연 방법이 예측성 향상을 도모하는데 효과가 있음을 밝혔다.

기상청 기후예측시스템은 대기-해양-해빙-지면모델이 결합된 앙상블 예측시스템으로 추계적 운동에너지 후방산란(Stochastic Kinetic Backscattering version2, SKEB2; Tennant et al., 2011)과 시간 지연 앙상블(time-lagged ensemble)을 통하여 과거 기후 예측장(hindcast)과 실시간 예측장(forecast)의 앙상블을 생산한다. 2021년 11월에는 기후예측시스템 GloSea5에서 6로 업그레이드하고, 2022년 2월부터 현업 운영 중에 있다(Kim et al., 2021a; Hyun et al., 2022). 따라서 본 연구에서는 계절내~계절 시간규모의 예측성능 개선을 위해 GloSea6의 앙상블 멤버 확대 효과를 면밀하게 살펴보고, 시간 지연기법에 대한 영향을 보고자 초기시간 변경 실험을 진행하여 예측 특성을 분석하고자 하였다.

2. 자료 및 분석 방법

2.1 자료

GloSea6에서는 SKEB2를 개선함과 동시에, 에너지와 수분이 보존되도록 추가적인 조치를 한 SPT (Stochastic Perturbation of Tendencies Scheme; Sanchez et al., 2016)를 채택하여 앙상블을 생산한다. 이전 버전보다 통계적으로 앙상블 평균의 에러가 증가하지 않고도 스프레드가 증가하는 효과가 있었고, 지역적으로는 적도에서 가장 긍정적인 영향이 있었다(Kim et al., 2021a). 기상청에서는 기후예측시스템을 GloSea6로 업그레이드 하면서 한 초기시간당 생산되는 추계물리 앙상블 수를 실시간 예측장은 4개에서 8개로, 과거기후 예측장은 3개에서 7개로 확대하여 시험생산해오고 있다. 이러한 앙상블 멤버 개수의 확대를 통해 부족한 앙상블 스프레드를 확보하고, 적은 최근의 초기시간만을 사용하여 시간 지연 기법의 앙상블 생산이 이루어져도 요구되는 앙상블 수를 충족할 수 있어 예측성 증가에도 유리하게 작용할 수 있다. GloSea6의 예측장(forecast)은 SKEB2와 SPT 기법을 함께 사용하여 한 초기 시간에 대하여 72일까지 4개, 240일까지 4개 총 8개의 앙상블로 생산되며, 과거기후 예측장(hindcast)은 한 초기 시간에 대하여 252일까지 7개의 앙상블이 생산된다.

한편, 초기시간 지연을 통한 시간 지연 앙상블 기법은 Fig. 1과 같이 예측장(forecast)의 경우 7일간의 초기시간으로 생산된 결과를 조합하여 60일(8주)까지의 예측에 활용하며, 21일간의 초기시간으로 생산된 결과가 6개월까지의 예측장 생산에 쓰이고 있다. 과거기후 예측장(hindcast)은 1993년부터 2016년까지의 기후기간 동안

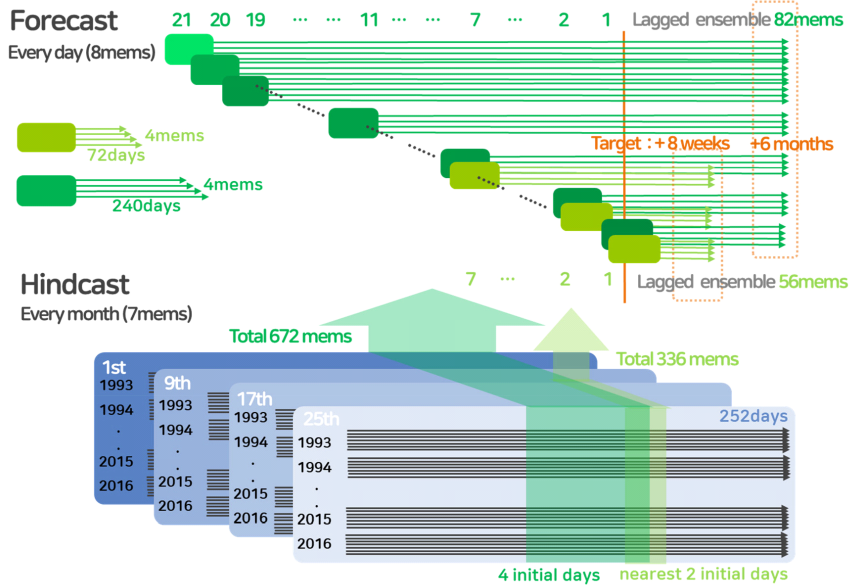


Fig. 1. Schematic diagram of the operational architectures for GloSea6 forecast and hindcast ensemble runs and time-lagged to produce one (six) month predictions.

1, 9, 17, 25일의 초기 시간에 대하여 60일, 6개월의 예측장을 생산한다. 4개의 날짜에 대해서만 자료를 생산하는 이유는 24년의 모든 날짜의 초기시간으로 앙상블을 생산하기에는 한계가 있어 전산자원량과 앙상블 수, 예측성능 사이의 적정 값을 설정하였다. 이렇게 생산된 과거 기후 초기시간 값은 현업에서의 실시간 예측 결과를 생산하는 날짜에 60일 예측에는 가까운 2개 초기값, 6개월 예측에는 가까운 4개 초기값에 가중치를 주는 방법으로 재생산하여 이용하고 있다(Park et al., 2021). 앞서 언급하였듯이 현재 GloSea6 앙상블 멤버 수를 예측장 8개, 과거기후 예측장 7개로 시험생산 중이지만, 현업에서는 각각 4, 3개만을 이용하여 장기전망 지원용 자료를 생산 중이며, 본 연구의 결과를 바탕으로 추후 생산되는 모든 앙상블 멤버를 사용하고 최적의 초기시간을 선택하여 사용할 예정이다.

분석에 이용된 변수는 1.5 m 기온, 해면기압, 500 hPa 지위고도, 해수면온도, 강수량, 850 hPa 바람장 예측자료이며, 검증을 위하여 동일한 기간에 대한 관측 자료로 1.5 m 기온, 해면기압, 500 hPa 지위고도, 850 hPa 바람장은 ERA-interim 재분석 자료(Dee et al., 2011), 해수면온도는 HadISST (Rayner et al., 2003), 강수량은 GPCP (Adler et al., 2018)를 이용하였다.

2.2 앙상블 확대 평가

3.1장에서는 GloSea6의 앙상블 확대 효과를 보기 위

해서 과거기후 예측장을 결정론적 검증과 확률론적 관점에서 정량적으로 평가해 보고자 한다. 60일, 6개월 예측에 이용되는 초기시간은 현업과 동일하게 각 2, 4개를 사용하였으며, 앙상블 멤버수를 3개에서 7개로 늘렸을 때의 확대 영향을 분석하였다. 따라서 과거재현 기후기간 24년(1993~2016년) 동안의 추계물리방법과 시간지연기법을 적용시킨 앙상블 멤버수는 60일 예측에 총 144개에서 336개, 6개월 예측에 총 288개에서 672개가 된다. 예측성능 검증은 먼저 앙상블 예측 시그널의 적정성 평가를 위해 예측가능요소비(Ratio of Predictable Components, RPC)를 살펴보고, 상관계수를 이용하여 예측성의 개선도를 평가하였으며, 확률론적 평가는 신뢰도 곡선(Weisheimer and Palmer, 2014; Hyun et al., 2020) 방법을 이용하였다.

RPC는 계절내-수년 규모에서 모델과 관측의 예측가능요소(Predictable Components, PC)를 비교하여 앙상블 예측 시그널 크기의 적정성을 평가하는 방법으로 PC는 총 분산에 대한 시그널 분산의 비($\sqrt{\sigma_{sig}^2/\sigma_{tot}^2}$)로 정의할 수 있다. 즉, 관측 대비 모델에서 노이즈 대비 상대적인 시그널 크기를 가늠하는 값으로, 관측만큼 앙상블 예측에서도 같은 크기의 시그널이 생산되어 RPC가 1일 때 가장 완벽한 예측성능을 가질 수 있다고 본다. RPC 값 분포는 지역별 예측성능에 따라 다르게 분포하겠지만, 앙상블 확대 전후의 RPC 비교를 통해 최선의 시그널의 확보를 위한 충분히 많은 앙상블이 생산되었는지 볼 수

있는 지수이다. RPC는 모델의 $\sqrt{\sigma_{sig}^2/\sigma_{tot}^2}$ 대비 상관계수 (r)의 비로 다음과 같이 계산한다(Eade et al., 2014; Dunstone et al., 2020; Smith et al., 2020).

$$RPC = \frac{PC_{obs}}{PC_{mat}} \geq \frac{r}{\sqrt{\sigma_{sig}^2/\sigma_{tot}^2}} \quad (1)$$

참고로 이렇게 계산된 값이 실제 RPC에 비해 작을 수 있다는 것은 향후 모델 개선 등으로 상관이 더 좋아질 수 있다는 의미를 가지고 있다.

확률 예측 신뢰도는 앙상블 예측에서의 3분위 발생 빈도(평년 이하, 평년 비슷, 평년 이상)가 관측에서의 빈도와 얼마나 일치하는지를 비교한다. 본 연구에서는 RD (Reliability Diagram) 검증 지수를 이용해 정량화하여 확률 예측 신뢰도를 평가하고자 하였다. RD를 나타내는 그래프에서는 앙상블 확률 예측(x축) 대비 실제 관측에서 나타난 빈도(y축)가 표시되며, 기울기가 1인 대각선과 정확하게 일치하면 완벽한 예측이 된다. 또한 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 BSS는 BS의 기후값을 reference로 비교하여, 모든 예측이 적중했을 때 1이 되도록 한 값이다. RD 그래프에서 가로 실선과 세로 실선은 기후적 발생 확률(33%)이며, 회색 음영은 BSS가 양의 값을 가지는 구간이고, 점선은 완벽한 예보를 나타내는 대각선과 기후적 발생 확률의 중앙 값이며, 기후적 발생 확률 대비 유용한 예측 인지의 여부를 판단하는 기준선(no skill line)이다. 즉, 확률 예측 결과가 회색 음영 구간(BSS > 0)에 들어가는 경우를 유용한 예측으로 판단한다. 이러한 방법은 그래프를 통해 가시적으로 예측 결과를 분석할 수 있는 장점이 있지만, 정량적으로 예측이 잘 맞았는지에 대한 검증

이 어렵다. 이러한 점을 보완하고자 RD를 통한 검증 결과를 5개 카테고리로 나누어 구간에 따른 예측성을 파악하고자 하였다. 먼저 기후값을 기준으로 하여 기온의 계절 예측 3분위를 구하고, 평년과 비슷한 범위는 연도별 자료를 순차적으로 배열한 후 표본의 수가 동일하게 되도록 3등분하여 중앙에 속하는 표본 값의 범위를 택하는 순위(rank) 방법을 이용하였다. 3분위의 각 분위별 예측 확률을 구하고, Bootstrap 기법을 이용해 무작위로 표본을 선택하여 선형회귀선(기울기)을 계산하는 과정을 1,000번 반복하였다. 이 과정을 통해 적정 선형회귀선과 오차 범위(75% confidence)를 구하고 앙상블 멤버의 스프레드를 고려하여 5개 신뢰도 등급(perfect, still useful, marginally useful, not useful, dangerously useless)으로 나누고(Table 1), 전지구의 지역별 신뢰도 지도를 제시하고자 하였다. 1등급 ‘Perfect Case’일 때 가장 높은 신뢰도를 가지며 5등급 ‘Dangerously Useless Case’일 때 가장 낮은 신뢰도를 갖게 된다.

2.3 앙상블 초기시간 평가

3.2장에서는 초기시간에 대한 예측성능을 평가하고자 과거기후 예측장의 경우에는 1-4개 날짜의 각 초기시간에 대한 예측성능을 편차와 상관계수를 이용하여 현업과 비교 분석하였다. Table 2와 같이 실험을 수행하였으며 최대 672개까지 멤버 수가 증가하였을 때 검증 결과를 분석하였다. 초기시간에 따른 분석은 앙상블 멤버 수는 과거기후 예측장 7개를 모두 사용하였다.

참고로 초기시간에 따른 예측성 평가에 앞서, 한달 간의 시간지연 초기시간을 앙상블 예측에 이용하는 것이 적절한지에 대해서는 선행연구에서 추계물리와 시간지연 간의 앙상블 스프레드 범위를 비교하여 적정성을 평가한 바 있어 이 연구에서는 평가하지 않았다(Park et al., 2021).

Table 1. 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 > 0.5
3	Marginally useful	The slope of the reliability line > 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

Adapted from Table 2 in Hyun et al. (2020) with permission of the Korean Meteorological Society. BSS, Brier Skill Score.

Table 2. Summary of ensemble experiment setting for the number of initial dates during GloSea6 hindcast period (1993–2016).

Hindcast	Operation	Experiment
1MON	2 Initial dates × 3 members × 24 years = 144 members	1 Initial date × 7 members × 24 years = 168 members (I1)
		2 Initial dates × 7 members × 24 years = 336 members (I2)
3MON (seasonal)	4 Initial dates × 3 members × 24 years = 288 members	3 Initial dates × 7 members × 24 years = 504 members (I3)
		4 Initial dates × 7 members × 24 years = 672 members (I4)

3. 결 과

3.1 앙상블 확대 효과 분석

3.1.1 앙상블 시그널 크기의 적정성 평가

Figure 2에서는 5월/11월의 4개 초기시간(1, 9, 17, 25 일)을 사용한 계절 예측에서 앙상블 확대에 따른 예측가 능요소비(RPC)의 비교 결과로 완벽한 예측을 의미하는 기준 값 1에 대한 차이의 절대값을 비교하였다. 앙상블 확대 시 전지구 RPC가 모든 변수에서 개선됨에 따라 비

교적 적정 시그널에 가까워짐을 확인할 수 있었으며, 개 선 효과는 주로 중고위도 중심으로 나타났다. 적도를 포 함하는 저위도에서는 RPC 측면에서 멤버 3개만을 이용 하여도 이미 충분한 예측능력이 확보되어 있기 때문에 앙상블 확대에 따른 영향이 크지 않은 것으로 보여진다.

영역별로는 중고위도에서 가장 크게 증가하고 주요 변수 중 기압장 변수(Z500, MSLP)들에 대하여 앙상블 확대에 따른 효과가 크게 증가하였다. 확대 전후의 앙상 블 스프레드를 RPC와 함께 비교한 결과, 앙상블 스프레 드가 증가한 영역에서 RPC가 개선됨을 알 수 있었다(그

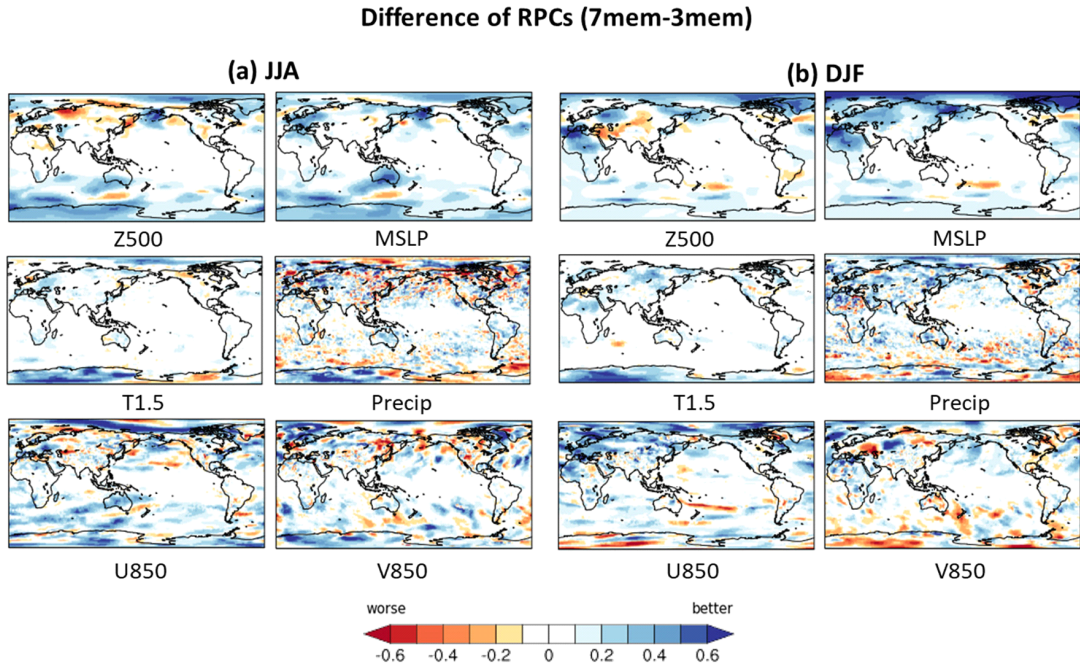


Fig. 2. Differences of Ratio of Predictable Components (RPC) between 7 ensemble members and 3 ensemble members of GloSea6 over global region during JJA (left) and DJF (right). The value represents the difference between the absolute values with 1 as the reference value.

림 생략). Park et al. (2021)에서도 GloSea5의 앙상블 수가 증가할수록 스프레드는 증가하며 특히 기압장 변수의 RPC가 1에 가까워짐을 확인한 바 있다. RPC가 더 나빠진 경우는 2가지로 RPC가 1보다 큰 지역에서 더 크게 벌어졌거나, 1보다 작은 지역에서 더 작게 감소한 경우이다. 그런데 현재 총 28개까지의 앙상블을 사용하였을 때는 RPC가 1보다 더 큰 경우는 나타나지 않아, 더 많은 앙상블이 사용되었으나, 분자의 상관성이 감소한 지역으로 볼 수 있다. 전지구적으로 RPC가 아직 1보다 작다는 것은 기후예측시스템에서 앙상블을 더 확대하여야

할 필요가 있다는 것을 의미한다. 상관 분포는 다음 절에서 자세히 살펴볼 것이다.

3.1.2 연변동 상관성 검증 결과

Figure 3은 현업과 같이 가까운 초기시간 2개(17, 25일)를 사용하여 6월/12월 한달을 예측했을 때와 초기시간 4개(1, 9, 17, 25일)를 사용하여 여름/겨울을 예측했을 때, GloSea6 과거기후 예측장(hindcast) 멤버 확대에 따른 관측과의 상관계수 차이를 나타낸다. 초록색이 진할수록 앙상블 멤버 확대 시 상관계수가 증가해 기후 변

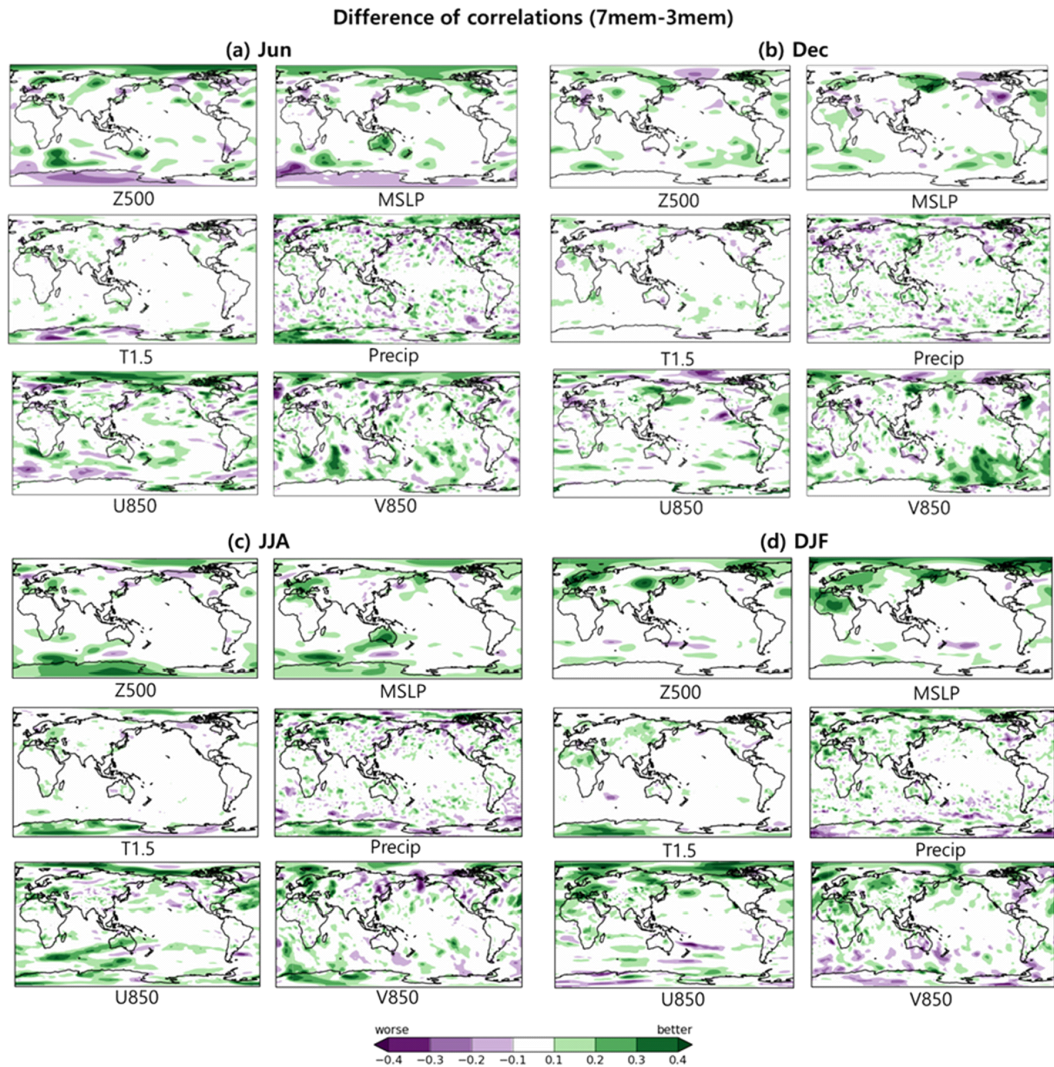


Fig. 3. Differences of correlation coefficients between 7 ensemble members and 3 ensemble members of GloSea6 over global region during June (upper left), December (upper right), JJA (bottom left), and DJF (bottom right). The correlation coefficients are calculated between GloSea6 hindcast and observation.

동성의 개선효과가 크게 나타남을 알 수 있다. 전반적으로 모든 변수에서 멤버 확대 시 상관계수가 증가하는 경향이 있었다. 주로 중고위도 중심으로 개선 효과가 뚜렷하게 드러나는데, 이는 평소 모델에서 예측 불확실성이 높았던 지역으로 앙상블의 확대가 이 지역의 기후 변동성 예측에 효과적임을 나타낸다. 6월 예측(Fig. 3a)에서 기압장 변수(Z500, MSLP)의 경우, 남극 대륙에서 상관계수가 다소 감소하는 패턴을 보이거나 호주와 북극 지역에서 0.2~0.4 정도의 상관도 증가를 보인다. 850 hPa 바람장 변수도 6월 예측에서 북극 지역의 개선효과가 크게 나타났다. 12월 예측(Fig. 3b)에서는 특히 850 hPa 남북바람의 남반구 해양에서 상관계수 0.3 이상의 증가가 크게 나타났다. 계절예측 그림(Figs. 3c, d)은 Hyun et al. (2022)의 Fig. 10에서도 같은 조건의 계절예측 결과를 4가지 변수에 대하여 살펴본 바 있으며, 여름철 남극대륙과 겨울철 북극 지역의 상관계수가 뚜렷하게 증가하는 패턴이 대부분 변수에서 공통으로 보였다. 특히 Z500, MSLP, U850 변수에서 개선되는 패턴이 강하게 나타났다. 이는 선행연구에서 GloSea5 앙상블 멤버를 14개로 확대하여 실험했을 때 여름과 겨울의 극지역에서 RMSE 결과가 크게 개선되었고, 500 hPa 지위고도와 해면기압

의 개선율이 가장 크게 나타난 것과 유사하다(Park et al., 2021). 그러나 선행연구 결과와 달리 저위도 강수량의 개선은 크지 않아 더 많은 앙상블 개수의 필요성을 시사한다. 기후 평균 오차에서는 전지구적으로 멤버 확대의 개선효과가 크게 나타나지 않는다(그림 생략).

Figure 4는 Fig. 3의 6월/12월 1개월 예측의 상관 차이를 동아시아 지역(10~55°N, 100~160°E)으로 확대하여 본 그림이다. 전지구적으로는 전반적인 개선 경향을 볼 수 있었으나, 동아시아 지역으로 영역을 좁혀 보았을 때에는 개선되는 지역과 그렇지 못한 지역이 혼합하여 존재한다. 그러나 12월 예측에서 앙상블 멤버 확대 효과가 더 잘 나타나며, 850 hPa 남북바람을 제외하고는 모든 변수에서 한반도 지역의 기후 변동성이 개선되는 모습을 보인다. 특히 기압장에서는 Z500의 강한 trough가 위치한 영역과 해면기압의 알루산 저기압이 위치한 지역에서 상관이 개선되는 효과를 보이고 있으며, 한반도 북쪽 대륙의 1.5 m 기온과 강수 또한 관측과의 상관계수가 증가하는 경향을 보여, 동아시아 겨울 몬순의 기후 변동성 개선에도 앙상블 확대의 효과가 나타남을 기대해 볼 수 있다. 6월 예측의 Z500에서는 앙상블 확대 시 한반도 부근의 지위고도가 편차 및 상관 측면에서 모두

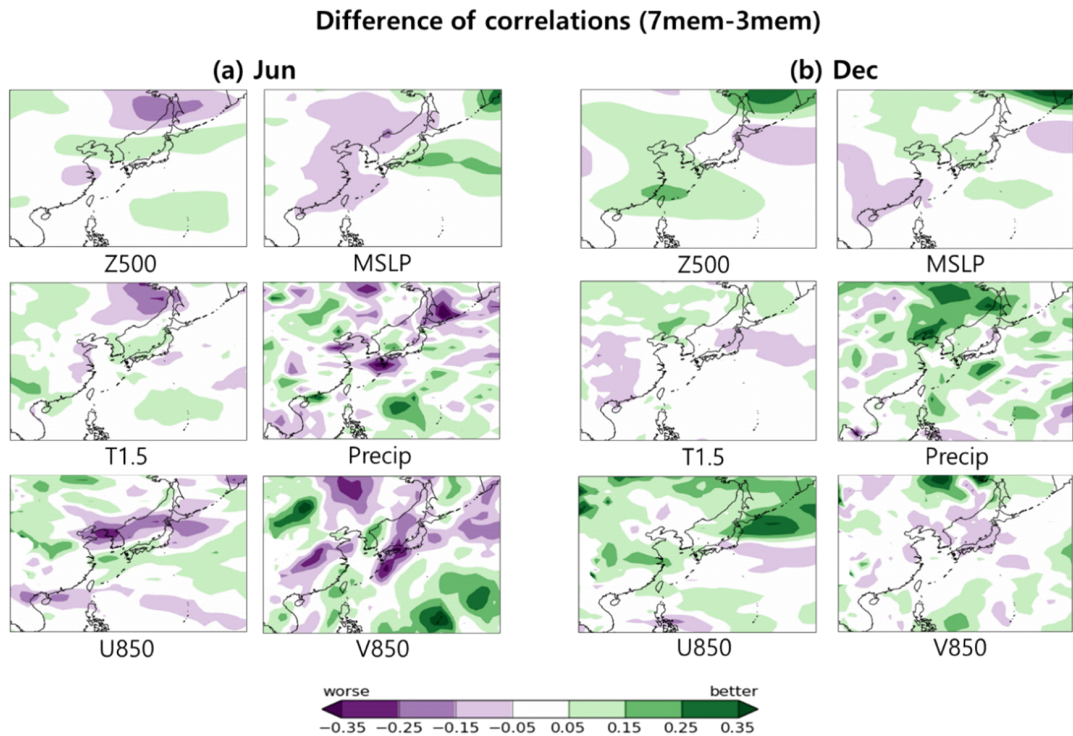


Fig. 4. Same as Fig. 3 except for East Asia region.

개선되었으나, 해면기압에서는 대륙 지역의 상관이 다소 감소하고, 일본 동쪽 해역의 상관은 증가하는 패턴을 보인다. 850 hPa 바람에서는 멤버 확대 시 한반도의 동서바람과 일본의 남북바람이 관측과의 상관계수가 감소하는 모습을 보인 반면, 북서태평양 해역의 850 hPa 남북바람은 기후변동성에서 큰 개선효과를 보인다.

계절예측의 상관 차이를 동아시아 지역(10~55°N, 100~160°E)으로 확대하여 본 Fig. 5는 여름철 기압장 변수(Z500, MSLP)들과 1.5 m 기온의 경우 한반도에서 0.1~0.2 정도의 상관도 증가를 보이고, 강수와 850 hPa 바람장 변수들의 경우 지역적으로 앙상블 확대 효과가 다르게 나타남을 알 수 있다. 특히 여름철 동아시아 몬순에 밀접한 영향을 주는 북서태평양 해역에서 850 hPa 바람장이 관측과의 상관이 증가하면서 앙상블 확대를 통해 북서태평양의 기후변동성 모의가 개선될 수 있음을 보인다. 겨울철 예측에서는 모든 변수에서 앙상블 확대 시 동아시아 지역의 상관계수 개선이 나타난다. 또한 겨울철 예측이 1개월 예측보다 더 뚜렷하게 앙상블 확대에 대한 개선 효과가 나타났다. Z500의 경우, 시베리아 고기압이 위치한 영역에서 0.4 이상의 상관도가 크게 증가하였고, 겨울철 한랭 이류에 중요한 850 hPa 바람

장도 대륙에서 0.2 이상의 개선효과를 보이며, 겨울철 기후 변동성 모의 개선의 가능성을 보여준다. 또한 Z500은 동일본 해역에서 평균 오차가 2 m 이상 개선되어(그림 생략) 앙상블 확대 시 겨울철의 강한 trough 발달을 모의하는 데도 효과가 있음을 알 수 있다.

3.1.3 확률 예측 신뢰도 분석

GloSea6의 hindcast 앙상블 확대가 확률 예측에 미치는 영향을 보기 위해 확률 예측 신뢰도 분석을 수행하였다(Fig. 6). 신뢰도 분석에서는 기후기간에 대하여 초기 시간당 앙상블 3개와 7개의 24년치 앙상블, 총 288개와 672개를 이용한 결과를 비교하였다. Figure. 6d는 RD를 나타내는 그래프로 2.2절에 자세히 설명하였다.

관측 기온이 평년보다 추운 것/더운 것으로 예측한 여름(cold/warm JJA)과 겨울(cold/warm DJF)에 대하여 21개 영역으로 나누어 신뢰도 등급을 산출한 결과(Figs. 6a, b), 전반적으로 겨울철에 비해 여름철 기온 예측의 신뢰도 등급이 높게 나타났고, 앙상블 확대 효과가 더욱 긍정적으로 나타났다. 이렇게 나타나는 이유는 유라시아 대륙의 겨울철 예측 신뢰도가 좋지 못한 것이 주요하다. 여름철의 경우 앙상블 확대 시 추운 여름 예측의 북

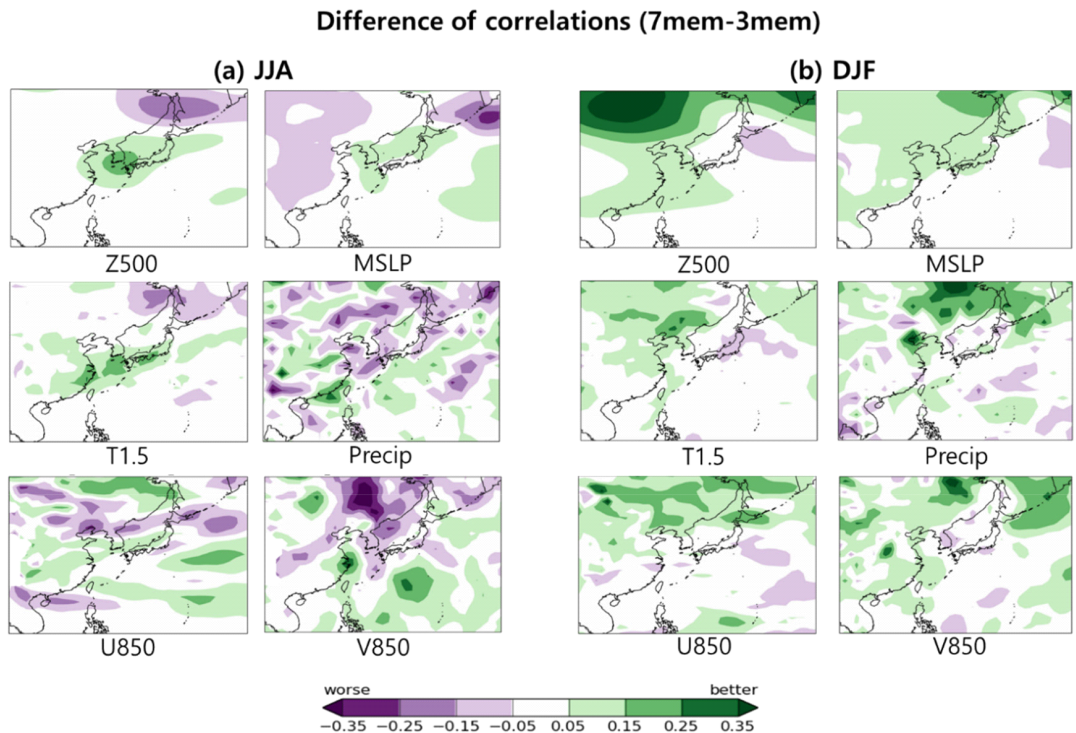


Fig. 5. Same as Fig. 3 except for East Asia region during JJA (left) and DJF (right).

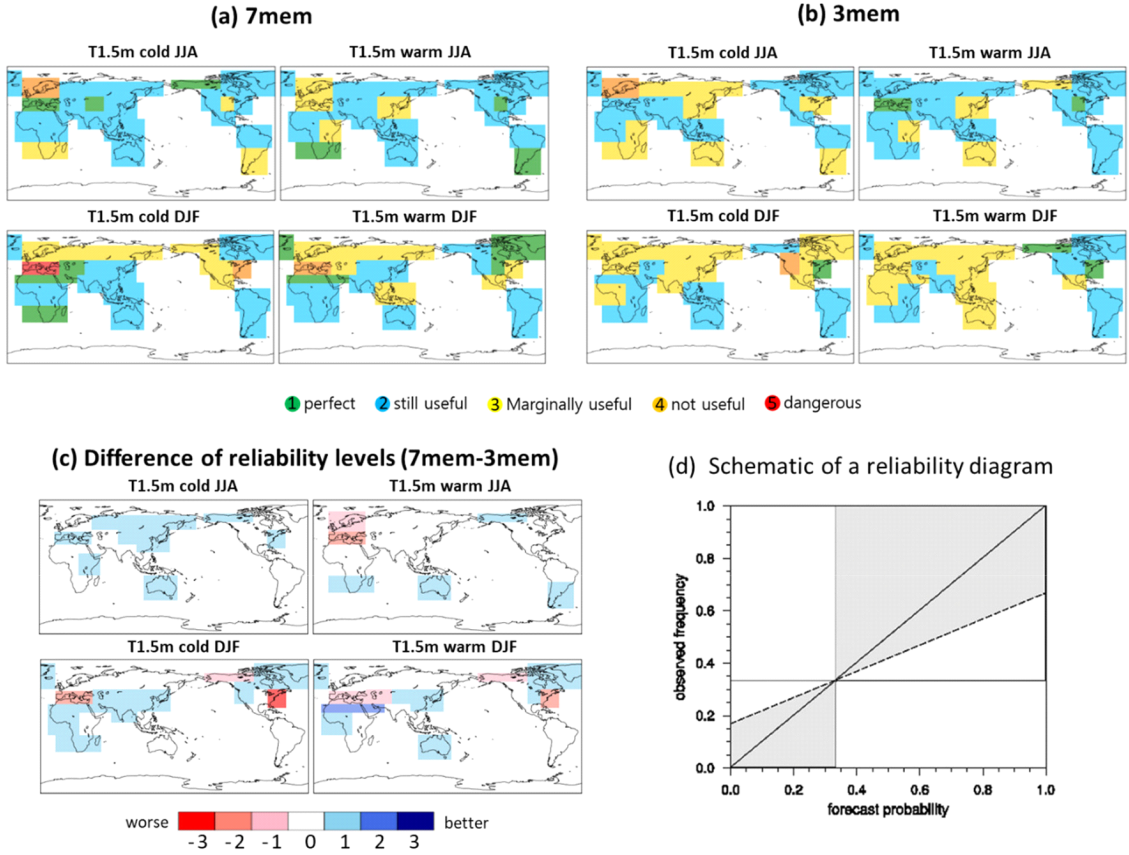


Fig. 6. The 1.5 m temperature reliability map of (a) 7 members and (b) 3 members on cold JJA (upper left), warm JJA (upper right), cold DJF (lower left) and warm DJF (lower right). Green, blue, yellow, orange and red colors are shaded based on reliability categorization from perfect to dangerous forecast. The difference of reliability levels between 7 members and 3 members is (c). The schematic of a reliability diagram is shown as (d).

유럽을 제외한 전 지역에서 3등급 이상의 좋은 신뢰도를 가지며, 예측 신뢰도 등급의 향상은 더운 여름에 비해 신뢰도가 좋지 못했던 추운 여름 예측에 두드러지게 나타났다. 추운 여름은 8개 영역에서 신뢰도가 향상됐으며, 더운 여름은 4개 영역에서 향상되었다(Fig. 6c). 선행 연구에서는 ECMWF의 30년 기후 기간에 대하여 9, 15, 51개의 앙상블을 이용하여 앙상블 수에 따른 확률 예측 신뢰도 분석을 수행한 결과, 겨울철에 비해 여름철의 신뢰도가 높게 평가됐으며, 앙상블 확대에 따라 신뢰도가 증가함을 보인 바 있다(Hyun et al., 2020).

겨울철은 일부 신뢰도가 감소하는 영역이 존재하였으나 많은 영역에서 신뢰도가 개선되었고, 앙상블 확대 효과는 추운 겨울과 더운 겨울이 유사하였다. 앞서 언급하였듯 유라시아 대륙의 신뢰도가 좋지 못한 편인데, 앙상블을 증가시켰을 때 내륙의 남쪽에 개선효과가 있었

다. 추운 겨울과 더운 겨울에서 7~9개 영역에서 신뢰도가 증가하고, 3~4개 영역에서 신뢰도가 감소하였다. 한편, 북유라시아 대륙에서는 앙상블 확대 전후의 신뢰도 등급이 모두 평균적으로 3등급을 나타내면서 확률 예측 신뢰도가 향상되는 효과가 나타나지 않았다. Hyun et al. (2020)에서는 기후예측시스템(GloSea5) 등 세계 현업 센터들의 운영시스템의 확률 예측 신뢰도를 평가한 결과, 기후 기간과 초기장 수, 앙상블 수로 구성된 표본이 클수록 신뢰도가 뚜렷하게 구분되고, 적을수록 중간 등급인 3등급 수준의 신뢰도로 편중되는 경향이 있음을 보인 바 있다. 이러한 결과를 고려해 본다면 동일한 기후 기간과 초기장 수를 이용한 여름철 확률 예측에 비해 겨울철 북유라시아 영역에서 더 많은 앙상블 수가 필요한 것으로 사료된다. 우리나라를 포함한 동아시아 지역도 더운 여름을 제외한 다른 경우에 모두 앙상블 확대에

따라 신뢰도가 개선되는 모습을 보였다.

3.2 초기시간 변화에 따른 예측성능 평가

Figure 7은 GloSea6 과거기후 예측장(hindcast) 앙상블 멤버 및 초기시간에 따른 비교를 변수별로 한번에 나타낸 그림이다. 원의 색깔은 7멤버와 3멤버의 절댓값 편차 차이를 영역 평균한 값이다. 변수마다 값의 절대적 크기가 다르기 때문에 모든 변수를 한 눈에 비교하기 위해 각각의 관측 표준편차로 표준화하였다. 원의 반지름은 7멤버와 3멤버의 상관 차이를 영역 평균한 값이며, 삼각형은 음의 값을 의미한다. 영역평균 지역은 전지구와 동아시아로 나누어 분석하였다. 도형의 초록색이 진할수록 7멤버가 3멤버보다 편차 측면에서 개선됨을 의미하며, 원의 반지름이 크고 삼각형이 없으면 7멤버가

상관 측면에서 개선됨을 의미한다. 사용한 초기시간은 I1부터 I4까지 있으며, x축에 나타냈다. I1은 예보시점에서 가장 가까운 초기시간 1개(25th)를 사용하고, I2는 2개(17, 25th), I3는 3개(9, 17, 25th), I4는 4개(1, 9, 17, 25th)의 초기시간을 사용하는 것을 의미한다. 앞의 4열은 1개월 예측, 뒤의 4열은 계절 예측을 나타낸다. y축은 분석에 사용한 7개의 주요 변수이며, 가장 오른쪽 3개의 원들은 상관차이 크기의 예시를 보여주어 가장 큰 원의 크기일 때 영역평균된 상관계수가 0.15 정도 개선됨을 말한다.

여름철 전지구 평균은 전체적으로 멤버 확대 시에 3 멤버보다 개선된 결과를 보였다. 1개월 예측에서는 예보 시점에 근접한 초기시간 1~2개(I1~I2)를 사용한 것이 편차와 상관 모두 개선된 결과를 보이며, 3개월 예측에서

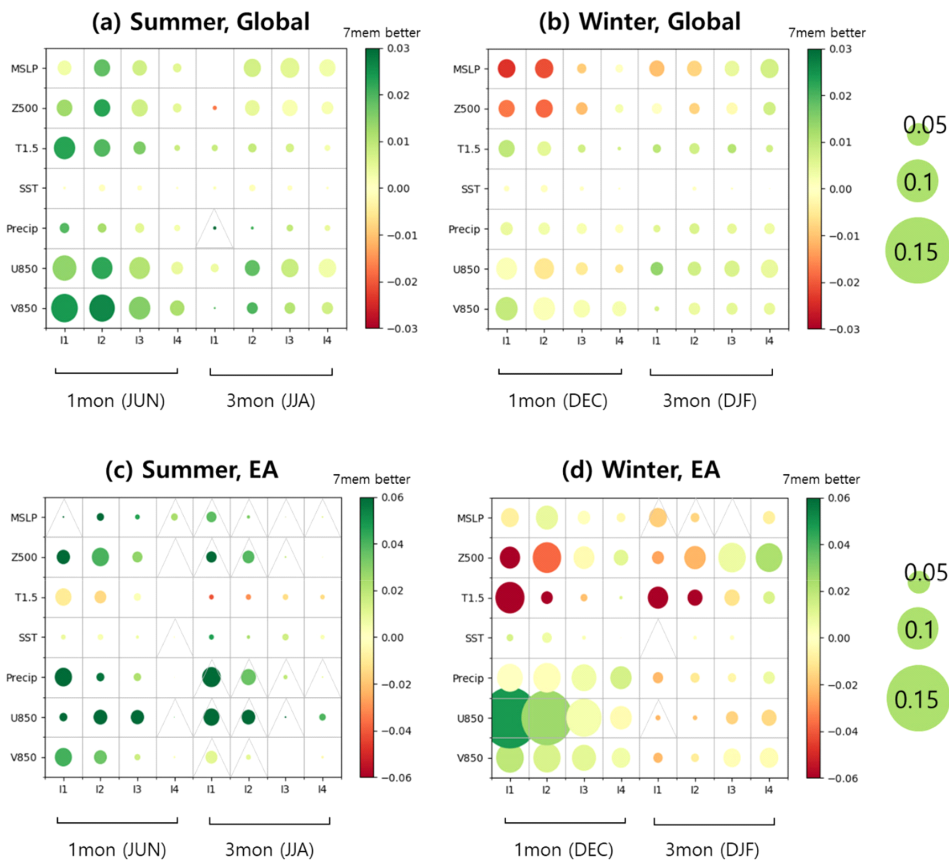


Fig. 7. Difference of correlation coefficients and normalized bias for GloSea6 hindcast ensemble expansion and initial dates over global (top) and East Asia (bottom) during (a,c) summer and (b,d) winter. Colors indicate area-mean value for difference of normalized bias between the results of 7 and 3 ensemble members. Radius of circle indicates area-mean for difference of correlation, and a triangle means negative value. I1~I4 on x-axis means initial condition as indicated in Table 1.

는 초기시간 2개(I2)를 사용했을 경우 가장 개선된 효과를 보인다. 특히 I1을 사용한 1개월 예측성능을 보았을 때 전반적으로 모든 변수의 여름 편차가 크게 줄어드는데, 앞서 상관계수가 중고위도 지역에서 주로 개선되었던 것처럼 편차 또한 북극 지역과 남반구 고위도 해양에서 감소하는 패턴이 주로 나타남을 확인했다. 겨울철 전 지구 평균 결과에서 MSLP 및 Z500은 멤버 확대 시 가까운 초기시간을 사용할수록 편차가 증가하는 경향을 보이며, 북반구 고위도 및 북극 지역에서 뚜렷하게 나타났다. I4일 때 1개월과 3개월 예측 모두 편차가 개선되는 모습을 보인다. 나머지 변수들은 대체로 1개월 예측에서 I1, 계절 예측에서 I4의 초기시간을 사용하였을 때 멤버 확대의 개선 효과가 가장 좋았다.

동아시아 지역(10~55°N, 100~160°E)을 영역 평균한 결과에서 1.5 m 기온을 제외한 여름철 변수들은 멤버 확대 시 편차가 개선된 효과를 보이며, 특히 강수와 850 hPa 동서바람에서 개선효과가 크게 나타났다. 상관 측면에서는 앞서 Figs. 4a, 6a에서 보았던 멤버 확대가 동아시아의 지역적 계절 예측에 개선효과가 뚜렷하지 못한 것으로 보인다. 전체적으로 여름 1개월 예측의 경우, I1~I2의 초기시간 사용 시 개선효과가 크고, 계절 예측에서 편차는 I1, 상관은 I4일 때 효과가 좋은 편이다. Hyun et al. (2022)은 예보시점과 가까운 25일 초기시간 하나만을 사용했을 때 여름철 1개월 예측에서 동아시아의 해면기압, 850 hPa 동서바람, 강수량의 편차가 개선되는 경향을 보인 바 있다. 본 논문에서도 가장 편차가 눈에 띄게 줄어드는 I1의 초기시간을 사용한 1개월 예측

의 동아시아 지역 편차 공간장을 살펴보았을 때, 여름 1개월의 경우 대부분 변수에서 동아시아 지역에 편차가 개선되는 패턴을 보이며, 선형연구와 유사한 결과를 나타냈다(Fig. 8a). 특히 북서태평양 지역에서 해면기압과 850 hPa 동서바람의 편차 감소가 두드러지는데, 이는 동아시아 여름 몬순의 대표적인 순환패턴인 북서태평양 고기압과 관련되어 동아시아 여름철 장마 발생에 매우 큰 영향을 미치는 것으로 알려져 있어(Kim et al., 2021b), 가까운 초기장을 썼을 경우 1개월 예측에서 여름철 동아시아 몬순 예측성의 개선이 기대된다. 겨울의 경우, 동아시아 영역의 500 hPa 지위고도장 및 기온은 예보시점과 가까운 초기시간을 사용했을 때 7멤버의 편차가 크게 나타나는 특징을 보여(Figs. 7d, 8b), I4의 초기시간을 사용하는 것이 효과적으로 보인다. 반대로 850 hPa 동서바람 및 남북바람은 1개월 예측에서 가까운 초기시간을 사용했을 때 편차 및 상관의 개선효과가 가장 크게 두드러졌다. 겨울철 동아시아의 3개월 예측은 모든 변수에서 대체로 I4의 초기시간 사용 시 개선효과를 보인다.

요약하자면, 과거기후 예측장의 멤버 확대는 전지구적으로 편차와 상관을 좋아지게 하나, 동아시아는 변수마다 차이가 있으며, 특히 850 hPa 바람에서 개선 효과가 크게 나타난다. 예보시점과 가까운 초기시간을 썼을 때 효과는 계절규모보다 계절내규모에서 좋은 경향을 보인다. Park et al. (2021)의 GloSea5 확대 실험 결과에서는 두 계절의 대부분 변수가 가까운 초기시간 2개를 사용하였을 때 예측성능이 좋았고, 4개월 때 가장 좋지 않게 모의되었는데, 본 연구에서는 여름철에 계절내규

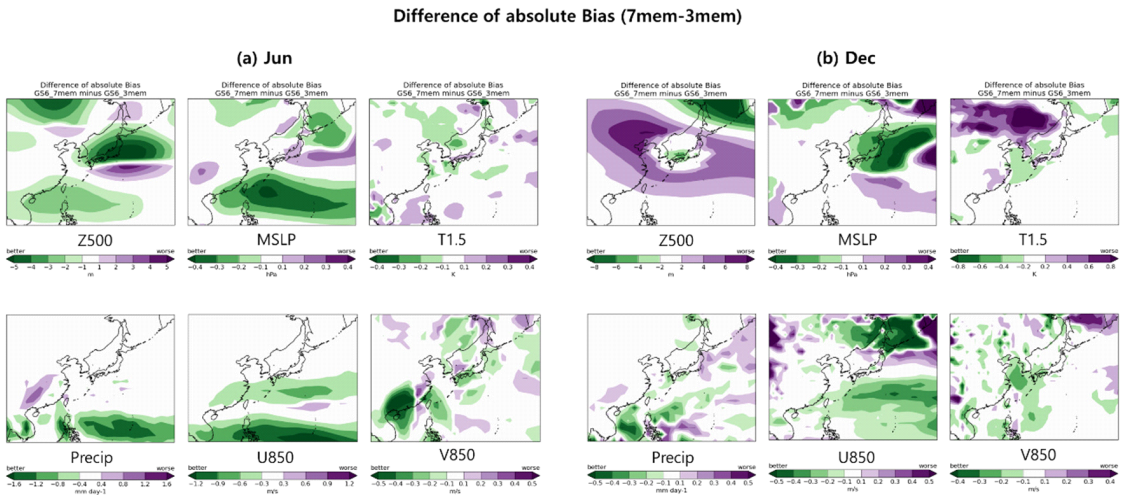


Fig. 8. Differences of absolute bias between 7 ensemble members and 3 ensemble members of GloSea6 over East Asia region during (a) June and (b) December when it used the 25th initial date (I1).

모와 계절예측 모두 가까운 초기시간을 사용했을 때 멤버 확대에서 개선되는 효과를 보이고, 겨울철은 계절예측 시 4개의 초기시간을 사용하는 것이 적절할 것으로 판단되어 모델과 계절에 따른 차이가 나타났다.

4. 요약 및 토의

본 연구에서는 기상청 기후예측시스템 GloSea6의 앙상블 확대와 초기시간에 대한 영향을 다각도로 분석하였다. 먼저 앙상블 확대에 따라 적정 시그널 크기에 가까워짐을 확인하였고, 상관 측면에서 보았을 때 전반적으로 모든 변수에서 기후변동성에 대한 예측성능 향상을 보였으며, 중고위도에서 효과가 크게 나타났다. 중고위도 지역은 평소 모델에서 불확실성이 높았던 지역으로 앙상블 확대를 통해 모델의 모의성능이 개선될 수 있음을 시사한다. 동아시아 지역에서는 특히 겨울철에 여름보다 상관계수 값이 크게 증가하여 앙상블 확대 시 겨울철 동아시아 기후변동성 모의의 개선 가능성을 보여준다.

확률 예측 신뢰도 분석에서는 여름철과 겨울철 모두 전반적으로 신뢰도가 향상됨을 보였고, 영역별로 차이가 있지만 겨울철에 비해 여름철 신뢰도의 개선 효과가 더 크게 나타났다. 이렇게 나타나는 이유는 유라시아 대륙의 겨울철 예측 신뢰도가 좋지 못한 것이 주요하였는데, 앙상블 확대에 따라 유라시아 내륙의 남쪽에 개선효과가 나타났으나, 앙상블 확대에도 시베리아 북쪽의 신뢰도는 여전히 개선되지 못하고 유지되었다. 여름철의 경우에도 추운 여름의 신뢰도는 이 지역에서 좋지 못하였지만, 앙상블 확대 영향으로 한 등급씩 향상되는 효과를 보였다. 우리나라를 포함한 동아시아 지역도 더운 여름을 제외한 다른 경우에 모두 앙상블 확대에 따라 신뢰도가 개선되는 모습을 보였다.

또한 확대 생산으로 확보한 앙상블 수를 가지고 초기시간의 변화를 주었을 때의 예측성 개선도를 평가하였다. 전반적으로 계절내 예측에서는 가까운 초기시간을 사용하는 것이, 계절 예측에서는 초기시간을 모두 사용하여 최대한 많은 앙상블이 확보되는 것이 좋은 결과를 보였다. 다만 초기시간을 줄였을 때 계절내 예측에 있어 특히 겨울의 기온과 기압 변수의 편차가 증가하는 경향이 있었다. 초기시간 변화에 따른 민감도는 850 hPa 바람장에서 가장 크게 나타났으며, 가까운 초기시간 1~2 개만을 사용했을 때 편차 및 상관계수에서 개선효과가 가장 좋게 나타났다.

종합적으로 판단해 보면 60일 예측에서는 앙상블 확대 후 초기시간을 줄여 예측성을 확보할 수 있을 것으로 보이고, 계절 예측에서는 확대된 초기장을 모두 사용하는 것이 가장 효율적일 것으로 판단되었으나, 모든 경우

좋아지는 것은 아니었으며, 앙상블이 좀 더 확대될 경우 예측성이 개선될 수 있는 여지가 남아 있었다.

WMO의 계절내~계절 예측성 향상 프로젝트에서는 앙상블 확대에 따른 영향을 서브 프로젝트로 하여 연구를 진행하고 있으며(<http://s2sprediction.net>), GloSea를 현업으로 같이 사용하는 영국기상청의 경우도 현재 앙상블을 3배 이상 증가시킬 계획에 있다. 호주기상청 또한 자체 자료동화시스템을 구축하여 이미 99개 앙상블을 생산하여 예보에 활용 중에 있다(<http://www.bom.gov.au/climate/ahead/about/model/access.shtml>). 한국 기상청 또한 2022년 8월부터 GloSea6 앙상블 멤버를 확대 적용시켜 기상청 장기전망에 활용하고 있으며, 분석을 위한 기간이 충분히 더 확보된다면 다양한 사례로부터 예측장(forecast)의 확대된 앙상블 효과를 추가적으로 확인할 수 있을 것이다. 또한 초기시간 선정에 따른 영향 평가를 위해 추후 가까운 초기시간으로 하루에 모든 앙상블을 생산하는 실험과 함께 현업 대비 RMSE뿐만 아니라, BS, Continuous Ranked Probability Score 등 결정적/확률적 측면에서 다양한 예측성능 분석을 진행할 예정이다.

감사의 글

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

REFERENCES

- Adler, R. F., and Coauthors, 2018: The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, doi:10.3390/atmos9040138.
- Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434-1456.
- _____, and _____, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503-2518.
- Dee, D. P., and Coauthors, 2011: The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, **137**, 553-597, doi:10.1002/qj.828.
- Dunstone, N., and Coauthors, 2020: Skilful interannual climate prediction from two large initialised model ensembles. *Environ. Res. Lett.*, **15**, 094083, doi:10.1088/1748-9326/ab9f7d.
- Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do season-

- al-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620-5628, doi:10.1002/2014GL061146.
- Hyun, Y.-K., J. Park, J. Lee, S. Lim, S.-I. Heo, H. Ham, S.-M. Lee, H.-S. Ji, and Y. Kim, 2020: Reliability assessment of temperature and precipitation seasonal probability in current climate prediction systems. *Atmosphere*, **30**, 141-154, doi:10.14191/Atmos.2020.30.2.141 (in Korean with English abstract).
- _____, and Coauthors, 2022: The KMA global seasonal forecasting system (GloSea6) - part 2: climatological mean bias characteristics. *Atmosphere*, **32**, 87-101, doi:10.14191/Atmos.2022.32.2.087 (in Korean with English abstract).
- Kim, H., J. Lee, Y.-K. Hyun, and S.-O. Hwang, 2021a: The KMA global seasonal forecasting system (GloSea6) - part 1: operational system and improvements. *Atmosphere*, **31**, 341-359, doi:10.14191/Atmos.2021.31.3.341 (in Korean with English abstract).
- Kim, J.-Y., Y.-K. Hyun, J. Lee, and B.-C. Shin, 2021b: Assessment on the East Asian summer monsoon simulation by improved global coupled (GC) model. *Atmosphere*, **31**, 563-576, doi:10.14191/Atmos.2021.31.5.563 (in Korean with English abstract).
- Kumar, A., and Hoerling M. P., 2000: Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction. *Bull. Am. Meteorol. Soc.*, **81**, 255-264.
- _____, A. G. Barnston, and M. P. Hoerling, 2001: Seasonal predictions, probabilistic verifications, and ensemble size. *J. Clim.*, **14**, 1671-1676.
- Mullen, S. L., and R. Buizza, 2002: The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system. *Wea. Forecasting*, **17**, 173-191.
- Palmer, T. N., and Coauthors, 2004: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.*, **85**, 853-872.
- Park, Y.-H., Y.-K. Hyun, S.-I. Heo, and H.-S. Ji, 2021: Assessment of the prediction performance of ensemble size-related in GloSea5 hindcast data. *Atmosphere*, **31**, 511-523, doi:10.14191/Atmos.2021.31.5.511 (in Korean with English abstract).
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407.
- Sanchez, C., K. D. Williams, and M. Collins, 2016: Improved stochastic physics schemes for global weather and climate models. *Q. J. R. Meteorol. Soc.*, **142**, 147-159, doi:10.1002/qj.2640.
- Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514-2519, doi:10.1002/2014GL059637.
- Smith, D. M., and Coauthors, 2020: North Atlantic climate far more predictable than models imply. *Nature*, **583**, 796-800, doi:10.1038/s41586-020-2525-0.
- Steven Tracton, M., and Kalnay, E., 1993: Operational ensemble prediction at the National Meteorological Center: practical aspects. *Wea. Forecasting*, **8**, 379-398.
- Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190-1206, doi:10.1175/2010MWR3430.1.
- Vitart, F., and Y. Takaya, 2021: Lagged ensembles in sub-seasonal predictions. *Q. J. R. Meteorol. Soc.*, **147**, 3227-3242, doi:10.1002/qj.4125.
- Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. R. Soc. Interface*, **11**, 1742-5662, doi:10.1098/rsif.2013.1162.