

기후예측모델 개선을 위한 예측성 평가 및 가이드라인 개발

**Forecast evaluation and development of
guidelines to improve climate prediction models**

김가영, 김해정, 이운영, 정여민, 함수련

기후사업본부 예측기술과

2020.12.

APEC기후센터

평문 초록

본 과제는 기상청 기후예측모형(GloSea5) 개선을 위한 계절내 시간 규모에 대한 현업 예측성 평가와 모형 특성 진단을 목적으로 기획되었다. 현업 예보에 대하여 객관적이면서도 일관되며 여러 측면을 검토하되 그것들을 종합적으로 아우를 수 있는 평가 시스템의 기틀을 마련하고자 우선, 기후값, 실시간 예측, 기후 리짐, 역학 커플링 네 부문에 대한 필수 평가 항목들을 발굴하고, 콘텐츠 구조화를 통해 평가프레임(CREDYT: Comprehensive and Relative Evaluating weekly mean DYnamical forecasT)을 정립하였다. 또한, 타 기관 대비 기상청 현업 예측의 상대적 성능 및 지구시스템간 연결 강도 등에 대한 기존과 다른 새로운 형태의 검증·진단 방법론을 고안하였다. CREDYT에 따르면, 2019년 52주 지위고도 주평균 예측의 경우, 5개 모형 중 2위 정도의 우수한 상대적 성능을 보이고, MJO를 비롯한 기후 지수도 적절하게 모의되고 있다. 하지만, 기온 평년값 오차가 북반구 대륙과 남극에서 뚜렷하고, 지구시스템간 접합강도가 계절별로 과소 혹은 과대 모의되거나 대기-해양, 대류권-성층권 간 역학적 커플링이 관측과 다른 형태로 재현되는 등 접합모형으로서의 구조적인 오차 특성을 확인할 수 있었다. 추가적으로, 기존 현업 기후값 생산방안을 변형한 민감도 평가를 수행하여 선행 날짜 가중치 없이 기후장 날짜수를 늘리는 개선안도 제안하였다.

초기 오차 성장 특성을 살펴보기 위해 과거재현기후 및 실시간예측장의 선행시간별 계절내 예측 성능의 분석을 수행하였다. 또한, 실시간 예측장에서의 초기장 차이가 계절내 예측 성능에 미치는 영향에 대해 살펴보기 위해 초기 조건을 고해상도 대기모형 예측장과 재분석 자료를 사용한 민감도 실험을 구성하여 분석하였다. 이를 통해, 대기모형 예측장을 초기장으로 사용할 때 보다 재분석자료를 초기장으로 사용할 때 동아시아 겨울철 기온 및 순환장의 계절내 예측 성능 및 계절내 열대 진동의 모의 성능이 효과적으로 개선됨을 밝혔다. 결론적으로, 계절내 규모에서의 오차의 감소를 위해 지면 오차 및 구름/복사와 연관된 초기 조건들의 개선이 필요함을 알 수 있다. 또한, 이 연구에서 사용된 계절내 열대 진동성 진단 방법은 준자동화가 가능한 틀로 구성하였으며, 추후 지속적이고 일관성 있는 기후예측모형 진단에 활용할 수 있을 것으로 기대된다.

정확도 높은 예보와 더불어 활용도 높은 예보 생산에 대한 요구가 지속적으로 증가되고 있지만 이는 자원과 비용에 대한 문제 또한 수반한다. 따라서 한정된 자원과 비용 안에서 가장 효율적인 기상청 장기예보 시스템을 운영할 수 있는 전략을 이 연구에서 발굴하였다. 예보 시스템은 앙상블을 생산하는 방식에 따라 예보의 정확도가 달라질 수 있는데 이 연구에서는 앙상블 수의 증가, 앙상블 생산 요일, 앙상블 생성 기법에 변화를 준 민감도 실험을 통해서 기존 현업 예보 시스템에 비해 더 나은 동아시아 기온을 생산할 수 있는 가이드라인을 도출하였다. 기본적으로 기상청 장기예보 시스템은 8주 이내의 동아시아 기온 예보에 있어서 일관되고 신뢰도가 높은 자료를 생산하고 있으나 그 중 3,4주 차는 개선이 필요하다. 이 기간에 대해서 앙상블의 수를 증가시키면 현업 시스템 대비 최대 약 3%의 예측 스킬을 증가시킬 수 있었다. 현재 현업 예보는 매일 4개씩 생산된 일주일치 앙상블 예측을 합치는 방식을 취하고 있는데 그보다 일주일에 한 번 몰아서 예보를 생산하고 그날이 월요일이라면 최대 6%의 정확도 개선을

가져올 수 있다. 이는 예보를 활용하는 예보 사용자에게는 기존 대비 약 5~9%의 경제적 수익, 예보관에게 있어서는 한파 발생에 대해 더욱 정확한 초기 가이드를 얻는데 도움이 되는 방안이었다. 그러나 매일이 아닌 주중인 목요일 하루에 예보를 몰아서 생산할 경우에는 개선 효과가 나타나지 않았다. 결국 같은 시간적 물적 자원을 사용할 때 선행시간을 줄일수록 더욱 신뢰할만한 예보를 생산할 수 있고 비용면에서 투자가 가능한 상황이라면 앙상블의 수를 늘리는 것도 더 나은 예보를 얻는데 도움을 줄 수 있다.

Executive Summary

This study is planned to evaluate the operational predictive skill of KMA's GloSea5 subseasonal forecast as well as to diagnose model's characteristics and sensitivity of the initial data and the ensemble treatment, thereby to give feasible guidelines for the enhancement of KMA's subseasonal prediction system.

First, the weekly mean dynamical forecast's evaluation frame, named CREDYT (Comprehensive and Relative Evaluating weekly mean DYnamical forecasT) is developed for coherent and systematic tracking of KMA's subseasonal forecast performance. CREDYT includes a series of evaluation information in terms of four aspects of the predictions: climatology, real-time forecast, climate regimes, dynamical coupling. CREDYT also devises the new methodologies in forecast verification and model diagnosis: rank-based verification metrics merging (RVMM), earth systems interaction measuring, and the climate regimes' predictive skill contribution estimation. Yearly produced CREDYT information is designed to be automatically documented and saved in PDF format. 2019 KMA-GloSea5 forecast/reforecast is evaluated in the frame of CREDYT. KMA's 52 weeks real-time forecast in 2019 shows moderate performance among 5 model systems and it even ranks 2nd place after ECMWF when it comes to Z500. From the CREDYT, major discrepancies of the global climate features in KMA-GloSea5 are identified: winter season Antarctic warm biases, altered earth-system interaction, disorganized dynamical coupling. Additionally, it is shown that current KMA's climatology estimation methodology can be improved by increasing reforecast samples with no time-distant weighting.

Next, the characteristics of systematic errors of the subseasonal prediction is investigated using an ensemble hindcast (1991-2010) and realtime forecast (2017-2018) produced by the KMA GloSea5 system. The KMA GloSea5 model is a global prediction system for the subseasonal-to-seasonal time scale, based on the fully-coupled atmosphere, land, ocean, and sea-ice model. This study focused on two important aspects: 1) Why are the realtime forecast (2017-2019) temperature bias patterns different from the hindcast (1991-2010) temperature bias pattern? 2) How much does the subseasonal forecast performance vary with different initial condition? The results of model experiments designed to test the sensitivity to initial data suggest that the forecasts using NWP data tend to have increased warm biases compared to the forecasts using ERA-Interim reanalysis data, especially over Arctic, East Asia, and North America. In addition, those forecasts show poorer MJO prediction skill that is derived from bivariate anomaly correlation coefficient. The large biases growth and very limited skill in subseasonal real-time forecast is directly linked to the NWP initial data that features the increased initial errors of soil temperature, cloud amount, radiation forcing, etc. The distinct initial data between hindcast and forecast may can affect the sign of the

anomaly field in real-time forecast, which bring about the opposite systematic biases between real-time forecast and hindcast..

Lastly, the strategy on an optimal ensemble system is proposed to improve subseasonal prediction in GloSea5. The basic ensemble characteristics of 8-weeks forecast for the East Asia temperature in operational hindcast are diagnosed and it is found the good availability of GloSea5 due to consistent and reliable ensemble distribution. The results of estimation of confidence suggest that there is a potential link between East Asia temperature forecast error and cooling in the land of higher latitudes and accompanying upper level high/low pressure system lying east and west. It implies the need of more realistic initialization for land surface or sea ice. By comparing the accuracy, usability and user-dependent benefit of three different ensemble sensitivity experiments, we derive guidelines of a ensemble treatment strategy for better temperature predictions. Basically, current KMA's operational system produces useful forecasts for up to eight weeks, but the forecasts for three and four weeks still have rooms to be improved in terms of accuracy and usability. For those lead weeks, increasing the number of ensemble members compared to the operational system could increase the forecast skill by up to about 3% and lengthen the predictable days by 6 days on average. To produce forecasts once a week as Monday rather than everyday could make up to 6% in skill improvements and the predictable days longer by 4 to 11 days. It could also bring about 5-9% economic benefit to forecast users and help forecasters obtain a more accurate initial guide to extreme weather events. After all, when using the same temporal and computing resource, having the shorter lead time obviously produces the more reliable forecast information. And increasing the number of ensembles by bearing expenses could also help obtain better East Asian temperature forecasts.

국문 요약

본 과제에서는 기상청(KMA) 기후예측모형 GloSea5의 계절내 시간 규모 예측에 대한 현업 예측성 평가와 모형 특성 진단 동시에 진행하였다. 주평균 예측 통합 평가프레임을 개발하여 2019년 기상청 연내 현업 예보에 대한 평가 정보를 생산하고, 초기 자료 처방에 따른 오차 성장 특성 및 앙상블 생산방안에 따른 예보 성능 개선 가능성 진단함으로써 모형 개선 방향에 대한 실질적인 제언을 도출하고 이를 문서화하여 모형 개발/수행 주체와 공유하였다.

기상청 계절내 현업 예보에 대한 객관적이고 일관되며 종합적인 평가 기틀을 마련하고자, 기본적인 기후값, 실시간 예측력, 기후 리짐, 역학적 커플링 네 부문에 대한 평가 컨텐츠 발굴하고, 이를 구조화하여 평가 프레임(CREDYT: Comprehensive and Relative Evaluating weekly mean DYnamical forecast)을 정립하였다. 또한, 실시간 예측 검증 및 모형 성능 진단에 대한 새로운 방법론(순위 기반 검증 지수 합산 평가 기법(RVMM), 지구시스템 간 연계성 진단법, 초기 기후 모드 상태에 대한 예측성 의존도 진단법 등)을 고안하였다. 2019년 52주 KMA-GloSea5 주평균 예측에 CREDYT 프레임을 적용해 평가한 결과, 타 기관 시스템 대비 KMA-GloSea5의 상대적 성능은 변수, 계절에 따라 다르지만 일부 지역에서 ECMWF에 이어 2위에 오르는 등 양호한 성능을 보인다. 실시간 예보 검증에 더해 2019년 버전 과거 재현 자료에 대한 평년값 오차, 계절성, 기후 리짐, 역학적 커플링 재현에 대한 구체적인 진단이 이루어졌다. 평년값 오차에서는 GloSea 계절 모형의 겨울철 남극 기온에는 warm bias가 존재하고, 특히 KMA에서 UKMO에는 나타나지 않는 아시아 대륙과 북미 일부에 큰 온난편차가 나타나는 점이 흥미롭다. 계절성에 있어서는 KMA-GloSea5가 여름 길이는 길게, ITCZ/SPCZ 강수 밴드 스윙은 다소 강하게 모의하는 특징을 확인할 수 있다.

10개 주요 기후 리짐의 지수 재현성과 이에 대한 지역 예보 영향을 평가한 결과, 가을철 동아시아 지위고도 예측력이 선행 1-2주에서 MJO, ART1 초기 조건 등에 민감한 것으로 나타났다. Chord 다이어그램 분석이 제시하는 KMA-GloSea5 선행 3주 수권-빙권-대류권-성층권 접합강도는 관측 대비 겨울철 과대모의, 가을철 과소모의가 나타났고, 전 계절에서는 수권 내부 연결 강도 약화와 대류권 내부 연결 강도 강화가 뚜렷하다. 또한, 강수-해수면 온도 관련성 극 지역 과대 모의, 동아시아 온난 편이에 선행하는 열대 대류활동 부재, 성층권 극 소용돌이에 대한 하층 기온 반응성 지연 등 다양한 역학과정에서의 모의 오차를 확인할 수 있다. 추가적으로, 현재 기상청에서 현업으로 사용하는 기후값 생산방안과 다양한 변경안을 검토한 결과, 기후값 생산을 위한 후처리에 있어 reforecast 앙상블 멤버간의 날짜에 따른 거리가증치를 주는 것보다 예측시점에서 다소 멀어지더라도 사용하는 reforecast 날짜 수를 확대하는 방향으로 리소스를 집중하는 것이 예측성능 향상에 도움이 될 것으로 판단하였다.

KMA-GloSea5 과거재현기후 및 실시간 예측장에 나타난 계절내 규모 초기 오차 성장 특성에 대해 살펴본 결과, 평균 10일을 전후로 예측 성능이 감소하며, 선행시간 3-4주를 전후로 그 오차가 수렴하고 있음을 확인하였다. 선행시간 1일의 지표온도 및 강수의 오차의 패턴은 선행시간 3-4주까지 거의 동일한 패턴으로 유지되고, 그 강도는 더 커지는 계통적 오차가 지배적이

다. 이러한 강수 오차 특성은 실시간 예측장에서도 동일하게 나타난다. 하지만, 과거재현기후에서의 지표 온도는 전반적으로 한랭 오차가 나타나는 반면, 실시간 예측장의 지표 온도 편차장은 온난 오차가 나타나는 특징을 보인다. 이러한 특성의 원인으로 초기장의 차이가 얼마나 기여를 하는지 알아보기 위해 초기 조건을 NWP 예측장과 재분석 자료를 사용한 실시간 예측장을 위한 민감도 실험을 구성하여 분석하였다. 그 결과, NWP 예측장을 초기장으로 사용할 때보다 재분석자료를 초기장으로 사용할 때 동아시아 겨울철 기온 및 순환장의 계절내 예측 성능 및 계절내 열대 진동의 모의 성능이 개선됨을 밝혔다. 또한, 가장 큰 초기 조건의 차이로는 지면 오차 및 구름/복사와 연관된 변수들의 오차로 분석되며, 계절내 규모에서의 오차의 감소를 위해 지면오차 및 구름/복사와 연관된 초기조건들의 개선이 필요하고 할 수 있다. 이 연구에서 사용된 MJO 분석 및 진단 방법은 기후예측모델의 개발/개선시 과거재현기후 및 실시간 예측장, 그리고 민감도 실험 등에 적용이 용이하도록 관측과 모델의 자료 전처리 방법과 분석 방법을 준자동화가 가능한 진단 틀로 구성하였으며, 추후 지속적이고 일관성 있는 기후예측모델 진단에 활용할 수 있을 것으로 기대된다.

계절내 예보에 대한 GloSea5의 예측성 개선을 위해 효율적인 앙상블 예측 시스템 구축과 개선 방향에 대해 제안하였다. 동아시아 기온의 8주 이내 예보에 대한 현업 후보실험의 앙상블의 기본 특성을 진단한 결과 GloSea5는 예보 2주 이후부터는 일관되고 신뢰도가 높은 앙상블을 생산하고 있어 활용성이 좋은 앙상블 분포를 가지고 있음이 확인되었다. 또한 GloSea5 앙상블에 대한 불확실성 진단을 통해 유라시아 대륙 고위도에서의 저온 아노말리와 그에 수반되는 상층 고/저기압의 동서 배치가 8주 이내 동아시아 기온 예보의 에어가 커지게 하는 요인으로 작용할 수 있다는 것을 확인하였고 따라서 더 나은 기온 예측을 위해서는 고위도 지면/해빙 처방에서의 개선이 필요하다고 제안하였다.

앙상블 개수, 생산 일자, 생성 기법에 변화를 준 민감도 실험들의 정확도, 유용성, 활용도를 기존 현업 예보와 비교함으로써 더 나은 동아시아 기온 예측을 가져다 줄 수 있는 앙상블 생산 방안에 대한 가이드라인을 도출하였다. 기본적으로 현업 실험의 경우 대체로 8주까지 유용한 예보를 생산해내고 있으나 3,4주 예보는 정확도와 유용성 측면에서 개선이 필요한 기간으로 판단되었다. 이 기간에 대해서 현업 시스템 대비 앙상블 멤버수를 증가시키면 최대 약 3%의 예측 스킬을 증가시킬 수 있었고 평균 6일의 예보일 개선을 기대할 수 있었다. 매일보다는 일주일에 한 번 예보를 생산하고 그날이 월요일이라면 최대 6%의 스킬 개선과 4~11일의 예보 일수에서 개선을 가져올 수 있다. 이는 또한 예보 사용자에게는 기존 대비 약 5~9%의 경제적 수익을 가져다 줄 수 있고 예보관에게 있어서는 극한 기상 발생에 대해 더욱 정확한 초기 가이드를 얻는데 도움이 되는 방안이다. 그러나 매일이 아닌 주중인 목요일 하루에 예보를 몰아서 생산할 경우에는 기온 예측에서 개선 효과가 나타나지 않으므로 추천하지 않는 방안이다. 결국 같은 시간적 물적 자원을 사용할 때 선행시간이 가까울수록 더욱 신뢰할만한 예측 정보를 생산할 수 있고 비용면에서 투자가 가능한 상황이라면 앙상블 수를 늘리는 것도 더 나은 동아시아 기온 예보를 얻는데 도움을 줄 수 있다.

목차

평문 초록	i
Executive Summary	iii
국문 요약	v
목차	vii
표 차례	viii
그림 차례	ix
1. 서론	1
2. 연구 자료 및 방법	4
가. 자료	4
(1) 계절내 모형 자료	4
(2) 관측 자료	7
나. 기후 지수들	8
다. 연구 전략 및 방법	9
(1) 주평균 예측 통합 평가프레임 개발	10
(2) 재분석 초기장 실험을 통한 초기 오차 영향 및 필수 개선 요소 파악	11
(3) 정확도, 유용성, 활용도 측면에서의 시간지연/후방산란 앙상블 기법의 최적 조합 제시	11
3. 결과	16
가. 기상청 계절내 예측 시스템 (GloSea5) 예측성 평가	16
(1) 현업 주평균 예측 통합 평가프레임 개발	16
(2) 2019년 KMA-GloSea5 예측 종합 평가	23
(3) GloSea5의 기후값 생산방안에 따른 재현/예측성능 분석	53
(4) KMA-GloSea5 기반 동아시아 겨울 기온 예측 성능 추정 알고리즘 개발	61
나. 초기장에 따른 오차 성장 특성 진단	65
(1) 최적 초기장 실험의 계절내 예측 오차 진단	65
(2) MJO 오차진단 틀 개발	95
다. 앙상블 생산 방안에 따른 계절내 기온 예측 민감도 평가	109
(1) 계절내 규모에서의 실시간 예측장 앙상블 특성진단	109
(2) 앙상블 생산방안별 기온 예측 민감도 평가	117
4. 요약 및 결론	134

사사	139
References	140
부록 A. 동아시아 겨울 기온의 GloSea5 선행3주 주평균예보 스킬추정을 위한 지침서	146
부록 B. KMA-GloSea5 기후값 생산 방안 가이드라인	149
부록 C. GloSea5 MJO 오차 진단 툴 매뉴얼	157
부록 D. Met Office-KMA Wiki science page 결과 공유	173

표 차례

Table	page
1. Description of subseasonal forecast models used in this study.	5
2. Validated target for the intercomparison of ensemble treatment.	6
3. Used forecast for intercomparison of forecast skill among sensitivity experiments.	7
4. Climate indices verified in this study.	9
5. Metrics for the diagnostic of ensemble characteristics in hindcast experiment and forecast skill in sensitivity test.	12
6. The meaning and consistency according to ensemble types.	13
7. Cost/Loss Decision Model.	15
8. Atmospheric initial condition for each S2S model used in this study.	25
9. The correlation coefficients of VPM1 and VPM2 for 2019 KMA real-time forecast.	46
10. Comparison of climatology production methods.	55
11. Verification of estimated ACC categories for 12 weeks during the 19/20 winter season utilizing three metrics (Proportion Correct, Hit Rate and Post Agreement).	65

그림 차례

Figure	page
1. Model configuration of KMA-GloSea5.	6
2. Overview of ensemble experiment.	7
3. Project structure and execution strategy.	10
4. Four categories and their components of CREDYT framework.	17
5. Schematics of (left) six verification metrics merging and (right) an example of relative performance of 5 models.	18
6. (diagonal) probabilistic distribution of individual climate indices for (green) the observation and (orange) 1- and (purple) 3-lead weeks GloSea5 forecasts.	19
7. Pie charts showing the relative contribution of ten climate regime strength in the 0-lead week (corresponding to -3 to +4 lead days) to weekly mean Z500 forecast ACC skill during SON season.	21
8. Chord diagram showing the strength of links among (upper) 11 climate indices and (lower) four earth systems for (the leftest) the observation, (second and third columns) 1- and 3-lead weeks forecasts.	22
9. Annual and seasonal mean biases for T2M of ECCC, ECMWF, KMA, NCEP, and UKMO for 1- to 4-lead week during 1999-2010.	24
10. Differences of ERA-Interim and ERA5 for annual and seasonal mean climatology during 1999-2010.	25
11. Annual and seasonal averaged relative RMSE difference of climatology of T2M over Globe, Tropics, and East Asia of ECCC, ECMWF, NCEP, and UKMO with KMA.	27
12. (left) Global averaged T2m seasonal cycle averaged for the period of 1999-2010 for 1- to 4-lead weeks forecasts.	28
13. GloSea5 long-term (1999-2010) averaged precipitation showing ITCZ/SPCZ pattern for (1st row) DJF and (2nd row) JJA season.	29
14. Annual and seasonal averaged relative ACC difference of 2019 real-time forecast of T2M over Globe, Tropics, and East Asia of ECCC, ECMWF, NCEP, and UKMO with KMA.	31
15. Annual and seasonal averaged overall MSSS of 2019 real-time forecast of T2M for 1- to 4-lead week over Globe, Tropics, and East Asia of KMA, ECCC, ECMWF, NCEP, and UKMO.	33

16. Annual and seasonal averaged relative MSSS difference of 2019 real-time forecast of T2M over Globe, Tropics, and East Asia of ECCC, ECMWF, NCEP, and UKMO with KMA.	34
17. (Left) daily ACC of 2019 real-time forecast for KMA, ECCC, ECMWF, NCEP, and UKMO for T2M, PREC, and Z500 over Globe, Tropics, and East Asia.	36
18. (Left) daily MSSS of 2019 real-time forecast for KMA, ECCC, ECMWF, NCEP, and UKMO for PREC over Globe, Tropics, and East Asia.	37
19. The cases that ECMWF fails to capture the extreme rainfall case over Southeast China on February 17, 2019.	38
20. The time series of overall MSSS value of ECMWF for PREC over East Asia for 2019 real-time forecast using 624 samples (52 weeks×12 years; blue line) and 4380 samples (365 days×12 years; orange line) of MSEC calculation.	39
21. The time series of MSE(blue) and MSEC(orange) of ECMWF for Z500 over Northern Hemisphere (north of 30° N).	40
22. Six verification scores for 1- to 4-lead weeks 52 weekly mean real-time forecasts in 2019 over the Globe, Tropics and East Asia.	41
23. (the leftest two columns) ranking of KMA-GloSea5 out of five models at each grid for 52 weekly mean forecasts in 2019 and (3rd column) regionally summarized ranking of KMA-GloSea5 for the Globe, Tropics and East Asia domain.	42
24. MJO statistics for 52 weeks in 2019 for observation and 1- to 4-lead weeks KMA-GloSea5 forecasts: (upper) phase and amplitude of individual weeks, (middle) three amplitude categories frequency of 8 phases for the observation and different lead weeks, (lower left) 8 phase frequency of 1- to 4-lead weeks for three amplitude categories, and (lower right) the frequency biases summary for different combination of phase, amplitude and lead weeks.	44
25. (Left) weekly VPM1 and VPM2 time series of observation and KMA for 1-, and 3-lead week for 2019 real-time forecast.	46
26. Bivariate correlation of daily MJO index of KMA, ECCC, ECMWF, NCEP, and UKMO for reforecast during 1991-2010 and 2019 real-time forecast.	47
27. Observed and forecasted timeseries of weekly mean climate index in 2019 and TCC score of 1- to 4-lead weeks forecasts: (upper left) NINO3.4, (upper right) ART1, (lower left) SH and (lower right) PV.	48
28. T2m predictive skill (ACC) variation with the phases and amplitude of MJO for different	

seasons and regions.	49
29. Stacked bars showing the strength of links among four earth systems for the observation, 1- and 3-lead weeks forecasts during all year round, and four seasons.	50
30. Air-Sea coupling: maps of lead-lag correlation coefficients between weekly mean SST and precipitation anomalies at each grid for (left) the observation, (middle) 1-, and (right) 3-lead weeks forecasts.	51
31. Tropics-midlatitude coupling: maps of lead-lag correlation coefficients between the weekly mean East Asia T2m (navy box) anomalies and velocity potential anomalies at 200 hPa at each grid for (left) the observation, (middle) 1-, and (right) 3-lead weeks forecasts.	52
32. Troposphere-stratosphere coupling: maps of lead-lag correlation coefficients between the weekly mean polar vortex (PV) anomalies and T2m anomalies anomalies at each grid for (left) the observation, (middle) 1-, and (right) 3-lead weeks forecasts.	53
33. Mean biases for both of T2M of KMA(UKMO), EXP1, EXP2, and EXP3 for 1-, and 3-lead week during DJF 1999-2010.	56
34. Climatology difference for T2M between EXP1, EXP2, and EXP3 and KMA(UKMO) for 1-, and 3-lead week during DJF 1999-2010.	57
35. Annual RMSE of climatology of T2M for 1-, 3-, 5-, and 7-lead week over Globe, Tropics, and East Asia of KMA(red) and EXP3(blue) during 1991-2010.	58
36. Annual and seasonal averaged relative ACC difference of 2019 real-time forecast for 3-lead week of T2M, PREC, and Z500 over Globe, Tropics, and East Asia of EXP1(green line), EXP2(orange line), and EXP3(gray bar) with KMA.	59
37. Annual ACC difference range between upper and lower 25% of 2019 real-time forecast for 1-, and 3-lead week of T2M, PREC, and Z500 over Globe, Tropics, and East Asia of EXP1, EXP2, and EXP3 with KMA.	61
39. Box-whisker of ACC for BN, NN and AN groups at different 4 lead weeks.	62
40. W3 mean ACC of 9 categories based on two different indices' tercile criteria of (top) W1, (middle) W2, and (bottom) W3.	63
41. (a) scatter plot between simulated PC1 stability and EA T2m ACC and (b) violin plot of EA T2m ACC for three PC1 stability categories for lead time of 3 weeks from 240 samples (12 weeks for 20 winter seasons).	64
42. Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) Blue shading indicates all ensemble member of hindcast data (1991-2010).	66

43. Root Mean Square Error of surface temperature over globe (left) and East Asia (right).	66
44. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) for initial month.	67
45. Temporal correlation coefficient (TCC) of the 2m temperature between ERA interim reanalysis data and GloSea5 hindcast with (a)-(d)1-4 week lead for DJF.	68
46. Temporal correlation coefficient (TCC) of the precipitation between ERA interim reanalysis data and GloSea5 hindcast with (a)-(d)1-4 week lead for DJF.	69
47. Distribution of averaged surface temperature from lead day 1 for JJA (left) and DJF (right) of hindcast (1991-2010).	70
48. Distribution of averaged surface precipitation from lead day 1 for JJA (left) and DJF (right) of hindcast (1991-2010).	70
49. Vertical profiles of temperature (left) and specific humidity (right) error difference between GloSea5 hindcast and ERA5.	71
50. Distribution of averaged surface sensible (left) and latent (right) heat fluxes from lead day 1 for JJA of hindcast (1991-2010).	71
51. Distribution of averaged total cloud amount (left) and 850 hPa specific humidity (right) from lead day 1 for JJA of hindcast (1991-2010).	72
52. Distribution of averaged surface temperature (left) and precipitation (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).	73
53. Distribution of averaged total cloud amount (left) and 850hPa specific humidity (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).	73
54. Distribution of averaged shortwave radiation fluxes (left) and longwave radiation fluxes (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).	74
55. Distribution of averaged sensible heat fluxes (left) and latent heat fluxes (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).	74
56. Vertical profiles of temperature (left) and specific humidity (right) error difference between GloSea5 hindcast and ERA5.	75
57. Ratio of large-scale precipitation and convective precipitation from total precipitation. ...	75
58. Histogram of precipitation for JJA over tropics (left) and East Asian (right) regions.	76
59. Root Mean Square Error of surface temperature over globe (left) and East Asia (right).	77
60. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) for initial month.	77

61. Distribution of averaged surface temperature from lead weeks 3-4 of hindcast (1991-2010) (left) and realtime forecast for 2018 (right).	78
62. Difference distribution of averaged surface temperature from lead weeks 3-4 between anomaly (left), original raw (right) data and ERA5.	78
63. Horizontal distribution of 500hPa geopotential height mean (solid line) and anomaly (shaded) for December to February from observation.	80
64. Anomaly distribution of surface temperature over Korean peninsula for 2017/18 winter. ·	80
65. 1st to 4th PC time series of EOF analysis from ERA5 reanalysis, GloSea5 lead weeks 1-2, and lead weeks 3-4.	81
66. Surface temperature regressed on ERA5, GloSea5 lead weeks 1-2, and lead weeks 3-4.	82
67. Mean biases (top) and Mean Square Errors (bottom) of zonal mean geopotential height in GloSea5 at lead weeks of 1-2 (left) and 3-4 (right) in DJF.	83
68. Difference of Mean Square Errors of zonal mean geopotential height between cold year and warm year cases in GloSea5 at lead weeks of 1-2 (left) and 3-4 (right) in DJF.	84
69. Anomaly Correlation Coefficient of 500hPa geopotential height over globe (left) and East Asia (right) Blue line is for warm year cases.	84
70. Sea level pressure distribution from 1st to 2nd mode of EOF analysis from ERA5 reanalysis, GloSea5 lead weeks 1-2, and lead weeks 3-4.	85
71. Difference distribution of sea level pressure at lead day 1 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).	86
72. Difference distribution of sea level pressure at lead weeks 3-4 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).	86
73. Horizontal distribution of 500hPa geopotential height at lead weeks 3-4 from ERA5 (top), NWPiF (middle), and ERAiF (bottom).	87
74. Difference distribution of surface temperature at lead weeks 3-4 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).	88
75. Difference distribution of surface temperature at lead day 1 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).	88
76. Difference distribution of surface precipitation at lead day 1 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).	89
77. Difference distribution of surface solar radiation fluxes (top-right), outgoing longwave radiation fluxes (top-left), surface longwave radiation fluxes (bottom-right), and cloud	

amount (bottom-left) at lead day 1 between ERAiF and NWPiF.	89
78. Difference distribution of surface solar radiation fluxes (top-right), outgoing longwave radiation fluxes (top-left), surface longwave radiation fluxes (bottom-right), and cloud amount (bottom-left) at lead weeks 3-4 between ERAiF and NWPiF.	90
79. Difference distribution of surface precipitation at lead weeks 3-4 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).	91
80. Difference distribution of surface temperature at lead weeks 3-4 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).	92
81. Improvement rate (*100%) of the ERAiF compare to the NWPiF using RMSE of surface temperature at lead day 1 (left), and lead weeks 3-4 (right).	92
82. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) at lead day 1.	93
83. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) at lead weeks 3-4.	93
84. Anomaly Correlation Coefficient of surface temperature (left) and 500hPa geopotential height (right) over globe (shaded bar) and East Asia (dashed bar).	94
85. Difference distribution of sea surface temperature at lead month 1-6 between ERAiF and NWPiF.	95
86. Spatial distributions of mean field (shading) and variance field (contour) for 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), and 200 hPa divergence (VP200) in observations (OBS) and GloSea5 model (GL5).	97
87. Power spectrum results of 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), and 200 hPa divergence (VP200) in observations (OBS) and GloSea5 model (GL5).	98
88. Wavenumber frequency results of 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), and 200 hPa divergence (VP200) in observations (OBS) and GloSea5 model (GL5).	99
89. MEOF (Multivariate Empirical Orthogonal Function) results of 850 hPa zonal wind (U850-black line), 200 hPa zonal wind (U200-red line), and 200 hPa divergence (VP200-blue line).	100
90. COR (Bivariate Anomaly Correlation Coefficient) for Velocity Potential MJO indices in GloSea5 HINDCAST (green line : 0.5).	101
91. RMSE (Bivariate Root Mean Square Error) for Velocity Potential MJO indices in GloSea5 HINDCAST (green line : $\sqrt{2}$).	101

92. Spatial distributions of mean field (shading) and variance field(contour) for 850 hPa zonal wind (U850) in observations (OBS) and initial sensitivity experiments (ERAiF) and controlled experiment (NWPiF).	102
93. Hovmoller diagram of 850 hPa zonal wind(U850) in observations (OBS) and initial sensitivity experiments (ERAiF) and controlled experiment (NWPiF) during the winter (Dec 2017~Apr 2018).	103
94. PCC skill of hovmoller diagram for 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), precipitation (PREC), and the mean of all variables in initial sensitivity experiments (ERAiF - filled bar) and controlled experiment (NWPiF - patterned bar).	104
95. Lag time-longitude correlation (i.e. MJO propagation diagram) of 850 hPa zonal wind (U850) in observations (OBS) and initial sensitivity experiments (ERAiF) and controlled experiment (NWPiF) during the winter (Dec 2017~Apr 2018) (60~180E, 10S~10N, 20~70 days filtered).	105
96. PCC skill of MJO propagation diagram (60~180E, 10S~10N, 20~70 days filtered) for 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), 200 hPa divergence (VP200), and the mean of the variable in initial sensitivity experiments (ERAiF - filled bar) and controlled experiment (NWPiF - patterned bar).	105
97. Distribution of specific humidity on middle atmosphere (850~500hPa) in observation (black) and ERAiF (red) and NWPiF (blue) experiments by the forecasting 1-2 weeks (left) and forecasting 3-4 weeks (right) (10S~10N, 20~70 days filtered).	106
98. Distribution of equivalent potential temperature on middle atmosphere (850~500hPa) in observation (black) and ERAiF (red) and NWPiF (blue) experiments by the forecasting 1-2 weeks (left) and forecasting 3-4 weeks (right) (10S~10N, 20~70 days filtered).	107
99. MEOF (Multivariate Empirical Orthogonal Function) results of 850 hPa zonal wind (U850-black), 200 hPa zonal wind (U200-red), and 200 hPa divergence (VP200) in initial sensitivity experiments (ERAiF) and controlled experiment (NWPiF) for the forecast 3-4 weeks.	108
100. Bivariate Correlation coefficient (BCOR) for Velocity Potential MJO indices in GloSea5 HINDCAST (grey), initial sensitivity experiments (ERAiF-red), and controlled experiment (NWPiF-blue).	109
101. Bivariate Root Mean Square Error (BRMSE) for Velocity Potential MJO indices in GloSea5 HINDCAST (grey), initial sensitivity experiments (ERAiF-red), and controlled experiment	

(NWPIF-blue).	109
102. Ranked histogram with 3 ensemble members of GloSea5 hindcasts and observed 2m temperature over the East Asia.	111
103. Same as Figure 102, but for each season, spring (green), summer (red), autumn (orange), winter (blue).	112
104. Tile map for ensemble type and reliability index in regard to 2m temperature over East Asia using GloSea5 hindcast.	113
105. Root mean square error (solide line) and spread (dashed line) of ensemble for 2m temperature in GloSea5 hindcast.	114
106. Error-spread relationship of winter(DJF) temperature hindcast with 1 to 8 weeks lead time in GloSea5.	115
107. Spatial pattern of ensemble spread for winter (DJF) temperature hindcast with 1 to 8 weeks lead time in GloSea5.	115
108. Uncertainty of winter temperature (upper panel) and geopotential height at 500hPa (below panel) with 8 weeks lead time in GloSea5.	116
109. Uncertainty of winter temperature (left panel) and geopotential height at 500hPa (right panel) in the largest spread section.	116
110. Reliability diagram for weekly probability forecast of 2m temperature over East Asia.	118
111. Brier score for daily probability forecast of EA 2m temperature in each category (below normal (short dashed line), normal (solid line), above normal (long dashed line).	118
112. Brier skill score of 2m temperature probability forecast in each category (below normal (blue bar), normal (yellow bar), above normal (red bar)).	120
113. Improvement ratio of brier score for 2m temperature probability forecast in each category (below normal (blue bar), normal (yellow bar), above normal (red bar)).	121
114. Difference of brier score between experiments regarding ensemble technique, LAF or LAF+burst.	122
115. Difference of brier score between experiments regarding issuing date, Thursday or Monday.	123
116. Same as Figure 112, but for ranked probability skill score (yellow bar).	124
117. ROC diagram of 2m temperature probability forecast for from lead 1 day to 49 day (sequence from left to right and top to down).	125
118. Improvement ratio for AUC of 2m temperature probability forecast (yellow bar) and the	

value of 0.5 subtracted from AUC itself for OPER (red asterisk).	126
119. Forecast accuracy (brier score, ranked probability score, improvement ratio of continuous ranked probability score) and useless (area under ROC curve, improvement ratio of area under curve) for 2m temperature probability forecast from lead 1 day to 57 day.	127
120. Improvement ratio for 2m temperature probability forecast in week3 and 4.	128
121. Economic value for 2m temperature probability forecast in 6 sensitivity experiment. ...	129
122. Experiment with the largest economic value for 2m temperature probability forecast from lead 1 day to 57 day and the value in that case.	130
123. Skill improvement and the number of days with improved skill of economic value within eight weeks forecast comparing to OPER.	130
124. Skill improvement and the number of days with improved skill of economic value for W3 and rate of return for W4 comparing to OPER.	131
125. Spatial pattern of temperature anomaly for a cold surge on the third week of December in 2017.	132
126. Scatter plot for hit rate and false alarm rate of temperature for the cold surge forecasts in each sensitivity experiment.	132
127. The violin plot for comprehensive intercomparison of all skill improvement on EA temperature forecast 3 and 4 weeks in advance.	133

1. 서론

2주에서 2달에 이르는 계절내-계절 (Subseasonal to seasonal; S2S) 시간규모는 초기장의 메모리 지속시간에 비해 다소 길고 경계조건의 영향을 받기에는 다소 짧기 때문에 양쪽 어디에서도 예측성을 보장받을 수 없는 태생적 한계가 있다. 이로 인해, 최근까지도 계절내 예보는 현업 예측의 “회색 영역”으로 여겨져 왔다. 특히, 기온과 강수 같은 기본 변수들의 예측에 있어 계절내 예측 모형의 예측 가능성이 2주보다 짧아 계절내 규모라는 목표에 미치지 못한다. 하지만, 이음새 없는 기상예보를 위한 계절내 예측자료의 중요성에 대한 인식이 증가하고 (Vitart et al. 2017), 농작물 및 수자원 관리, 악기상의 조기 대응 등 다양한 섹터로부터의 수요가 증가함에 따라 계절내 예측기술 향상을 위한 연구들이 국제적으로 활발히 수행되고 있다 (S2S 프로젝트 (WMO 2013) 등). 한국 기상청 역시 WMO S2S 연구 프로젝트에 참여함으로써 이러한 국제적 연구 흐름에 적극적으로 부응하고 있다.

전지구 기후 모형(Global Climate Models; GCMs)은 계절내 예측에서 계절 예측에 이르기까지 기본적으로 사용되고 있는 필수적인 존재이다. 이에 전세계 현업 기관들 내에서 기후예측모형은 중장기 예측 성능 향상을 목표로 끊임없이 발전해 오고 있다. 한국 기상청에서도 1999년부터 대기대순환 모형에 근거한 장기예보 생산체계를 구축하고 역학적 방법에 의한 장기 예측 자료를 생산하였다. 국립기상과학원에서도 2000년대에 들어서 기후예측모형을 구축, 역학적 방법에 의한 장기예측자료를 생산하였다. 최근에는 한국 기상청과 국립기상과학원은 계절 규모에서 한영 공동 기후예측시스템을 정착하고 현업으로 이용하기 위해 영국 기상청과 협력해왔고, 그 결과 2010년부터 영국 기상청의 전지구 대기-해양-해빙 결합모형 (HadGEM3)을 기반으로 한 계절예측 시스템 Global Seasonal Forecasting System version 4 (GloSea4)를 도입하고 구축하였다. 또한, 기상청은 향상된 계절 예보를 위해 2014년부터 GloSea4에서 모델 물리 과정과 수평 분해능을 개선한 고해상도 기후모형인 GloSea5 (GloSea version 5)를 현업에 도입하였다.

GloSea 모형은 영국 MetOffice의 HadGEM (Hadley Center Global Environment Model)에 기반한 앙상블 계절 예측 시스템이지만, 단기 및 중기 예측에서의 성능이 상당히 높아 (Lee et al. 2016) 계절내 시간규모 예측에서도 그 가능성에 대해 높이 평가되고 있다. 따라서, 이음새 없는 기상 정보 생산에 대한 GloSea5의 가치를 계절 예측 뿐만 아니라 계절내 규모 예측까지 아울러서 “통합적”으로 진단할 필요가 있다. 2010년 도입 이래 KMA-GloSea 모형은 지속적으로 개선을 거듭해 왔고 향후 GloSea Version 6 출시를 앞두고 있으므로, “지속적”이고 “일관된” 형태의 예측 특성 및 성능 추적 연구가 요구되는 시점이다. 특히, 영국, 호주의 GloSea 계열 모형들 그리고 타 기관 모형들과의 비교를 통한 기상청 현업 기후예측모형, GloSea5의 오차에 대한 체계적인 분석, 이를 바탕으로 한 모형 개선 가능성 및 방향에 대한 제언이 필요한 시점이다. 현재까지 GloSea5 모형 자체 개선 및 예측성 평가에 대한 의미 있는 연구 활동이 학계를 통해 다양하게 이루어져 왔고 현재도 수행 중이지만, 개별 과제의 주제가 파편적으로 흩어져 있고 단발성에 그치는 경우가 많았다. 기상청 기후예측시스템의 기본 요소에 대한 “통합적”인 평가를 외부의 “객관적” 시각으로 장기적인 프레임 안에서 진행할 수 있도록 기획하였으며, 이러한 연구 활동은 현재의 KMA-GloSea5 모형뿐만 아니라 향후 개발될 차세대 중장기 예측 시스템을 일관된 프레임 안에서 평가하고 이를 개선하는데 밑거름이 될

것으로 판단된다.

계절 예측의 경우 대기 초기장의 영향을 크게 받는 NWP와는 달리 해양 강제력 등의 천천히 움직이는 기후 조건들의 영향만 중요하게 여겨져 왔다. 최근 들어 전 세계적으로 여러 현업 센터에서는 이음새 없는 예보를 위하여 중단기 날씨-계절예보를 비롯하여 긴 시간 규모를 가지는 기후 예측까지 할 수 있는 시스템을 개발해오고 있다. 이에 따라, 계절내-계절 시간 규모 예측에서도 지면, 대기 등의 짧은 시간 규모에서 중요하게 여겨지던 초기 조건들의 영향이 주목받고 있다. 계절예측시스템에서의 초기 조건은 크게 대기, 지면, 해양/해빙으로 나눌 수 있다. 여기서 대기 조건은 날씨 규모에서 가장 중요한 강제력이며, 계절내 예측 특히 열대 진동 및 성층권 순환 등에 영향을 줄 수 있는 것으로 알려져 있다. 그에 비해 지면 초기조건은 토양 수분 및 눈 상태 등을 통해 계절내 예측에서 계절 예측까지 영향을 미치며, 해양/해빙 등의 조건은 해수면 온도 및 해빙 상태 등을 통해 계절 예측 및 좀 더 긴 규모의 기후예측에서 영향을 미치는 것으로 알려져 있다. 따라서, 대부분의 계절 예측 시스템에서는 해양의 초기장에 대한 중요성이 잘 알려져 왔으며, 최근 들어, 지면 및 해빙의 초기장의 중요성이 언급되어 오고 있다. 특히, 기상청 기후예측시스템인 GloSea5는 계절예측을 타깃으로 개발되었기 때문에 최근 계절내 예측에 대한 그 중요성이 언급되기 시작하였다. 그에 대한 예로, 날씨부터 계절내 예측에 중요하다고 알려진 지면 과정에 대한 초기화 기법이 GloSea5의 2019년 현업 버전부터 적용된 바 있다. 지면 초기화는 특히 계절 예측에서 예측성이 아주 낮게 나타나는 지면 온도에 영향을 미치는 것으로 알려져 있으며 (Koster et al. 2011), 서은교 등 (2016)에 따르면, GloSea5에 토양 수분 초기화에 대한 영향을 살펴본 결과, 지면 온도의 1개월 예측에 있어서 특히 중위도 지역에서 예측 성능의 뚜렷한 개선을 보인 바 있다. 다만, 지면 초기화 효과에 따른 예측 성능 개선 효과는 각각의 모델마다, 그리고 지역별로 차이를 보이고 있어 다양한 모델 및 계절/지역별로 검증되어야 할 것으로 예상된다. 정영운 등(2016)에 따르면, GloSea5의 현재 시스템에서 대기-해양 접합이 단기 기후예측에 긍정적으로 작용하고 있으며, 우리나라 근해에서 더 많은 관측 자료를 자료 동화에 사용한 결과 더 현실적인 예측값을 가져올 수 있음을 나타낸 바 있다. 이처럼 현재 GloSea5 및 차후 현업 버전에 적용하기 위한 해빙 및 해양 초기화 등 여러 초기 조건들의 개선을 위해 학계 및 현업 기관에서 많은 노력을 하고 있다. 다만, 초기 조건의 개선과는 별개로 현재 시스템의 특성을 이해하고 알맞게 분석하는 것도 상당히 중요하다. 또한, 지면 초기장 및 해양/해빙 초기장의 이해 및 개선에 비해 대기 초기 조건에 대한 관심은 다소 부족하다 할 수 있다. 따라서 본 과제에서는 대기 초기장에 따른 계절내 예측장 차이에 대해 살펴보려고 한다.

기상·기후 분야에서 가장 선진 예보 기술인 앙상블 예측 시스템은 가장 가능성이 높은 예측 결과를 추정할 수 있어 수치 예보 분야에서 패러다임의 전환을 가져왔다. 앙상블 예보의 시작으로 정확도 높은 예보와 더불어 활용도 높은 예보 생산에 대한 요구가 지속적으로 증가되고 있으나 자원과 비용에 대한 문제 또한 수반되고 있다. 따라서 최근 한정된 자원과 비용 안에서 가장 효율적인 앙상블 예측 시스템을 구축할 수 있는 다양한 전략을 발굴하려는 노력들이 이루어지고 있다. 특히 모델 해상도, 앙상블 생성 기법, 예보 생산 일자, 앙상블 크기 등에 따라 앙상블 예측 시스템의 예보 성능이 좌우될 수 있기 때문에 이들은 효율적인 시스템을 운영하는데 있어서 필수적인 전략 요소로 다루어진다. 기상청에서 현업에 활용중인 GloSea5는 고해상도 전지구 결합 앙상블 예측시스템으로 확률 예보 생산을 통해 단기 및 중기, 장기예보

를 지원하고 있다. 2010년 도입 이래 GloSea5는 지속적으로 개선을 거듭해왔고, 이음새 없는 기상 정보 생산을 위해 계절내 예보에 대해서도 최적의 앙상블 예측 시스템을 구축하고 개선할 필요가 있다. 2018년 국립기상과학원은 “현업 장기예측시스템 운영체계 개선을 위한 최적 앙상블 생성법 개발 및 모니터링 개선” 사업을 통해 다양한 민감도 실험을 수행하였고 효율적인 앙상블 방안을 찾기 위해 결정론적 검증을 수행한 바 있다. 그러나 확률 예보 생산을 통해 현업 장기예보를 지원하는 GloSea5의 실질적인 개선을 위해선 확률론적 검증 또한 필요하다. 따라서 이 연구에서는 GloSea5의 계절내 확률 예보를 개선할 수 있는 앙상블 생산방안에 대해 제안하고자 한다.

이 보고서의 2장은 과제에서 사용된 자료 및 분석 방법론, 특히 각 세부 과제 목표를 달성하기 위한 추진 전략에 대해 서술하였다. 3장은 기상청 GloSea5 계절내 규모 예측과 관련하여 (가) 필수 평가 요소 정립 및 현업 성능 다중 모형 비교, (나) 초기 자료 처방에 따른 모형 오차 성장 특성, (다) 앙상블 생산 방안에 따른 예보 성능 개선 가능성에 대한 연구 결과를 포함한다. 4장은 연구 결과들을 요약하고, 기상청 기후예측 시스템 개선과 관련한 시사점을 제시한다.

2. 연구 자료 및 방법

가. 자료

(1) 계절내 모형 자료

(가) 주평균 현업 예측 자료 (S2S)

이 연구에서 사용된 계절내 예측자료는 S2S prediction project database (Vitart et al. 2017)에서 수집하여 준 실시간으로 제공되고 있다. S2S prediction project에 참여하는 총 11개의 모형 중 다섯 개 모형(ECCC, ECMWF, KMA, NCEP, UKMO)의 자료가 이 연구에 사용되었다 (Table 1). 이 중 세 모형(ECCC, ECMWF, NCEP)은 모두 매주 목요일에 예측을 수행하고, 과거 재현자료(reforecast) 역시 예측날짜와 동일한 날짜에 제공하고 있다. 반면 GloSea5를 기반으로 한 모형인 KMA와 UKMO의 예측 스케줄은 나머지 세 모델과는 다른데, 과거재현자료를 한 달에 4회 고정된 날짜(1일, 9일, 17일, 25일)에 제공하고 있으며 예측자료는 매일 생산, 제공한다. 그러나 실시간 예측을 위한 앙상블 멤버의 수가 4개로 적어서 확률예측 계산 등을 위해서는 일주일 정도의 예측 결과를 합쳐서 사용할 필요가 있다. 기상청에서도 실시간 예측을 위해 KMA 모델에서 생산된 최근 7일의 예측자료를 사용한다. 이 연구에서는 KMA와 UKMO에 대해서도 예측 시작 날짜를 다른 모델과 동일하게 맞추기 위해, 예측 시작 날짜로부터 6일 전인 전주 금요일부터 예측 시작 날짜인 목요일까지 총 7일의 예측자료를 이용하여 시간 차이에 따른 앙상블을 구성하고 실시간 예측자료를 생산하였다. 각 모형의 수평 해상도 역시 저마다 다르므로, 모두 $1.5^{\circ} \times 1.5^{\circ}$ 로 내삽하여 일관된 비교가 가능하도록 하였다. 또한, 계절내 예측성능을 살펴보기 위해 일주일 단위로 평균하여 선행시간 1주부터 4주까지의 예측자료를 생산하였다. 예측 스케줄이 동일한 ECCC, ECMWF, NCEP의 기후값은 예측 시작 날짜와 동일한 날짜에 해당하는 과거재현자료와 일주일 전/후 과거재현자료의 총 3개의 평균으로 구하였다. 또한 KMA와 UKMO의 기후값은 기상청의 기후값 생산방법을 참고하여 계산하였는데(Ham et al. 2017), 예측 시작 날짜와 가장 가까운 2개 과거예측자료에 날짜에 따른 거리가중치를 주어 평균하는 방법을 사용하고 3.가.(3)(가)㉠절에 기후값 생산 방법에 대해 자세히 서술하였다. 실시간 예측 아노말리는 이렇게 구해진 기후값을 해당 날짜의 실시간 예측에서 제거하였다.

Table 1. Description of subseasonal forecast models used in this study.

Center	Time Range	Resolution	Ens. Size	Fcst. Frequency	Refcst Period	Refcst Ens. Size	Refcst. Frequency
ECCC	32 days	0.45x0.45, L40	21	weekly	1998-2017	4	weekly
ECMWF	46 days	Tco639/319 L91	51	2/week	past 20 years	11	2/week
KMA	60 days	T216 L85	4	daily	1991-2010	3	4/month
NCEP	44 days	T126 L64	16	daily	1999-2010	4	daily
UKMO	60 days	T216 L85	4	daily	1993-2016	7	4/month

(나) KMA-GloSea5 초기장 치환 실험 자료

우리나라 기상청은 2010년 영국 기상청(Met Office)과 대기-해양 결합 모델을 이용한 계절 예측시스템의 공동구축 및 운영에 관한 협정을 체결하여 단기간 날씨부터 계절내 및 계절 예측, 기후 예측을 포함하는 포괄적인 계절예측시스템인 Global Seasonal Forecasting System version 5 (GloSea5) (MacLachlan et al. 2015) 을 도입하였고, 2014년부터 현재까지 현업 계절예측에 활용 중이다. 영국 기상청 기후 모델인 HadGEM3 (Hadley Centre Global Environment Model version 3) 에 기초하여 구축된 GloSea5는 Met Office의 Unified Model (UM) 대기 모델 (Global Atmosphere model; Walters et al. 2017) 과 Nucleus for European Modeling of the Ocean (NEMO) (Madec, 2008) 해양모델, Los Alamos sea ice model (CICE) (Rae et al. 2015) 해빙 모델, Joint UK Land Environment Simulator (JULES) (Walters et al. 2017) 지면 모델로 구성되며, 지면-대기-해양 플럭스 교환을 위해 OASIS 커플러로 결합되어 있다 (Kim et al. 2018). GloSea5의 대기모델 수평격자 해상도는 N216으로 위도 0.56° 와 경도 0.83° 이며 (약 ~60km), 연직으로 85개의 층을 가지고 있다. 해양모델의 수평 및 수직 해상도는 각각 $0.25^\circ \times 0.25^\circ$ 와 75층이며, 대기-해양 플럭스는 3시간 간격으로 교환 된다 (Figure 1). 특히, 계절내 예측을 위한 GloSea5 과거재현기후 (Hindcast) 실험은 1991년부터 2010년까지 총 20년 동안 매달 1일, 9일 17일, 25일 ECMWF의 ERA-Interim 자료를 초기장으로 이용하여, 3개 앙상블 멤버의 60일 적분으로 수행된다. 이 연구는 GloSea5 계절내 시간 규모 분석을 위해 Hindcast version 2018을 사용하여 기본적인 검증을 수행하였다. 또한, 실시간 예측장(Realtime Forecast)의 경우 2017년 10월에서 2018년 12월까지 NWP로부터 생산된 실시간 예측장을 초기장으로 사용하며, 과거재현기후와 마찬가지로 매달 1일, 9일 17일, 25일의 자료를 사용하였다.

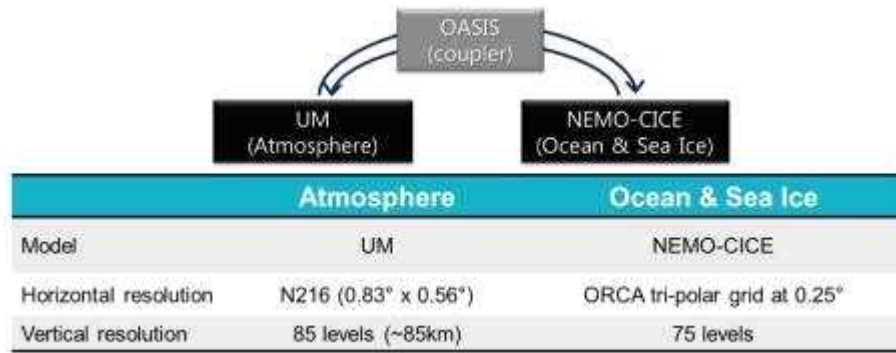


Figure 1. Model configuration of KMA-GloSea5.

(다) KMA-GloSea5 앙상블 민감도 실험 자료

본 연구에서는 GloSea5의 동아시아 기온에 대한 계절내 예보를 대상으로 현업 앙상블 분포의 특성을 살펴보고 다양한 앙상블 생산방안을 비교/평가하고자 하였다. 대상지역은 한반도를 포함하는 동아시아 지역이고 이 지역의 8주 이내 기온 예측성에 대해 평가하였다(Table 2).

Table 2. Validated target for the intercomparison of ensemble treatment.

Variable	Region	Forecast Period
Temperature	East Asia [30° N-55° N, 105° E-145° E]	1-8 weeks/ 3,4 week

① 현업후보실험자료

GloSea5 앙상블의 기본적인 특성을 살펴보기 위하여 현업 후보실험 자료를 활용함으로써 장기예측자료를 확보하였다. 후보실험자료의 생산 연도는 1991~2010년이고 매월 1일, 9일, 17일, 25일의 초기 자료가 있으며 각 초기장은 3개의 앙상블 멤버를 가지고 있다.

② 현업예보실험자료

현업 실험은 시간 지연 앙상블 기법(Time lagged ensemble)과 추계적 운동에너지 후방산란(Stochastic Kinetic Energy Backscattering version2; SKEB2)기법을 이용하여 앙상블 멤버를 구성하고 있다. 매일 4개 멤버를 생산하므로 주별 28개 앙상블 멤버를 활용할 수 있다.

③ 민감도예보실험자료

동아시아 계절내 기온 예측에 있어서 앙상블 개수, 생산 일자, 생성 기법의 변화 중 현업 실험 대비 가장 효율적인 앙상블 예측 방안 제시하고자 현업 예보 자료 외 다섯 가지 민감도

실험 자료(Figure 2)를 사용하였다.

앙상블 생산 일자에 대한 민감도 실험은 주중(목) 1일 28개, 주말(월) 1일 28개, 매일 4개(주별 28개, 현업)의 앙상블 생산 일자에 대한 실험을 말한다.

앙상블 크기에 대한 민감도 실험은 주중 1일 42개, 주말 1일 42개, 매일 6개(현업 4개 포함, 주별 42개)의 앙상블 생산 수에 대한 실험을 말한다.

민감도 실험 간 예측성 비교를 위하여 4개의 예측이 사용되었고 자세한 내용은 Table 3과 같다.

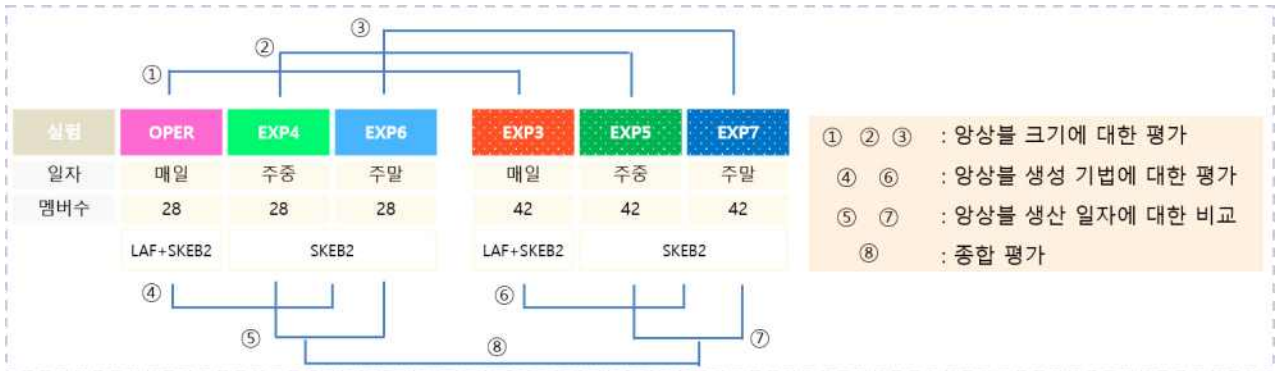


Figure 2. Overview of ensemble experiment.

Table 3. Used forecast for intercomparison of forecast skill among sensitivity experiments.

Season	Forecast day		
	Weekday	Weekend	Everyday
Spring	2018.02.15	2018.02.19	2018.02.19
Summer	2018.05.17	2018.05.21	2018.05.21
Fall	2017.08.17	2017.08.21	2017.08.21
Winter	2017.11.16	2017.11.20	2017.11.20

(2) 관측 자료

이 연구에서 모형의 예측을 평가하기 위한 비교 대상으로 ECMWF에서 제공하는 최신 재분석 자료인 ERA5를 사용하였다. ERA5는 2019년 7월에 제공 종료된 ERA-Interim을 대체하는 ECMWF의 최신 재분석자료로, 모델 물리 과정, 코어 역학 및 자료 동화 기법이 ERA-Interim보다 상당 수준 개선되었고 수평 해상도 또한 향상되었다고 알려져 있다(Hersbach et al. 2020). 사용된 변수는 기온(T2M), 500hPa 지위고도(Z500), 200hPa 바람 자료(U200, V200)이다. 200hPa 바람 자료는 200hPa velocity potential(VP200)을 사용한 MJO 지수의 검증에 활용되었다. 강수

(PREC)는 Global Precipitation Climatology Project(GPCP) v1.3(Huffman et al. 2001)을 사용하였고, 해수면 온도(SST)는 Optimum Interpolation Sea Surface Temperature(OISST; Reynolds et al. 2007)를 사용하여 검증하였다. ERA5의 공간해상도는 $0.25^{\circ} \times 0.25^{\circ}$ 로 중위도 기준 약 30km 격자 간격을 가지고 있고 다른 자료들 역시 제각각의 해상도를 가지고 있으나, 모델과의 비교를 위하여 모델과 동일한 격자인 $1.5^{\circ} \times 1.5^{\circ}$ 로 내삽한 후 일별 자료를 주 평균하여 사용되었다. 관측자료의 기후값 역시 모델과 마찬가지로 예측 시작 날짜와 동일한 날짜에 해당하는 과거재현자료와 일주일 전/후 과거재현자료의 총 3개의 평균으로 구하였고, 실시간 예측 아노말리 또한 실시간 예측에서 기후값을 제거하여 계산하였다.

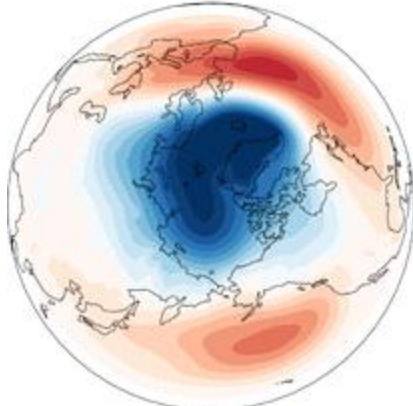
나. 기후 지수들

이 연구는 계절내 예측 모형이 구현하는 주요 기후 모드에 대한 검증 및 특성 분석을 수행하였다. 특히, 적도 지역 계절내 진동의 대표적인 현상인 매든-줄리안 진동(Madden-Julian oscillation, MJO)의 예측성능을 진단하기 위한 지수로 속도 포텐셜 MJO 지수(Velocity Potential MJO index, VPM)를 분석에 사용하였다. 지수는 200 hPa 속도 포텐셜(VP200)과 200 hPa 동서 바람장(U200), 850 hPa 동서 바람장(U850)의 다변량 경험직교함수(Multivariate Empirical Orthogonal Function, MEOF) 분석을 통해 정의한다. 각 데이터는 계절내 시간규모 진동만을 고려하기 위해 경년 변동 및 계절 변동을 제거한 아노말리값을 120일 이동평균하고 $15^{\circ} \text{N}-15^{\circ} \text{S}$ 위도 평균값을 표준화하여 생산한 일자료를 MEOF 분석에 활용한다. VPM 지수는 장파 복사(OLR)와 200 hPa 동서 바람장(U200), 850 hPa 동서 바람장(U850)으로 산출하는 실시간 다변량 MJO 지수(Real time Multivariate MJO index, RMM)에 비해 여름철 MJO 현상 및 인도양과 적도 태평양의 대류 감지, 대서양의 열대저기압 예측력이 상대적으로 높다고 알려져 있다(Ventrice et al. 2013). 예측 선행 일수에 따른 예측성능을 평가하는 이변량 상관 계수(Bivariate Correlation Coefficient, BCOR)는 RMM 지수의 예측력이 상대적으로 더 길게 유지되지만, 예측 선행 일수에 따른 오차를 평가하는 이변량 평균 제곱근 오차(Bivariate Root Mean Square Error, BRMSE)는 VPM 지수의 성능이 상대적으로 더 길게 유지되는 것으로 알려져 있다(Green and Sun, 2017).

MJO 이외에 각 지구 시스템에서 발생하는 대표 기후 모드 분석을 위해 사용된 지수들은 Table 4와 같다.

Table 4. Climate indices verified in this study. * indicates normalized value (anomalies).

Index		Definition	Reference
Hydrosphere	NINO3.4	Sea Surface Temperature (SST*) average ($5^{\circ} \text{S}-5^{\circ} \text{N}$, $170^{\circ} \text{W}-120^{\circ} \text{W}$)	
	Atlantic Equatorial Mode (AEM)	SST* average ($3^{\circ} \text{S}-3^{\circ} \text{N}$, $30^{\circ} \text{W}-0^{\circ} \text{W}$)	
Cryosphere	Arctic Temp. index 1	T2m* average	Kug et al.

	(ART1)	(70° N-80° N, 30° E-70° E)	(2015)
	Arctic Temp. index 2 (ART2)	T2m* average (65° N-80° N, 160° E-160° W)	Kug et al. (2015)
Stratosphere	Polar Vortex (PV)	50hPa zonal wind (U50*) average within 60° N-80° N	
	Quasi-Biennial Oscillation (QBO)	50hPa zonal wind (U50*) average within 3° S-3° N	Naujokat (1986) Baldwin et al. (2001)
Troposphere	East Asia Winter Monsoon (EAWM)	200hPa zonal wind (U200) gradient $\{[(30^\circ -35^\circ \text{ N}, 90^\circ -160^\circ \text{ E})-(50^\circ -60^\circ \text{ N}, 70^\circ -170^\circ \text{ E})]+[(30^\circ -35^\circ \text{ N}, 90^\circ -160^\circ \text{ E})-(5^\circ \text{ S}-10^\circ \text{ N}, 90^\circ -160^\circ \text{ E})]\}/2$	Li and Yang (2010)
	Siberia High (SH)	1000hPa geopotential height (Z1000)* average (40° -65° N, 80° -120° E)	Panagiotopoulos et al. (2005)
	Arctic Oscillation (AO)	PC1 of Z1000* poleward of 20° N, its loading pattern as below 	Thompson and Wallace (1998)

다. 연구 전략 및 방법

본 과제는 KMA-GloSea5 계절내 시간 규모 예측에 대한 현업 예측성 평가와 모형 특성 진단의 두 흐름으로 나누어 병렬적으로 진행되었다(Figure 3). 첫 번째는, 주별 예측에 대한 통합 평가프레임 개발하고 다중 모형간 비교를 통한 상대적이고 종합적인 KMA 현업 성능을 확인하는 것이다. 두 번째는, 예측을 위한 자료 세팅 및 후처리에 관한 모형 민감도를 파악하는 것이 골자이다. 특히, 초기 자료 처방에 따른 모형 오차 성장 및 예보 성능에의 영향을 진단하고, 앙상블 생산 방안별 비교를 통해 최적의 앙상블 처방 방안을 제안하고자 한다. 각 세부 과제에 대한 구체적인 수행 전략은 아래에 설명하였다.

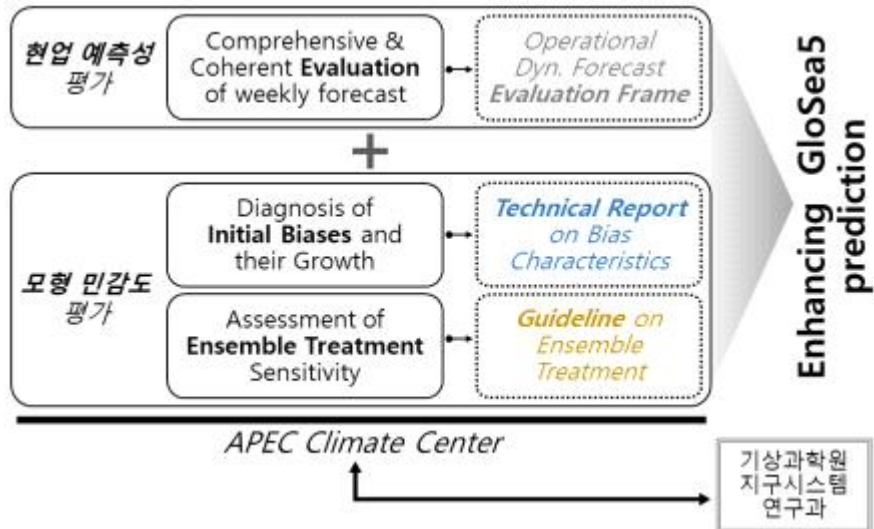


Figure 3. Project structure and execution strategy.

(1) 주평균 예측 통합 평가프레임 개발

KMA 1개월 예보에 대한 필수 메트릭 검증은 실시간으로 이루어지고 있고, 주요 모형 이슈에 대한 평가 및 진단은 연구 과제를 통해 깊이 있게 다뤄지고 있다. 하지만, 한 해 동안 예보의 질이 어떠했는지, 다른 기관 대비 얼마나 우수한지, 작년에 비해서 성능 개선이 이루어졌는지에 대한 총체적이고 정기적인 정보 수집이 이루어지지 않는 실정이다. 이를 위해 이 과제에서는 기상청 계절내 예측시스템 (KMA-GloSea5) 현업 주평균 예측을 연례적으로 평가하는 통합프레임(CREDYT:Comprehensive and Relative Evaluating Weekly Mean DYnamical Forecast)을 개발하였다. 이를 통해, 세계 우수 예측 기관 대비 KMA 현업 예측의 상대적 비교우위를 판단하고, 향후 예보 개선 도모는 물론 그 방향성을 도출할 수 있을 것으로 기대할 수 있다. 구체적인 CREDYT 개발은 아래 4개 질문에 대한 “Yes “라는 답변을 이끌어낼 수 있는 목표 및 방향으로 이루어졌다.

- ▶ KMA 현업 예보 성능에 대한 타기관 대비 상대적 검증 및 매년 지속적 추적은 이루어지고 있는가?
- ▶ 성능 변화의 원인 추정을 위한 예측 모형의 기본 역학 특성에 대한 일관되고 정기적인 진단이 이루어지고 있는가?
- ▶ 그렇다면, 현업 예보 성능 검증 및 모형 역학 특성 진단을 포괄하는 현업 주평균 예측 시스템 평가의 기본 요소는 정립되어 있는가?
- ▶ KMA 예보 역량 및 그 우수성에 대해 공감대를 이룰만한 현실적이고 객관적 정보가 생산/공유되고 있는가?

또한, CREDYT 프레임을 통한 양질의 평가 정보 제공을 위해 기존에 비해 효율적이고 새로운 형태의 검증·진단 방법론을 고안하고자 하였다. 또한, 개발된 CREDYT 프레임과 방법론을 2019년 주평균 예보에 적용해서 평가 정보를 생산하였다.

(2) 재분석 초기장 실험을 통한 초기 오차 영향 및 필수 개선 요소 파악

이 연구에서는 초기 오차에 대한 영향을 살펴보기 위해 과거재현기후와 실시간 예측장에서의 초기 오차의 특성에 대해 살펴보았다. 이를 통해, 과거재현기후와 실시간 예측장에서의 초기장 불일치로 인한 초기 오차의 영향이 계절내 예측 성능에 어떠한 영향을 주는지 살펴보고자 한다. 우선 과거재현기후에서의 초기 오차 및 계절내 예측에서의 오차 특성들을 진단하여, 시스템이 가지고 있는 초기장 및 예측장에서의 계통적 특성에 대해 살펴보았다. 또한, 과거재현기후에서의 초기장을 ERA-interim 재분석자료를 사용하고 있는 반면, 실시간 예측장은 초기장을 NWP예측장을 사용하고 있는 차이점에 주목하였다. 예측장에서의 계통적 오차의 제거를 위한 방법으로 실시간 예측장의 분석을 과거재현기후의 기후값을 이용한 편차장을 통해 예보를 하고 있음을 고려할 때, 과거재현기후와 실시간예측장에서의 편차의 특성이 다름이 초기장의 차이에서 올 수 있음을 예상할 수 있다. 이를 증명하기 위해 다음과 같은 민감도 실험을 구성하여 초기장의 차이가 계절내 예측의 편차장에 미치는 영향에 대해 알아보았다. 현재 실시간 예측장의 초기장으로 사용되고 있는 NWP 예측장과 과거재현기후의 초기장으로 사용되고 있는 ERA-interim 재분석자료를 각각의 초기장으로 설정하여 실시간 예측 시스템에 적용하여 두 실험을 구성하였다. 초기장 확보의 제한적 문제로 인해 초기 오차는 선행시간 1일의 자료로 분석하며, 계절내 예측장은 선행시간 3-4주를 기준으로 분석한다.

(3) 정확도, 유용성, 활용도 측면에서의 시간지연/후방산란 앙상블 기법의 최적 조합 제시

현업 계절내 예측 앙상블 분포의 기본 특성을 살펴보고 정확도, 유용성, 활용도 측면에서 현업 대비 가장 효과 있는 앙상블 전략 개선 방안을 제시하고자 하였다 (Table 5).

Table 5. Metrics for the diagnostic of ensemble characteristics in hindcast experiment and forecast skill in sensitivity test.

		Metric	
Diagnostics	Ensemble Characteristics	Ensemble Type	
		Ensemble Reliability	
		Error-Spread Relationship	
		Ensemble Uncertainty Estimation	
	Forecast Skill	Accuracy	Brier Score
			Ranked Probability Score
		Usefulness	Continuous Ranked Probability Score
			Relative operating characteristics
			Area under ROC curve
			Economic Value
	Return of Ratio		
	Extreme Forecast Index		
Observation	ERA5		

(가) 앙상블 특성

GloSea5의 계절내 확률 예보를 개선할 수 있는 최적의 앙상블 전략에 관한 가이드라인을 도출하기 위하여 현업 앙상블의 특성을 이해하고 민감도 실험 간 비교를 통해 정확도, 유용성, 활용성 측면에서의 개선 정도를 파악하였다.

먼저 동아시아 기온 예측에 대한 앙상블 특징을 살펴보기 위하여 현업 후보실험 결과로부터 ranked histogram, 신뢰도 지수, 앙상블 스프레드 및 flow-dependent spread를 통한 불확실성 추정 분석을 수행하였다.

① Ranked histogram

앙상블 유형은 주로 ranked histogram을 통해 진단할 수 있다. ranked histogram 분석을 통해 앙상블 예측(멤버들)과 관측이 서로 통계적으로 구별될 수 있는지 여부를 볼 수 있는데 앙상블 예측의 일관성(consistency)과 신뢰도(reliability)를 측정할 수 있다. 즉, 모형에 대한 bias를 보여주는 진단 방식으로 앙상블 예측이 정확하면 ranked histogram이 평편해진다. 앙상블이 관측의 값을 예측하기에 충분한 스프레드가 없을 때 강한 U자형 히스토그램이 만들어지고 이때 우리는 앙상블 멤버들이 underdispersion하는 특징을 가진다고 말한다. ECMWF나 GloSea5 등 현재 널리 쓰이고 있는 모형들은 대체로 앙상블이 underdispersion 하는 특징이 있다 (e.g.

Monache et al. 2006). 앙상블 유형은 크게 네 가지로 구분되고 유형별 특징이 Table 6에 정리되어 있다.

Table 6. The meaning and consistency according to ensemble types.

Type	Meaning	Consistency
Flat	ensemble spread correctly represents forecast uncertainty	consistent
U-Shape	too small spread	underdispersion (overconfident)
Dome-Shape	too large spread	overdispersion (underconfident)
Asymmetric	ensemble contains bias	negative/positive bias

② 신뢰도 지수

신뢰도 지수(Reliability Index)는 가장 이상적인 앙상블 형태인 평편한 모양의 ranked histogram으로부터 실제 ranked histogram이 얼마나 가까운지 그 정도를 나타낸 것이다 (Monache et al. 2006). 신뢰도 지수는 아래와 같이 정의한다.

$$RI = \frac{\frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \left| \frac{count_i}{N_{point}} - \frac{1}{N_{bin}} \right|}{\frac{1}{N_{bin}}} \times 100 = \sum_{i=1}^{N_{bin}} \left| \frac{count_i}{N_{point}} - \frac{1}{N_{bin}} \right| \times 100$$

③ 앙상블 스프레드

앙상블 스프레드는 예측의 분산을 가리키고 평균으로부터 앙상블 예측의 편차를 측정하여 계산한다(Zhu, 2005; 김상욱 2019). 스프레드는 아래와 같이 정의한다.

$$spread = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (\bar{f} - f(n))^2}$$

여기서, $\bar{f} = \frac{1}{N} \sum_{n=1}^N f(n)$ 은 앙상블 평균이고, f 는 앙상블 예보이다.

④ Flow-dependent spread (FDS)

대기 순환의 움직임은 보통 무질서하고 (Lorenz, 1963) 대기 순환의 발달은 초기 상태에 민감하기 때문에 수치 예보의 예측성은 초기 상태가 어떠한지에 따라 달라진다. 어떤 특정 대기 흐름 구조로부터 예측이 시작될 때 더 정확한 예보를 생산하기 쉽다. 다시 말하면, 예보는 초기 에러에 민감할 뿐만 아니라 초기 에러의 성장률은 초기 에러 자체에 비례하는데 특정 유

형의 대기 구조가 상당히 안정적이면 그로 인해 예측 가능한 부분이 많아지고, 반면 안정적이지 않으면 예측하기 힘들다는 것이 밝혀졌다. 초기 오류와 마찬가지로, 모델에 있어서 불확실성은 대기 흐름에 의존적인 역학 불안정성과 관련된 불확실성이 반영되어 순환 패턴에 영향을 미친다. 이러한 예보 불확실성의 척도를 제공하기 위해 앙상블 방안을 사용하는데 앙상블 멤버 간의 스프레드가 예측에 대한 신뢰도의 추정치를 제공한다. 어떤 날에는, 스프레드가 작을 수 있고 이는 대기가 매우 예측 가능하다는 것을 암시한다. 또 어떤 날에는 예보 앙상블이 상당히 넓게 분포할 수 있는데 이는 대기가 덜 예측 가능하다는 것을 의미한다(Ferranti et al. 2015). 고로 앙상블 스프레드는 예보 에러를 예상할 수 있는 좋은 인자이다(MacLeod et al. 2017). 각각 스프레드 구간 안에서의 모든 앙상블 예보의 평균 에러를 계산함으로써 FDS를 평가할 수 있는데 FDS 분석은 대기 흐름의 어떤 구조가 더 예측 가능한 상태를 유도하는지 그래서 더 정확한 예보를 만들어내는지 미리 알 수 있는 방법이다(Ferranti et al. 2015).

(나) 민감도 실험비교를 위한 검증 방법들

동아시아 기온 예보에 대한 GloSea5 현업 확률 예측의 품질을 평가하고 다양한 민감도 실험 결과들의 비교를 위해 통계적 검증 기법들을 사용하였다. 크게 확률 예보의 정확도, 유용성, 활용성, 그리고 극한 사상 예보에 대한 적용성 측면에서 현업 대비 가장 효과 있는 계절내 예측의 개선 방안을 찾고자 하였다.

① 정확도

Brier score는 특정 사건이 관측에서 나타나는 빈도와 그것을 예측한 확률이 일치하는 정도를 계산한 수치로써 예보의 정확도를 의미한다. 즉, 결정론적 예보에서의 평균제곱근오차를 확률 예보에 적용하였다고 볼 수 있어 완벽한 예보의 경우 그 값이 0이 된다. 식은 아래와 같다.

$$BS = \frac{1}{n} \sum_{i=1}^I N_i (y_i - \bar{o}_i)^2 - \frac{1}{n} \sum_{i=1}^I N_i (\bar{o}_i - \bar{o})^2 + \bar{o}(1 - \bar{o})$$

여기서 I 는 앙상블 크기, i 는 사건을 예보하는 앙상블 멤버의 수(확률 카테고리), N 은 검증 영역(domain) 내의 전체 지점 수, N_i 는 i 번째 멤버에 의해 예보되는 사건이 일어나는 지점의 수, \bar{O}_i 는 $\sum_{i=1}^{N_i} O_i$ 로 N_i 에서의 사건의 빈도, \bar{O} 는 사건의 전체 빈도(샘플 기후값)이다.

확률 예보 정확도를 예보 카테고리 전체에 대해서 종합적으로 살펴보기 위해 ranked probability score (RPS)를 계산하였다. RPS는 brier score를 확장시킨 기법으로 예측 확률과 관측 빈도 사이의 거리에 민감한 스코어로 식은 아래와 같다.

$$RPS = \sum_{m=1}^J (Y_m - O_m)^2 = \sum_{m=1}^J \left[\left(\sum_{j=1}^m y_j \right) - \left(\sum_{j=1}^m o_j \right) \right]^2$$

Continuous ranked probability score는 예보의 누적분포함수와 관측의 누적분포함수 사이의

차이를 진단함으로써 어떤 특정 구간보다는 예측 분포 전체에서의 정확도를 평가한다. 가능한 모든 임계값에 대해 Brier core를 적분하므로 RPS의 확장으로 해석할 수 있다.

② 유용성

예보가 얼마나 유용한가라는 측면에서 예측성 진단을 위해 예보의 적중률(hit rate)과 오보율(false alarm rate)을 통해 예보의 분포가 사건과 무사건 사이에서 잘 분류되어 있는지를 평가하였다. 이 적중률과 오보율의 관계를 그림으로 도식화한 것이 Receiver Operating Characteristic Curve이고 만약 완벽한 예보라면 이 곡선 아래 면적(AUC, Area under ROC curve)이 1이 되며 예보가 기후학적 확률과 견주어 이득을 볼 수 없다면 즉 쓸모없는 예보라면 0.5의 값을 가진다.

③ 활용성

사용자 중심의 시각에서 더 나은 예측성을 가져다 줄 수 있는 앙상블 방안은 연구자의 관점에서 진단한 그것과 다를 수 있다. 이를 찾기 위하여 사용자의 결정에 좌우되는 이득이 얼마인지 평가할 수 있는 두 가지 방법을 적용하였다.

간단한 의사 결정 모델인 비용-손실 모델을 통해 유도된 가치 진단치를 Economic Value라 한다 (Buizza et al. 2003). 사용자는 발생 가능한 손실(L)을 방지하기 위해 비용(C)을 지출하기로 결정할 수 있는데, 따라서 사건 발생 여부에 따라 사용자의 C 혹은 L의 지출이 정해진다 (Table 7). 가치인 V는 가진 비용과 손실의 비율(C/L)로 예보 사용자가 의사결정 과정에서 절약한 비용을 상대적으로 측정한 값이며, 미래 날씨에 대한 완벽한 지식이 있으면 최대값인 V=1을 얻을 수 있으며, V=0은 예측 값이 기후 정보에 대해 아무런 가치가 없음을 나타낸다. 각 사용자는 특정 날씨에 대해 서로 다른 민감도를 가지며, 이는 0과 1 사이의 값인 C/L에 의해 달라진다. 낮은 C/L 값은 주로 악기상에 민감도가 높은 사용자를 나타내며, 손실을 피하는 비용에 비해 잠재적인 경제적 손실이 높다. 사용자의 C/L 분포는 잘 알려지지 않았으나 낮은 C/L로 집중될 가능성이 있다 (Roebber and Bosart, 1996).

Table 7. Cost/Loss Decision Model.

	Event occurs	Event does not occur
User protects	C	C
User does not protect	L	0

Return of Rate (ROR)은 두 가지 다른 확률 예측의 상대적인 가치를 평가하는 진단 도구 (Hagedorn and Smith, 2009)로 예보에 대한 지식을 가진 사용자가 주어진 정보에 기인하여 어

떠한 확률로 예측에 베팅할 경우 기후값 정보 혹은 다른 예측 정보를 가진 경우에 비해 얻을 수 있는 수익률을 나타내는 즉, 도박 수익률의 측면에서 해석될 수 있는 메트릭이다. ROR은 정보 이론, ignorance라 불리는 메트릭을 사용하는데 연구용으로 흔히 쓰는 skill score에 익숙하지 않은 광범위한 예보 사용자가 이해하기 쉽고 (Good, 1952) 불확실성 속에서 실질적인 의사 결정을 위하 마련된 틀이라 여겨지고 있다. ignorance는 RPS와 달리 기준 예보인 기후값과 예보 사이의 유사성을 관측에 해당하는 확률 카테고리에서만 계산한다. RPS는 세 개의 카테고리에 대한 확률 예보 모두를 사용하고 관측에 인접한 카테고리에 더 높은 확률이 할당될 때 더 큰 점수를 얻게 된다.

$$ROR = 100 \times (2^{\langle I_g \rangle - \langle I_{g_{climo}} \rangle} - 1)$$

여기서 I_g 는 $-\log_2 P(\text{observed category})$ 이다.

④ 극한 사상 예보에 대한 적용성

Extreme forecast index는 모델의 기후 분포와 확률 예보 사이의 차이를 통해 예보관에게 잠재적인 극한 기상 사건에 대해 초기 가이드를 제공하는 도구로서 ECMWF에서 개발되었고 현업에 활용되고 있다.

3. 결과

가. 기상청 계절내 예측 시스템 (GloSea5) 예측성 평가

(1) 현업 주평균 예측 통합 평가프레임 개발

(가) 필수 평가 콘텐츠 발굴 및 구조화

기상청 계절내 예측 시스템 (KMA-GloSea5) 현업 주평균 예측 평가프레임, CREDYT의 개발은 예측 시스템 진단·평가에 있어 가장 기본이 되는 요소를 발굴하는 것으로부터 시작되었다. 실시간 예보의 다방면 성능 검증과 더불어 모형 평균 오차, 기후 리짐, 기본 역학 과정 재현성 평가 등을 다루고자 하였다. 그 결과, Figure 4처럼 네 부문에 대한 현업 예측 평가 요소들이 아래와 같이 선정되었다.

- 가. 기후값: 평년값 오차장, 시·공간적 계절성, 여름 기간, ITCZ/SPCZ 등에 대한 모형 비교 평가
- 나. 실시간 예측 검증: 기온/강수/지위고도(500hPa)에 대한 단정/확률 예측 스킬의 모형 간 비교 (개별 검증 지수 & RVMM)
- 다. Climate regime: KMA-GloSea5 내 주요 기후 지수의 연내 시계열 검증, 초기 기후 지수 위상 에 따른 예측스킬 의존도 분석, 기후 지수간/지구 시스템간 연결성 진단 등

라. Dynamical coupling: KMA-GloSea5 내 열대-중위도, air-sea, 대류권-성층권 dynamical coupling 분석 및 관측과 비교

이렇게 선정된 콘텐츠 구조화를 통해 평가프레임(CREDYT: Comprehensive and Relative Evaluating weekly mean DYnamical ForecasT)을 정립하였고, 이를 2019년 1년 52주 예보에 적용, 평가 정보를 생산함으로써 CREDYT_2019를 완성하였다.

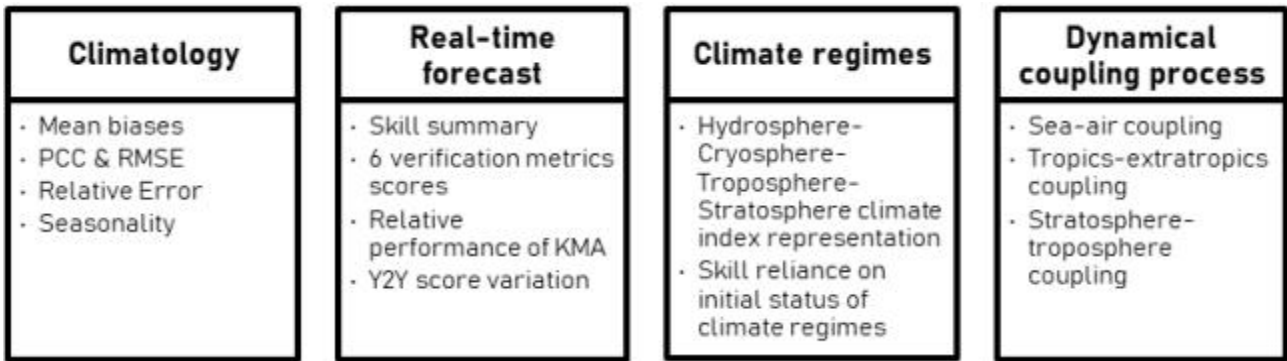


Figure 4. Four categories and their components of CREDYT framework.

(나) 새로운 실시간 예측 평가법 및 모형내 역학 특성 진단법 개발

① 순위 기반 검증 지수 합산법 (RVMM: Rank-based Verification Metrics Merging)

기후 예측 모형 기반의 역학 예측은 예보 표출 방식에 따라 앙상블 기댓값인 결정 예보와 앙상블 분포인 확률 예보로 나눌 수 있다. 또한, 이 두 방식의 예보의 정확도를 재는 다양한 검증 지수들이 존재한다. CREDYT은 평균값 기반의 결정 예보와 삼분위 기준 확률 예보값을 6개 검증지수로 평가하였다. 그 중 결정 예보 검증 지수는 ACC, MSSS이고, 확률 예보 검증 지수는 ROC, HSS, LEPS, CRPSS이다. CREDYT은 개별 모형별로 선행 주수, 계절, 도메인에 대한 검증값을 제공한다. 하지만, 개별 지수들은 각각 다른 예측의 측면을 평가하기에 적절하나, 하나의 모형 성능이 지수별 일관되지 않고 들쭉날쭉 할 수 있다. 개별 시스템의 종합적인 성능을 제시하기 위해서는 예보의 다양한 측면에 대한 평가를 포괄하도록 위의 지수들을 적절하게 합산할 필요가 있다 (Figure 4좌). 하지만, 제시한 지수들은 기준값이나 그 분산정도가 제각각이라 표준화가 쉽지 않다는 맹점이 있다. 본 연구에서는 6개 검증 지수를 합리적으로 합산하는 방법을 제시하였다. 평균이나 표준편차를 이용하여 정규화하는 에너지를 쓰는 대신, 모형 그룹에서 순위를 구하고 6개 검증 지수 순위를 합산하는 방법을 고안하였고, 이를 RVMM(Rank-based Verification Metrics Merging)이라고 명명하였다. 특히 이때, 상대적 성능을 표현하기 위해서 정수 형태가 아닌 검증 수치 기반으로 보정된 실수 순위를 선 생산하게 된다. 이렇게 함으로써 2019년 버전 CREDYT에서 비교하고자 한 5개 모형의 상대적 순위를 제시할 수 있게 된다. (Figure 5우)

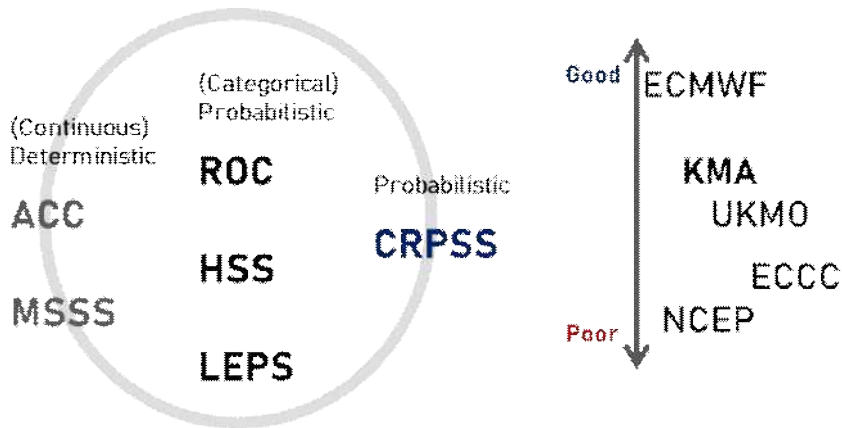


Figure 5. Schematics of (left) six verification metrics merging and (right) an example of relative performance of 5 models.

② 기후 리짐 초기 상태에 따른 예보 성능 기여도 평가법

CREDYT은 MJO를 포함한 총 10개 기후 리짐에 대한 실시간 지수 예측성 주평균 시계열 평가 정보를 기본적으로 제시한다. 하지만, 시계열 정보 뿐만 아니라, 모형 내 각 리짐의 재현 분포나 리짐 간 관련성 역시 중요한 검증 요소이기에, 이를 손쉽게 확인할 수 있는 시각화 방법을 개발하였다 (Figure 6). 과거 재현 자료를 기반으로 주평균 지수들의 분포가 선행 주수별로 얼마나 달라지는지, 관측에 비해 얼마나 차이가 나는지는 대각선 패널의 그래프를 통해 확인할 수 있다. 두 지수 분포 및 지수간 선형 회귀선은 좌하단 패널들에서 산포도 형태로, 두 지수간 상관관계값은 우상단 패널들에서 막대 그래프와 텍스트 값 형태로 표출하였다. 통계적으로 유의한 상관관계가 있을 경우 막대와 값은 진하게 표시된다.

Connectivity among Climate Regimes (1999-2010, SON)

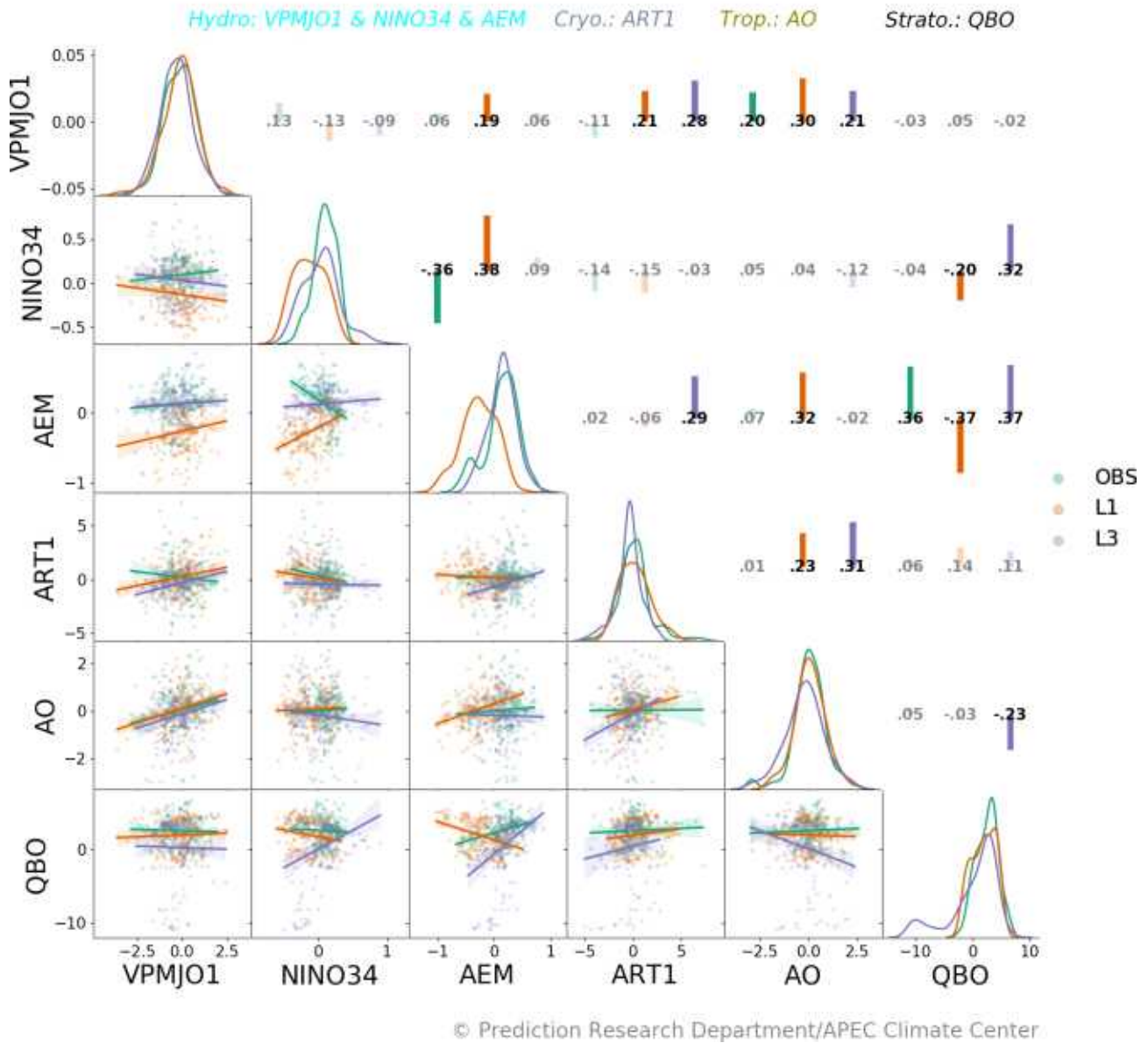


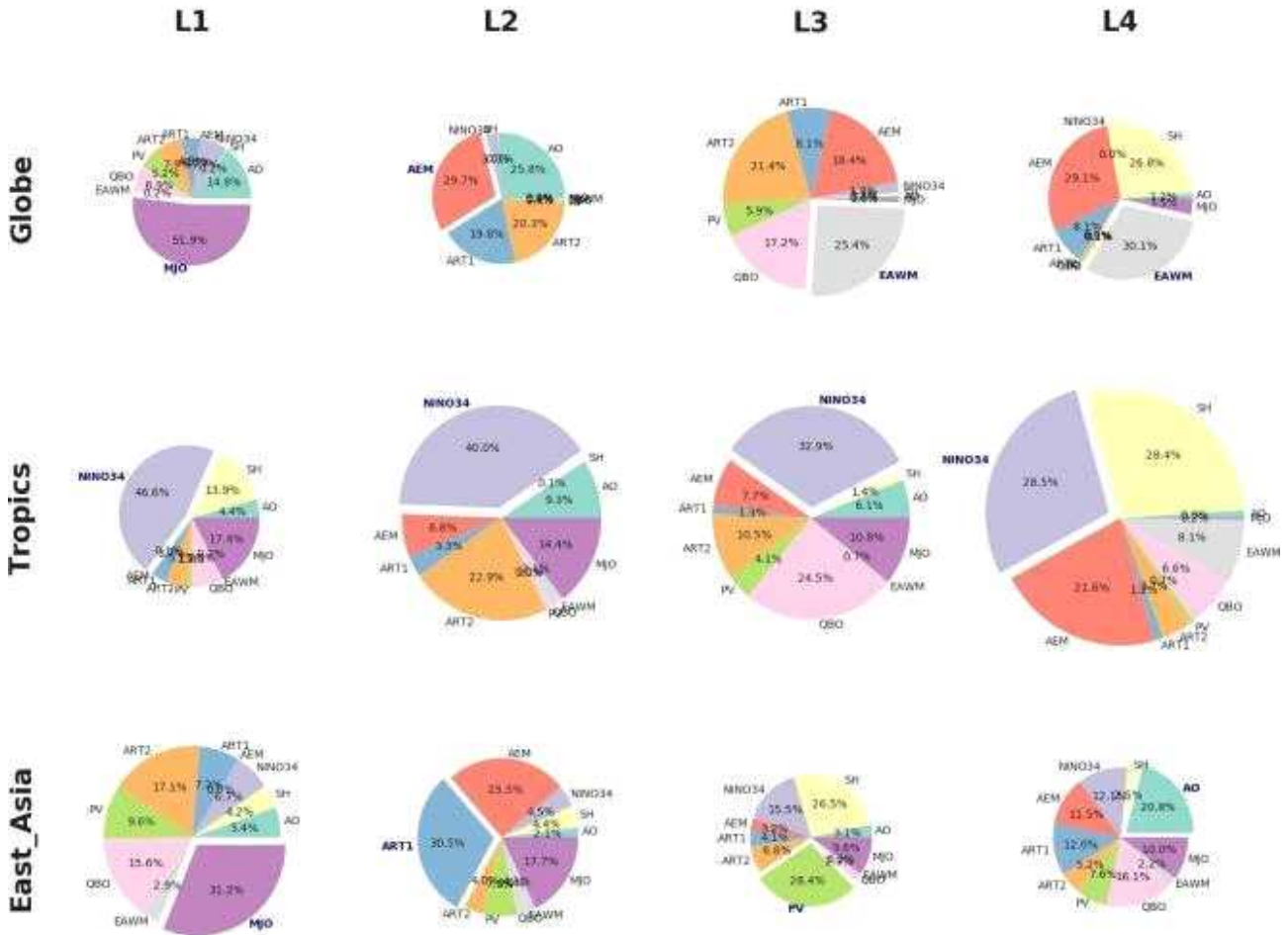
Figure 6. (diagonal) probabilistic distribution of individual climate indices for (green) the observation and (orange) 1- and (purple) 3-lead weeks GloSea5 forecasts. Two types of diagrams showing the relationship between two different climate modes: (lower left) scatter/regression line and (upper right) correlation coefficients. Dark color bar with black color value indicates the correlation coefficient is confidence with 95% level.

계절 예측 모형의 경우, 해양 강제력, 대표적으로 ENSO 상태에 따라 그 예측성이 매우 예민하다. 예를 들어, 해양 초기 자료가 강한 엘니뇨나 라니냐 상태를 포함하면, 지역별 예측은 ENSO 원격 상관을 거의 따르게 되고, 이 때 예측성능도 높은 경향을 보인다. 그렇다면, 계절 내 주평균 예측에서는 과연 어떤 기후 리짐들이 초기 조건에서 예측성을 좌지우지 하게 될까? 사실, 계절내 규모의 경우 해양 강제력이 완전히 지배한다고 하기엔 그 메모리가 매우 짧고, 날씨 예보처럼 대기 초기장 자체에 의존하기에는 그 메모리가 길다. 지면 강제력이 중요하다고

알려져 있으나, 최신 모형조차 이러한 (지면 등) 예측성의 원천을 제대로 제시하는 것 같지는 않다. 본 연구에서는 다양한 기후 리짐들의 초기 조건이 얼마나 기본 전구 예측 변수들의 예보 스킬에 상대적으로 얼마만큼 기여하는지를 산출하고 가시화하는 방안은 개발하였다 (Figure 7). 먼저 각 지수에 대해서 예보 시작일 직전 1주일 관측 지수와 도메인별 ACC 시계열의 R제곱 값을 계산하고, 이 지수들의 상대적인 비율을 파이 그래프로 표출하는 방식이다. 이 때 파이의 면적은 R제곱 전체 총합에 비례하도록 설계하였다. 각 계절에서 주요 기후 리짐을 파악할 수 있도록, 해당하는 수치를 선행 주수별 도메인별로 패널링하였다. 북반구 가을철(SON)에는 선행 주수에 관계없이 열대 지역 계절내 예측성의 주요 원천은 ENSO임을 확인할 수 있다. 동아시아의 경우, 선행 1주는 MJO, 선행 2주는 ART1의 기여도가 주요한 것으로 나타나, 기존 연구들의 결과와 일관된 면을 보인다. 하지만 열대와 달리 선행 주수에 따라 최대로 기여하는 기후 리짐이 들쭉날쭉 한데, 계절내 시간 규모임을 감안하면 이해할 수 있을 법하다. 또한, 중위도의 경우 다양한 강제력의 한꺼번에 작용하고 각각을 분리하는 것이 쉽지 않는 특성을 있기 때문에, 더더욱 이러한 변동은 당연해 보인다.

Initial predictability source of Z500 for SON

* area: proportional to R2 [ACC_Z500 & clim. index]



© Prediction Research Department/APEC Climate Center

Figure 7. Pie charts showing the relative contribution of ten climate regime strength in the 0-lead week (corresponding to -3 to +4 lead days) to weekly mean Z500 forecast ACC skill during SON season. Top contributor is displayed with a spaced pie and navy color text.

③ 지구 시스템 간 연결강도 진단법

최신 기후 예측 모형은 대기, 해양뿐만 아니라 이들과 상호 작용하는 여러 시스템을 포함하는 지구 시스템 모형의 형태로 진화하고 있다. 특히 계절내 규모 예측에 있어서 지면 과정, 담수 순환, 식생 작용 등 새로운 요소들의 중요성이 부각되고 있고, 이들의 결합을 적용했을 때의 시스템 개선 가능성에 대한 다양한 연구들이 이루어지고 있다. CREDYT은 가장 기본이 되는 대류권-성층권-수권-빙권간의 연결상태를 KMA-GloSea5 모형에서 평가하고자, 앞서 제시한 10개 기후 리짐의 지수값을 활용할 수 있는 진단법을 개발하였다. 기본적인 매트릭은 지수간의 상관 계수의 절대값이며, 이 값에 해당하는 Chord의 개수를 세팅하고 이를 바탕으로 아래와 같이 11개 지수 사이의 연결강도를 표출하는 Chord 다이어그램을 생산할 수 있다 (Figure 8). 이 때 Chord의 너비가 넓고 개수가 많을수록 두 지수간의 연결성이 강함을 의미한다. 선행

1주와 3주의 연결 상태가 관측과 얼마나 유사한지를 다이어그램 비교를 통해 확인할 수 있다. 또한 각 지수에서 출발하는 모든 chord의 총합을 누적 막대로 우측에 표시하였는데, 이를 통해 각 지수들이 타 지수와 가지는 연결 정도를 상대적으로 비교할 수 있다.

지구 시스템 간 연결성 진단을 위해서, 11개 기후 지수를 해당하는 대류권, 성층권, 대류권, 수권에 할당시키고, 각 시스템간의 연결 조합에 따라 chord 컬러를 세팅하면 Figure 8 하단 패널과 같은 그림을 얻을 수 있다. 예를 들어, 진한 초록색 Chord는 대류권-성층권 조합에 해당하며, 이 색에 해당하는 chord만 확인하여 관측과 모형간의 양상 비교가 가능하다. 이 때 각 컬러에 해당하는 코드 개수 총합으로 지구 시스템간의 연결 상태를 쉽게 비교할 수 있는데(우측, 누적 막대그래프), 북반구 가을철의 경우, 모형 선행 3주에서 시스템 간 접합 강도가 관측 대비 매우 약한 것을 확인할 수 있다. 특히, 하늘색에 해당하는 수권 내부 연결 강도가 눈에 띄게 줄어든 것이 보이는데, 이는 수권에 해당하는 NINO3.4, AEM, VPMJO1, VPMJO2 사이의 연결 강도 자체가 약화된 것일 수도 있고 혹은 관련성이 뒤집어졌을 가능성도 배제할 수 없다. 개발된 진단법이 기후 지수 간 상관 관계 부호를 인지하지는 못하기 때문에, 상관성이 marginal한 경우들이 합쳐지면서 접합 강도가 약화된 것으로 보일 수도 있기에 해석에 유의해야 한다. 접합 강도의 약화는 지구 시스템 모형의 개선의 출발 포인트가 될 수 있기에, 향후 수권 내부 연결성 약화에 대한 자세한 분석이 필요할 것으로 판단된다.

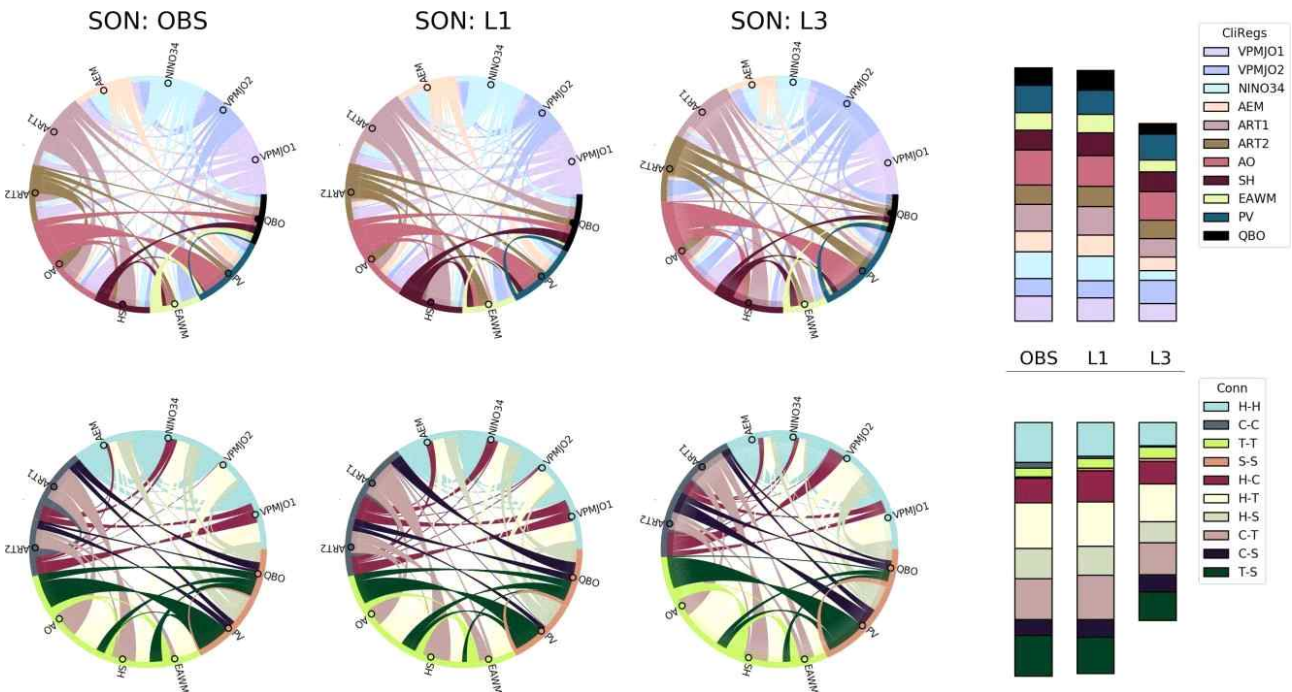


Figure 8. Chord diagram showing the strength of links among (upper) 11 climate indices and (lower) four earth systems for (the leftest) the observation, (second and third columns) 1- and 3-lead weeks forecasts. Stacked bars on the rightest column show the total number of chords (upper) ending to individual climate indices and (lower) connecting earth systems of 10 different combinations.

(2) 2019년 KMA-GloSea5 예측 종합 평가

(가) 2019년 주요 평가 정보 요약

① 기후값 재현성능 평가

㉞ Mean biases

Mean bias는 예측자료와 관측자료의 편차에 대한 시간 평균으로 아래 수식과 같이 계산되고, 기상청 계절내 예측 시스템 예측성 평가에는 실제 예보 활용도를 고려하여 목요일에 생산된 자료를 그다음 주 월요일부터 7일씩 나누어 주 평균하여(월요일-일요일) 총 4주의 주별 예측성능을 살펴보았다.

$$Mean\ bias = \frac{1}{N} \sum_{t=1}^N (\bar{F} - \bar{O})$$

Figure 9는 각 모델의 연평균 및 계절 평균 기온에 대한 선행시간 1주부터 4주까지의 mean bias이를 나타낸다. 같은 GloSea5 모델인 KMA와 UKMO에서는 겨울철에만 남극에서 warm bias가 나타나고 4주간 지속된다는 점이 두 모델에서 공통으로 나타나는 특징이다. 겨울을 제외한 나머지 계절과 연평균에 대해서는 반대로 남극에서 cold bias가 나타난다. ECCC의 경우도 KMA나 UKMO와 비슷하게 겨울철에 남극 warm bias가, 다른 계절에는 약한 cold bias가 나타나지만, KMA나 UKMO처럼 극단적인 차이를 보이지는 않는다. 반면 ECMWF에서는 겨울철 남극에 warm bias가 나타나지 않고, 전구에 전반적으로 cold bias가 나타난다. NCEP은 계절과 상관없이 전반적으로 남극에 warm bias가 존재하며 해양에는 warm bias, 대륙에는 cold bias가 약하게 존재한다.

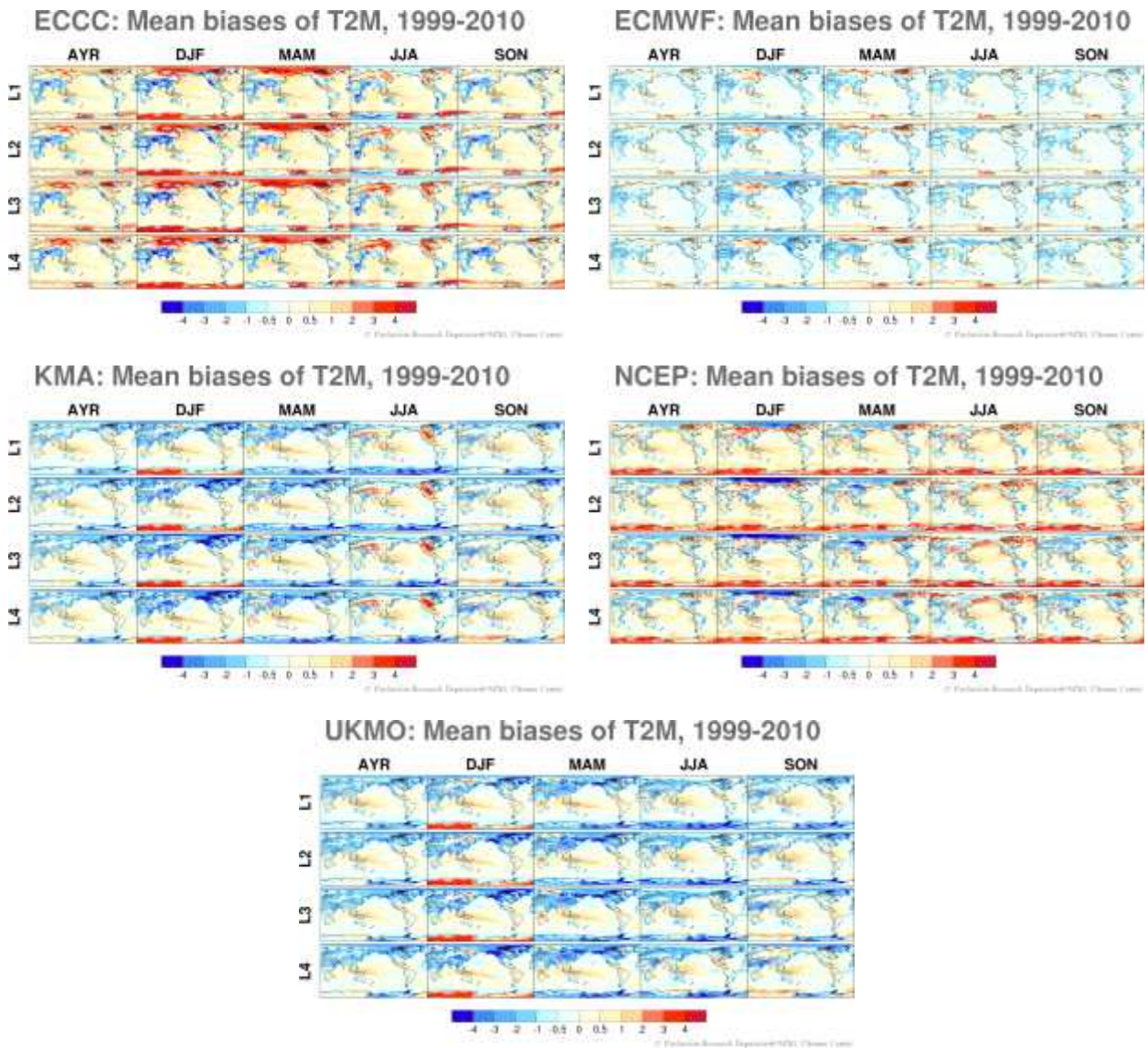


Figure 9. Annual and seasonal mean biases for T2M of ECCC, ECMWF, KMA, NCEP, and UKMO for 1- to 4-lead week during 1999-2010.

겨울철에 KMA와 UKMO에서 공통적으로 나타나는 남극 warm bias의 원인을 찾기 위해 이 연구에서 사용된 계절내 예측 모델들의 대기 초기조건을 살펴보았다(Table 8). ECMWF는 ERA5 재분석자료를 대기 초기조건으로 사용하는데, 이는 이 연구에서 사용하는 관측자료와 동일하다. KMA와 UKMO에서는 대기 초기조건으로 검증에 사용한 ERA5가 아닌 ERA-Interim 재분석자료가 사용된다는 공통점이 있다. Gossart et al. (2019)은 다양한 재분석자료에 따른 남극 지역 지표변수들의 차이점을 살펴보았는데, 겨울철 남극 기온이 ERA-Interim이 ERA5에 비해 약 5°C 정도 따뜻한 특징이 있었다(Figure 3 in Gossart et al. 2019). 이 연구에서 사용된 자료 기간(1999-2010년)에 대한 ERA-Interim과 ERA5의 기온 차이에서도 역시 겨울철 남극 기온은 ERA-Interim이 더 따뜻하다는 점을 알 수 있다(Figure 10). 따라서 ERA-Interim을 대기 초기조건으로 사용한 KMA와 UKMO의 겨울철 남극 기온이 관측자료인 ERA5에 비해 따뜻하게 모의될 수밖에 없다. 이 두 모델 외에 ERA-Interim을 대기 초기조건으로 사용하는 모델은 ECCC로,

이 모델 역시 겨울철 남극 기온에 warm bias가 존재한다. 나머지 모델인 NCEP은 대기 초기조건으로 NCEP에서 제공하는 CFSR이 사용되었는데, CFSR은 계절에 상관없이 ERA5에 비해 남극 기온이 따뜻하고, 특히 여름철에 가장 큰 차이를 보여주었다. 이에 따라 NCEP의 남극 기온은 전반적으로 warm bias를 가지고, 여름철에 가장 큰 warm bias를 보여준다. 따라서 어떤 재분석자료를 관측 참고자료로 사용하느냐에 따라 mean bias는 달라질 수 있다. 재분석자료에 따른 mean bias의 차이는 단순히 남극 기온에서만 국한되지 않고, 북반구 전반적으로도 재분석자료에 따라 오차의 크기가 달라질 수 있다.

Table 8. Atmospheric initial condition for each S2S model used in this study.

Model	KMA	ECCC	ECMWF	NCEP	UKMO
atmospheric initial condition (source)	ERA-Interim (ECMWF)	ERA-Interim (ECMWF)	ERA5 (ECMWF)	CFSR (NCEP)	ERA-Interim (ECMWF)

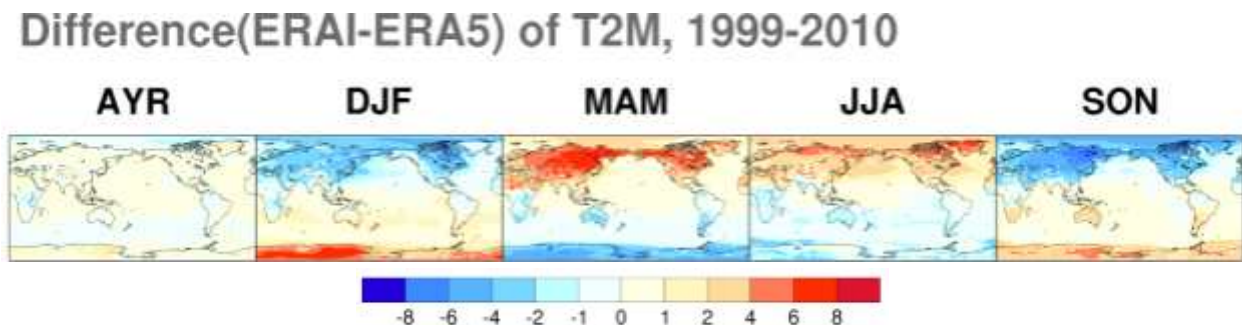


Figure 10. Differences of ERA-Interim and ERA5 for annual and seasonal mean climatology during 1999-2010.

㉔ Relative Error

이 연구의 가장 큰 목적은 앞서 언급한 대로, KMA 모델을 다른 모델과 비교하여 상대적으로 우수한지 아닌지를 비교하는 것이다. 기후값 재현성능은 예측성능에 직접적인 영향을 미치는 것은 아니다. 그럼에도 불구하고, 기후값을 재현하는 모델의 능력은 예측성능에 상당한 영향을 미칠 수 있다(Kang et al. 2004; Wang et al. 2009). 따라서 타 모델 대비 KMA 기후값 재현성능 오차의 비교우위를 살펴보기 위해 Root Mean Square Error(RMSE) 관점에서의 상대적 재현성능 오차 차이를 살펴보았다. 이는 2019년 APCC 연구과제에서 제안했던 방법으로 아래와 같은 수식을 사용하여 계산된다. 변수별, 모델별로 가로축은 계절(연평균, 봄/여름/가을/겨울 평균)을 세로축은 지역(전구, 동아시아) 평균을 나타낸다. 위쪽 삼각형에서부터 반시계방향으로 진행되면서 선행시간이 증가하며 위, 왼쪽, 아래, 오른쪽 삼각형은 각각 선행시간 1, 2, 3, 4주

를 의미한다. 상대적 RMSE가 양의 값을 가질 때는 타 모델과 비교해 KMA의 재현능력이 좋을, 음의 값이면 KMA의 상대적인 재현능력이 나쁠을 의미한다(김가영 등 2019). 또한, 상대적 RMSE 차이가 신뢰할 만한 값인지 판단하기 위해, 상대적 RMSE 차이가 KMA 기후값 RMSE 변동성(표준편차)의 ± 1.5 배를 초과할 때 신뢰할 만한 차이가 있는 것으로 하였다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (F - O)^2}$$

$$Relative\ diff.\ of\ RMSE(\%) = \frac{RMSE_{other} - RMSE_{KMA}}{RMSE_{KMA}} \times 100$$

Figure 11은 계절별, 지역별로 기온에 대한 KMA 기후값의 선행시간에 따른 상대적 RMSE 차이를 보여준다. ECMWF 기온은 계절과 지역, 선행시간에 상관없이 KMA에 비해 약 20~60% 정도 우수한 기후값 재현능력을 보여준다. 반면 KMA는 ECCC, NCEP, UKMO에 비해 전반적으로 비슷하거나 약간 우수한 성능을 보여준다. KMA가 모의하는 기온의 기후값 재현성능은 가을철 동아시아에서 ECCC, NCEP, UKMO에 비해 우수하며, ECCC 대비 KMA의 기후값 재현성능은 특히 선행시간 4주에 약 40~60% 정도 우수하다.

RMSE of T2M, 1999-2010

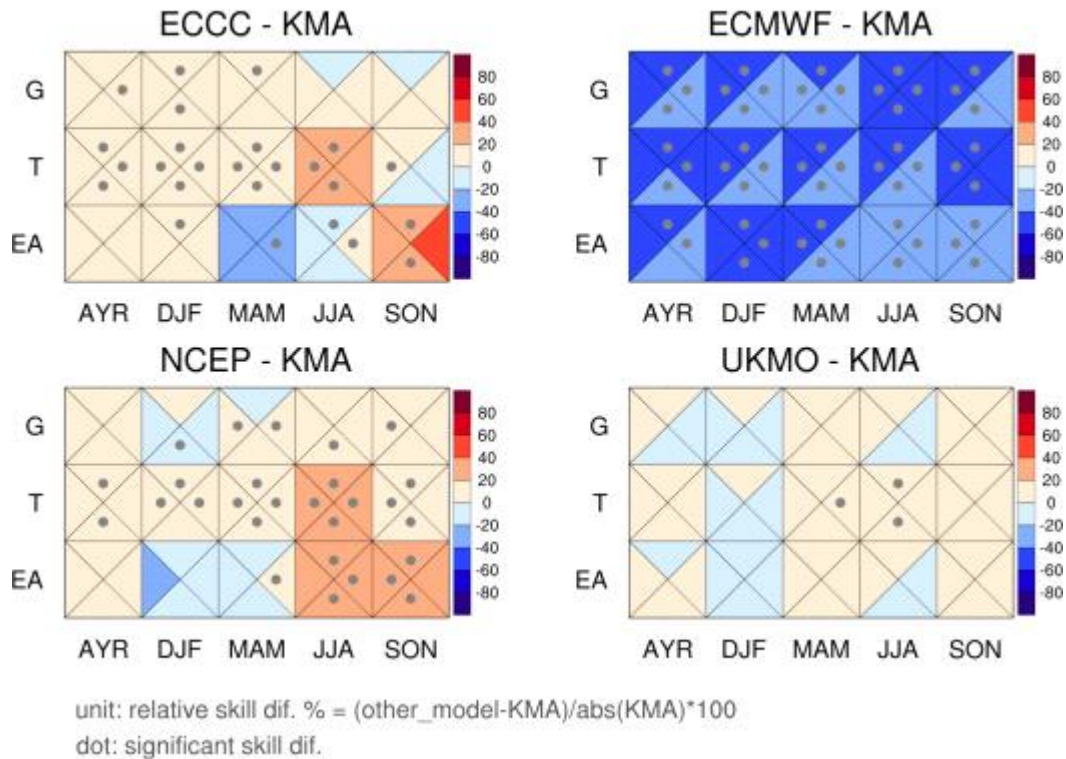


Figure 11. Annual and seasonal averaged relative RMSE difference of climatology of T2M over Globe, Tropics, and East Asia of ECCC, ECMWF, NCEP, and UKMO with KMA. The top, left, bottom and right triangles represent for 1-, 2-, 3-, and 4-lead week, respectively. Gray dot indicates that relative difference is significantly large.

㉔ 여름 기간 연장 및 ITCZ/SPCZ 과대모의

선행 1-4주 모형이 재현하는 시·공간적 계절성을 확인하기 위해, 전구, 열대, 동아시아 세 지역에 대해서 지표 평균 기온의 시계열을 관측과 비교하였다 (전구 케이스, Figure 12). 특히 관측 상위 25%에 해당하는 값을 기준으로 여름 날짜를 세었을 때, KMA-GloSea5를 포함한 세계 모형이 여름 기간이 연장된 것으로 파악된다. 특히 NCEP의 경우는 선행 1-2주에서 무려 16주에 해당하는 매우 긴 여름 기간을 보인다. 하지만 반대로, ECMWF는 전체 선행 주에서 9주의 짧은 여름 기간을, UKMO-GloSea5는 관측과 동일한 기간을 보였다. 여름 기간이 연장된 세 모형과 ECMWF를 비교하면, 모형별 전구 평균 오차가 이러한 여름 길이 변화에 대부분을 기여하는 것으로 보인다. 하지만, ECCC의 최고 온도 값과 비대칭성 등을 봤을 때 반드시 그렇지는 않음을 알 수 있다. 계절내 예측 성능에서 독보적인 ECMWF의 계절 변동이 관측 및 타 모형보다 약한 것을 주목할 만한 정보라고 생각된다. 또한, 같은 GC3 시스템인 KMA와 UKMO의 계절성이 크게 다른 점은 꼭 짚고 넘어가야할 포인트이다. 앞서 평균 오차 분석에서 여름철 (JJA)만 북반구 아시아, 아메리카 대륙에서 두 시스템이 크게 다른 것과 직접적으로 연관된 것으로 보인다. 향후, 무엇이 이러한 차이를 불러왔는지에 대해서 대기/해양 초기자료 처방 등

reforecast setting을 직접 확인하는 것으로부터 분석이 시작될 수 있을 것이며, 2020년 버전에서 이러한 차이가 재생산되는지도 확인할 필요가 있다.

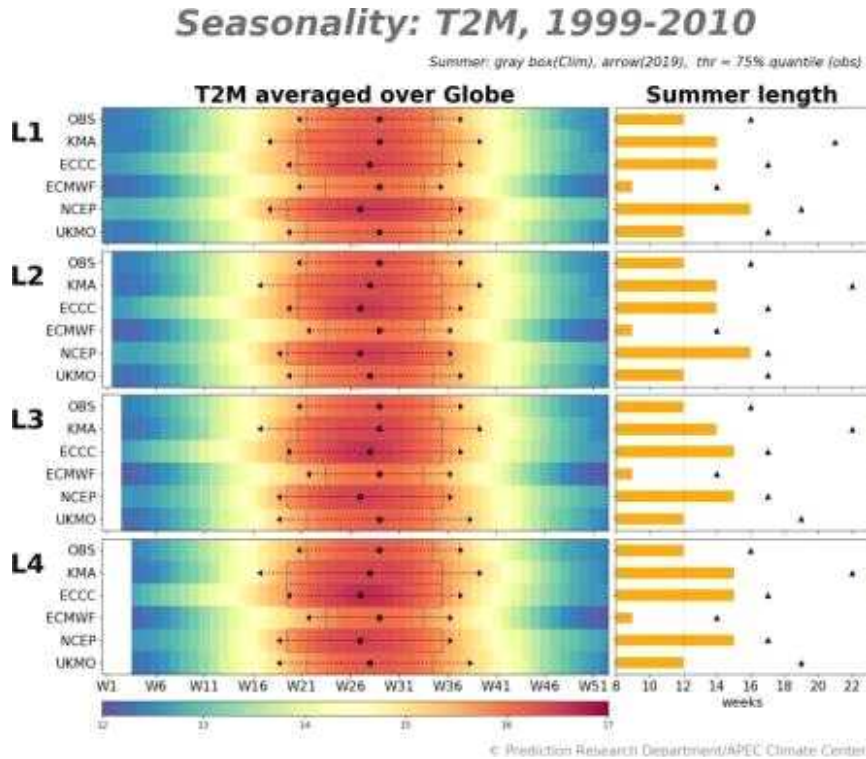


Figure 12. (left) Global averaged T2m seasonal cycle averaged for the period of 1999-2010 for 1- to 4-lead weeks forecasts. The summer is defined the period whose temperature exceeding top 75% temperature in the observation. The beginning and ending days of the summer and its length is displayed as gray box in the left and orange horizontal bar in the right. Navy arrow shows the value for the year 2019.

모형이 재현하는 전구 강수 시스템의 시·공간 변동성은 Inter-Tropical Convergence Zone(ITCZ)/Southern Pacific Convergence Zone(SPCZ)의 계절 swing의 평균 상태에서부터 확인 가능하다(Figure 13). 관측에서 나타나는 북반구 여름철 ITCZ 확장/SPCZ 축소와 겨울철 ITCZ 축소/SPCZ 확장의 시·공간적 스윙이 GloSea5에서도 적절히 모의되고는 있으나, 그 강도가 매우 강하다(3행, Figure 13). 이는 GloSea 계열 모형의 전반적인 wet bias 경향과도 일관된다. 각 모형들에서 ITCZ/SPCZ 최성기의 주 수와 위도를 확인해 관측과 비교해 보았다(4행 좌, Figure 13). ECCC를 제외하고 선행 주수별로 큰 차이를 보이지는 않는다. non-coupled 모형인 ECCC는 선행 3-4주에서 ITCZ 최성기가 매우 이르게 고위도에서 나타나는 오차를 보인다. 그 외 다른 모형들은 관측과 유사한 시기/위도를 보이나, SPCZ의 경우 뚜렷하게 수 주 지연되어 남쪽에 나타나는 것을 확인할 수 있다. 특징적인 것은, ECMWF의 강수밴드 시·공간 변동 강도가 극도로 작게 나타나는데, 이는 80° E-80° W 경도 평균을 했을 때 SPCZ의 최성기가 구분되지 않고 오히려 북반구에 극값이 인식되는 문제가 나타난다는 점이다. 실시간 계절내 예측 성능면

에서 최고인 ECMWF 모형이 SPCZ 계절 스윙에 있어서는 다른 모형보다 오차가 큰 것이 흥미롭다. 하지만, 경도 평균 패턴에 대해서는 ECWFM 모형이 가장 작은 오차(RMSE)를 보이는데(4행 우, Figure 13), 강수 밴드의 남북 방향 이동은 뚜렷하지 않지만 전반적인 강수 강도를 관측과 유사하게 재현하기 때문으로 판단할 수 있다.

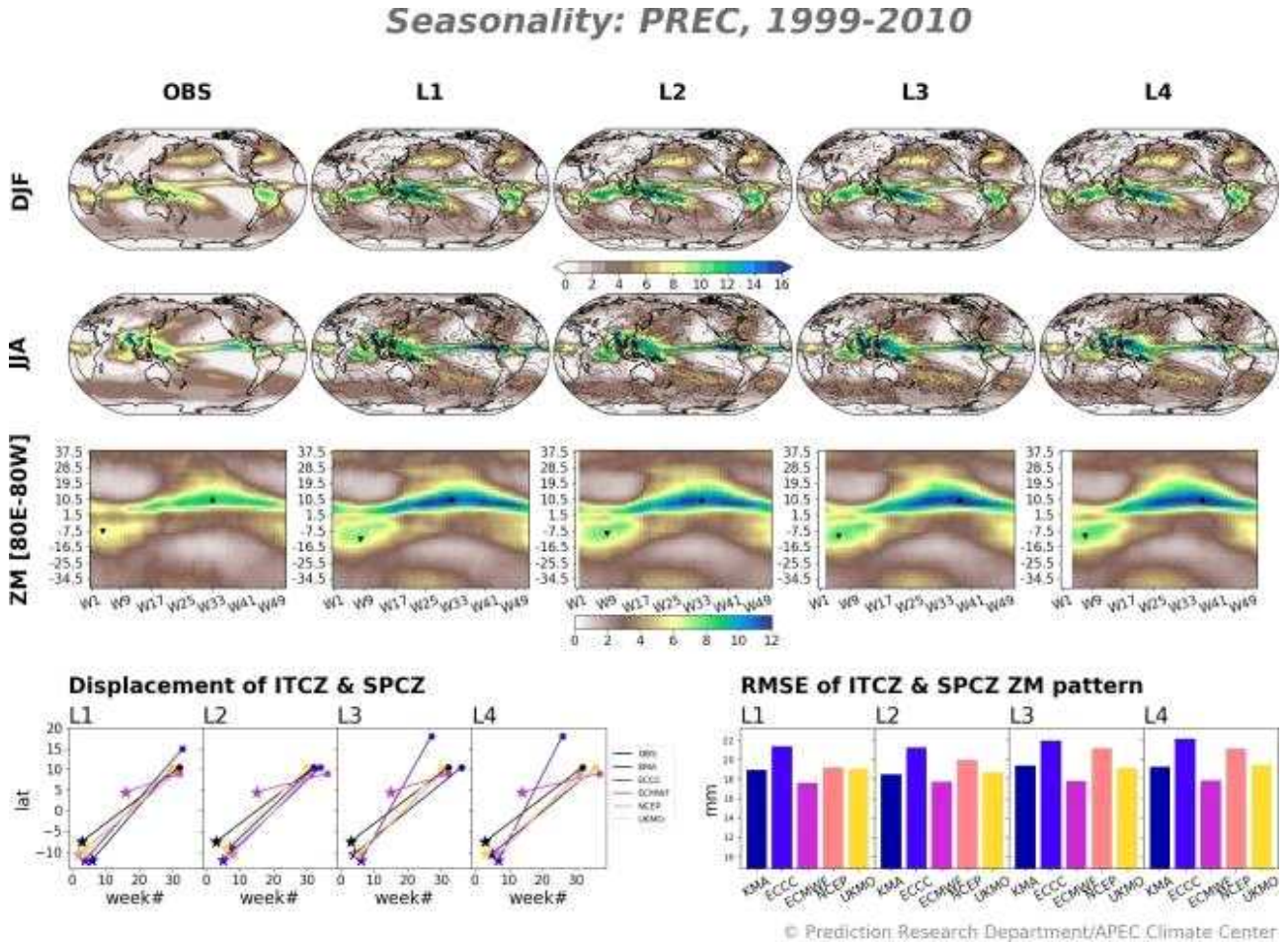


Figure 13. GloSea5 long-term (1999–2010) averaged precipitation showing ITCZ/SPCZ pattern for (1st row) DJF and (2nd row) JJA season. (3rd row) temporal and latitudinal migration of Pacific basin [80 degE to 80 degW] averaged precipitation and (bottom right) the pattern error (RMSE) of five models relative to the observation. (bottom left) timing and latitude of the maximum ITCZ/SPCZ strength.

② 2019년 52주 실시간 예측 스킬 평가

KMA의 2019년 실시간 예측자료의 주별 예측에 대해 타 모형 대비 비교우위를 살펴보고자 한다. Anomaly Correlation Coefficient(ACC)는 결정론적 예측성능을 판단하기 위해 가장 널리 쓰이는 검증 기법으로, 아래와 같은 수식을 사용하여 변수별 ACC의 상대적인 차이를 계산하였다. 이 또한 2019년 연구과제에서 제안했던 방법으로(김가영 등 2019), 앞서 살펴본 기후값 RMSE의 상대적 차이에서와 동일한 방법으로 상대적 ACC 차이가 KMA 실시간 예측에 대한 ACC 변동성(표준편차)의 ± 1.5 배를 초과할 때 신뢰할 만한 차이가 있는 것으로 판단하였다.

$$Relative\ diff.\ of\ ACC(\%) = \frac{ACC_{KMA} - ACC_{other}}{ACC_{KMA}} \times 100$$

Figure 14에는 기온에 대한 KMA 실시간 예측의 상대적인 ACC 차이를 나타내었다. 계절내 예측에서 주로 활용이 기대되는 선행시간 3-4주를 중심으로 살펴보면, ECMWF의 계절내 예측 성능은 타 모델보다 우수한 것으로 알려져 있는데(문자연 등 2018; 김가영 등 2019), 2019년의 실시간 예측성능 역시 ECMWF는 KMA에 비해 거의 모든 계절과 지역에서 비교우위에 있음을 알 수 있다. 이는 앞서 제시했던 기후값 재현성능 오차와도 일치하는 결과이다. 같은 GloSea5 모델인 UKMO와 비교해보면, KMA의 2019년 실시간 예측성능은 다소 낮은 경향이 있다. 그러나 가을철 동아시아의 예측성능은 3-4주 모두 UKMO에 비해 약 25~50% 정도 우수한 성능을 보여준다. 또한 KMA는 ECCO와 NCEP과는 거의 비슷하거나 다소 우수한 예측성능을 보인다. 특히 ECCO 대비 KMA의 동아시아 기온 예측성능은 겨울철을 제외하고 높은 수준의 예측성능 비교우위를 보여준다. 특히 가을철의 경우는 선행시간 4주에서 100% 이상 성능이 우수하다. 전반적으로 KMA의 선행시간 3-4주 가을철 동아시아 예측성능은 ECMWF를 제외한 다른 모형들에 비해 우수하다. 이는 기후값 재현성능 오차와도 유사한 결과라고 할 수 있겠다.

2019: ACC of T2M

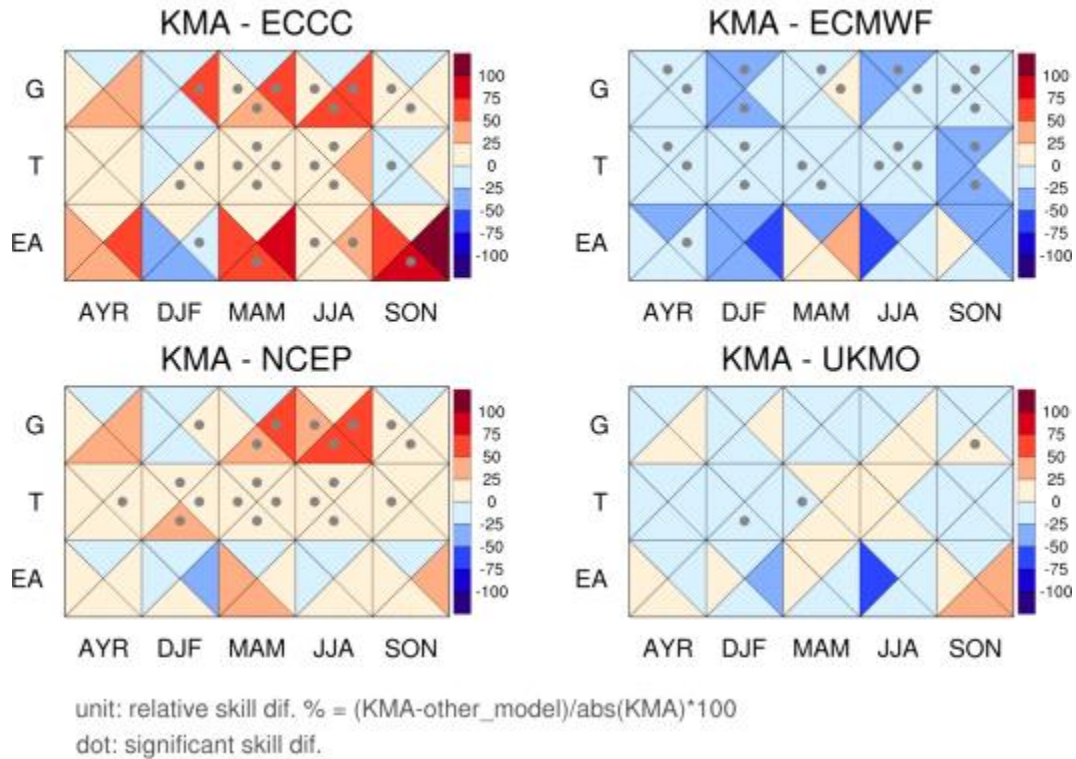


Figure 14. Annual and seasonal averaged relative ACC difference of 2019 real-time forecast of T2M over Globe, Tropics, and East Asia of ECCC, ECMWF, NCEP, and UKMO with KMA. The top, left, bottom and right triangles represent for 1-, 2-, 3-, and 4-lead week, respectively. Gray dot indicates that relative difference is significantly large.

㉓ 주별 MSSS 검증

2020년 연구과제에서는 주별 결정론적 예측성능을 판단하기 위한 검증 기법으로 ACC 외에 Mean Square Skill Score(MSSS)를 추가하였다. MSSS는 World Meteorological Organization/Commission for Basic System(WMO/CBS) 회의를 통해 표준화된 검증체계로서, 계절 규모의 예측성능 검증에 자주 사용되는 검증 기법이며(Choi et al. 2016), 계절내 예측성능 검증에도 활용된다(Jung et al. 2016; Lim et al. 2018; Ham et al. 2017). 이는 예측값이 가지고 있는 관측값과의 오차 특성을 종합하여 파악하는 평가 지표로써, 각 격자값으로 정확도의 분포를 파악하거나 영역 평균 시계열로부터 정확도 추이 분석이 가능하다. $-\infty$ 에서 $+1$ 사이의 값을 가지며, MSSS가 1이면 완벽한 예보임을 뜻한다.

MSSS는 Mean Squared Error(MSE)를 사용하여 계산한다. 아래 수식에서 MSE_j 는 예측값이 관측값으로부터 가지는 오차를 제공하여 평균한 것이고, MSE_{c_j} 는 기후값을 예측값으로 이용할 경우의 MSE를 의미한다. 이 연구에서는 실시간 예측능력에 대한 MSSS를 구하고자 하므로, MSE_j 는 실시간 예측자료와 이와 동일한 기간에 대한 관측자료로부터 구해지고, MSE_{c_j} 는

reforecast 기간(1999-2010년)에 대한 관측값으로부터 구해진다. 관심있는 영역에 대해 각 격자점의 MSSS를 평균하면 MSSS 총점(overall MSSS)를 구할 수 있는데, 격자점별 위도(θ)에 따른 면적 가중($\cos\theta$)을 고려하여 계산한다. 완벽한 예보일 때 MSSS는 1의 값을 갖고, 그 값이 낮을수록 예측성이 떨어짐을 의미한다. 관측의 변동성에 비해 MSE_j 가 커지면 MSSS는 하한 없이 낮아진다는 점에서 ACC와 차이를 보인다. MSSS가 실시간 예측자료의 오차와 관측의 변동성이 같아지는 값인 0 이상일 때 예측성이 있다고 할 수 있다(Kim et al. 2018).

$$MSSS_j = \frac{MSE_{c_j} - MSE_j}{MSE_{c_j}} = 1 - \frac{MSE_j}{MSE_{c_j}}$$

$$MSE_j = \frac{1}{N} \sum_{i=1}^N (F_{ij} - O_{ij})^2$$

$$MSE_{c_j} = \left(\frac{N}{N-1}\right)^2 \frac{1}{N} \sum_{i=1}^N (OH_{ij} - \overline{OH_j})^2$$

$$MSSS = 1 - \frac{\sum_{j=1}^n \cos\theta_j MSE_j}{\sum_{j=1}^n \cos\theta_j MSE_{c_j}}$$

Figure 15에는 기온에 대한 모델별 연평균 또는 계절 평균된 MSSS의 지역별 평균값을 나타내었다. 계절이나 지역에 상관없이 ECMWF는 선행시간 1주에서 양의 값을 보이기 때문에 예측성능이 있다고 판단할 수 있고, 다른 모델들은 대부분 음의 값을 가지므로 예측성능이 선행시간 1주에서도 다소 떨어진다고 할 수 있다. 이후 선행시간 2주부터는 모든 모델의 예측성능이 계절이나 지역에 상관없이 음의 값이 되며 선행시간에 따라 점차 예측성능이 감소한다. ECCO를 사용한 동아시아 예측의 경우, 계절내 예측에서 관심 있는 선행시간 3-4주에서 MSSS가 -2 이하로 분석에 사용한 모델 중 가장 낮은 예측성능을 보인다.

2019: MSSS of T2M

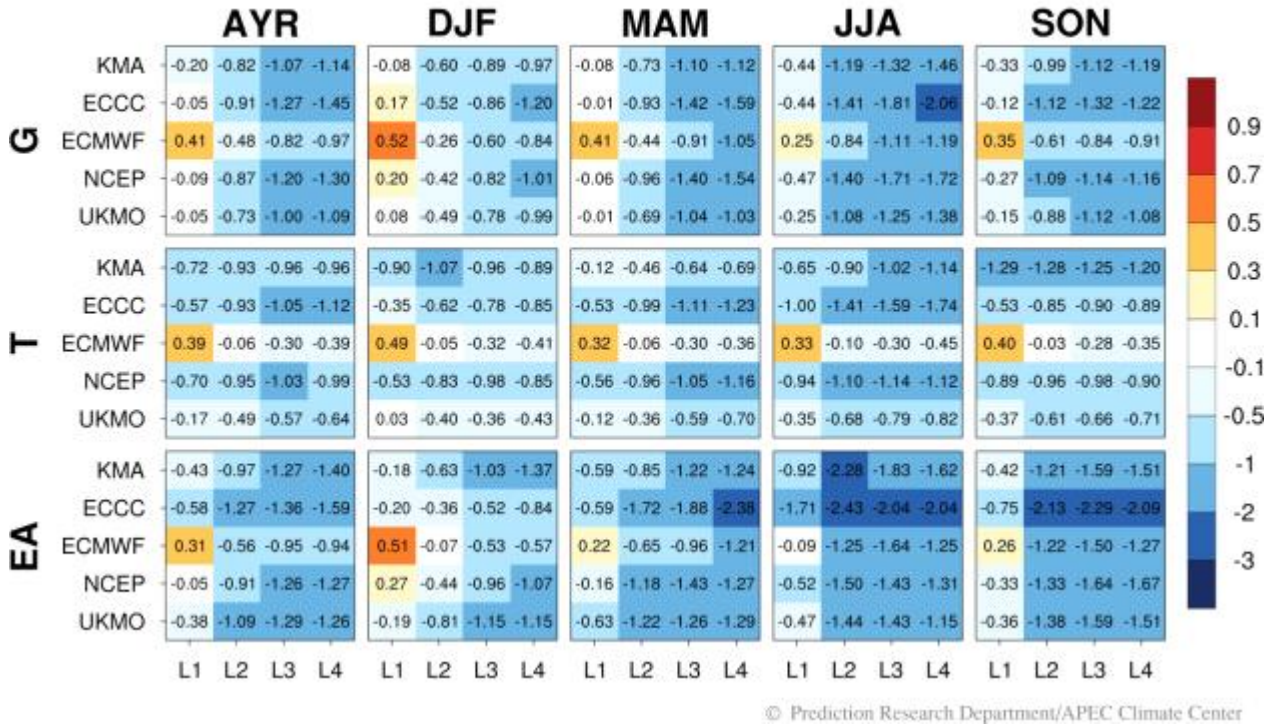


Figure 15. Annual and seasonal averaged overall MSSS of 2019 real-time forecast of T2M for 1- to 4-lead week over Globe, Tropics, and East Asia of KMA, ECCC, ECMWF, NCEP, and UKMO.

Figure 16에는 기온에 대한 KMA 실시간 예측의 상대적인 MSSS 차이를 제시하였다. 앞서 보인 바와 같이 계절과 변수, 선행시간에 상관없이 KMA의 MSSS는 대부분 음수이므로 부호가 바뀌어 반대로 해석되는 것을 방지하기 위해 분모에 절대값을 취한 후 계산을 수행했다. 자세한 수식은 아래에 나타났다.

$$Relative\ diff.\ of\ MSSS(\%) = \frac{MSSS_{KMA} - MSSS_{other}}{|MSSS_{KMA}|} \times 100$$

MSSS 역시 ACC와 마찬가지로 ECMWF는 KMA보다 선행시간 1주 차에는 월등히 좋은 성능을 보여주고, 특히 전구에서는 이 차이가 신뢰할 만큼 크다고 판단된다. 선행시간이 지남에 따라 차이는 점차 줄어들게 된다. KMA는 ECCC에 비하면 약간 우수한 성능을 보여주는 편이다. 특히 겨울철을 제외하면 동아시아 기온 예측에 있어서 선행시간에 상관없이 KMA가 ECCC보다 예측성능이 좋다. NCEP에 비하면 KMA는 동아시아 3-4주 기온 관점에서 봄과 가을철에 예측성능이 우수함을 알 수 있다. UKMO와 비교하면 전반적으로 KMA의 예측성능이 낮은 편이나, 신뢰할만한 차이는 아닌 것으로 판단된다. 그럼에도 불구하고, 동아시아 봄철의 기온은 KMA가

UKMO보다 예측성능이 약간 우수하다.

2019: MSSS of T2M

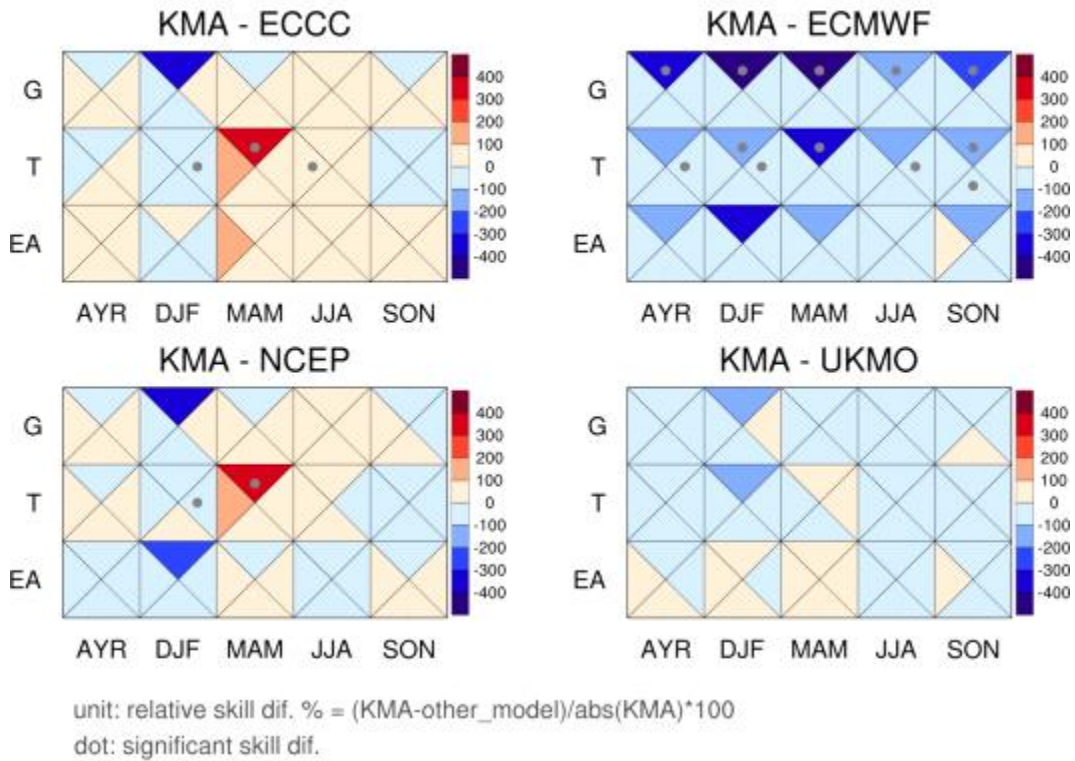


Figure 16. Annual and seasonal averaged relative MSSS difference of 2019 real-time forecast of T2M over Globe, Tropics, and East Asia of ECCC, ECMWF, NCEP, and UKMO with KMA. The top, left, bottom and right triangles represent for 1-, 2-, 3-, and 4-lead week, respectively. Gray dot indicates that relative difference is significantly large.

㊤ 일별 실시간 예측성능

앞서 언급한 내용은 모두 2019년 실시간 예측에 대한 주별 예측성능을 보여주었다면, 이 절에서는 일별 예측성능을 보여주고자 한다. 이때 주별 예측성능에 사용된 ACC와 MSSS를 일별 예측성능에 적용해보고자 했다. 그러나 MSSS는 극한 강수 등의 이유로 일별 실시간 검증에 적합하지 않다고 판단되어 ACC만 사용하였고, MSSS가 일별 예측성능 검증에 부적합한 이유에 대한 자세한 설명은 이 절 마지막에 기술하겠다.

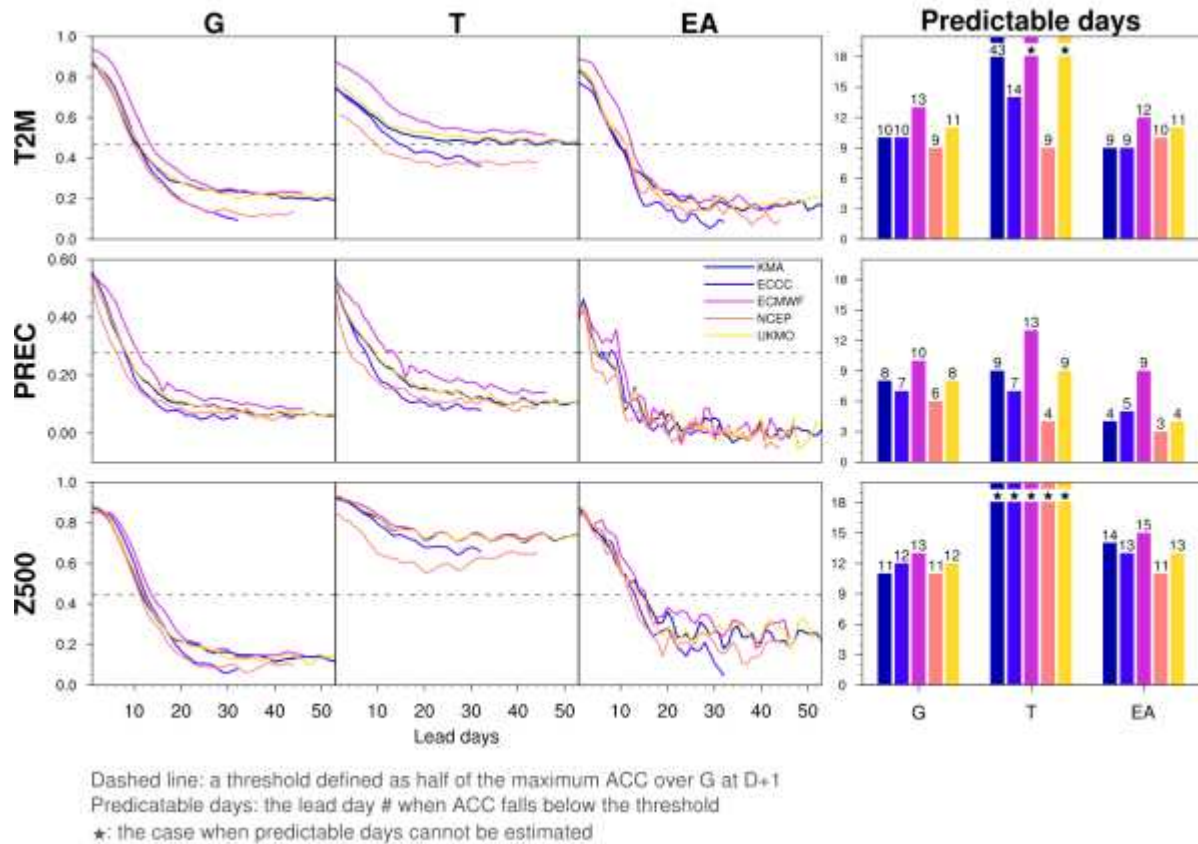
ACC를 이용한 2019년 실시간 예측 검증 결과를 Figure 17에 나타내었다. 왼쪽 9개의 시계열은 각 모델의 지역별, 변수별 선행시간에 따른 ACC를 보여준다. 여기서 예측 가능일(predictable days)를 판단하기 위해서는 특정한 임계값이 필요하다. MSSS의 경우에는 주로 예

측성능이 있다고 판단되는 0을 기준으로 MSSS가 0 이하가 되는 선행시간을 예측 가능일로 정의하는데(Kim et al. 2018), ACC는 예측 가능일을 정의하는 기준이 따로 있지는 않았다. 따라서 이 연구에서는 변수별로 선행시간 0일의 ACC 최대값의 절반에 해당하는 값을 임계값으로 정의하고, 이 임계값의 이하가 될 때를 예측 가능일로 정의했다. 이렇게 계산된 각 모델의 변수별, 지역별 예측 가능일을 오른쪽에 나타냈다. ★은 ACC가 임계값 이하로 떨어지지 않아서 예측 가능일을 정의할 수 없는 경우를 나타낸다. 전구 예측 가능일은 기온과 500hPa 지위 고도가 거의 비슷한 수준을 보이고, 9~13일의 범위를 보여준다. 이 중 ECMWF의 예측 가능일이 이 변수들에서 모두 13일로 가장 길다. 전구 강수의 경우는 두 변수에 비해 예측 가능일이 약간 짧는데 6~10일의 범위를 나타내고, 역시 ECMWF가 예측 가능일이 10일로 가장 길다. KMA의 전구에 대한 예측 가능일은 기온 10일, 강수 8일, 500hPa 지위고도 11일이다.

적도 기온의 예측 가능일은 NCEP이 9일로 가장 짧고 KMA가 43일로 가장 길다. ECMWF와 UKMO의 적도 기온은 선행시간 내에 임계값 이하로 떨어지지 않기 때문에 예측 가능일을 정의할 수 없었다. 적도 500hPa 지위 고도는 모든 모델이 모두 임계값 밑으로 떨어지지 않기 때문에 예측 가능일을 나타낼 수 없다. 적도 강수는 4~13일의 범위를 보이며 KMA는 UKMO와 함께 두 번째로 긴 예측 가능일을 보여준다.

동아시아 기온의 경우에는 KMA가 ECCC와 함께 가장 짧은 9일의 예측 가능일을 보여주었다. 동아시아 강수 역시 ECMWF가 다른 모델에 비해 2~3배 긴 9일의 예측 가능일을 나타냈고, KMA는 예측 가능일이 4일로 UKMO와 같다. 동아시아 500hPa 지위고도 역시 ECMWF가 가장 긴 15일의 예측 가능일을 보여주고, KMA와 UKMO는 13일로서 중간 정도의 성능을 기록했다.

2019 daily ACC



© Prediction Research Department/APEC Climate Center

Figure 17. (Left) daily ACC of 2019 real-time forecast for KMA, ECCO, ECMWF, NCEP, and UKMO for T2M, PREC, and Z500 over Globe, Tropics, and East Asia. (Right) ACC predictable day for each models that is the lead day number when ACC falls below the threshold. ★ indicates the case when predictable days cannot be estimated.

ACC에 이어 MSSS를 사용한 일별 실시간 예측성능을 살펴보았다. Figure 18에는 각 모델에서 예측한 2019년 강수에 대한 지역별 일별 MSSS와 MSSS의 예측 가능일을 나타내었다. MSSS의 예측 가능일은 모형의 오차와 관측의 변동성이 같을 때인 0 이하로 내려갈 때로 정의하였다(Kim et al. 2018). 전구나 적도 강수의 MSSS를 살펴보면 ECCO와 NCEP을 제외하면 일반적으로 예측성능이 떨어지는 모습이 보인다. ECCO의 경우는 모든 지역에서 약 24일에 급격한 예측성능 저하를 보이고, NCEP은 다른 모델들과 반대로 예측성능이 약 3일까지 급격히 감소하다가 3일 이후로는 다시 증가한다. 반면 동아시아 강수의 MSSS는 모든 모델에서 약 일주일 주기로 진동하는 주기성이 보이는데, 이는 ACC와는 상당히 다른 모습이다.

2019: Predictable days (MSSS) of PREC

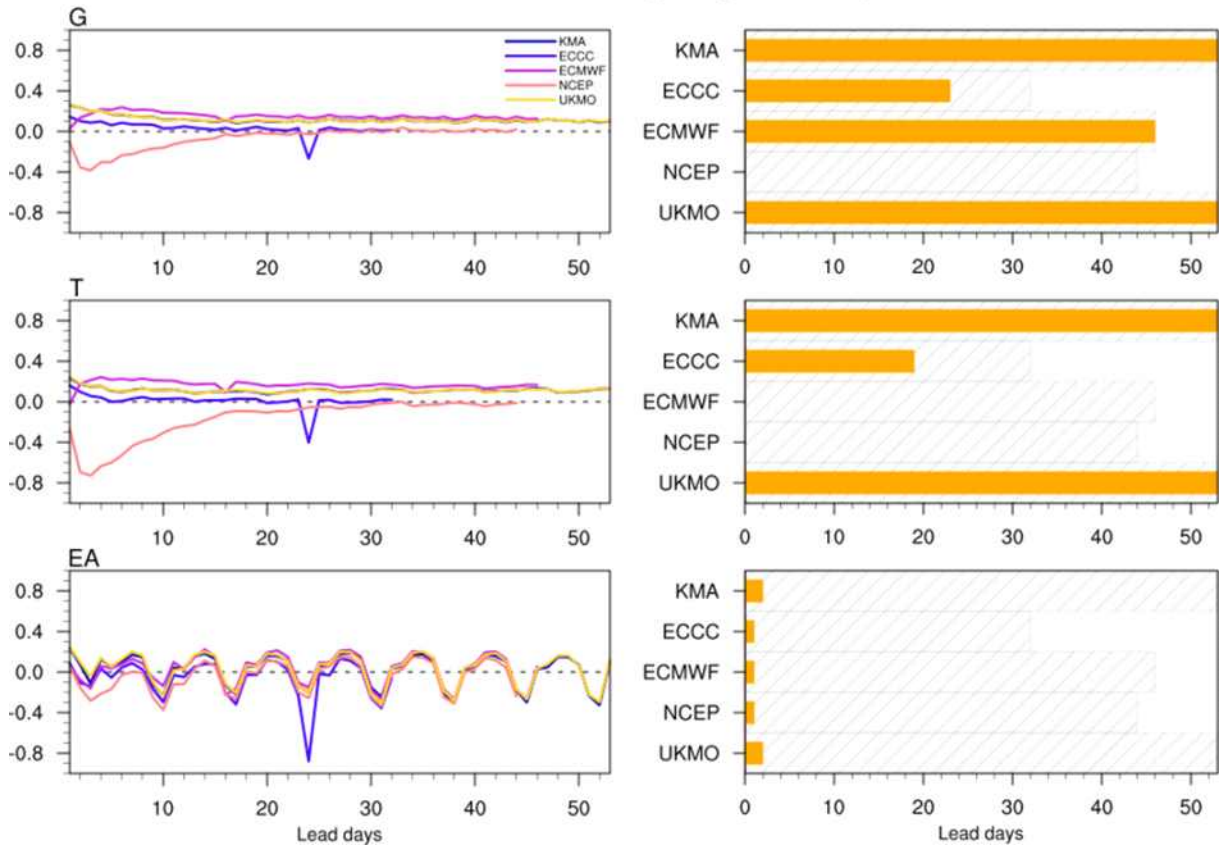


Figure 18. (Left) daily MSSS of 2019 real-time forecast for KMA, ECCC, ECMWF, NCEP, and UKMO for PREC over Globe, Tropics, and East Asia. (Right) ACC predictable day (orange) for each models that is the lead day number when ACC falls below zero. Gray dashed bar indicates the lead day for each models.

그렇다면 MSSS에서 주기성이 나타나는 이유는 무엇일까? MSSS의 계산에 사용되는 실시간 예측 오차(MSE_j)와 관측 오차(MSE_{c_j})에서 그 이유를 찾아보았다. 첫 번째로 관측의 극단적인 강수를 모의하지 못하기 때문에 생기는 에러와 관련된다고 볼 수 있다. Figure 19에는 2019년 2월 17일의 중국 남동부 폭우 사례를 나타냈다. 이 날 관측(GPCP)에서는 중국 남동부에서 폭우 사례가 뚜렷하게 보인다. 계절내 예측성능이 우수하다고 알려진 ECMWF가 모의한 강수 패턴을 살펴보면, 중국 남동부 폭우가 일어나기 46일 전인 2019년 1월 3일에는 물론이고 18일 전인 2019년 1월 31일에도 폭우 패턴은 전혀 예측하지 못했다. 폭우 사례가 나타나기 11일 전인 2019년 2월 7일에 중국 남부지역에 약한 강수 패턴이 모의되기는 하지만 폭우라고 할 수 있는 정도는 아니고, 4일 전인 2019년 2월 14일에도 실제 관측에서 보여주는 폭우 사례보다는 강수량이 적다. 이렇게 지속기간이 짧은 폭우를 모델에서는 거의 예측하지 못하였다. 따라서 MSSS 계산 시 모델의 오차(MSE_j)가 관측(MSE_{c_j})에 비해 매우 커지게 되어 MSSS는 매우 큰 음의 값을 가지게 되고, 모델의 초기날짜가 돌아오는 일주일 주기로 음의 값이 반복된다.

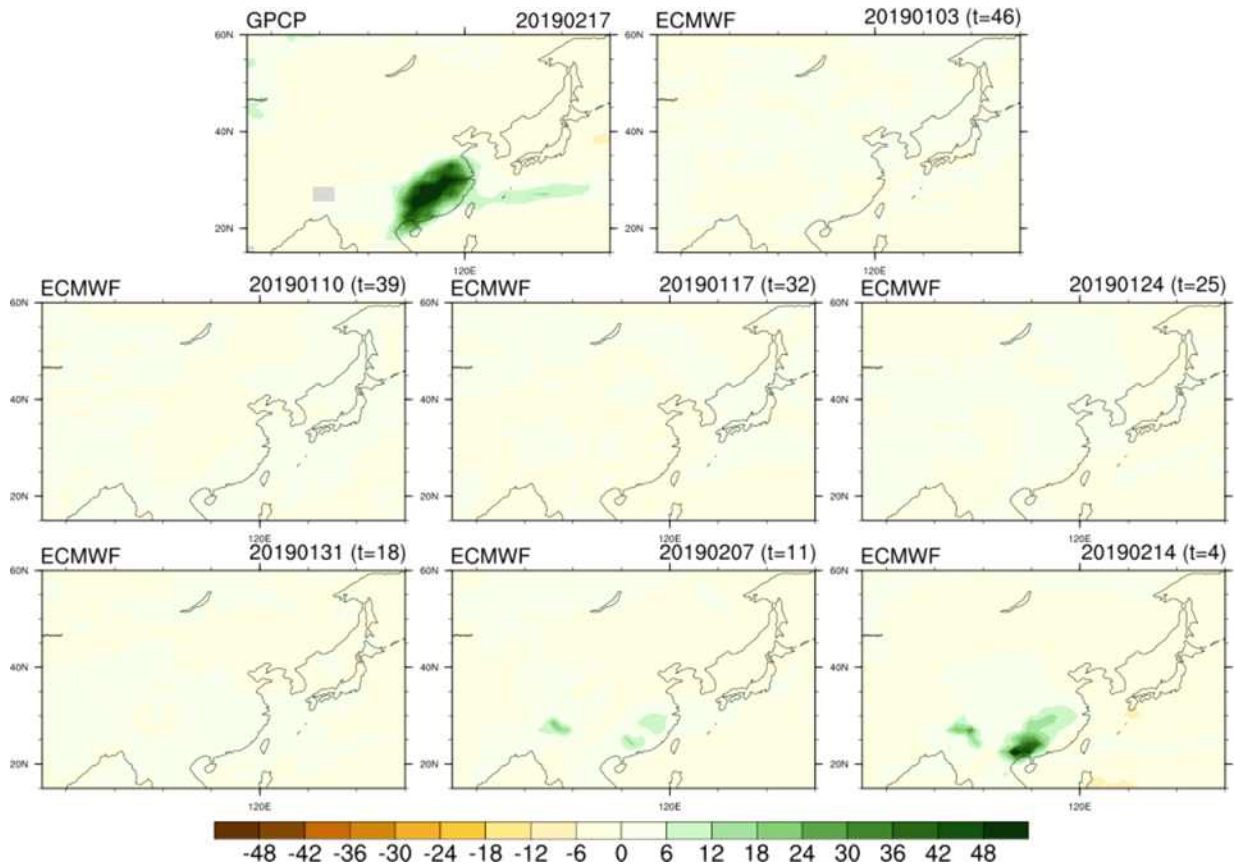


Figure 19. The cases that ECMWF fails to capture the extreme rainfall case over Southeast China on February 17, 2019.

두 번째 이유는 관측 오차를 계산하기 위해 선행 주 수별 각 격자점에 대해 총 12년 동안의 52주의 샘플(52주×12년=624개; 이하 old 버전)을 사용했기 때문으로 판단된다. 관측 오차를 모델과 같은 형태인 일주일 단위로 만들게 되면서 주기성이 생겼을 것으로 추측된다. 관측 오차의 주기성을 없애기 위해 선행 주 수 개념 없이 각 격자점에 대해 총 12년 동안의 365일의 샘플(365일×12년=4,380개)을 사용한 MSSS(이하 new MSSS)를 계산하여 기존의 MSSS(이하 old MSSS)와 비교해보았다(Figure 20). Old MSSS는 7일 주기로 최대값이 나타나는데 비하여 new MSSS는 old MSSS에서 나타났던 가장 큰 peak은 사라지기는 하지만 여전히 주기성은 남아 있다. 따라서 MSSS는 극한사례가 빈번하게 일어나는 강수 중에서도 특히 좁은 지역(예를 들어 동아시아)의 검증에는 적합하지 않다는 결론을 내릴 수 있다.

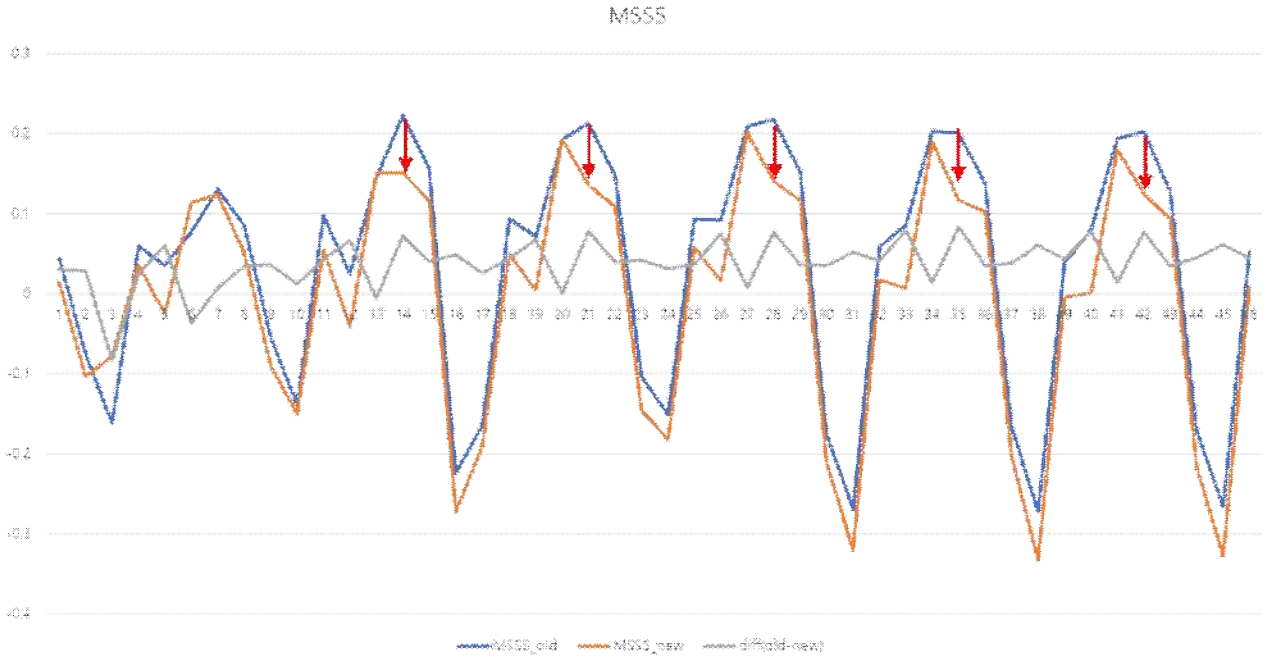


Figure 20. The time series of overall MSSS value of ECMWF for PREC over East Asia for 2019 real-time forecast using 624 samples (52 weeks×12 years; blue line) and 4380 samples (365 days×12 years; orange line) of MSEc calculation. The gray line indicates the difference between blue and orange line.

또한, 2019년 ECMWF의 500hPa 지위고도 실시간 예측에 대한 MSSS(그림 없음)는 reforecast를 이용한 MSSS를 제시한 Son et al. (2020)의 결과와는 다소 차이가 있었다. ECMWF에서 예측한 2019년 북반구(30-90°N) 500hPa의 모델 오차(MSE)와 관측 오차(MSE_o)를 살펴보면, 관측 오차는 주기성이 나타나지 않고 모델 오차에서는 후반부로 갈수록 주기성이 나타난다. 또한 선행연구(Figure 4 in Son et al. 2020)에 비해 모델 오차의 시작과 성장 속도가 다른데, 선행연구에서 사용한 reforecast에 비해 초반에 오차가 게 시작하여 점차 천천히 성장한다(Figure 21). 이는 관측 오차를 설정할 때 Son et al. (2020)과 다르게 하였거나, 검증에 사용한 관측자료가 다르기 때문으로 추측할 수 있다.

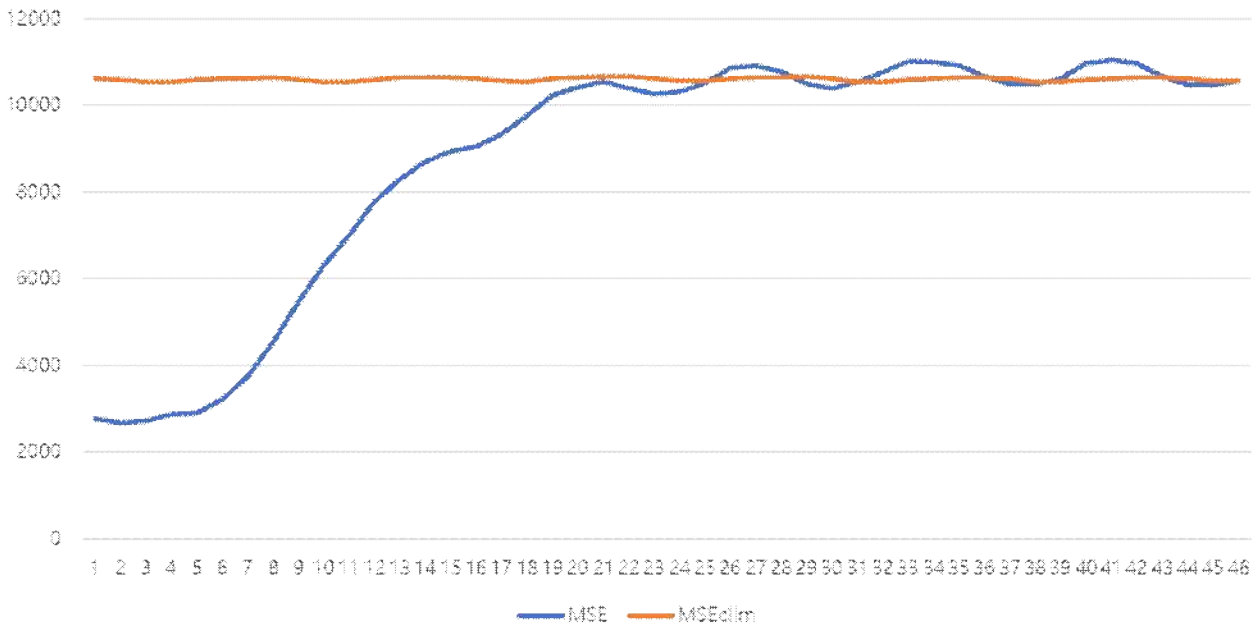


Figure 21. The time series of MSE(blue) and MSE_c(orange) of ECMWF for Z500 over Northern Hemisphere (north of 30°N).

㉔ 타기관 모형 대비 종합 성능 (RVMM)

CREDYT은 역학 모형이 제시하는 2019년도 결정론적 예보와 확률론적 예보에 대한 종합적인 평가를 위해, 우선 개별적으로 계산한 검증 지수들을 변수별, 도메인별, 모형별로 한 눈에 비교할 수 있게끔 요약하였다(Figure 22). 예상할 수 있듯이 선행 주수가 길어질수록 대부분의 스킬들이 감소하는데, 지역적으로는 열대, 메트릭에서는 HSS, 모형별로는 NCEP에서 그렇지 않은 경우가 종종 발생한다. 특히, 열대지역 HSS는 선행 주수가 길어질수록 오히려 스킬이 증가하는 것처럼 보이는데, 이는 열대지역의 기온, 지위 고도 변동성이 크지 않은 분포 특성 때문으로 판단된다. 초기 주수에서 예보 분포가 최적화되지 못해서 삼분위 카테고리 예보가 marginal하게 실패하는데 반해, 오히려 선행 주수가 증가할수록 관측과 유사한 좁은 분포로 최적화되기 때문으로 예상할 수 있다. 결정론적 예보와 확률론적 예보 각각의 스킬 스코어가 제시하는 의미가 조금 다를 수 있다. MSSS(결정론적)는 선행 초반을 제외하고는 거의 무작위 예보 보다 못하다고 제시하는데 반해, CRPSS(확률론적)는 4주까지 무작위 예보보다 나은 스킬을 유지한다. 아마도, 확률 분포의 오차를 재는 방식이 절대값 오차를 재는 방식에 비해 다소 느슨한 기준이 적용된 것이 아닌가 판단된다. 하지만 강수의 경우는 두 메트릭에서 모두 스킬이 거의 없는 것으로 확인된다. NCEP 모형의 경우, 선행주수별 성능 변화가 다른 모형과 상이한 경우가 종종 나타난다. 검증의 기준이 되는 CFS 모형이 재현하는 과거 세상 즉, reforecast 분포를 앙상블을 뿌리는 방식(6시간 간격으로 perturbation없이 4번 수행해서 합치는 방식)과 연관시켜 살펴볼 필요가 있다. 변수별 차이를 살펴보면, 상대적으로 작은 시·공간 규모에서 국지적 과정을 통해 발생하는 강수가 다른 두 변수에 비해서 예측 스킬이 현저히 낮고, 이는 계절내 규모 예측이 극복하지 못하고 있는 여러 한계 중 하나이다.

2019 skill summary

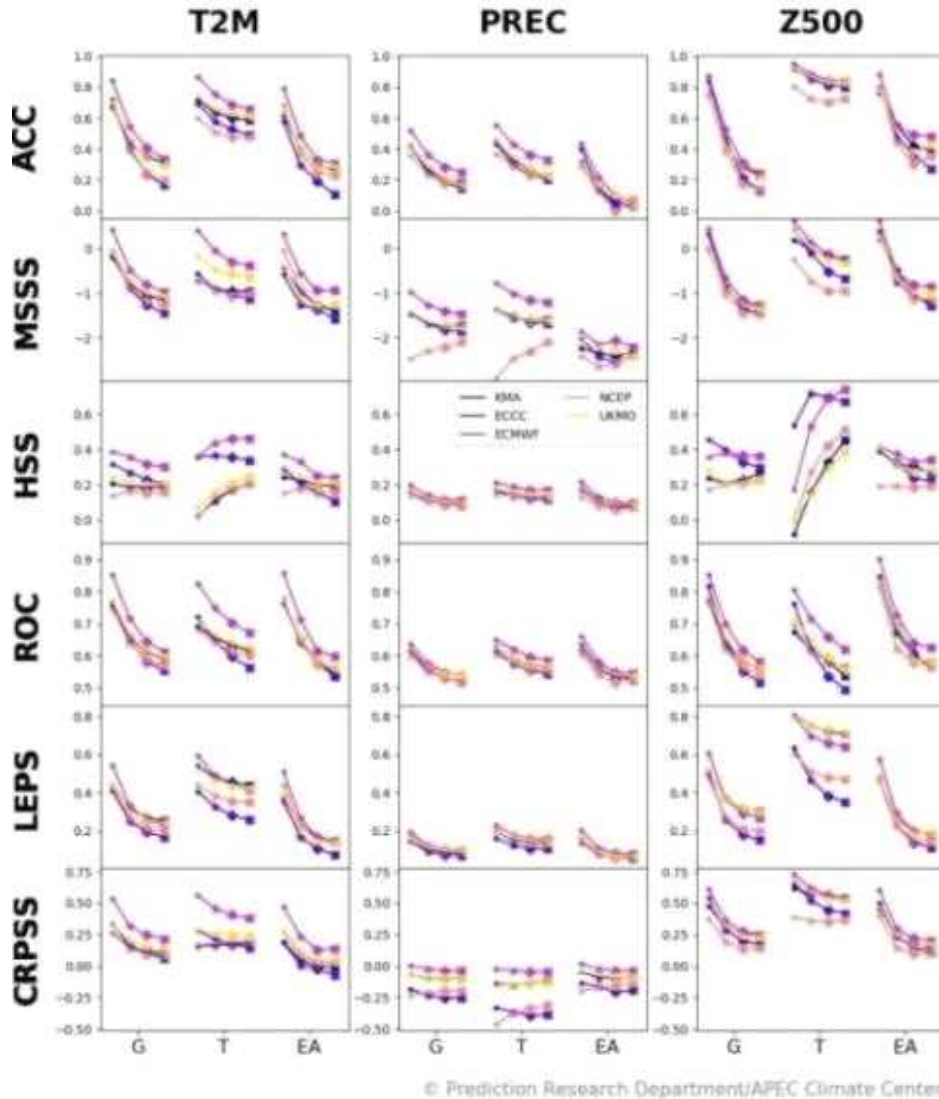


Figure 22. Six verification scores for 1- to 4-lead weeks 52 weekly mean real-time forecasts in 2019 over the Globe, Tropics and East Asia. Five models including KMA-GloSea5 are displayed together with distinct colors.

앞서 언급한 RVMM 방법론을 5개 모형에 대해 동아시아 도메인 지위고도에 적용한 결과는 Figure 23와과 같다. KMA-GloSea5의 순위는 예보 선행 3-4주 한반도 주변 많은 격자에서 1위에 근접한 성능을 보이고, 봄철 선행 3주, 가을철 선행 4주에서는 한반도 역시 상위에 랭크됨을 알 수 있다 (Figure 23좌). 지역 평균 순위에서(우측, 라인 플랏), 1-2주에 낮았던 성능이 3-4주에서는 상위권으로 올라서는 모습을 볼 수 있지만, 1위와는 다소 격차가 있다. 각 격자에서 순위 기반 검증 지수를 합산에서 최고 성능을 보인 모형을 표시했을 때(Figure 23우), 선행 3-4주에서 ECMWF 대신 GloSea 계열 모형들 즉, UKMO, KMA의 빈도가 일부 계절에서 높아진 것을 확인할 수 있다. 하지만, 여전히 한반도 국지적인 격자에서는 ECMWF의 성능이 압도적이다. top-ranker로 확인된 격자의 도메인 내 fraction을 살펴보면, 예상할 수 있듯이 ECMWF가

독보적으로 높다. 하지만, 선행 주수가 늘어날수록 fraction이 감소하고, UKMO, KMA의 비율이 늘어나는 것을 알 수 있다. 이를 통해, 계절내 시간 규모 3-4주에서 GloSea 계열 모형들이 경쟁력을 확인했으며, 향후 모형 개선을 통해 충분히 ECMWF에 도전할 여지가 있다고 판단된다.

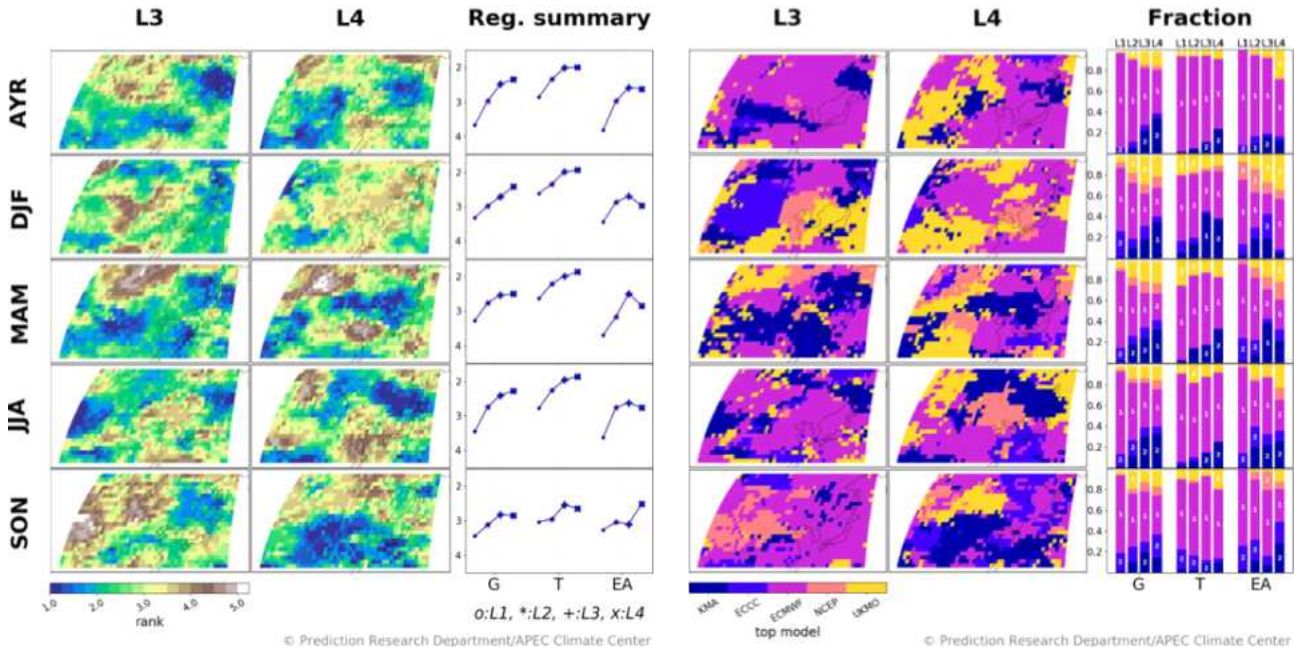


Figure 23. (the leftest two columns) ranking of KMA-GloSea5 out of five models at each grid for 52 weekly mean forecasts in 2019 and (3rd column) regionally summarized ranking of KMA-GloSea5 for the Globe, Tropics and East Asia domain. (4th and 5th columns) top performing model at each grid and (the rightest column) the grid fraction of individual models getting the first place over three regional domains.

③ GloSea5 기후모드 재현 특성

㉓ 2019년 52주에 대한 주요 기후 지수 통계 및 예측성

- 2019년 MJO 발생 통계 및 모형 오차

CREDYT은 매해 주평균 MJO 지수의 변동 추이와 위상별 발생 빈도를 KMA-GloSea5 모형이 얼마나 적절하게 모의했는지를 한눈에 파악하고자 하였다(Figure 24). 물론, 월-일의 단위로 단절된 분석이 MJO behavior의 연속성을 고려하지 못한다는 단점이 있을 수 있으나, 전체 시스템의 일관성을 위하여 “주평균” 지수 검증에 집중하였다. 상단 패널은 관측, 선행 1-4주에 대한 2019년 52주 MJO 위상과 강도를 나타내었다. 적도 대류활동의 시·공간 변동을 모형이 제대로 모의하지 못함을 알 수 있다. 특히, 관측되지 않은 강수 3이상의 지수가 모형에는 자주 나타나 강도 분산이 넓은 것으로 판단된다. 위상별 강도별 빈도는 가운데 열에서 확인할 수 있다. 선행 2-3주에서 위상3의 빈도가 과다 모의되고, 선행 4주에서는 위상2/4의 빈도가 과다 모의되고 있음을 알 수 있다. 하지만 전반적으로 서쪽 반구의 활동 위상7의 빈도 과소 모의, 위상5의 빈도는 선행 주수가 늘어날수록 감소하는 특성을 확인할 수 있다. 하단 패널에서는 강도 별로 MJO 위상별 빈도를 확인할 수 있는데, 강도에 따른 전체 빈도는 3-4주 예보가 1-2주 예보에 비해 관측과 비슷하게 재현하고 있음을 알 수 있다. 하단 우측의 타일맵은 빈도의 관측 대비 오차를 나타내는데, 앞서 언급했듯이 위상2의 빈도 과다 모의가 뚜렷하고, 중간 강도에서는 위상2-4의 과소 모의가 뚜렷하다.

2019: MJO freq. amp. diagnostics

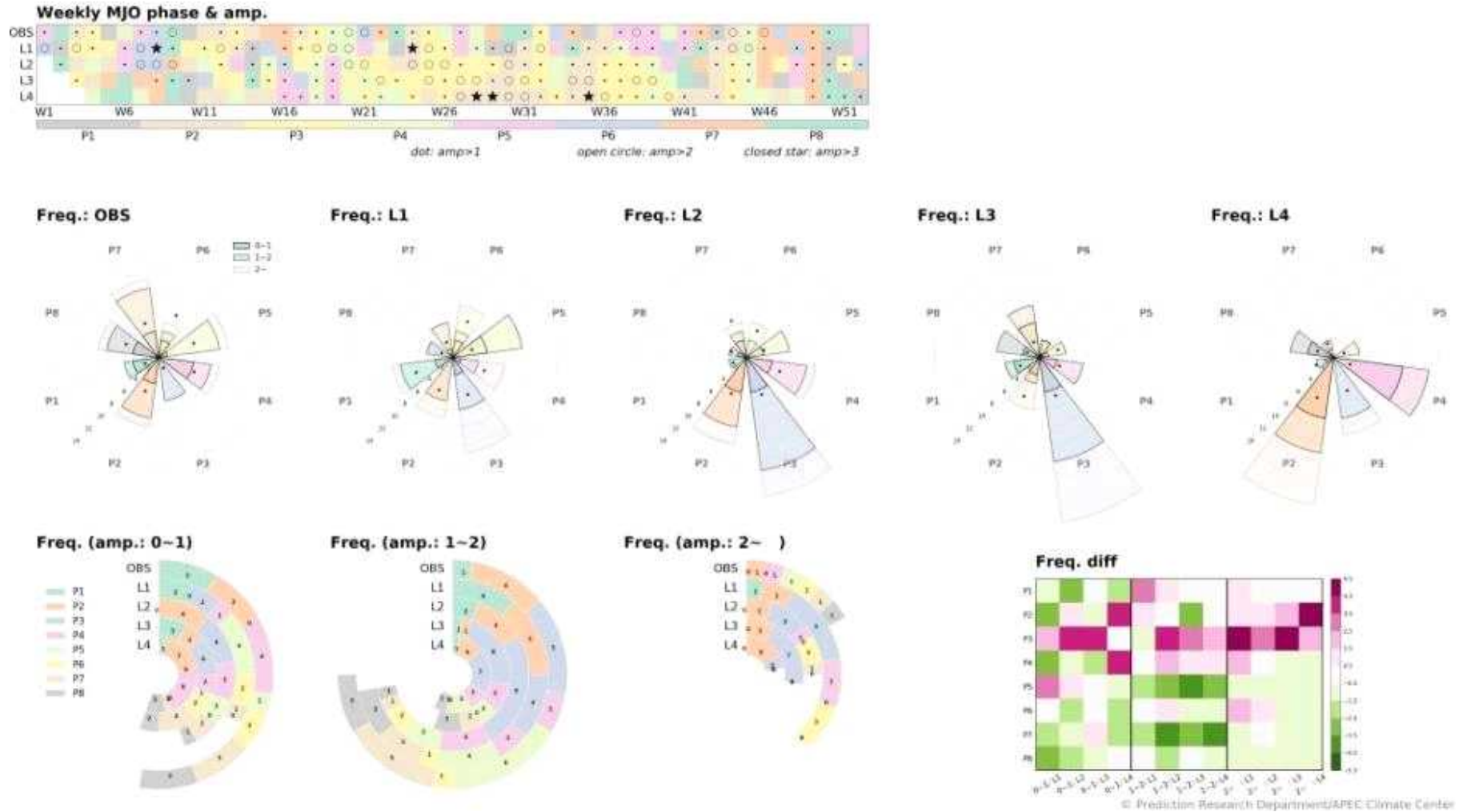


Figure 24. MJO statistics for 52 weeks in 2019 for observation and 1- to 4-lead weeks KMA-GloSea5 forecasts: (upper) phase and amplitude of individual weeks, (middle) three amplitude categories frequency of 8 phases for the observation and different lead weeks, (lower left) 8 phase frequency of 1- to 4-lead weeks for three amplitude categories, and (lower right) the frequency biases summary for different combination of phase, amplitude and lead weeks.

- 2019년 주별 · MJO 지수 검증

MJO 지수는 보통 OLR(Outgoing Longwave Radiation)를 변수로 사용하는 RMM(Real-time Multivariate MJO)가 널리 쓰이나, KMA에서는 OLR이 제공되지 않기 때문에 OLR 대신 VP200을 사용하는 VPM(Velocity Potential MJO)을 MJO지수로 사용하였다. 일반적으로 MJO 지수는 일별 자료를 사용하여 계산되나, 이 연구에서는 KMA 모델의 주별 예측성능 분석에 초점을 두었기 때문에 MJO 지수 또한 주별 예측자료를 이용하여 계산되어야 한다. 따라서 앞서 일별 예측성능을 검증하기 위해 생산된 일별 예측자료를 사용하여 MJO 지수를 계산하고 이를 주 평균(월요일-일요일)한 값과 주별 예측자료를 사용하여 직접 계산한 MJO 지수의 값을 비교해보았다. 테스트 결과 실시간 예측과 과거재현자료에서 모두 PC1, PC2, MJO 지수의 주 평균 값이 상관계수 0.95 이상으로 차이가 거의 없었다. 따라서 주별 예측자료를 사용하여 MJO 지수를 계산하고 이를 검증하였다.

MJO 지수의 검증에는 bivariate correlation(Lin et al. 2008; Rashid et al. 2011)가 사용되었고 계산식은 아래와 같다.

$$Bivariate\ correlation(\tau) = \frac{\sum_{t=1}^N [a_1(t)b_1(t,\tau) + a_2(t)b_2(t,\tau)]}{\sqrt{\sum_{t=1}^N [a_1^2(t) + a_2^2(t)]} \sqrt{\sum_{t=1}^N [b_1^2(t,\tau) + b_2^2(t,\tau)']}}$$

또한 일별 MJO 지수를 추가로 계산하여 KMA의 MJO 예측 가능일을 판단했다. 예측 가능일은 bivariate correlation이 0.5 이하가 되는 날짜로 정의되는데, 이는 climatological forecast (예측 MJO 아노말리가 0으로서 $b_1=b_2=0$)일 때의 bivariate RMSE($=\sqrt{2}$)일 때의 선행시간과 거의 유사하다(Lin et al. 2008; Rashid et al. 2011). Figure 23의 왼쪽에는 KMA가 예측한 2019년 52주 실시간 예측에 대한 선행시간 1주와 3주의 VPM1과 VPM2를 관측과 함께 제시하였다. 선행시간 3주에서는 VPM1과 VPM2에서 모두 선행시간 1주에 비해 관측의 VPM 지수를 잘 따라가지 못한다. VPM 모드의 예측성능이 계절에 따라 바뀌는지를 살펴보기 위해 Table 9에 선행시간에 따른 연간, 북반구 겨울(NDJFMA), 북반구 여름(MJJASO)에 대한 VPM 모드의 상관계수를 나타내었다. 선행시간 1주에서는 계절에 상관없이 VPM1과 VPM2 모두 95% 이상 신뢰수준에서 유의한 상관계수를 보여준다. 그러나 선행시간 3주에서는 전반적으로 선행시간 1주에 비해 상관계수가 낮은 특징이 있다. 그리고 상관계수는 계절에 따라 상이한데, 북반구 겨울철에는 VPM1은 -0.17의 매우 낮은 상관계수를 보이는 데 반해, VPM2가 95% 이상 신뢰할 만한 수준인 0.70의 값을 보여준다. Figure 25의 오른쪽에는 공통기간에 대한 과거재현자료와 2019년 실시간 예측에 대한 일별 VPM 지수의 bivariate correlation을 나타낸다. Reforecast와 실시간 예측에 대한 bivariate correlation 모두 선행시간에 따라 감소하는 모습을 보여주지만, 실시간 예측 기간은 reforecast 기간보다 샘플 수가 상당히 적기 때문에 매끄럽게 감소하지는 않는다. Reforecast 기간에는 17일의 예측 가능일을 보여주고, 2019년 실시간 예측 기간에는 14일의 예측 가능일을 보여주어 reforecast에 비해 3일 짧다.

Table 9. The correlation coefficients of VPM1 and VPM2 for 2019 KMA real-time forecast. * indicates that the correlation coefficient is statistically significant at the 95% level using two-tailed Student's *t*-test.

	All year		NDJFMA		MJJASO	
	VPM1	VPM2	VPM1	VPM2	VPM1	VPM2
L1	0.78*	0.81*	0.70*	0.87*	0.85*	0.80*
L3	0.03	0.12	-0.17	0.70*	0.12	-0.23

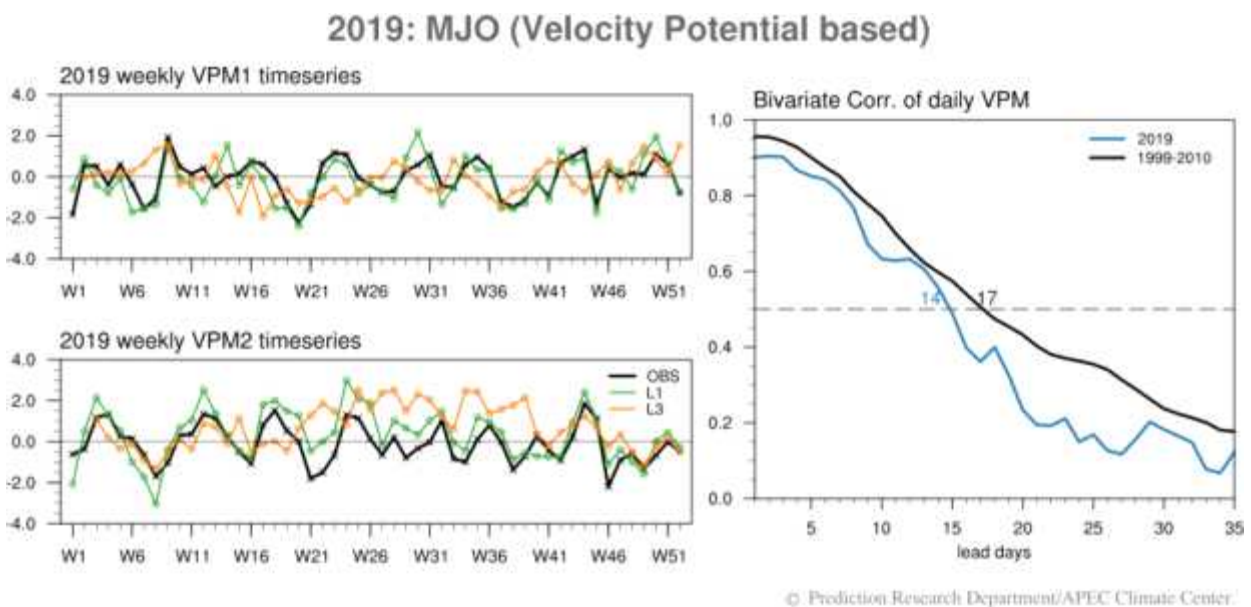


Figure 25. (Left) weekly VPM1 and VPM2 time series of observation and KMA for 1-, and 3-lead week for 2019 real-time forecast. (Right) bivariate correlation of daily VPM for 2019 real-time forecast and reforecast during 1991-2010. The numbers indicate the predictable days of MJO, respectively.

- KMA-GloSea5와 타 기관 일별 MJO 지수 성능 비교

Figure 26에는 KMA의 일별 MJO 지수와 타 기관 MJO 지수의 예측성능을 reforecast 기간과 2019년 실시간 예측기간에 대해 각각 비교하였다. Reforecast와 실시간 예측 기간에서 예측 가능일이 가장 긴 모델은 ECMWF로, 20일 이상의 예측 가능일을 보여준다. 뒤를 이어 KMA가 UKMO와 함께 거의 비슷한 성능을 보인다. 이는 KMA와 UKMO가 GloSea5를 공동운영하기 때문에 나타나는 당연한 결과로 생각할 수 있다. ECCC는 다른 모델과 다르게 상당히 낮은 예측 성능을 보여준다. 이는 2019년 7월 이후 자료에서 MJO 생산에 필요한 U200, V200, U850에서 특정 위치에 존재하는 특정 bias에 의한 것으로(그림 없음), ECCC의 현업 모델 변경 기간과 비

늦은 시기부터 bias가 생기는 것으로 보아 현업모델 변경과 관련된 오류인 것으로 판단된다.

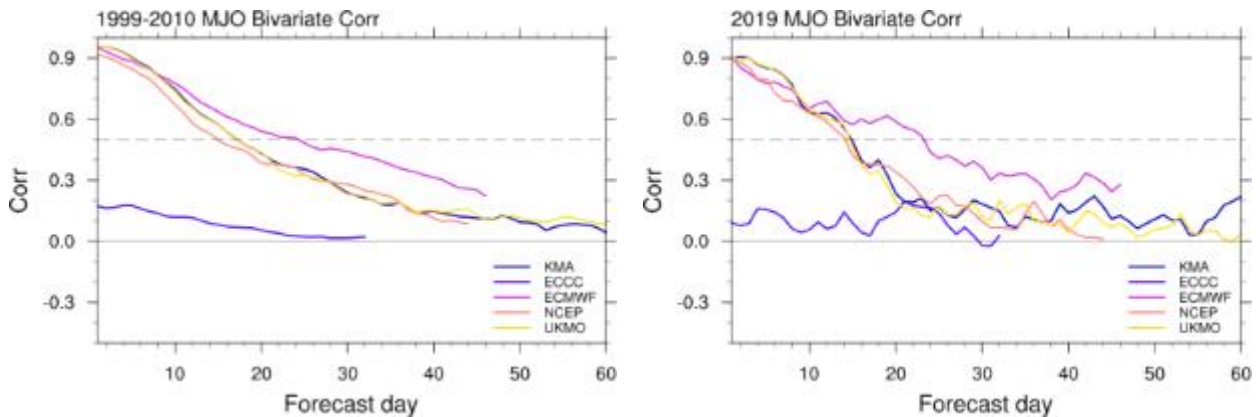


Figure 26. Bivariate correlation of daily MJO index of KMA, ECCO, ECMWF, NCEP, and UKMO for reforecast during 1991-2010 and 2019 real-time forecast. Gray dashed line indicates the predictable days of MJO (0.5) that is usually being used as the threshold of skill(Rashid et al. 2011).

- MJO 외 기후 지수 시계열 검증

CREDYT은 MJO 이외에도 9개 기후 리짐에 대한 지수 검증을 수행하였다. 특히 한 해 동안의 week-to-week 시계열 및 상관관계 분석을 통해 주별 성능 추이를 확인하였다(Figure 27). 수권에 해당하는 NINO3.4, 성층권에 해당하는 PV 지수처럼 메모리가 길고 천천히 변하는 기후 리짐의 경우는 TCC 스킬이 매우 높다. 하지만 2019년의 경우, NINO3.4 지수의 변동성이 매우 커서 과거(99-10) 평균대비 스킬이 낮다. PV의 경우 모든 선행 주수에서 거의 관측과 유사하게 2019년 지수를 예측하고 있고, 과거 평균대비 스킬이 월등히 높다. ART1과 SH는 선행 주수가 늘어날수록 스킬 감소가 뚜렷하고, 특히 ART1의 3-4주 성능은 과거 평균보다 현저히 낮다. 기후 리짐 재현에 있어 역학 모형 스킬의 기준으로 persistent 예보를 설정하였으며, 아래 제시한 지수들은 NINO3.4 선행 1주를 제외하고 대부분 그 기준을 통과한 것을 알 수 있다.

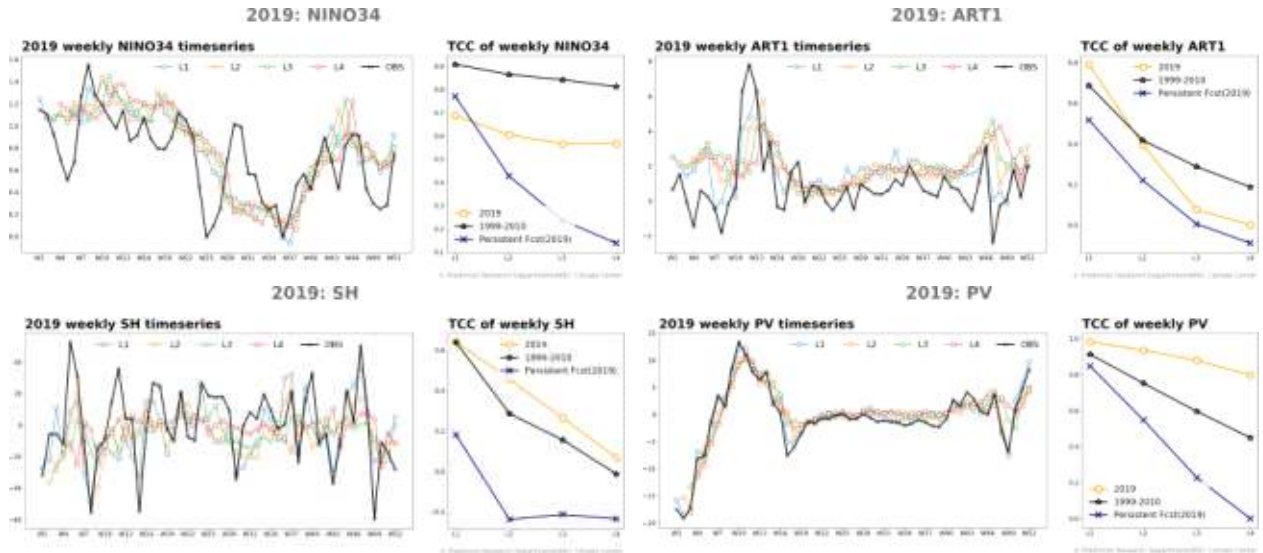
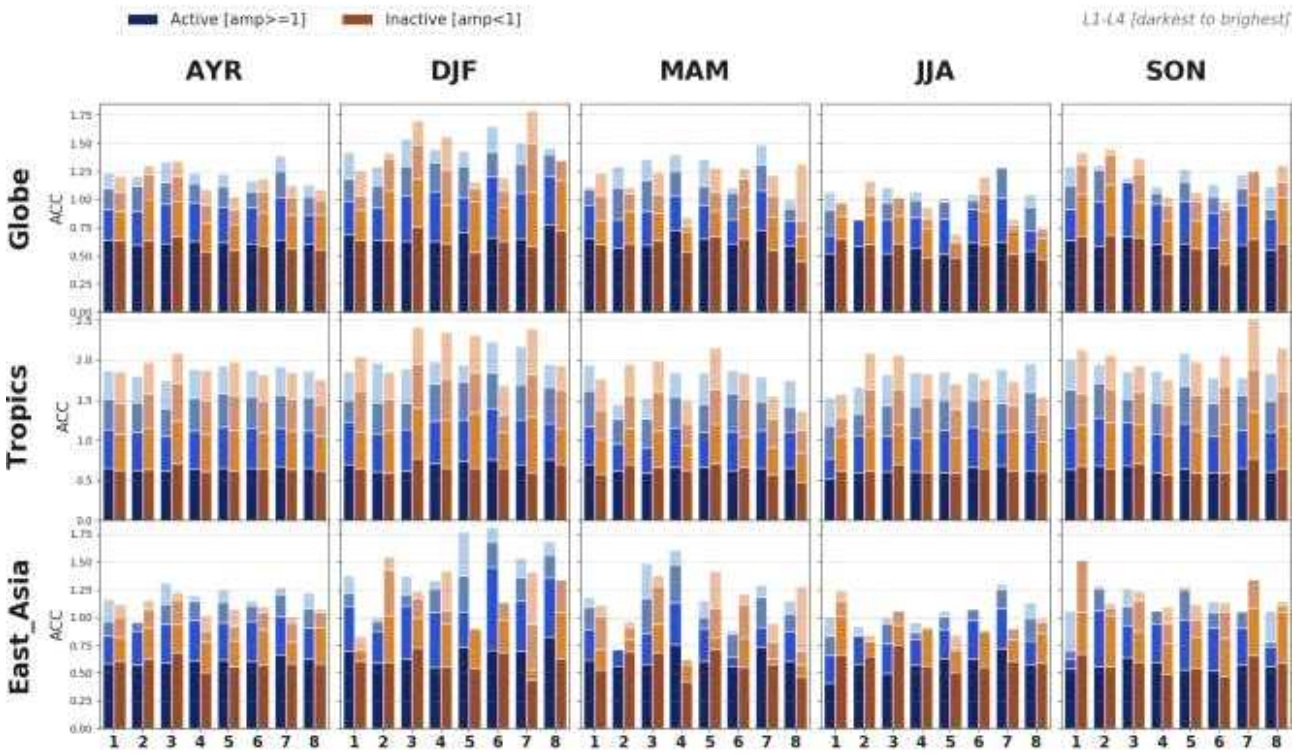


Figure 27. Observed and forecasted timeseries of weekly mean climate index in 2019 and TCC score of 1- to 4-lead weeks forecasts: (upper left) NINO3.4, (upper right) ART1, (lower left) SH and (lower right) PV. As the reference, TCC scores of persistent forecast and the values averaged for 12 years (1999-2010) are also displayed.

㉔ 기후모드 초기 위상 및 강도에 대한 예측성 민감도

기후 리짐의 위상/강도에 따라 모형 예측력이 달라질 수 있음이 잘 알려져 있다. 계절 예측의 경우, 해양 강제력이 강할 때 즉 엘니뇨나 라니냐 위상일 때가 중립 해양 상태일 때 보다 그 예측력이 훨씬 높은 경향을 보인다. 계절내 규모에서도 열대 매든-줄리안 진동의 위상에 따른 예측성능 의존성에 대한 연구들이 이루어져 왔다 (Baggett et al. 2017; DelSole et al. 2017; Li and Robertson 2015; Rodney et al. 2013). CREDYT은 계절내 규모에서 중요한 MJO의 위상/강도에 따라 전구 예보 기본 성능이 어떻게 변하는지를 확인하였다(Figure 28). 예상과 달리, 강한 강도의 대류활동 시기가 약한 강도 시기 성능은 오히려 낮아져 전반적으로 우열을 가리기 힘들다. 하지만, 봄철에서 강한 강도의 2-4위상 시기에 전구 기온 예보 성능이 상대적으로 높고, 이러한 양상이 가을철에는 4-6위상으로 바뀌는 것을 확인할 수 있다. 동아시아 도메인의 경우는 계절별로 차이는 나지만 다소 일관된 격차를 확인할 수 있다. 북반구 겨울철과 여름철에 위상5-8일 경우, 강한 대류활동일 때가 약한 대류활동일 때보다 전반적으로 기온 예보 성능이 높다. 특히 겨울철 위상5-6의 경우는 그 격차가 매우 뚜렷해서 매뉴얼 예보에서 활용할 수 있을 정보로 판단된다.

ACC [T2M] reliance on MJO phase



© Prediction Research Department/APEC Climate Center

Figure 28. T2m predictive skill (ACC) variation with the phases and amplitude of MJO for different seasons and regions.

㊤ 지구 시스템간 상호 연결 강도 진단

앞서 제시한 지구 시스템 간 연결강도 진단법 적용 결과에 따르면, KMA-GloSea5 모형에서 시스템 간 접합의 과대/과소 오차 경향은 선행 주수별, 계절별로 다르다(Figure 29). 전체 계절 평균적으로 봤을 때, 선행 3주 예보는 세부적으로 조금의 차이는 있지만 관측과 매우 유사한 접합 강도를 보이는데 반해, 초기 1주에서는 다소 과소 모의되는 것을 알 수 있다. 특히, 선행 주수가 늘어날수록 수권 내부 접합 강도는 증가하고, 반대로 대류권 내부 접합강도는 증가하는 뚜렷한 경향을 확인 할 수 있다. 수권 내부 접합 강도 과소모의는 북반구 여름철과 가을철에서 뚜렷하고, 대류권 내부 접합 강도 과대모의는 겨울철에 매우 뚜렷하다. 선행 3주 예보 전 계절 접합 강도는 관측과 유사하지만, 계절에 따른 변동량은 매우 크다. 특히 겨울철 과대모의와 가을철 과소모의가 매우 뚜렷하고, 봄/여름은 관측과 유사하다. 하지만, 선행 1주 예보는 봄/여름 과소모의가 나타나 반대 경향을 보인다. 지구 시스템 간 연결성은 전체적으로 한쪽으로 치우치기 보다는, 총합이 유지되는 선에서 계절적 분포 오차가 나타나는 것으로 판단된다. 또한, 이런 계절적 경향은 아무래도 현재의 예측 모델이 ENSO와 그와 관련된 원격 상관의 seasonal locking 상태에 다소 간혀있는 형국임을 짐작할 수 있다.

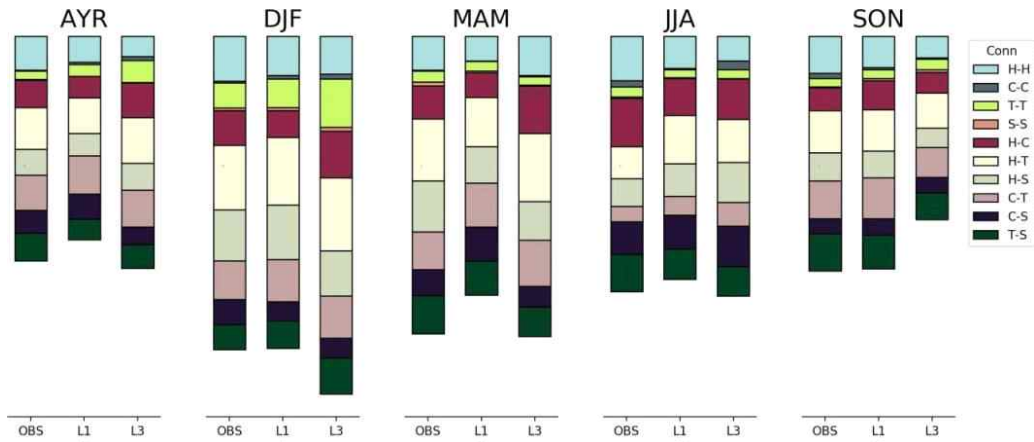


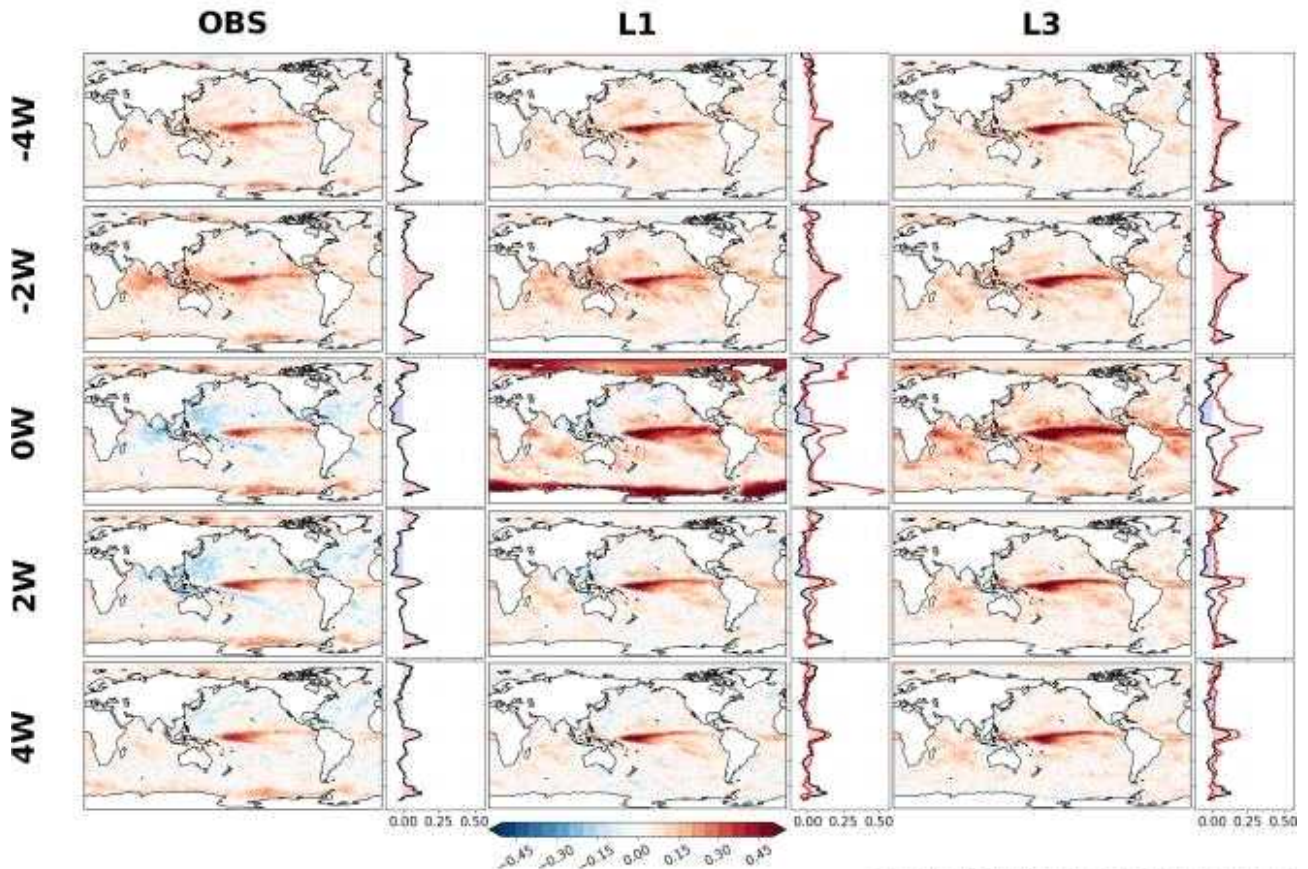
Figure 29. Stacked bars showing the strength of links among four earth systems for the observation, 1- and 3-lead weeks forecasts during all year round, and four seasons. Inter-(intra-)systems connectivity is derived from the total number of chords connecting two different systems (within an earth system) as in Figure 6.

④ GloSea5 대기-해양, 열대-중위도, 대류권-성층권 역학 커플링 재현 특성

CREDYT은 기후 리짐 상관성 기반의 모형 접합성 진단 외에 KMA-GloSea5에서의 기본적인 대기-해양, 열대-중위도, 대류권-성층권 커플링 진단 역시 수행하며, 총 12년 과거 재현 주 평균 자료를 활용하였다. SST와 강수 사이의 자연 상관성 분석에 기반한 대기-해양 커플링 패턴은 Figure 30와과 같다. 열대 워풀 지역에서는 강한 강수가 해수면 온도를 내리는 기작이 작동하지만, 선행 3주 역학 모형에서는 이러한 기작이 전혀 작동하지 않는다. 대신 태평양 적도 주변 coupling은 선행 주수, 자연 주수에 관계없이 과대 모의되고 있다. 또한, 동시 상관에 있어서 남극 북극 주변의 매우 강한 coupling이 확인되는데, 이는 관측에는 없는 기작이라 향후 모형 개선 방향에서 검토가 필요한 포인트라고 생각된다.

Weekly SST-PREC Corr. during 1999-2010

Line: 80E to 80W zonal avg., -4W & -2W: SST leading, +2W & +4W: SST lagging



© Prediction Research Department/APEC Climate Center

Figure 30. Air-Sea coupling: maps of lead-lag correlation coefficients between weekly mean SST and precipitation anomalies at each grid for (left) the observation, (middle) 1-, and (right) 3-lead weeks forecasts. Line plot next to the map displays the zonal average [80degE - 80degW] of the map.

열대-중위도 coupling의 일종인 열대 대류활동과 동아시아 기온 사이의 관련성에 대한 KMA-GloSea5의 재현 특성은 Figure 31와과 같다. 한반도를 포함한 좁은 동아시아 영역 평균 기온이 상승하기 4주 전에 열대 서태평양 워플 지역에 약한 대류활동이 관측되고, 기온 상승 후 4주 정도가 지나면 다시 비슷한 형태의 대류활동을 보인다. 안타깝게도 선행 1, 3주에서 모두 이러한 자연 상관을 확인할 수 없다. 그리고 동서 태평양 간의 대규모 시소 패턴이 아닌, 다소 산만한 형태의 상관 관계 패턴이 나타나는데, 이는 모형 세상에서 동아시아 기온 변동에 있어 열대 대류 활동이 주요한 기작으로 작동하지 못함을 의미한다.

Weekly VP200-EA(T2m) Corr. during 1999-2010

Box: East Asia, Line: 10S to 10N merid. avg., -4W & -2W: VP200 leading, +2W & +4W: VP200 lagging

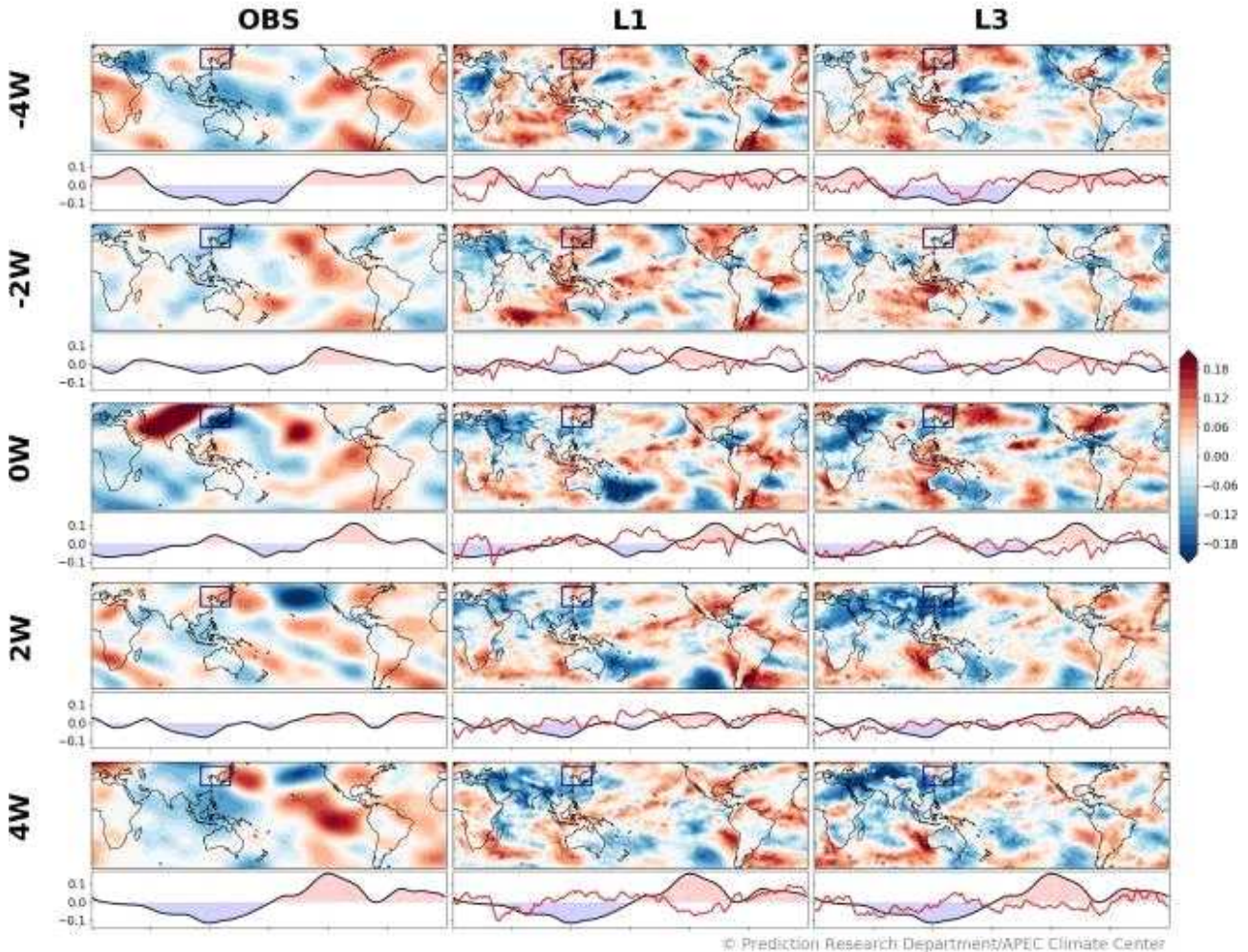


Figure 31. Tropics-midlatitude coupling: maps of lead-lag correlation coefficients between the weekly mean East Asia T2m (navy box) anomalies and velocity potential anomalies at 200 hPa at each grid for (left) the observation, (middle) 1-, and (right) 3-lead weeks forecasts. Line plot beneath the map displays the meridional average [10degS - 10degN] of the map.

성층권 극 소용돌이 세기는 파동 전파를 통해 지표 기후에 영향을 줄 수 있다. CREDYT은 상층 PV 지수와 지표 기온 사이의 상관관계 패턴 분석을 통해 KMA-GloSea5의 성층권-대류권 coupling 재현성을 진단하였다(Figure 32). 기본적으로 극 소용돌이가 강해지면 유라시아 북부는 온난해지고 반대편 캐나다 아이슬란드 주변은 차가워진다. 반대로, 극 소용돌이가 약해져서 북극이 확장하면, 북부 유라시아는 물론 한참 남쪽 한반도까지 한파가 발생하게 된다. 모형의 예측 선행 1주에서 이러한 상호 작용은 거의 관측과 유사하게 모의되고 있다. 선행 3주에서도 특성이 잘 유지되고 있으나, 상호 작용이 다소 지연되어 나타나는 것을 알 수 있다. -2W에서 시그널이 잘 보이지 않는 대신 +4W까지 동시 상관 패턴이 유지되고, 하지만 중위도/한반도까지 확장하는 시그널은 다소 약화된다. 천천히 변동하는 성층권 변동성과 초기 성층권 상태가 예측 선행 3주까지 유지되는 모형 특성 때문에, 선행 3주에서의 coupling 패턴이 다소 지연되

는 결과가 나온 것으로 추측된다.

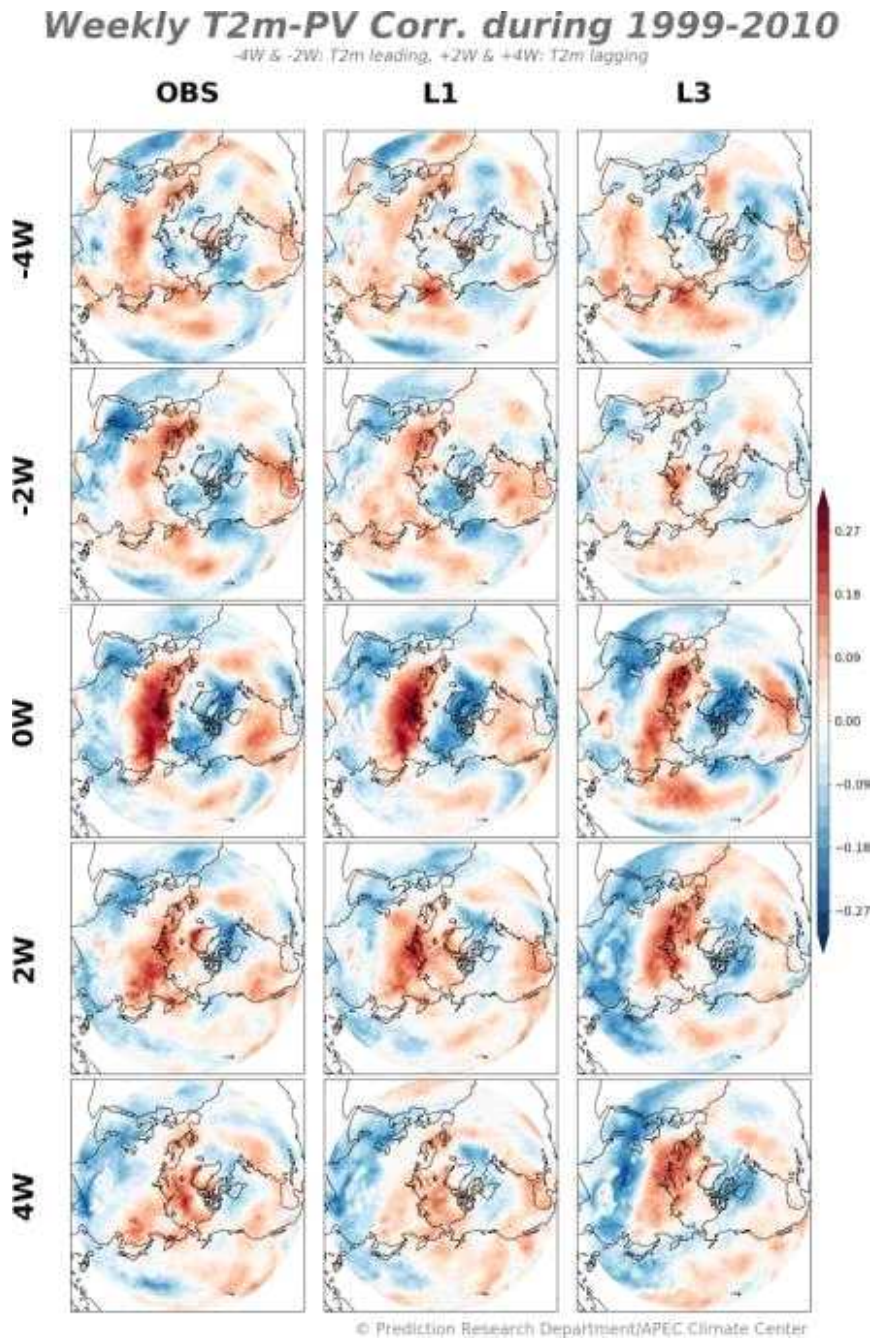


Figure 32. Troposphere-stratosphere coupling: maps of lead-lag correlation coefficients between the weekly mean polar vortex (PV) anomalies and T2m anomalies anomalies at each grid for (left) the observation, (middle) 1-, and (right) 3-lead weeks forecasts.

(3) GloSea5의 기후값 생산방안에 따른 재현/예측성능 분석

이 절에서는 현재 기상청에서 현업으로 사용하는 기후값 생산방안을 적용한 KMA 모델의 기후값 재현성능 및 실시간 예측성능을 통해 기존 기후값 생산방안의 한계를 파악하고자 한다. 이와 더불어 KMA 모델 성능 개선을 위한 다양한 기후값 처방방안을 제안하고, 이를 적용했을 때 실제로 기후값 재현성능과 실시간 예측성능이 개선되었는지를 살펴보고자 한다.

(가) 기후값 생산방안별 기후값 재현성 민감도

① 기후값 생산방안

㉓ KMA 기후값 생산방안

KMA의 실시간 예측자료는 매일 생산되는 초기장에 대해서 4개의 앙상블 멤버로 구성되어 있다. 그러나 reforecast는 3개의 앙상블 멤버가 매달 특정 날짜(1일, 9일, 17일, 25일)을 시작으로 적분 된다. 매일 생산되는 예측자료와 달리 특정일에 생산되는 reforecast 때문에 예측자료에 준하는 매일의 reforecast 자료가 필요하다. 따라서 예측자료의 날짜를 기준으로 전/후의 가장 가까운 reforecast 날짜 2개를 선택하고 예측날짜와 reforecast 날짜 간의 거리에 따른 가중평균을 하여 예측자료의 날짜에 준하는 reforecast 자료를 생산한다. 현재 기상청 장기예측 시스템은 안정적인 운영을 위해 1달 선행하여 reforecast를 생산하고 있다.

Reforecast의 가중평균은 예측자료와 k 개 중 n 번째 reforecast 날짜 사이의 거리(d_n)를 각각 구하여 아래 수식과 같은 가중치(ω_n)를 구한 후 정규화하여 구해진다. 만약 예측자료 날짜와 reforecast 날짜가 같은 경우에는 ± 0.5 를 취하여 거리의 차이가 0이 되지 않도록 하였다.

$$\omega = e^{-d^2/100}$$

$$\text{Normalized weight} = \omega_n / \sum_{n=1}^k \omega_n$$

기상청 주별 앙상블 예측은 최근 7일의 예측자료인 28개의 앙상블 멤버(4개×7일)를 수집하여 시간 차이에 따른 앙상블을 구성하며, 기후장 역시 예측장에 준하는 날짜의 가중평균 기후장을 이용하여 주별 앙상블 예측자료를 생산한다(Ham et al. 2017).

㉔ 기후값 생산을 위한 다양한 실험방법

이 연구에서 제안하고자 하는 기후값 생산방안은 총 3가지로, 모두 KMA에서 생산하는 기후값과 비교할 수 있도록 전반적인 생산 과정은 기본적으로 KMA의 기후값 생산방안을 따르도록 하고 거리가중치를 적용하지 않거나 reforecast 날짜 수를 증가하는 등의 변형을 주어 민감도 실험을 구성하였다. 이에 대한 자세한 내용은 Table 10에 나타내었다. EXP1의 경우, KMA

기후값 생산방안에서 거리가중치를 적용하지 않고 reforecast 날짜 수를 그대로 사용하여 거리가중치에 대한 민감도를 판단할 수 있는 실험이다. EXP2는 KMA의 거리가중치는 그대로 두되, reforecast 날짜 수를 기존 reforecast 날짜에 앞, 뒤로 총 2개의 날짜를 추가하여 기후값을 만들 때 더 큰 smoothing 효과가 나타날 수 있도록 구성하였다. 예를 들어, 2019년 5월 2일 예측장에 준하는 기후값을 만드는 데 사용하는 reforecast의 날짜는 기존에 사용하는 5월 1일과 5월 9일에 4월 25일과 5월 17일을 추가했다. EXP3은 거리가중치를 주지 않고 reforecast 날짜를 4개를 사용하여 EXP1과 EXP2의 효과를 동시에 볼 수 있도록 한 실험이다.

Table 10. Comparison of climatology production methods.

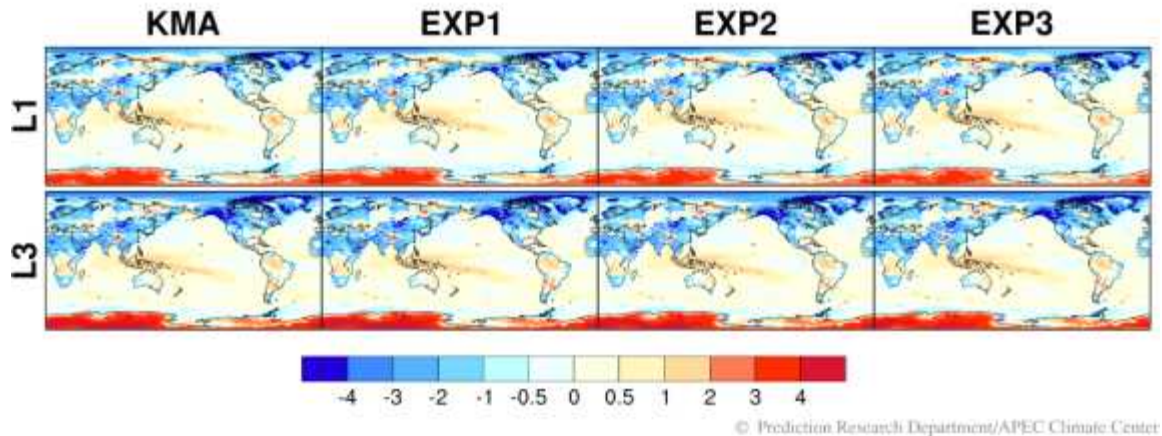
EXP	Weighting	Reforecast sample size
KMA	$e^{(-d^2/100)}$	$(3 \times 2) \times 7\text{days}=42$
1	×	$(3 \times 2) \times 7\text{days}=42$
2	$e^{(-d^2/100)}$	$(3 \times 4) \times 7\text{days}=84$
3	×	$(3 \times 4) \times 7\text{days}=84$

② 재현성능

㉞ 겨울철 남극 warm bias 개선

앞서 Figure 9에서 제시한 바와 같이, KMA 모델이 모의하는 남극의 겨울철 기온은 warm bias, 북반구 대륙에는 cold bias가 존재한다. 이 절에서는 다양한 실험을 통해 생산된 KMA와 UKMO의 mean bias를 기존의 mean bias와 비교하여 살펴보았다(Figure 33). EXP1의 남극 겨울철 기온의 warm bias는 KMA와 비슷한 수준이고, EXP2 역시 겨울철 남극 warm bias가 크게 개선되지는 않았다. EXP3은 KMA보다 남극 기온의 warm bias가 어느 정도 줄어든 모습을 볼 수 있다. KMA의 겨울철 mean bias와 비슷한 패턴을 보인 UKMO에도 같은 방법을 적용하여 비교하면 3가지 기후값 생산방안에서 모두 남극의 warm bias가 줄어들었음을 확인할 수 있다.

KMA: Mean biases of T2M, DJF, 1999-2010



UKMO: Mean biases of T2M, DJF, 1999-2010

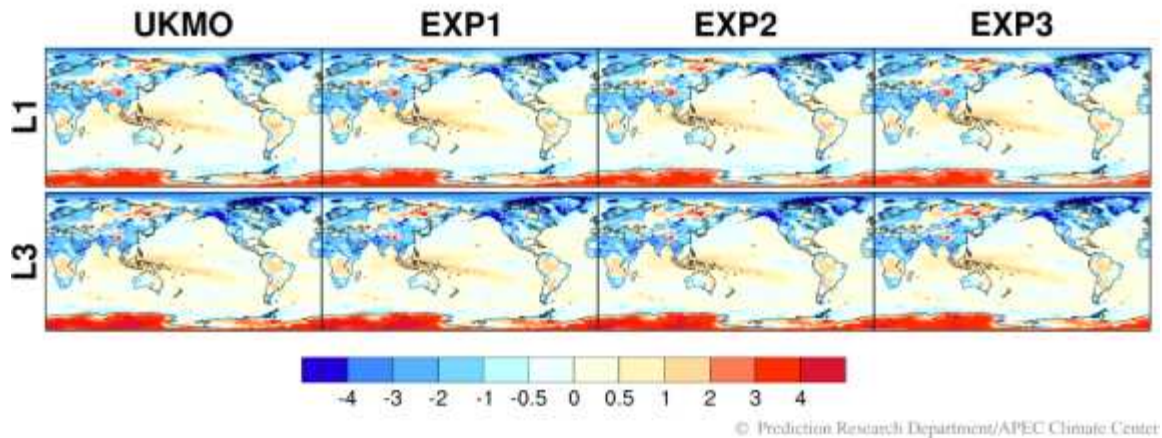
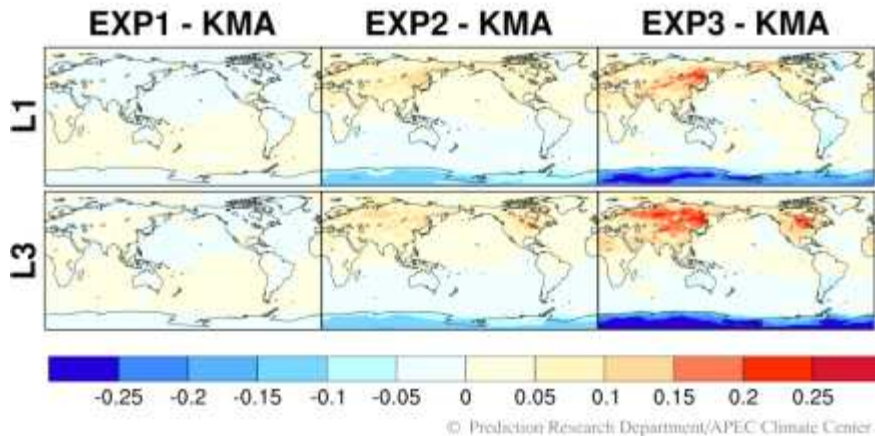


Figure 33. Mean biases for both of T2M of KMA(UKMO), EXP1, EXP2, and EXP3 for 1-, and 3-lead week during DJF 1999-2010.

기후값 실험에 의한 mean bias가 KMA의 mean bias에 비해 어느 정도 개선되었는지 비교하기 위해 각 기후값 실험에서 KMA의 기후값을 뺀 결과를 선행시간 1주와 3주에 대해 Figure 34에 제시하였다. EXP1는 겨울철 남극 기온의 warm bias나 북반구 대륙의 cold bias가 KMA에서 나타나는 모습과 크게 차이가 없다. EXP2의 경우, 남극 기온의 warm bias와 북반구 대륙 cold bias가 약간 개선되었고, EXP3는 EXP1와 EXP2에 비하면 상당 수준 보정되었고, 선행시간 1주보다 3주에서 더 많이 개선되었다. 따라서 KMA에 나타나는 mean bias가 가장 많이 개선된 실험은 EXP3으로 판단된다. 그림은 제시하지 않았지만 다른 계절에 대해서도 EXP3 실험에서 mean bias가 가장 크게 보정되었다. UKMO 모델에도 KMA와 비슷한 모습의 mean bias가 존재하는데, 동일한 실험을 UKMO에 적용하였을 때 역시 EXP3 방법을 통해 가장 많은 보정이 이루어졌다.

KMA: Clim diff of T2M, DJF, 1999-2010



UKMO: Clim diff of T2M, DJF, 1999-2010

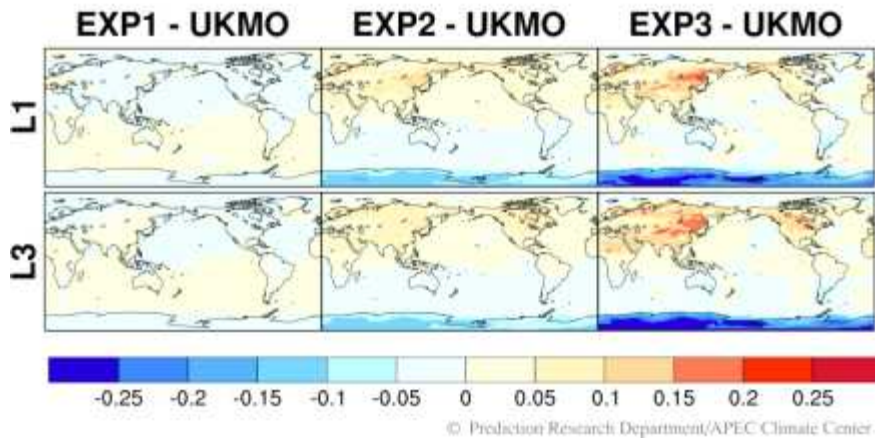


Figure 34. Climatology difference for T2M between EXP1, EXP2, and EXP3 and KMA(UKMO) for 1-, and 3-lead week during DJF 1999-2010.

㉔ 기후값 실험 간의 RMSE 분포 비교

앞서 mean bias 관점에서 보정이 가장 많이 이루어진 EXP3의 연간 기온의 기후값 RMSE의 분포를 선행시간 1, 3, 5, 7주에 대해 KMA와 비교해보았다(Figure 35). KMA와 EXP3 기후값 모두 선행시간에 따라 RMSE가 전반적으로 증가하는 모습을 보인다. 이는 선행시간에 따라 오차가 증가하는 당연한 모습이다. 그러나 오차의 증가 속도 관점에서 살펴보면 EXP3의 RMSE 증가 속도는 KMA에 비해 느린 것을 볼 수 있다. 따라서 RMSE 관점에서도 EXP3의 기후값이 KMA보다 더 나은 기후값 재현성능을 보여주는 것으로 판단할 수 있다. 동아시아에서는 다른 지역에 비해 KMA와 EXP의 RMSE 성장 속도의 차이가 크지는 않지만, RMSE 자체가 큰 지역임에도 불구하고 기존 기후값에 비해 RMSE가 천천히 증가한다. 계절내 예측에서 가장 많이 활용될 것으로 기대되는 선행시간 3주의 예측성능을 중심으로 살펴보면 EXP의 최대/최소값 뿐만 아니라 상/하위 25퍼센타일, 중앙값, 평균값이 전 지역에서 KMA에 비해 낮다. 따라서 EXP3의

재현성능은 현업에서 사용하는 기후값에 비해 전반적으로 향상되었다고 할 수 있다.

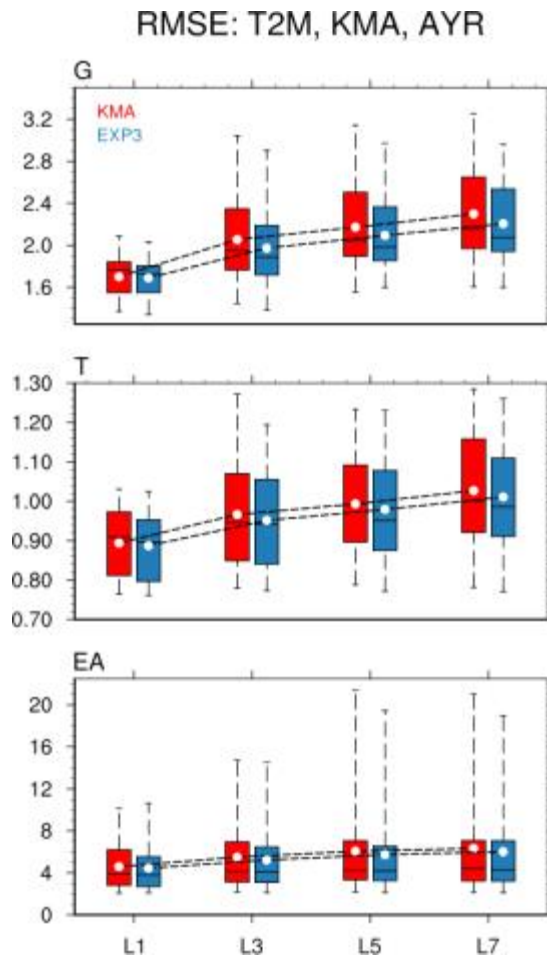


Figure 35. Annual RMSE of climatology of T2M for 1-, 3-, 5-, and 7-lead week over Globe, Tropics, and East Asia of KMA (red) and EXP3 (blue) during 1991-2010. White dot indicates the average value for each case.

(나) 기후값 생산방안에 따른 실시간 예측성능에 미치는 영향 평가

① KMA 실시간 예측 대비 상대적 ACC 차이

KMA에서 생산하는 실시간 예측자료에서 각 기후값 실험을 통해 생산된 기후값을 제거하여 실시간 예측 아노말리를 생산하고 기후값 생산방안에 따른 실시간 예측성능을 살펴보았다. Figure 36에는 지역별, 계절별, 변수별 KMA 기후값을 사용한 실시간 예측성능 대비 각 기후값 생산방안의 실시간 예측성능을 상대적으로 비교하여 통합적으로 나타내었다. 변수별 계절별 차이는 있으나 전반적으로 각 실험의 예측성능은 기존의 현업에서 사용하는 방법에 비해 향상되었다. 이들 중 EXP3의 예측성능은 동아시아를 제외하고 다른 기후값 실험에 비해 크게 향상되

었다. 그러나 가을철에는 EXP3을 이용한 동아시아 예측성능이 다른 실험보다 크게 향상되는 것을 볼 수 있다. 따라서 실시간 예측에서도 기후값 재현에서와 마찬가지로 EXP3의 예측성능이 우수함을 알 수 있다.

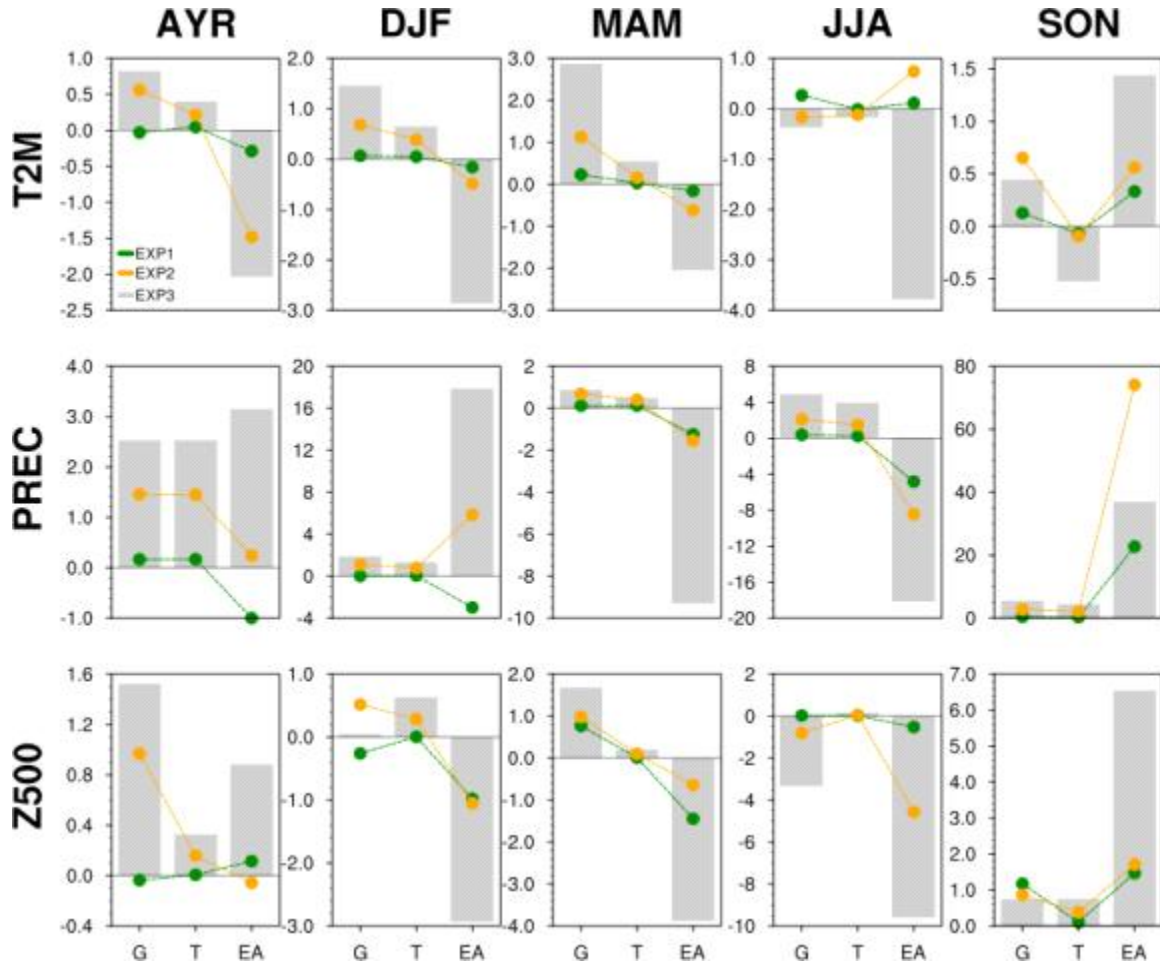


Figure 36. Annual and seasonal averaged relative ACC difference of 2019 real-time forecast for 3-lead week of T2M, PREC, and Z500 over Globe, Tropics, and East Asia of EXP1(green line), EXP2(orange line), and EXP3(gray bar) with KMA. The unit is %.

② KMA 실시간 예측 대비 ACC 차이의 분포

KMA와 각 기후값 실험의 연간 실시간 예측 간 ACC 차이를 아래와 같이 계산한 후, 상/하위 25퍼센타일, 중간값, 평균값의 분포를 Figure 37에 나타내었다.

$$ACCdiff = ACC_{other} - ACC_{KMA}$$

EXP1은 전반적으로 KMA보다 약간 우월한 예측성능을 나타낸다. 이는 기후값을 생산하는 데 있어서 거리가중치를 주지 않는 것이 예측성능을 조금이나마 향상하는 데 도움이 될 수 있다고 할 수 있다. 기존 방법에서 reforecast 수를 확장하는 EXP2 방법은 KMA와 EXP1에 비해 더 향상된 예측성능을 보여주고, 계절내 예측의 타깃이라고 할 수 있는 선행시간 3주에서도 전반적으로 예측성능이 우수하다. 기후값 생산 시 거리가중치 제거와 reforecast 수 증가 중 하나를 선택해야 한다면 reforecast 수 증가를 선택하는 방향이 조금 더 예측성능 향상에 기여할 수 있을 것으로 판단된다. EXP1과 EXP2의 장점을 모두 취하는 EXP3의 예측성능은 동아시아 기온을 제외하고는 EXP1과 EXP2에 비해 예측성능이 눈에 띄게 향상되었다. 예측성능이 낮다고 알려진 동아시아 강수의 선행시간 3주 예측도 EXP3은 KMA나 다른 실험에 비해 미미하지만 더 좋은 예측성능을 보여준다. 전반적으로 판단했을 때, 예측성능을 향상하기 위해서는 기후값을 생산할 때 거리가중치를 주는 대신 예측 시점에서 멀어지더라도 사용하는 reforecast 수를 늘리는 방향으로 리소스를 집중하는 편이 더 효과적이라고 할 수 있다.

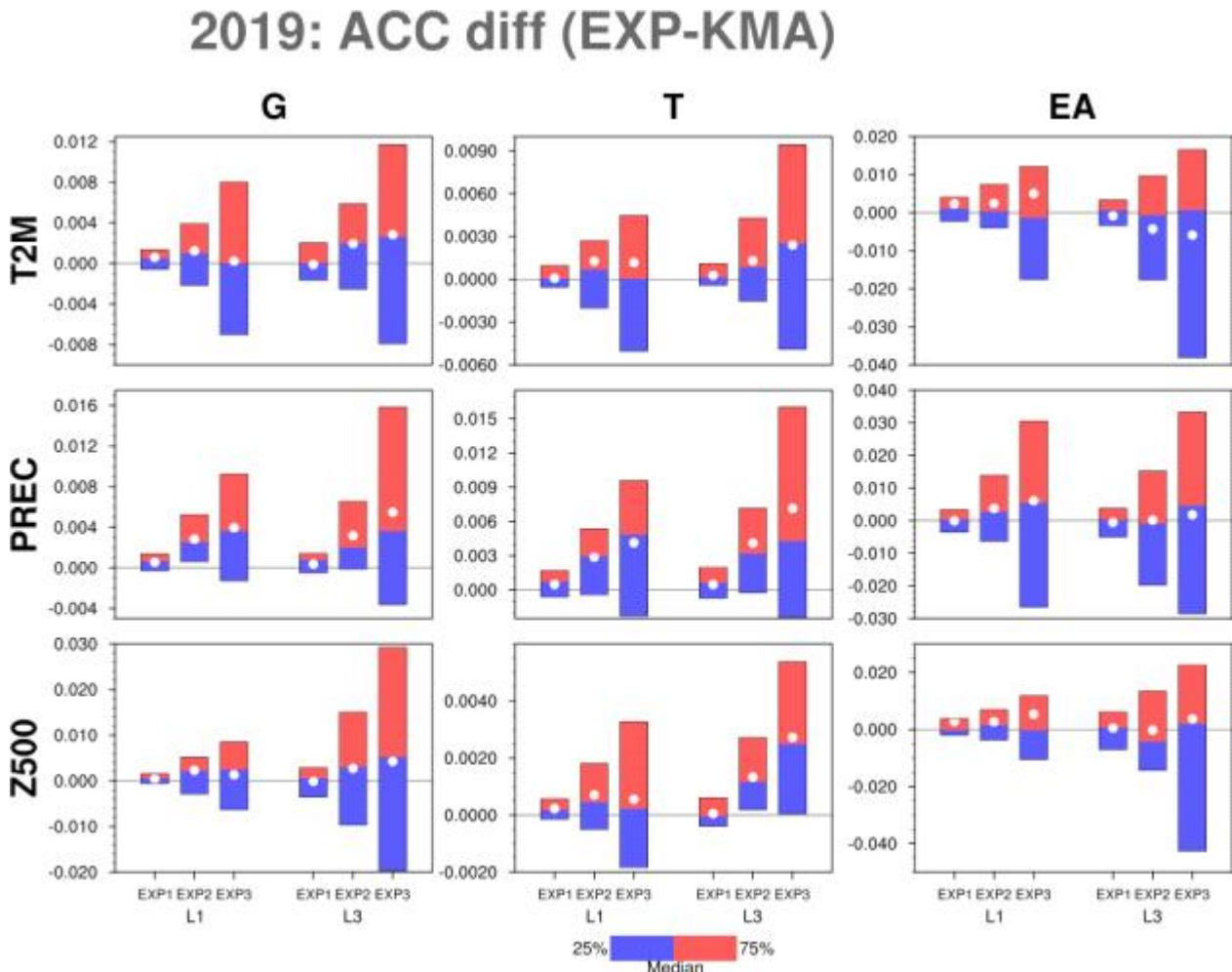


Figure 37. Annual ACC difference range between upper and lower 25% of 2019 real-time forecast for 1-, and 3-lead week of T2M, PREC, and Z500 over Globe, Tropics, and East Asia of EXP1, EXP2, and EXP3 with KMA. White dot indicates the average value for each case.

(4) KMA-GloSea5 기반 동아시아 겨울 기온 예측 성능 추정 알고리즘 개발

(가) 기후지수 교차 조건 및 예측 안정도 기반 주평균 예측 성능 의존성 파악

기후 예측 서비스는 미래의 대기 상태뿐만 아니라 그 불확실성에 대한 정보 역시 제공함으로써 예측 정보에 대한 사용자의 의사 결정을 돕는 역할을 수행하게 된다. 대부분의 현업 기관 및 기후 센터들은 과거 재현 자료의 성능 검증표나 혹은 최근 예측 성능 평균값을 예보 정확도의 근거로 제시하고 있다. 하지만, 이들은 해당 예측 시스템의 전반적인 성능 즉 기대값을 의미하기 때문에 매 실시간 예측에 최적화시켜 그 성능의 범위를 좁혀주지는 못한다. 계절내 예측, 특히 동아시아 겨울 기온 선행 3주차 주평균 예측은 그 중요성에 비해 스킬이 매우 낮다. 매 주 예보를 내보내는 시점에 그 성능의 추정치도 함께 제공할 수 있다면, 계절내 예측 자료의 사용성이 좀 더 높아질 수 있을 것이다. 이러한 배경 하에, 2019년 “다중모델 기반 계절내 예측 활용 가이드라인 개발” 과제에서는 MME 예측을 활용한 겨울철 동아시아 주평균 기온 예측 성능 추정 가이드라인을 개발하였다. 하지만, 현재 계절내 MME를 현업적으로 생산하지 않기 때문에 기존의 가이드라인을 즉시 자료로 활용하기 어렵다는 한계가 있다. 따라서, 실시간으로 활용가능한 예보 정보인 KMA-GloSea5 자료에 기반하여 기존의 가이드라인을 수정할 필요가 있었다. 본 과제에서는 실시간 예보 성능 추정을 위한 도출된 기존 세 가지 조건들이 KMA-GloSea5 모델 자료에서도 적절한지 분석하고, 그에 따라 알고리즘을 수정하였으며, 실제 2019/20년 겨울철 예보에 대해서 새로운 알고리즘의 적용성을 평가하였다. 또한, 기존 MME기반 알고리즘의 경우 과거 재현 자료 공통 기간이 1999-2010 총 12년으로 짧았지만, KMA-GloSea5 단일 모형을 활용함에 따라 1991-2010 총 20년으로 샘플 수와 기간을 늘릴 수 있었다.

동아시아 도메인에서 겨울 기온의 주간 변동을 경험적 직교 함수 분석으로 나타내면 아래 (Figure 38)와 같다. 첫 번째 모드는 한반도를 포함한 도메인 전체가 진동하는 형태이고, 두 번째 모드는 한반도를 포함한 남쪽과 고위도가 나뉘어서 시소 형태로 진동하는 형태이다. 두 모드 역시 다른 기후 리짐들과 함께 예보 성능 추정의 인자로 검토되었다.

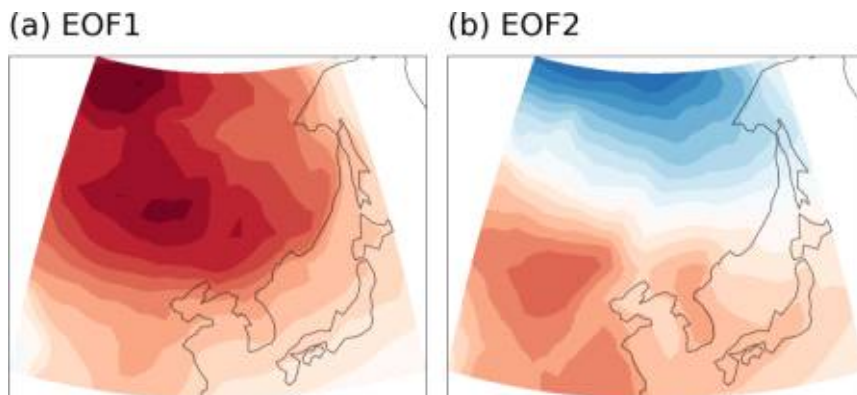


Figure 38. Spatial patterns of two leading modes of T2m over East Asia region (30° - 65° N, 105.5° - 145.5° E).

㉞ 동아시아 겨울 기온 예측력의 지속 특성

계절내 모형의 성능은 대체로 선행 1주에서 제일 높고, 선행 시간이 길어짐에 따라 급격히 감소하고 그 폭이 커지는 특징이 있다. 초기 예측 성능이 긴 선행시간에도 일정 정도 유지되는 특성을 기존 MME 기반 분석에서 확인한 바가 있다. 단기 예보 기간보다 조금 길면서 활용성 측면에서 수요가 높은 예보 3-4주차의 성능이 초기 선행 1주차 성능을 따르는지 여부를 파악하고자 각 예보들을 선행 1주 성능을 기준으로 Above Normal (AN), Near Normal (NN), Below Normal (BN)의 삼분위로 구분하고, 각 카테고리 예보들의 매 선행 주수 성능을 박스-휘스커 그래프로 분석해 보았다 (Figure 39). 선행 1주의 경우 정의에 따라 완벽히 3분위로 나뉘고, 대부분의 예보들이 높은 성능을 보여 AN과 NN의 범위가 매우 좁고 BN 영역이 매우 넓게 나타난다 (Figure 39a). 하지만, 예보 2주차부터 이러한 1주차 성능 구분 양상을 완전히 잃어버린다 (Figure 39bc). 이는 KMA-GloSea5 예측력을 가늠하는 데 있어 선행 1주 성능 정보를 활용하는 무리가 있음을 의미한다. 따라서 새로운 알고리즘에서는 “예측 스킬 지속성”은 성능 추정 인자에서 제외시켰다.

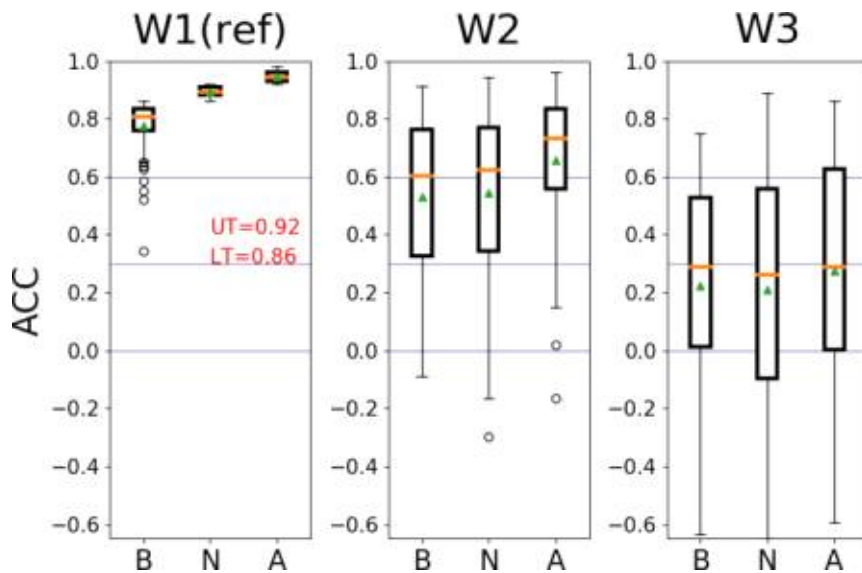


Figure 39. Box-whisker of ACC for BN, NN and AN groups at different 4 lead weeks. The three categories are based on lower (33.3%) and upper (66.6%) terciles of ACC at lead week 1 (displayed with texts). Green triangles indicates the mean of ACC.

㉟ 기후 리짐 위상 교차 조건에 따른 예측력 의존성

동아시아 겨울 기온은 다양한 기후 리짐과 역학적으로 관련되어 있기 때문에, 특정 기후 모드 위상에 따라 그 성능이 뚜렷이 월등한 조건이 존재할 것이다. 이런 조건들을 파악함으로써 실시간 예측 성능을 좀 더 구체적으로 추정할 수 있겠다. 본 연구에서는 동아시아 평균 기

온값, 경험 직교 함수에서 얻어진 2개의 변동모드 시계열, 그리고 8개 기후 지수를 포함 총 11개 지수를 활용하였다. 지수 각각에서 삼분위 값을 찾고, 매 두 개 지수 조합에 대하여 3by3 교차 조건에 따른 9개 카테고리에서의 예측 성능 기대값을 확인한다. 모든 지수 조합들 (부록 A) 중에서, 9개 카테고리 내에서 샘플 개수가 충분하면서 스킵 차이가 뚜렷한 카테고리가 존재하는 조합만을 Figure 40에 나타내었다. 또한 기존 알고리즘과 달리, 선행 3주 동시 조건뿐만 아니라, 선행 1-2주에서의 조건도 적용하여 성능 산출을 위한 샘플을 늘릴 수 있도록 하였다. 총 14개의 다양한 교차 조건에서 유의미한 예측성 구분이 이루어졌다. 이 파트는 실시간 예측성 추정을 위한 MOS (model output statistics) 관점으로 접근하였기 때문에, 조건 구분을 위해 사용된 지수들은 모두 예측장에서 얻어진 지수들이다. 예측장 안에서의 역학관계가 관측과 정확히 같지 않을 것이므로, 관측 지수 조건을 적용한다면 상이한 결과를 얻을 수도 있을 것이다. 부록 A에서 제시한 결정트리에서 확인할 수 있듯이, 예보장에서 얻은 기후 지수 교차 조건에 따라 순차적으로 삼분위로 성능을 추정을 하고 그것들을 합산하여 최종 예측력의 삼분위 확률을 추정할 수 있다.

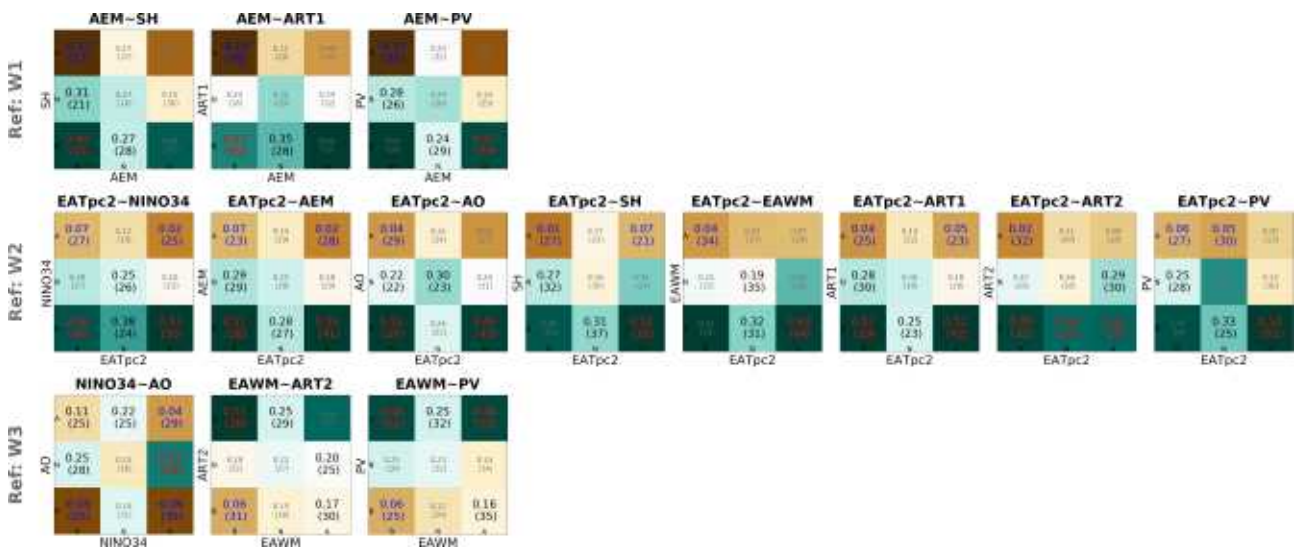


Figure 40. W3 mean ACC of 9 categories based on two different indices' tercile criteria of (top) W1, (middle) W2, and (bottom) W3. For the tercile categorization, anomaly of each index is calculated then it is normalized by its standard deviation. Upper and lower tercile is -0.43σ and 0.43σ , respectively. ACC average value and sample size within parenthesis are displayed in each box. Bold blue (red) color denotes the category of distinguished large (small) ACC average value greater (less) than 0.43σ (-0.43σ) of all ACCs, having the sample size 20 and over and the sample spread less than 1σ of all ACCs. Bold black color denotes the category of ACC average between -0.43σ and 0.43σ , having the sample size 20 and over and the sample spread less than 1σ of all ACCs. Here, only the cross-conditions whose 9 categories include at least one of all three blue, black, and red categories are displayed.

㉔ 1-4주 예측 안정도에 따른 성능 민감도

계절내 예측 현업 운영자의 경험에 따르면 선행 주수 증가에 대해 안정된 예보일수록, 전반적인 성능이 높고, 반대로 안정도가 떨어질수록 성능이 낮은 경향이 있다고 한다. 예를 들어,

전체 기간 안에서 기온의 등락이 거의 없이 안정된 상태로 예측된 경우에 전반적인 성능이 높았고, 반대로 기온의 등락이 심하게 예측된 경우는 전반적인 성능이 낮다는 의미이다. 이는 극한 기온 현상에 대한 단주기의 소규모 현상을 제대로 모의하지 못하는 모형의 한계라고 볼 수 있겠다. 기존 MME 기반 성능 추정 알고리즘에서 동아시아 기온 EOF 첫 번째 모드의 안정도에 대해 선행 3주 동아시아 겨울 기온 ACC는 대략적으로 비례관계가 나타났다. KMA-GloSea5에 대한 두 값의 산포도와 안정도의 3분위 기준에 따른 성능의 아메바 그래프는 아래와 같다 (Figure 41). 여기서 안정도는 (표준 편차/평균)의 역으로 정의하였다. 산포도에서 안정도가 클 경우는 성능이 낮은 경우가 거의 없고, 시간상관계수도 통계적으로 유의한 0.254 값을 보인다. 아메바 그래프에서 성능 분포를 확인하면 평균과 최빈값이 AN에서 높고 범위도 좁아 불확실성이 줄어든 것을 알 수 있다. 기후 리짐 교차 조건과 함께 안정도 조건 역시 실시간 예보 성능 추정 인자로 선택되었다.

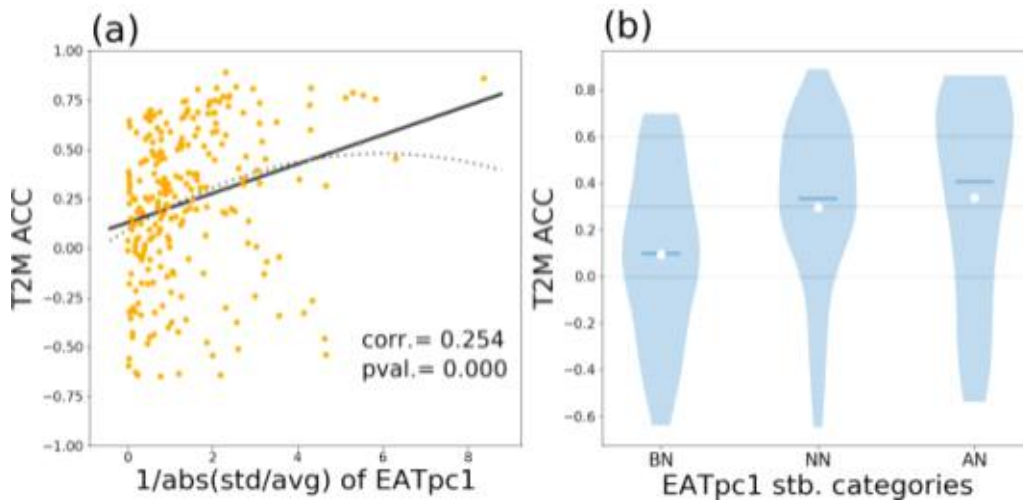


Figure 41. (a) scatter plot between simulated PC1 stability and EA T2m ACC and (b) violin plot of EA T2m ACC for three PC1 stability categories for lead time of 3 weeks from 240 samples (12 weeks for 20 winter seasons). X-axis displays the stability which is defined as an inverse of PC1 standard deviation for 4 lead weeks divided by absolute value of their average. The first and second order least square fitted lines are delineated (black solid and gray dash) in (a). The three categories in (b) are based on Upper (0.43) and lower (-0.43) tercile of standardized PC1 stability. Horizontal lines and white circles in (b) indicate median and mean values of samples for three stability categories.

(나) 19/20년 겨울철 13주에 대한 알고리즘 적용성 평가

부록 A에 제시된 KMA-GloSea5 기반의 새로운 성능 추정 결정트리를 2019/2020 겨울철 12개 주평균 예보에 대해 적용했을 때의 결과는 아래의 표와 같다. 기본적으로 ACC 스킴의 3분위 카테고리를 추정하게 되고, 3주 이후에 반영되는 실제 예측값 카테고리와의 비교를 통해 알고리즘의 적용성을 판단할 수 있었다. 삼분위 성능 카테고리 일치성은, 기후 리짐 교차 조건을

적용했을 때 50% 확률을 보였다. 하지만 기온 안정도 인자를 적용했을 때는 17% 정도에 불과했다. 관측 카테고리를 기준으로 한 Hit Rate 검증 결과, 관측 성능이 낮거나 보통인 카테고리에서는 기후 리짐 교차 조건을 적용할 때 hit 확률이 높은 것을 알 수 있다. 반대로, 관측 성능이 높은 카테고리에 대해서는 기온 안정도 조건을 적용하는 것이 훨씬 높은 적중도(67%)를 보였다. 추정 카테고리를 기준으로 한 검증값 Post Agreement (PA)를 적용했을 때는 샘플 수와 적중도가 다소 변하지만, 비슷한 양상을 보이는 것을 알 수 있다. 한 해, 12주라는 적은 샘플에 적용한 결과이기에 다소 신뢰도가 떨어질 수 있으나, 그럼에도 불구하고 의미 있는 성능을 보여주었다고 생각된다. 실시간 예보 추가 테스트를 통해 알고리즘 활용에 대한 확신을 얻을 수 있으리라 본다.

Table 11. Verification of estimated ACC categories for 12 weeks during the 19/20 winter season utilizing three metrics (Proportion Correct, Hit Rate and Post Agreement). Numbers in brackets indicate the sample size considered in the score calculation.

Ingredient	Proportion Correct (PC)	Hit Rate (HR)			Post Agreement (PA)		
		ACC_O: Low	ACC_O: Moderate	ACC_O: High	ACC_F: Low	ACC_F: Moderate	ACC_F: High
Clim. Regimes cross-condition	0.50 [12]	0.60 [5]	0.50 [4]	0.33 [3]	0.75 [4]	0.50 [4]	0.25 [4]
EATpc1 stability	0.17 [12]	0.00 [5]	0.00 [4]	0.67 [3]	0.00 [4]	0.00 [2]	0.33 [6]

나. 초기장에 따른 오차 성장 특성 진단

(1) 최적 초기장 실험의 계절내 예측 오차 진단

(가) 초기 오차 특성 및 오차 성장 진단

① 과거재현기후에서의 초기 오차 특성 진단

㉠ 과거재현기후의 예측 특성

초기장에 따른 오차 성장 특성을 살펴보기 위해 먼저 기후값으로 사용되는 과거재현기후의 예측 특성에 대해 살펴보았다. 지표 기온에 대한 선행 시간(1-60일)에 따른 전지구 및 동아시아 예측 성능을 figure 42와 43에 나타내었다. 참고 자료로는 ERA5자료가 사용되었으며 예측 성능은 Anomaly Correlation Coefficient (ACC), Root Mean Square Error (RMSE)로 나타내었다. ACC(RMSE)가 1(0)일 때를 perfect prediction으로 볼 수 있으며, 편의를 위해 ACC가 0.5일 때를 점선으로 표기하였다. 1991-2010의 모든 경우의 ACC를 shading으로 표기하였으며, 그 평균을 파란 실선으로 표기하였다.

Figure 42에서, 전지구, 동아시아 지역 모두 선행시간이 증가할수록 급격하게 예측 성능이 감소하고 있음을 확인할 수 있다. 특히 평균적으로 10일을 전후로 예측 성능이 0.5 아래로 떨어지며 3주를 전후로 예측성능이 0.1-0.2 정도에서 수렴하고 있음을 볼 수 있다. 동아시아 지역의 경우 자료에 따라 편차가 전구보다 조금 더 크게 나타나지만, 그 특징은 동일하게 나타나고 있다. Figure 43에서, 전지구, 동아시아 지역 모두 선행시간이 증가할수록 급격하게 예측 오차

가 증가하고 있음을 확인할 수 있다. ACC와 마찬가지로 선행 시간 3-4주를 기준으로 그 오차는 수렴되고 있음을 알 수 있다. 즉, 계절내 예측 규모인 3-4주 기간에서 지표 기온의 예측 성능은 지역에 따라, 조건에 따라 변할 수 있다고 해석할 수 있다. 또한, 상대적으로 낮은 계절내 및 계절 예측 성능을 극복하기 위한 다양한 시도가 필요한 상황임을 확인할 수 있다.

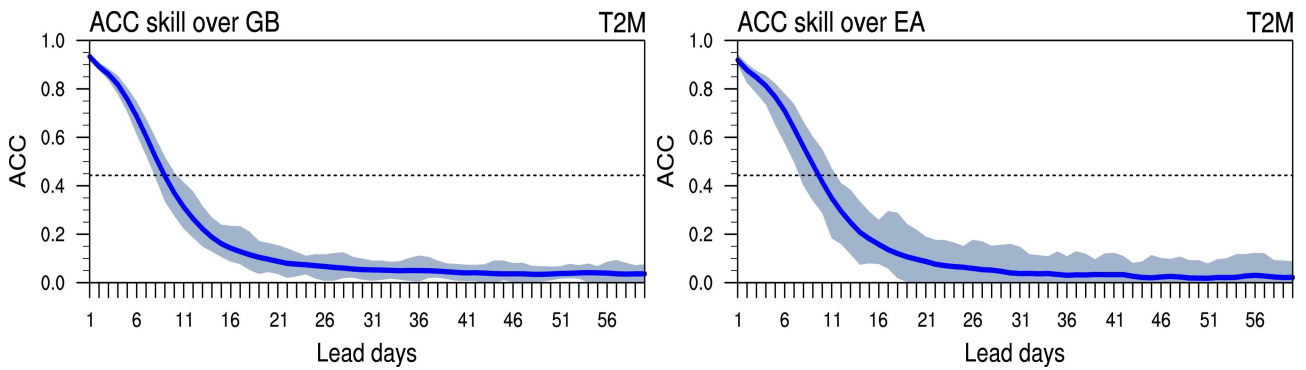


Figure 42. Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) Blue shading indicates all ensemble member of hindcast data (1991-2010). Blue solid line means averaged all members.

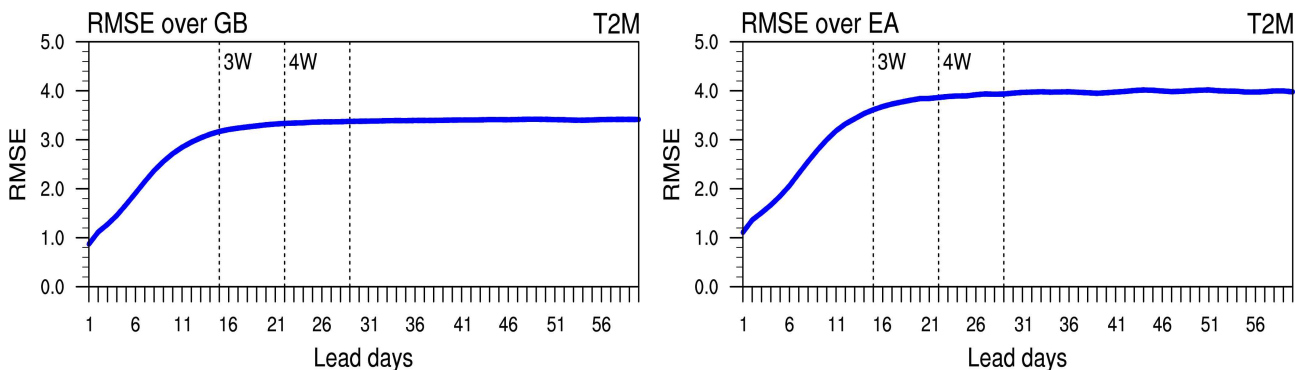


Figure 43. Root Mean Square Error of surface temperature over globe (left) and East Asia (right).

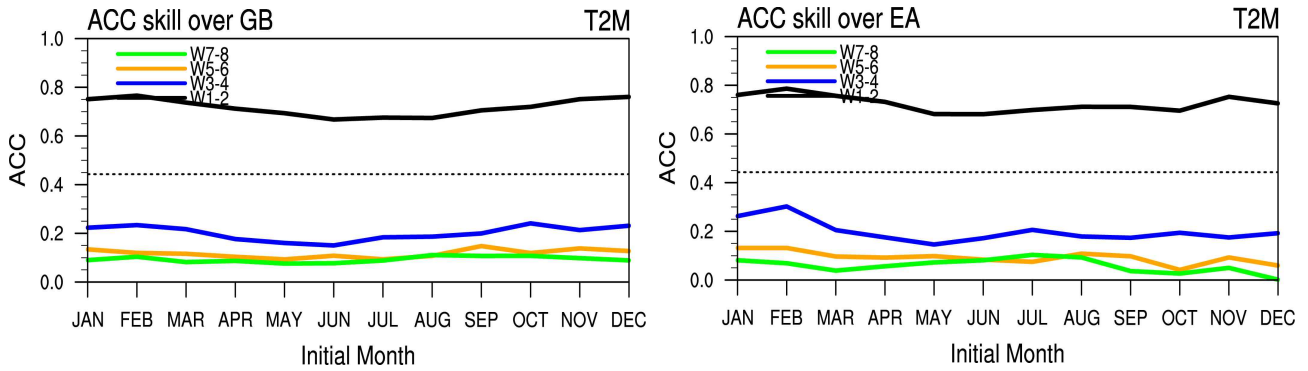


Figure 44. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) for initial month. Black, blue, orange and green lines indicate averaged lead weeks 1-2, 3-4, 5-6, and 7-8, respectively.

Figure 44은 월별 평균된 ACC를 선행 시간에 따라 나타낸 그림이다. 예상할 수 있듯이 전 계절에 동일하게 선행 주수 1-2주의 예측 성능이 0.7 정도로 가장 높게 나타나며, 북반구 겨울철에 비해 여름철의 예측 성능이 다소 떨어지는 경향이 있다. 그에 비해 선행 주수 3-4주부터는 예측 성능이 0.3이하로 확연하게 떨어지는 것을 알 수 있다. 또한 예측 주수 5-6주, 7-8주의 예측 성능은 0.2 이하로 거의 동일하게 나타나고 있다. 동아시아 지역의 경우 예측 주수 3-4주의 경우 2월 초기장의 예측 성능이 가장 높게 나타나고 있으며 봄철의 예측 성능이 가장 낮게 나타나고 있다. 또한, 예측 주수 1-2주에서 겨울철 성능이 여름철보다 높게 나타나는 반면, 예측 주수 5주 이상에서는 12월의 예측 성능이 매우 낮게 나타나는 특징이 있다. 특히 동아시아 겨울철의 경우, 가장 중요하게 고려되는 변수인 500hPa 지위고도의 예측성능 역시 1월 초기장에서 낮게 나타나고 있어(나타내지 않음) 겨울철 분석의 경우 주의를 요할 필요가 있다.

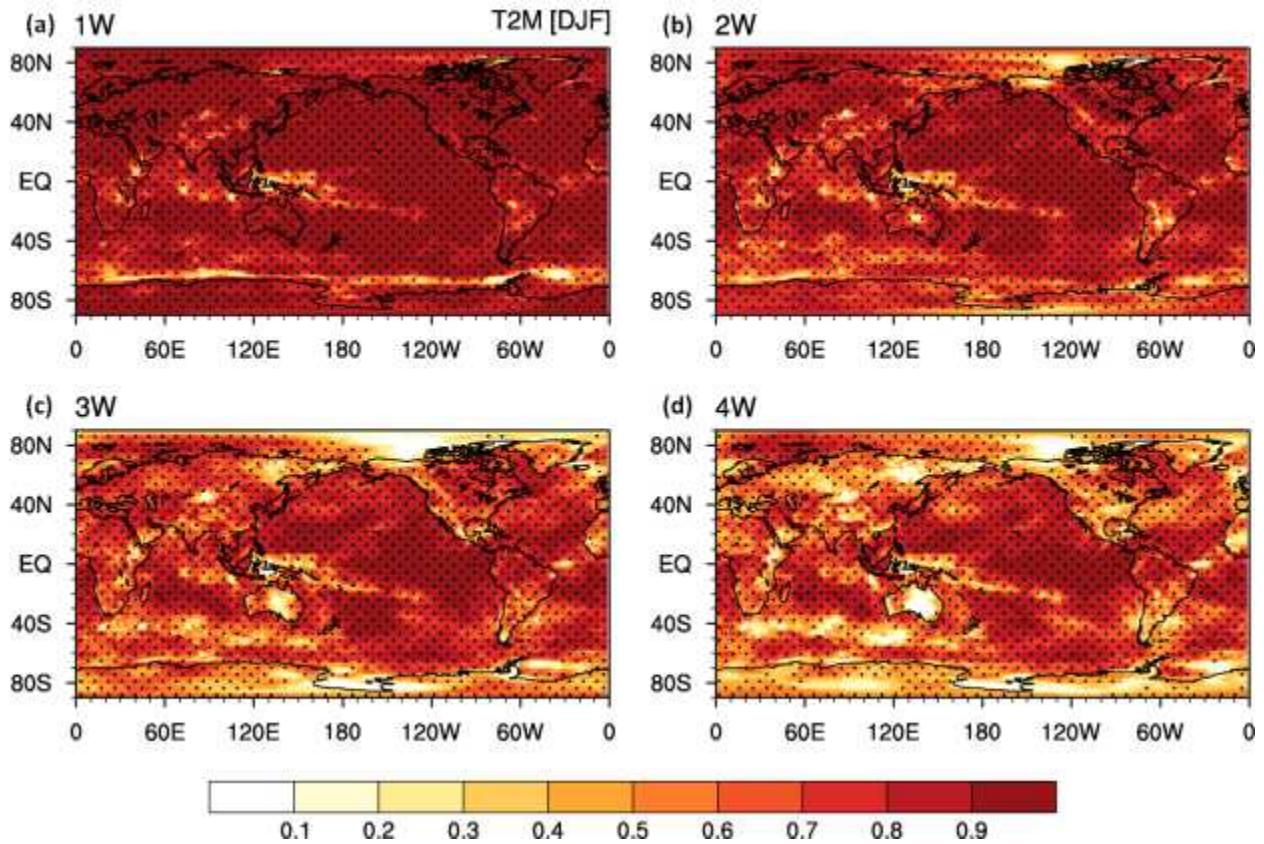


Figure 45. Temporal correlation coefficient (TCC) of the 2m temperature between ERA interim reanalysis data and GloSea5 hindcast with (a)-(d)1-4 week lead for DJF. Dots in (a)-(d) indicates the area statistically significant at the 95% level.

지표 온도 및 강수의 주별 선행시간의 시간상관계수 분포를 figure 45과 figure 46에 각각 나타내었다. 주별 지표 온도의 예측성능을 살펴보면 선행시간 4주까지 0.6이상의 높은 시간상관계수를 보이고 있음을 알 수 있다. 선행시간 3-4주에서는 극지역 및 Maritime Continent 지역 일부, 대륙 지역을 중심으로 상관계수가 다소 낮아진다. 지표 온도와는 달리 강수의 주별 평균 시간 상관 계수의 분포는 적도 태평양 강수대를 제외하고 매우 낮게 나타나고 있다. 이는 대부분의 기후 예측 모형에서 나타나는 특징이다. 선행시간 1주에는 대부분의 지역에서 높은 상관 계수를 보이지만, 선행시간이 길어질수록 상관 계수는 급격히 감소한다. 특히 선행 시간 3-4주에 육지지역 및 중위도, 극지역 대부분에서 상관계수는 0에 가깝게 낮다. 하지만, 적도 태평양 및 대서양 인도양 일부까지 높은 상관계수를 유지하고 있다.

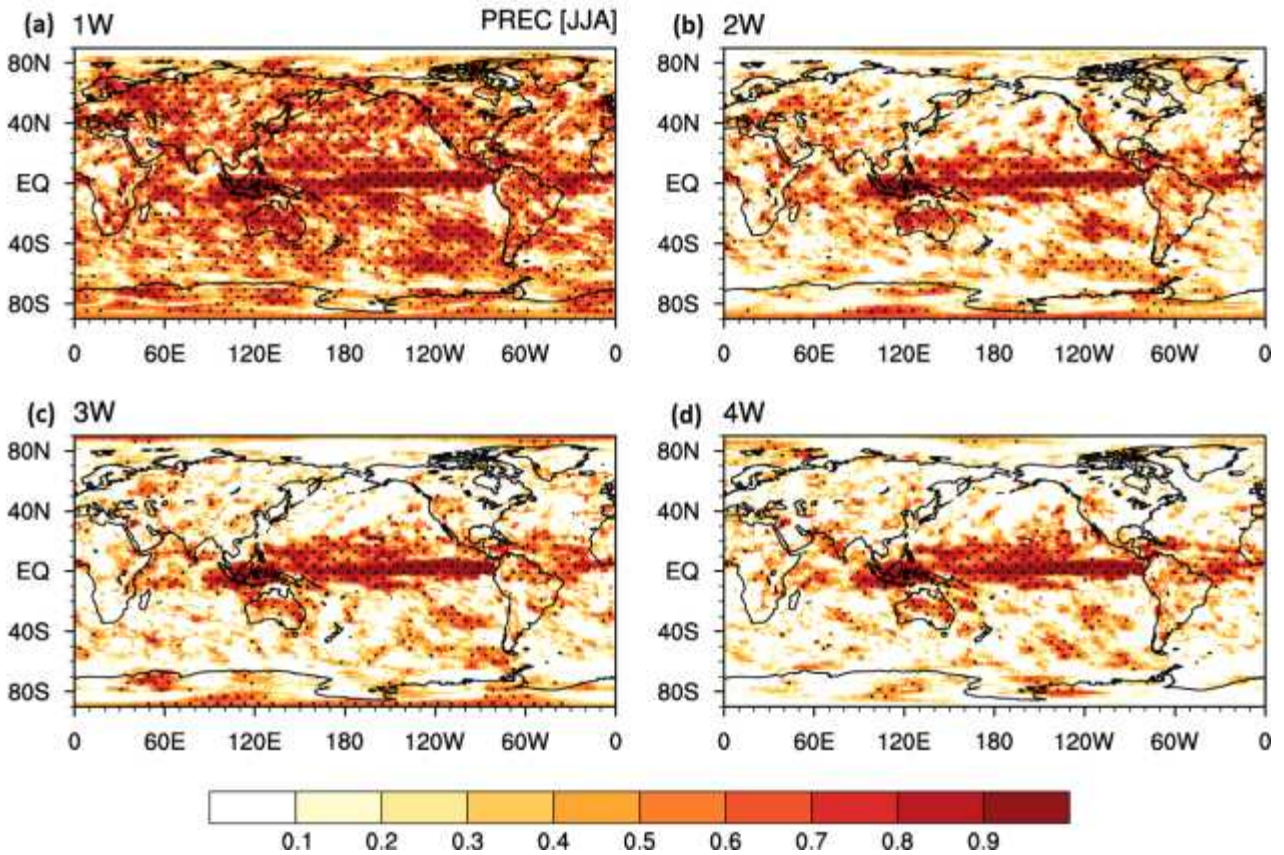


Figure 46. Temporal correlation coefficient (TCC) of the precipitation between ERA interim reanalysis data and GloSea5 hindcast with (a)-(d)1-4 week lead for DJF. Dots in (a)-(d) indicates the area statistically significant at the 95% level.

㉠ 선행시간 1일에서의 오차 특성

초기장 확보의 어려움으로 초기 오차는 선행시간 1일의 평균장으로 대신하여 분석하였다. Figure 47은 선행시간 1일에서의 여름철/겨울철 지표 온도의 오차를 나타낸 그림이다. 전체적으로 여름철/겨울철 모두 지표 온도의 초기 오차는 크지 않게 나타나고 있다. 선행시간 1일의 지표 온도의 오차는 전반적으로 음의 오차를 나타내고 있음을 알 수 있다. 여름철의 경우 동아시아, 북미, 남극 지역에서 양의 오차를 나타내며, 겨울철의 경우 북극 지역을 비롯한 고위도 지역을 중심으로 양의 오차를 나타내고 있다. Figure 48은 선행시간 1일에서의 여름철/겨울철 지표 강수의 오차를 나타낸 그림이다. 강수는 전반적으로 음의 편차가 지배적으로 나타나고 있다. 여름철의 경우 적도 태평양에서 동아시아 연안에 이르기까지 음의 편차가 나타나고 있으며 ITCZ를 중심으로 위/아래 지역에서 양의 편차가 나타나고 있다. 특히 동아시아 몬순과 연관된 한반도 지역 및 일본지역 전반에 걸쳐 음의 편차가 지배적이다. 겨울철의 경우, 서태평양을 중심으로 북서태평양 적도 지역 등에서 양의 편차가 나타나고 있다.

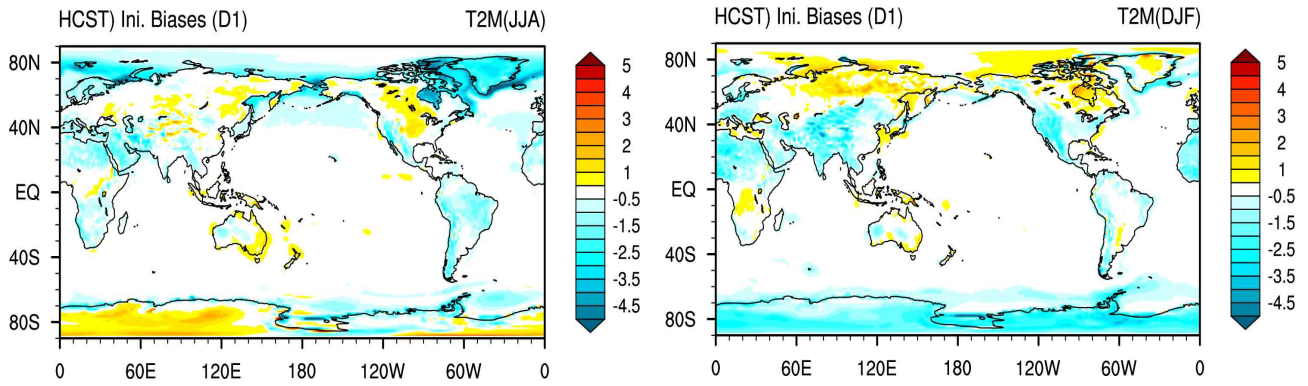


Figure 47. Distribution of averaged surface temperature from lead day 1 for JJA (left) and DJF (right) of hindcast (1991-2010).

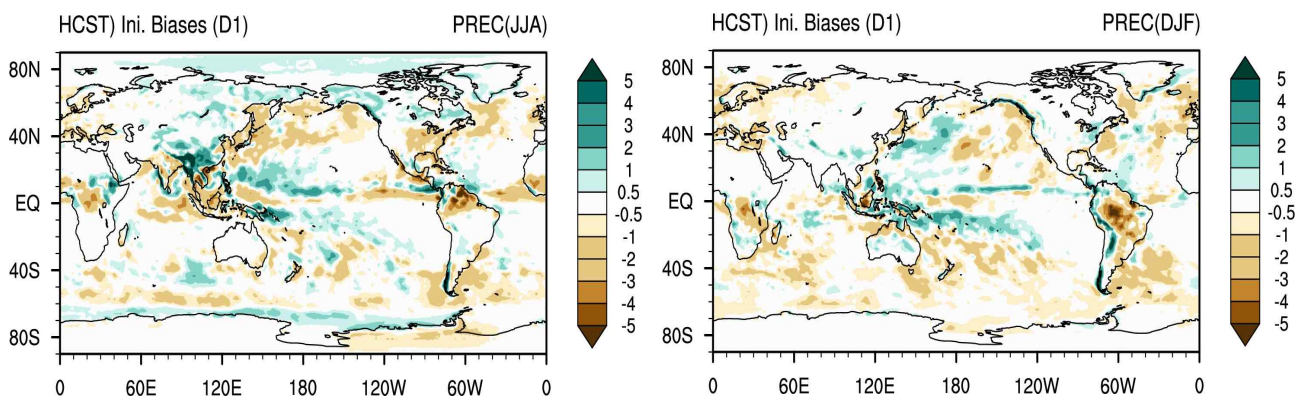


Figure 48. Distribution of averaged surface precipitation from lead day 1 for JJA (left) and DJF (right) of hindcast (1991-2010).

초기 오차의 특성을 조금 더 살펴보기 위해 여름철의 선행시간 1일의 결과를 조금 더 살펴보았다. Figure 49는 GloSea5의 여름철 과거재현기후와 ERA5 자료로부터 동서 평균된 연직 온도 및 습도의 오차를 나타낸 그림이다. 중위도 하층에서의 약한 양의 온도 오차가 있긴 하지만, 연직 온도 오차는 매우 작게 나타나고 있다. 다만 남반구 극지역에서의 강한 양의 오차가 나타나고 있지만, 이는 ERA5의 특성과도 연관이 있어, 관측의 특성을 더 살펴볼 필요가 있다. 그에 비해 습도의 오차의 경우 하층에서 넓은 지역에서 음의 오차를 보이고 있음을 확인할 수 있다. 특히 북반구 중위도에서는 약 500hPa까지도 음의 오차가 확장되어 있음을 볼 수 있다. 하층 건조 오차는 과도한 강수와 연관되어 설명될 수 있다. 이러한 특성을 조금 더 살펴보기 위해 하층 열속 및 복사 플럭스 등을 조사해보았다.

Figure 50은 과거재현기후의 현열과 잠열의 오차를 나타낸 그림이다. 현열의 오차 분포는 현열의 오차분포와 극지역을 제외하고 거의 유사하게 나타난다. 잠열의 오차 분포는 적도 지역을 중심으로 양의 오차가 나타나고 있으며, 중위도 대륙 지역을 중심으로 음의 오차가 나타나고 있다. 즉, 적도 지역을 중심으로 과도한 증발이 발생하고, 증발된 수증기는 강수 물리 과정을 통해 많은 강수로 전환되어 오히려 건조한 대기를 만들었다고 할 수 있다. 최근 이러한 적도 지역의 건조 대기는 열대 진동의 모의 성능과도 연관이 되어 있음을 언급되고 있으며, GloSea5의 건조 특성 역시 열대 진동의 모의 성능과 연관될 수 있다. 이에 대해서는 다음 세션

에서 더 논의하고자 한다.

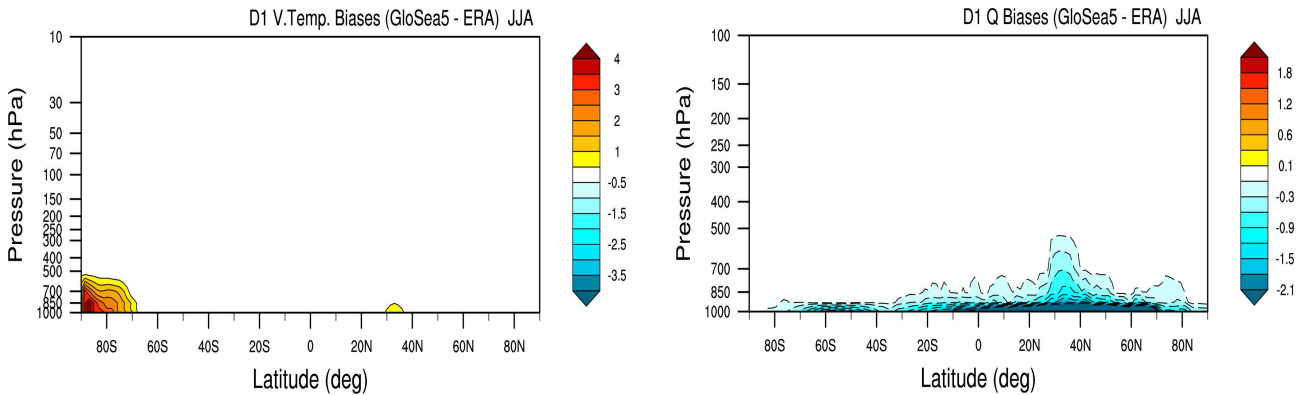


Figure 49. Vertical profiles of temperature (left) and specific humidity (right) error difference between GloSea5 hindcast and ERA5.

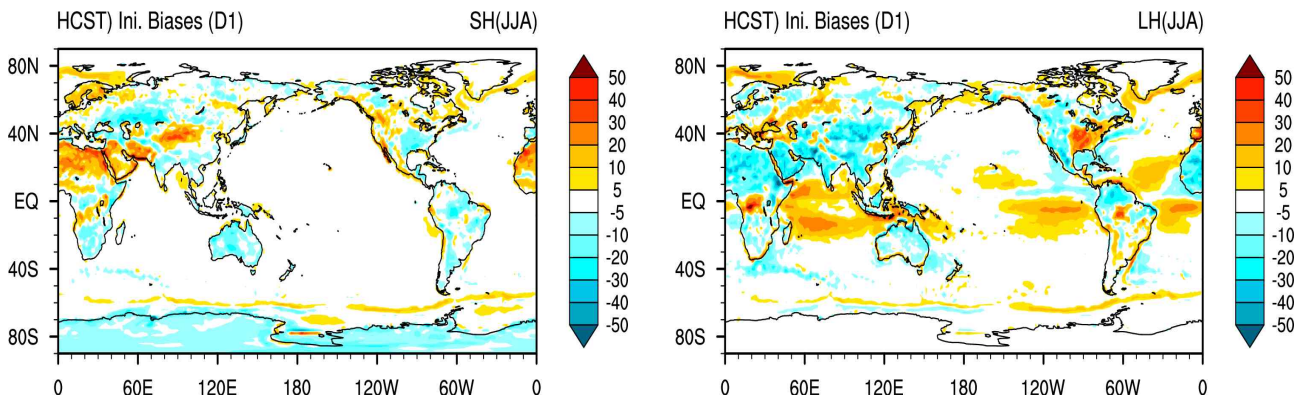


Figure 50. Distribution of averaged surface sensible (left) and latent (right) heat fluxes from lead day 1 for JJA of hindcast (1991-2010).

Figure 51은 과거재현기후의 선행시간 1일의 총운량 및 하층 습도의 오차 분포를 나타낸 그림이다. 관측과 비교하였을 때, 총운량의 초기 오차는 매우 크다. 선행시간 1일의 평균에는 모델의 첫 타임의 값이 포함되어 있기 때문에, 이는 구름 양의 초기 값의 설정에 현실적인 값이 처방되지 않았을 것으로 예상할 수 있다. 대부분의 구름 예단 및 진단 모듈에서 초기값이 임의의 값으로 처방되어 그 다음 타임스텝에서 계산되는 형식이 사용되기 때문에 이 값은 초기값 설정의 문제라고 생각할 수 있지만, 정확한 진단을 위해서는 구름 진단 모듈의 설정값을 분석해 볼 필요가 있다. 그에 비해 하층 습도의 오차의 경우, 산악지역을 제외하고는 크지 않은 오차를 보이고 있기 때문에 모델의 초기 입력 변수인 습도의 경우 적절한 값들이 잘 입력되었다고 판단할 수 있다.

이러한 결과를 바탕으로 과거재현기후의 경우 입력변수인 바람, 습도, 온도의 경우 자료동화를 통해 오차가 거의 없는 값이 적절히 처방되었다고 할 수 있다. 하지만, 입력변수들이 아

변수들은 선행시간 1일부터 오차들이 보여지며, 이러한 오차가 계통적으로 나타나는지 혹은 초기에만 나타나는지 알아볼 필요가 있다.

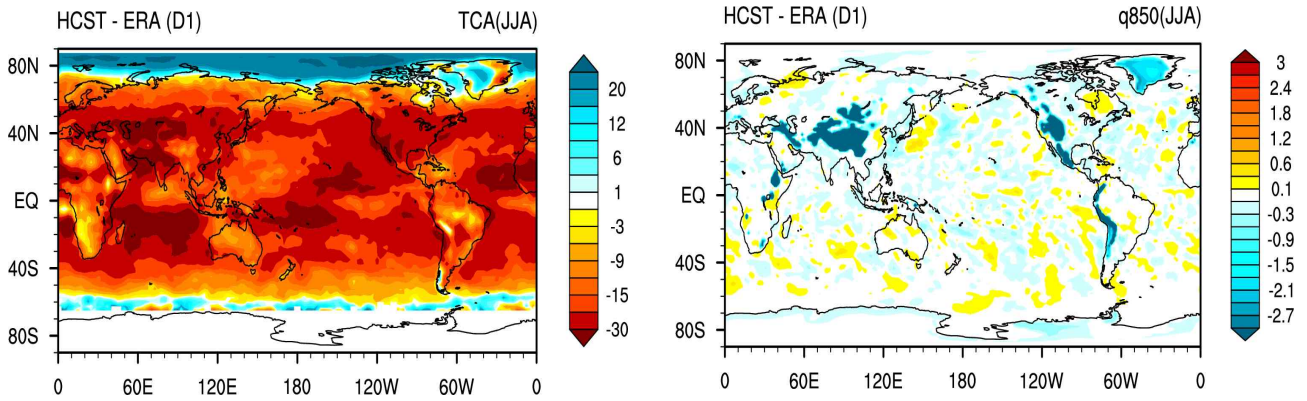


Figure 51. Distribution of averaged total cloud amount (left) and 850 hPa specific humidity (right) from lead day 1 for JJA of hindcast (1991-2010).

㉔ 선행시간 3-4주에서의 오차 특성

선행시간 1일에서 나타난 오차 패턴이 계통적으로 동일하게 나타나는지, 오차가 증가하는지, 혹은 초기에만 나타나는지를 알아보기 위해 선행시간 3-4주의 오차 특성을 살펴보았다. Figure 52는 과거재현기후에서 선행시간 3-4주의 지표 온도와 강수의 오차 분포를 나타낸 그림이다. 지표 온도의 오차 분포의 경우, 선행시간 1일과 거의 유사한 패턴을 보이고 있다. 다만 오차의 크기가 더 증가된 것을 확인할 수 있다. 북반구 중위도 대륙 지역을 중심으로 온도의 양의 오차가 두드러지게 나타나며 서태평양으로 약한 양의 오차도 발견된다. 극 지역으로는 음의 오차가 나타나고 있음을 알 수 있다. 강수의 오차의 경우 역시 초기오차의 패턴과 크게 다르지 않다. MC 지역에서 건조한 오차가 계속해서 나타나고 있고, 인도차이나반도에서 건조 오차가 나타나고 있다. 한반도 남쪽 해안으로도 약한 건조 오차가 나타난다. 다만 서태평양과 인도양 주위로 나타나던 초기오차는 건조한 형태였지만, 계절내 규모에서는 양의 오차가 크게 나타나고 있다. 다만 서태평양과 인도양 주위로 나타나던 초기오차는 건조한 형태였지만, 계절내 규모에서는 양의 오차가 크게 나타나고 있다. 다시 말해, 적도 서태평양의 양의 온도 편차와 연관된 과도한 대류 활동이 양의 강수 편차와 연관 있음을 알 수 있다.

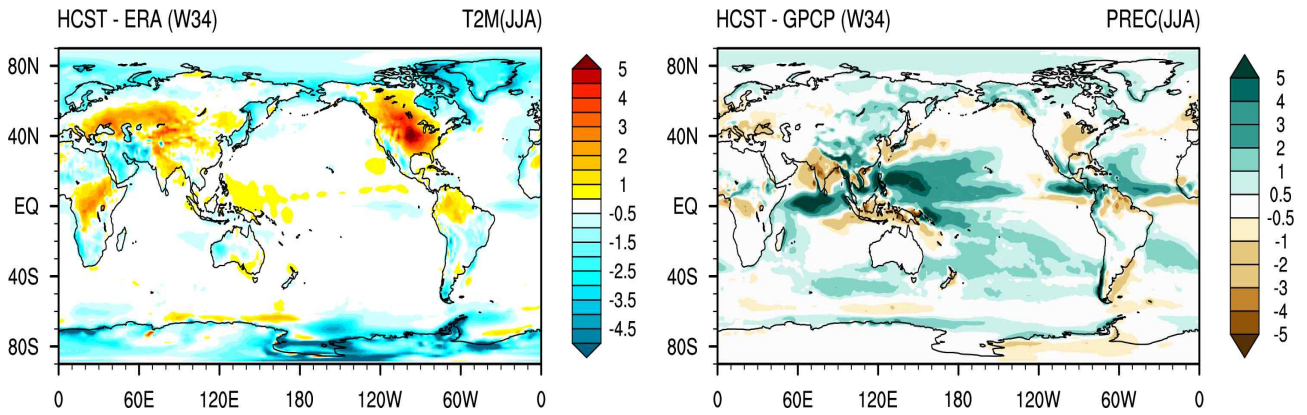


Figure 52. Distribution of averaged surface temperature (left) and precipitation (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).

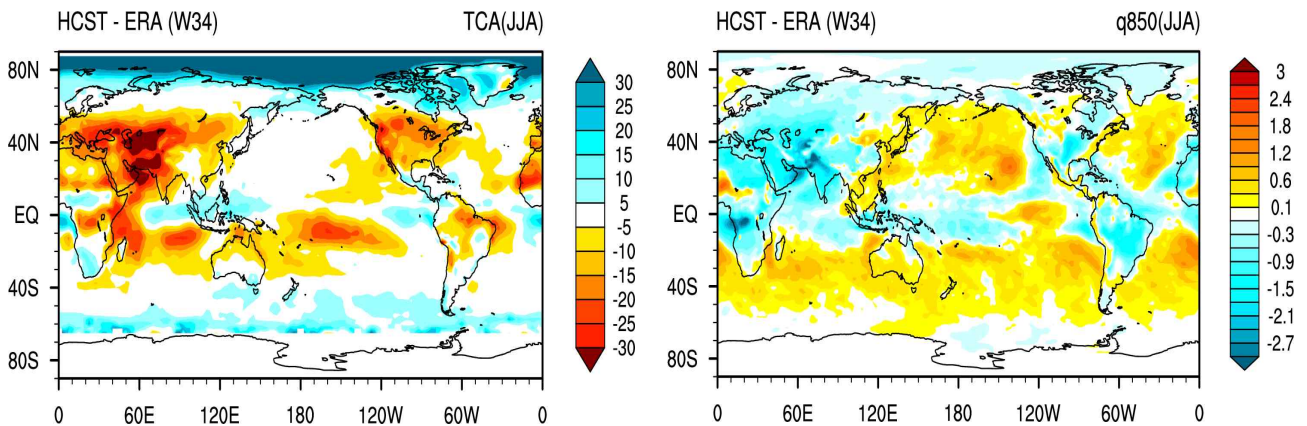


Figure 53. Distribution of averaged total cloud amount (left) and 850hPa specific humidity (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).

강수 오차의 특성을 더 살펴보기 위하여 운량 및 절대 습도의 오차 분포를 figure 53에 나타내었다. 총운량의 오차 분포를 보면 MC지역을 적도 지역 일부에서 양의 오차가 나타나고 있으며 중위도 지역에서는 매우 적은 운량을 나타내고 있음을 알 수 있다. 또한 북반구 극지역에서 과도하게 많은 운량을 보이고 있는 것이 특징이다. 중위도의 과도하게 적은 운량은 특히나 구름-복사 상호작용의 오차와도 크게 연관이 있기 때문에 중위도 대륙에서의 온도의 양의 편차와 연관된다 할 수 있다. 중위도의 과소모의된 운량에 의해 상대적으로 많은 태양복사 열속이 지표로 도달하게 되고 이로 인해 온도의 양의 편차를 가져올 수 있게 된다. 또한 극 지역에서의 많은 구름 양은 태양복사의 반사를 통해 지표에 도달하는 열속을 줄여 온도의 음의 편차를 가져온다. 그와 비슷한 패턴으로 중위도 대륙 지역에서 매우 건조한 오차를 보이고 있다. 또한, 적도 지역에서의 수증기 역시 매우 적게 모의하고 있다. 그해 비해 운량은 관측보다 많게 모의하고 있으며, 온도는 높게 모의하고 있다. 즉, 높은 온도에 의한 대류 활동에 의해 증발한 수증기는 공기중에서 구름물리 과정에 의해 과도하게 강수로 전환되어 오히려 건조 대기를 유지하고 있으며, 강수는 관측에 비해 많이 발생하고 있는 특징을 보이고 있다.

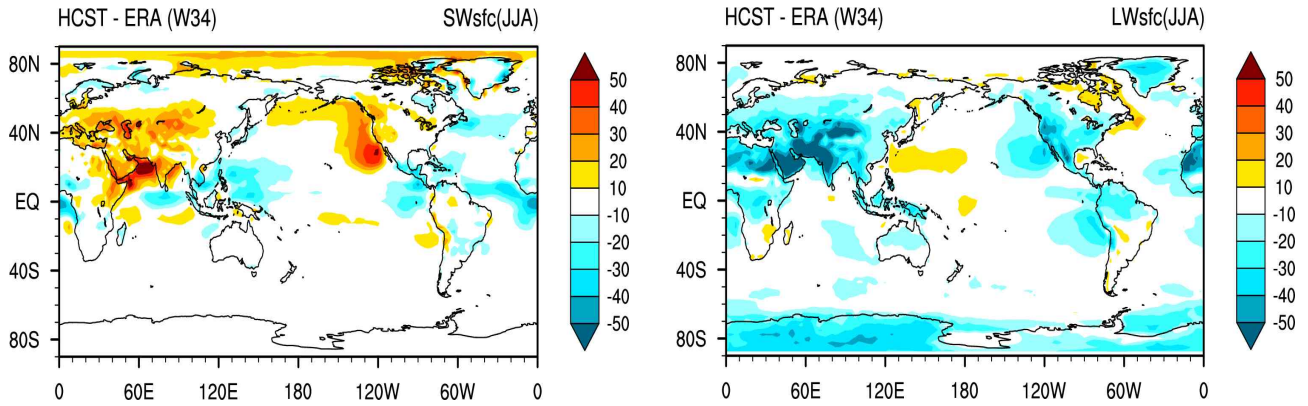


Figure 54. Distribution of averaged shortwave radiation fluxes (left) and longwave radiation fluxes (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).

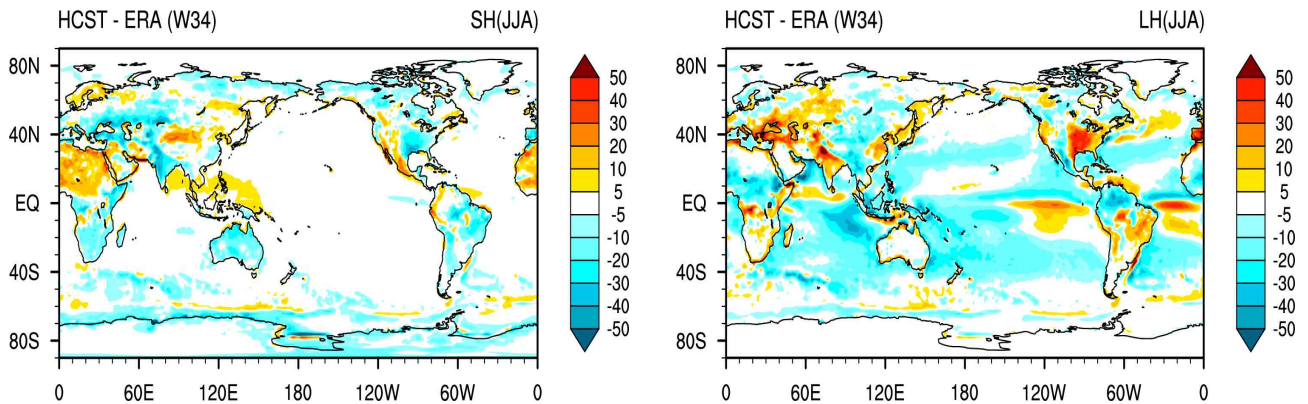


Figure 55. Distribution of averaged sensible heat fluxes (left) and latent heat fluxes (right) from lead weeks 3-4 for JJA of hindcast (1991-2010).

Figure 54에서 복사플럭스의 오차 분포를 통해 위에서 설명했던 구름-복사 상호작용의 분포를 한번 더 확인할 수 있다. 운량이 작게 나타났던 중위도 대륙을 중심으로 단파복사 플럭스가 과도하게 입사하고 있음을 확인할 수 있으며, 그에 비해 강수가 많게 나타났던 적도 지역에서는 단파복사가 적게 입사하고 있음을 알 수 있다. 또한, 지표 온도와 크게 관련있는 장파복사의 경우 전체적으로 한랭오차가 나타나는 이유로 관측에 비해 적은 양이 방출되고 있음을 확인할 수 있다. 전체적으로 증발이 강하게 이루어지고 있어 순열속의 경우 지표 온도와 마찬가지로 음의 오차를 가지는 것이 특징이다(나타내지 않음). 이러한 지표 열속의 오차를 figure 55에서 확인할 수 있다. 지표 온도와 연관된 현열의 경우 온도 및 장파복사의 오차와 마찬가지로 음의 오차를 나타내고 있으며, 잠열의 경우 과도하게 방출되어 지표 온도를 냉각시키는 방향으로 나타나고 있음을 확인할 수 있다.

강수 오차의 특성을 더 살펴보기 위하여 온도와 습도의 연직 오차 분포를 figure 56에 나타내었다. 가장 주목할 만한 오차로 중위도 지역으로의 한랭 오차가 대부분의 층에서 두드러지게 나타나며, 특히 적도 지역의 상층에서 온난 오차가 중위도까지 넓게 나타나고 있다. 특히 이러한 중층 대기의 온난 오차는 특히 운량과 관계가 있다고 할 수 있다. 적도 지역의 과도한

운량 및 강수에 의해 상층에서 입사하는 태양복사의 반사 혹은 흡수가 일어나기 때문에 중층에서는 오히려 온도가 상승하고 하층대기는 상대적으로 차가워진다. 따라서 중위도 대륙을 제외한 대부분의 지표 한랭 오차는 이러한 대기-복사 상호작용에 의해 설명될 수 있다.

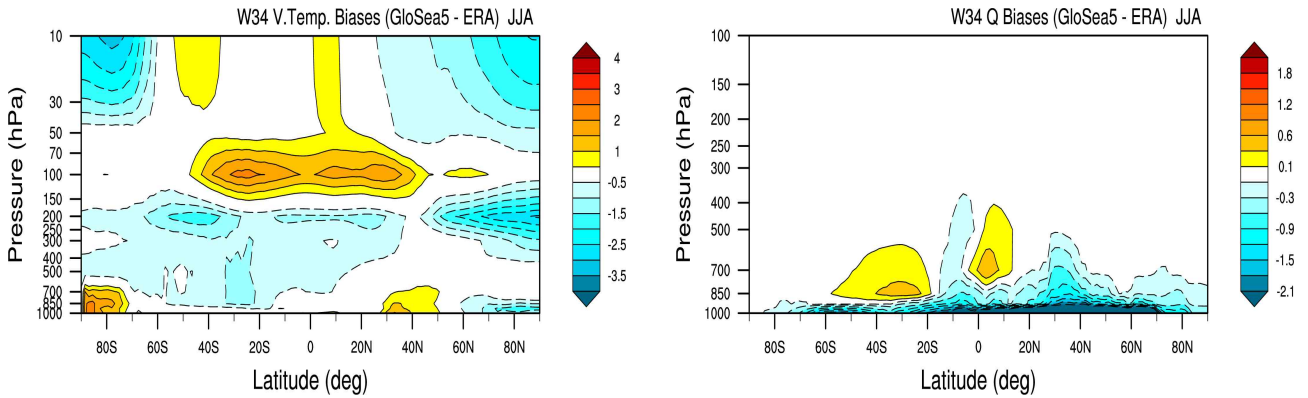


Figure 56. Vertical profiles of temperature (left) and specific humidity (right) error difference between GloSea5 hindcast and ERA5.

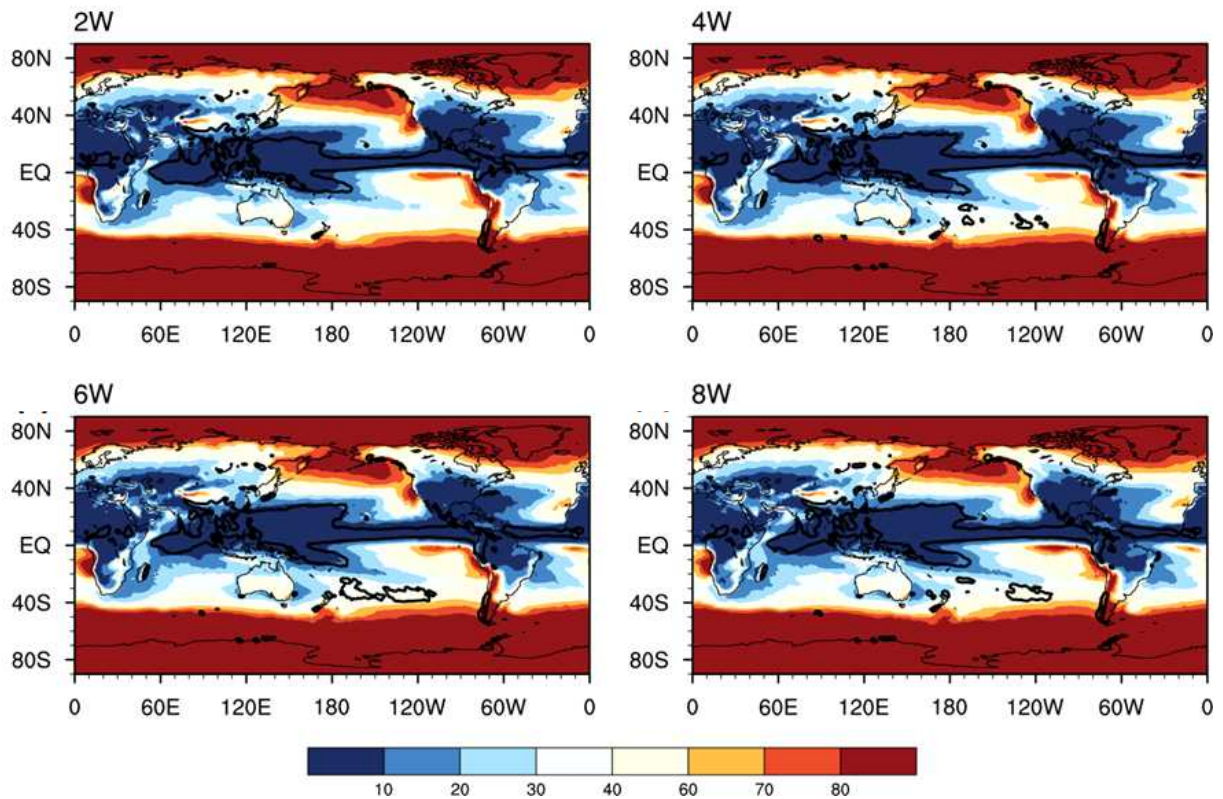


Figure 57. Ratio of large-scale precipitation and convective precipitation from total precipitation. Red (Blue) indicates that large-scale (convective) precipitation is dominant. Bold black line means precipitation biases (over 5mm/day).

강수의 오차 특성을 확인하기 위해 미세물리과정에 의한 강수와 적운 모수화 과정에 의한

강수의 비율을 figure 57에 나타내었다. 그림에서 알 수 있듯이 적도에서 중위도 대부분의 지역까지도 적게는 70%, 많게는 90% 이상의 비율로 적은 모수화 과정에 의한 강수로 이루어져 있다. 또한, 중위도에서 고위도로 갈수록 미세물리과정에 의한 강수가 늘어난다. 또한 그림에서 실선으로 나타나는 양의 강수 오차 (5mm/day) 이상의 지역은 가장 푸른색으로 나타나는 적은 모수화 과정 90% 이상의 지역임을 알 수 있다. 즉, 대부분의 오차를 설명하는 적도 태평양의 양의 오차는 대류과정에 의한 강수에 의해 발생한다고 생각할 수 있다.

이러한 특징은 figure 58의 강수의 히스토그램 분포에서도 확인할 수 있다. 적도 강수를 살펴보면, 약한 강수를 약 10% 정도 과소 모의하고 있음을 알 수 있다. 그에 비해 10mm/day 이상의 강한 강수의 빈도는 관측에 비해 과대 모의하고 있음을 알 수 있다. 이러한 특징은 앞서 설명한 적도지역의 건조 대기 및 과도한 강수 모의와 연관되어 설명할 수 있다. 즉, GloSea5 과거재현기후에서 강한 대류활동에 의해 증발한 수증기는 과도하게 강수로 전환되어 오히려 대기는 건조해진다고 할 수 있다. 동아시아 지역은 적도 지역과는 조금 다르다고 할 수 있다. 중위도 지역에서는 대류 활동에 의한 강수 보다 미세물리과정에 의한 강수 과정이 중요하다 할 수 있는데, 모델에서는 이러한 현상을 잘 설명하지 못하고 대부분의 강수를 대류 과정에 의한 강수로 모의하고 있다. 그 결과 약한 강수는 과소 모의하고 있으며, 중간 강도의 강수를 과대 모의하고 있는 것을 알 수 있다. 특히, 한반도 지역은 구름의 양도 작게 모의가 되며, 대기 중의 수증기도 관측보다 더 많이 남아있는 것을 알 수 있다. 이러한 오차를 개선하기 위해서는 적도지역에서 중요한 강수 과정인 강수모수화의 특성과 더불어 중위도 지역의 강수와 큰 연관이 있는 미세물리과정 역시 자세한 분석이 필요하다 할 수 있다.

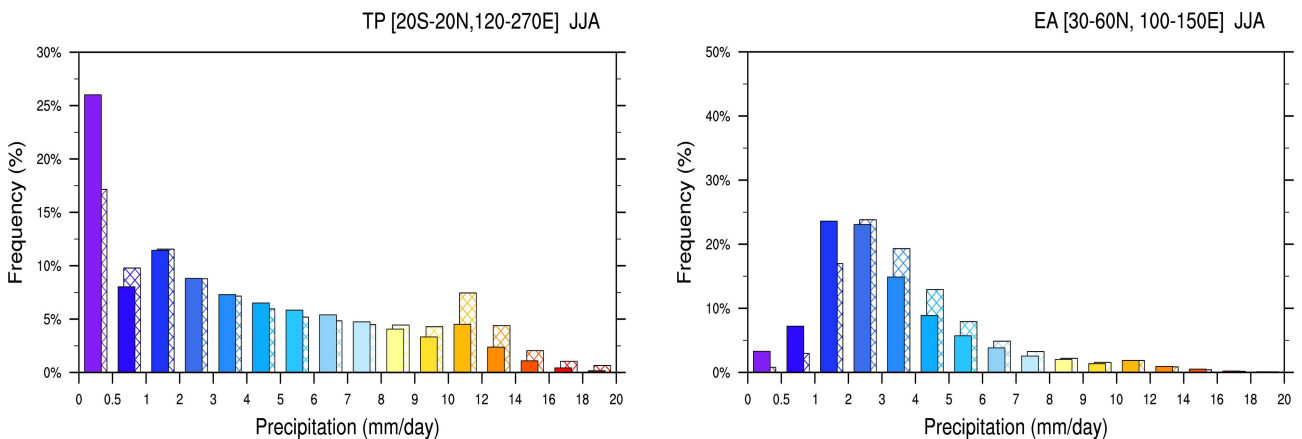


Figure 58. Histogram of precipitation for JJA over tropics (left) and East Asian (right) regions. Dashed (shaded) bar indicates values from GloSea5 (observation).

② 과거재현기후와 실시간 예측장의 초기 오차 특성 차이

실시간 예측장의 계통적 오차가 과거재현기후와 동일하게 나타나는지 살펴보기 위해 2018년 실시간 예측장과 과거재현기후의 오차 분포를 비교하였다. 최근 대부분의 예측 시스템의 과거재현기후의 예측 성능은 상당히 높게 나타나는 반면, 여전히 실시간 예측 시스템의 예측 성

능은 사례별, 계절별로 편차가 크게 나타난다. 이러한 이유로는 우선 과거재현기후에 비해 훨씬 작은 샘플 수를 들 수 있다. 기후예측 성능은 엘니뇨 등 강한 강제력이 있는 경우 예측 성능이 높아지는 특성이 있지만, 이상 기후 등 짧은 규모의 강제력이 중심이 되는 사례의 경우 예측 성능이 매우 낮아지기도 한다. 하지만, 이런 경우에도 동일한 예측시스템내에서 발생하는 계통적 오차의 특성의 경우에는 비슷하게 나타난다고 예상할 수 있다. 두번째로, 과거재현기후와 실시간 예측장에서의 다른 초기조건에 대한 문제도 제기할 수 있다. 초기 조건의 차이는 선행시간이 길어질수록 많은 차이를 줄 수 있다. 하지만, 대부분의 현업시스템에서는 과거재현기후 시스템과 실시간 예측시스템의 작은 차이는 크게 고려할 수 없는 환경 때문에 오차 차이에 있어서 초기 조건의 아주 작은 차이에 대한 영향은 지금까지 크게 고려되지 않았다.

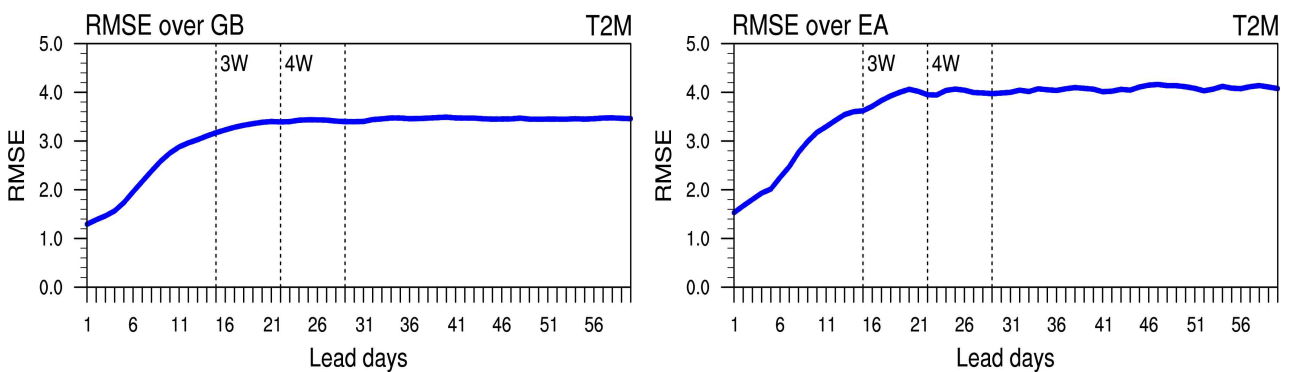


Figure 59. Root Mean Square Error of surface temperature over globe (left) and East Asia (right).

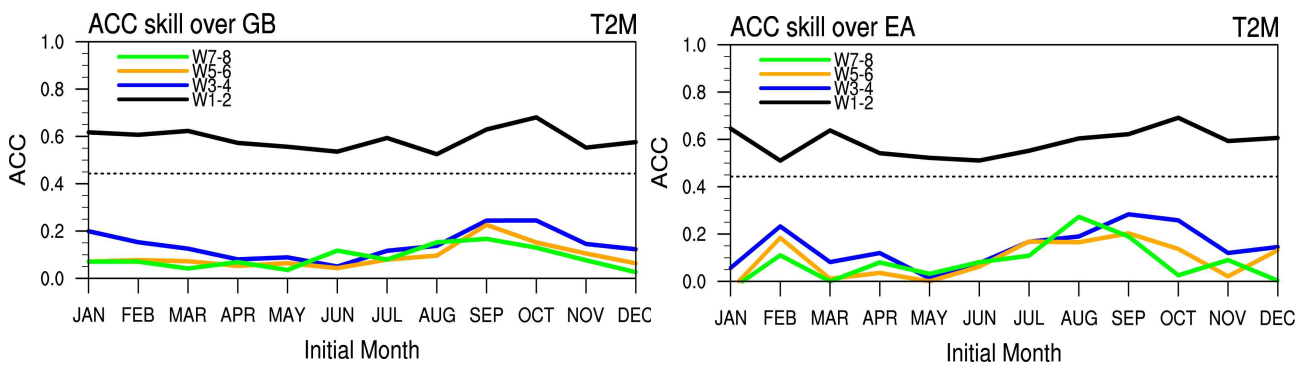


Figure 60. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) for initial month. Black, blue, orange and green lines indicate averaged lead weeks 1-2, 3-4, 5-6, and 7-8, respectively.

Figure 59에서 실시간 예측장의 예측 성능을 RMSE로 나타낸 결과를 살펴보면, 과거재현기후보다 오차가 다소 크긴 하지만 안정적인 예측 성능을 보이고 있음을 알 수 있다. Figure 60에서 나타난 월별 ACC의 경우 샘플수가 확연히 줄어들기 때문에 예측 성능은 더 큰 변동성을 보이고 있다. 특히 1월, 5월의 초기장을 사용한 동아시아 예측 성능의 경우 매우 낮게 나타나

고 있다. 따라서, 과거현재기후에 비해 실시간예측장의 예측성능은 다소 낮게 나타나고 있음을 확인할 수 있으며, 이는 특정해에 대한 예측 성능이기 때문이라고 어느 정도는 설명할 수 있다.

예측 성능과는 별개로 오차의 특성이 어떠한 형태로 나타나는지를 알아보기 위해 지표 온도의 오차 분포를 다음과 같이 살펴보았다. Figure 61은 과거재현기후와 실시간 예측장의 ERA5와의 지표 온도의 차이를 나타낸 그림이다. 앞 장에서 살펴본 바와 같이, 과거재현기후에서는 적도지역을 제외하고 전반적으로 한랭 오차가 지배적이다. 그에 비해 극 지역 일부를 제외하고 온난 오차가 지배적으로 나타나고 있다. 이러한 오차의 패턴이 다르게 나타나는 것은 예보장을 분석할 때 오류를 범할 수 있기 때문에, 그 이유에 대해 알아보는 것은 매우 중요하다 할 수 있다.

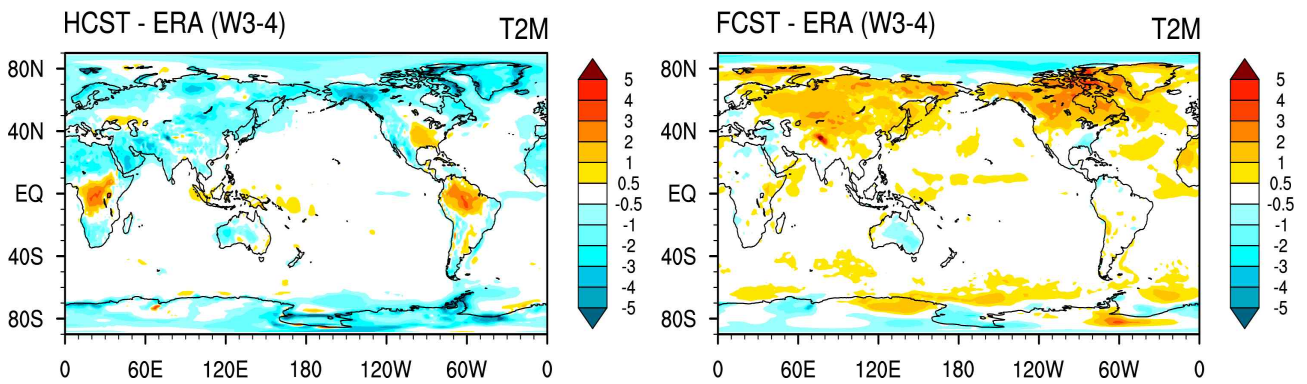


Figure 61. Distribution of averaged surface temperature from lead weeks 3-4 of hindcast (1991-2010) (left) and realtime forecast for 2018 (right).

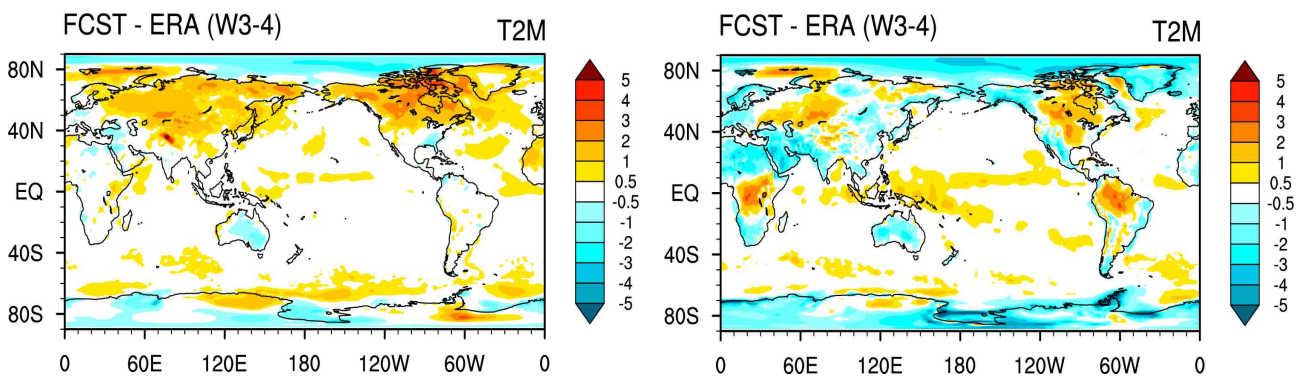


Figure 62. Difference distribution of averaged surface temperature from lead weeks 3-4 between anomaly (left), original raw (right) data and ERA5.

Figure 62는 실시간 예측장에서 흔히 사용되는 아노말리의 오차 분포와 원자료의 오차 분포를 나타낸 그림이다. 원자료의 오차를 보면, 과거재현기후에서 나타났던 오차와 마찬가지로 중위도 대륙을 중심으로 고위도에 이르기까지 한랭 오차가 나타나고 있으며 적도 태평양에서

온난 오차가 두드러지게 나타나고 있다. 또한 과거재현기후에서 나타났던 계통적 오차 역시 제거되지 않고 함께 나타나고 있음을 알 수 있다. 기후값을 통해 계통적오차를 일부 제거한 아노말리의 오차 분포를 살펴보면 대륙 일부에서 나타나던 계통적 오차가 효과적으로 제거되고 있음을 알 수 있다. 하지만, 기후값이 한랭오차를 나타냄에 따라 편차장의 오차는 반대로 온난 오차가 두드러지는 것을 확인할 수 있다. 이러한 특성의 차이의 원인에 대해 알아보기 위해, 본 연구에서는 2017/18년 겨울철 예측 성능 및 2018년 계절내예측에서 서로 다르게 처방된 초기 조건이 미치는 영향을 민감도 실험을 통해 설명해 보고자 한다.

(나) 초기장 차이에 따른 계절내 예측 성능 기여도 평가

① 실시간 계절내 예측 특성 진단

초기장 차이에 따른 계절내 예측의 영향에 대해 살펴보기 위해 민감도 실험을 구성하여 결과를 비교하였다. 앞 장에서도 설명했듯이, 과거재현기후와 실시간 예측장의 초기 조건이 다르게 처방되어있다. 현재 현업 시스템에 사용되는 실시간 예측장의 초기 조건은 NWP 시스템에서 계산된 예측장을 사용하고 있다. 이렇게 생산된 계절내 예측장을 NWPIF 실험 결과로 명명하기로 한다. 또한, 과거재현기후에서 사용되는 초기장인 ERA-interim 재분석장을 사용하여 실시간 예측장을 생산하는 ERAiF 실험을 수행하였다. NWPIF와 ERAiF 실험 결과를 서로 비교하여 대기 조건 혹은 대기 오차에 따른 계절내 예측 성능이 어떻게 달라지는지, 오차가 계통적으로 성장하는지를 알아보려고 하였다.

㉞ 2017/18년 겨울철 예측 특성

실시간 예측 특성을 살펴보기 위해 먼저 실험 사례인 2017/18년 겨울철에 대한 특성을 알아보았다. 2017/18 겨울철 한반도는 전체적으로 평년 대비 낮은 기온을 보이는 특징을 보였다. 특히 우랄산맥 부근 강한 고기압성 편차, 풍하측인 시베리아 및 동아시아에서의 기압골 강화, 북태평양 및 베링해 지역의 고기압성 편차가 지배적으로 나타나 우리나라가 추울 수 있는 전 지구 순환장을 보였던 것이 가장 큰 특징이라 할 수 있겠다. 또한, 약한 라니냐가 계속됨에 따라 라니냐와 연관된 중태평양 대류 활동 억제 및 서태평양 및 동인도양 대류 활동 강화가 나타났던 해이다(나타내지 않음). 특히 엘리뇨때에 비해 라니냐때, 북태평양 및 베링해 지역 블로킹 빈도 및 지속기간이 많은 경향이 알려져 있다. 실제로 2017/18 겨울철 순환장에 대해 다음과 같이 살펴보았다.

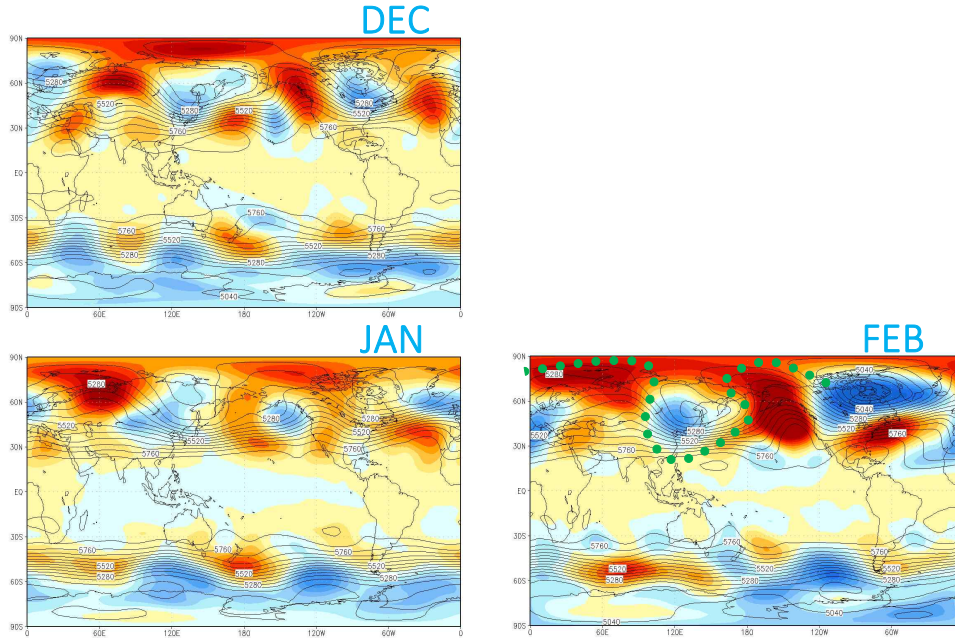


Figure 63. Horizontal distribution of 500hPa geopotential height mean (solid line) and anomaly (shaded) for December to February from observation. *Figures from KMA.*

Figure 63는 500hPa 평균 지위고도 및 편차장을 월별로 나타낸 그림이다. 12, 1, 2월 모두 우랄산맥 부근 강한 고기압성 편차 및 동아시아 기압골, 베링해 지역의 고기압성 편차 등의 추운 겨울의 요소가 지속적으로 나타나고 있음을 알 수 있다. 한반도가 추울 수 있는 대표적인 블로킹의 패턴을 월평균장에서도 확인할 수 있다. 실제적으로 우리나라의 기온 편차장을 살펴 보면, 평년에 비해 아주 낮은 기온을 보였음을 알 수 있다 (figure 64). 1월 일부 지역에서 평년과 비슷한 기온을 보이기도 하였으나, 전체적으로 아주 추운 겨울이었음을 알 수 있다.

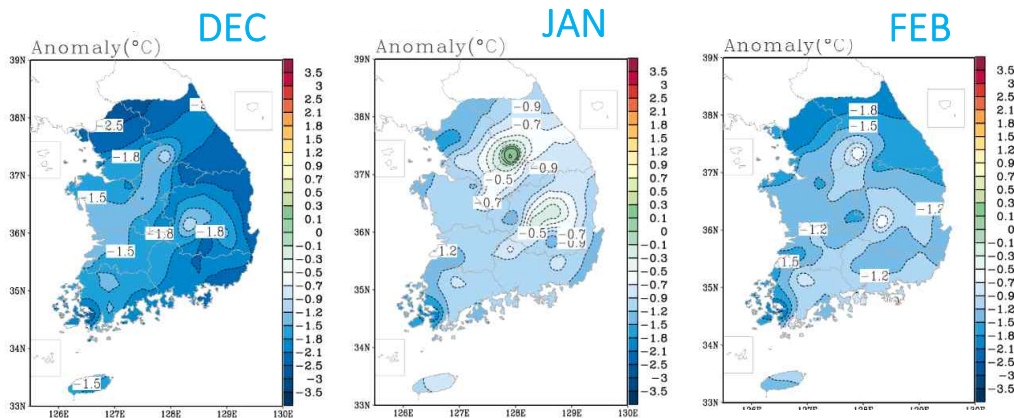


Figure 64. Anomaly distribution of surface temperature over Korean peninsula for 2017/18 winter. *Figures from KMA.*

㉔ 과거재현기후에서 추운 겨울 모의 특성 진단

실제 GloSea5의 겨울철 모의 성능을 살펴보기 위해 과거재현기후 자료를 이용하여 겨울철 예측 특성을 분석하였다. 동아시아 겨울철 기온의 변동성에 대한 분석을 위해 EOF 분석을 통해 ERA5와 함께 비교하였다. Figure 65에서 ERA5와 GloSea5의 과거재현기후 자료를 사용한 EOF PC time series를 나타내었다. 전체 50% 정도의 변동률을 나타내는 1, 2번째 모드에서 GloSea5는 전반적으로 관측과 매우 유사하게 모의하고 있다. 특히 선행주수 1-2주의 결과는 ERA5와 거의 동일하다. 선행주수 3-4주에서는 전반적으로 잘 모의하지만 몇몇해에서는 관측의 패턴과 다르게 나타나고 있다. 전체 15%정도를 설명하는 EOF 모드 3-4번 모드에서는 ERA5에서 나타난 패턴을 선행주수 1-2주에서도 모의하지 못하고 있다.

Figure 66에서는 EOF PC모드를 지표기온에 회귀한 패턴을 통해 실제 동아시아 겨울철 패턴을 얼마나 잘 모의하는지 살펴보았다. 첫 번째 모드는 극에서의 온도 증가와 관련하여 동아시아 대륙으로 차가운 온도가 나타나는 패턴을 설명하고 있다. 선행주수 1-2주의 경우 재분석 자료에서 보이는 패턴을 매우 잘 모의하고 있다. 선행주수 3-4주에서는 극에서의 온난 패턴은 매우 잘 모의하지만 대륙에서의 차가운 온도는 매우 작게 모의하고 있다. 두 번째 모드에서 대륙고기압에 의해 차가운 온도가 크게 확장된 패턴을 설명하고 있다. 선행주수 1-2주에서는 그 패턴을 잘 모의하고 있지만 선행주수 3-4주에서 다소 약하게 모의하고 있다. PC time series와 마찬가지로 세 번째, 네 번째 모드에 회귀된 온도 패턴은 재분석 자료에서 나타난 패턴을 거의 모의하지 못하고 있음을 알 수 있다.

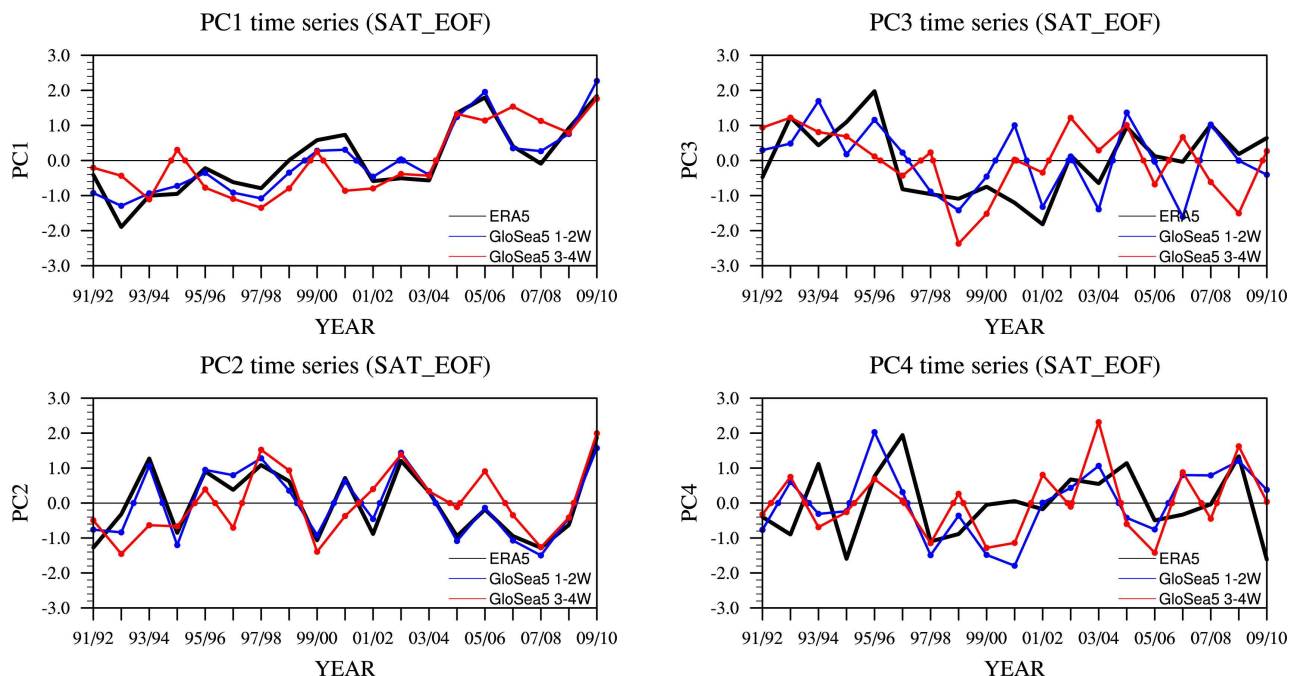


Figure 65. 1st to 4th PC time series of EOF analysis from ERA5 reanalysis, GloSea5 lead weeks 1-2, and lead weeks 3-4.

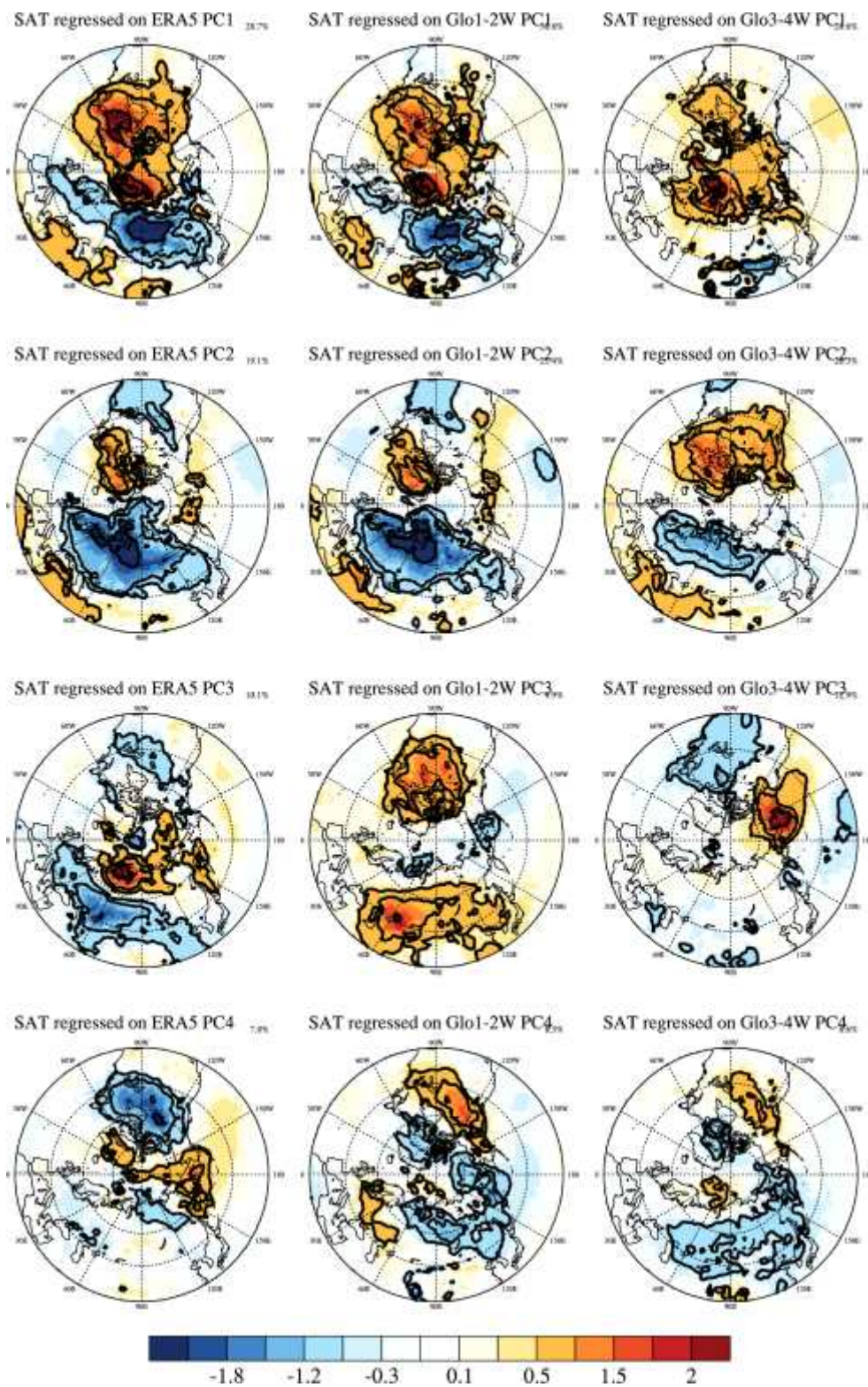


Figure 66. Surface temperature regressed on ERA5, GloSea5 lead weeks 1-2, and lead weeks 3-4.

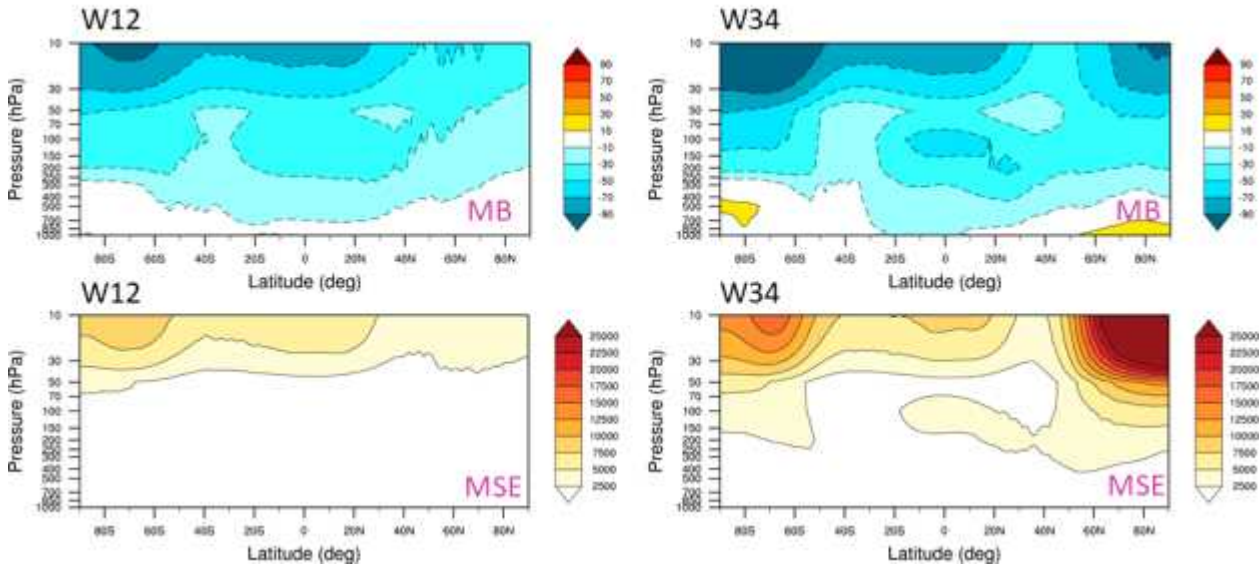


Figure 67. Mean biases (top) and Mean Square Errors (bottom) of zonal mean geopotential height in GloSea5 at lead weeks of 1-2 (left) and 3-4 (right) in DJF.

연직 순환장의 구조적 오차를 살펴보기 위해 동서평균된 지위고도의 평균오차(MB) 및 평균제곱오차(MSE)를 figure 67에 나타내었다. 선행주수 1-2주의 결과에서는 상부 성층권에서 가장 큰 음의 MB에서 존재하고 있으며 선행주수 3-4주로 갈수록 상부 성층권에서 하부 성층권 및 상부 대류권까지 음의 MB가 강하게 나타나고 있다. 선행주수 1-2주에서는 상부 성층권에서 MSE가 크게 나타나고 있지만, 선행시간이 증가하면 상부 성층권, 특히 극지역을 중심으로 MSE가 매우 증가하고 있음을 알 수 있다.

동아시아 지역 기온의 예측 성능의 특성을 살펴보기 위해 한반도 지역의 기온의 시간 변화에서 0.5*표준편차로 구분하여 기온이 높을 때와 낮을 때의 연도를 구분하여 분석하였다. 높은 온도의 연도로는 1997, 1998, 2001, 2006, 2008년이 선택되었으며, 낮은 온도의 연도로는 1995, 1999, 2000, 2004, 2009년이 선택되었다. 추운 겨울과 따뜻한 겨울 사례의 MSE 연직 오차의 차이를 figure 68에 나타내었다. 저위도 및 중위도 상부 성층권에서의 오차는 추운 겨울에 더 크게 나타나고 있으며, 고위도 성층권 및 상부 대류권에 이르기까지 넓게 퍼져있는 MB의 경우 따뜻한 겨울에서 크게 나타나고 있음을 알 수 있다. 특히 선행주수 3-4주에서 중위도 상부 성층권에서의 오차는 추운겨울에 더욱 크게 나타나고 있다. 북반구와 동아시아 지역의 선행 시간에 따른 500hPa 지위고도 예측성능을 figure 69에 나타내었다. 과거재현기후의 모든 사례의 예측 성능과 비교하였을 때 추운 겨울, 따뜻한 겨울 사례의 성능 모두 비슷하게 선행시간에 따라 감소하는 패턴을 보이고 있다. 특히 3주전까지 예측 성능이 거의 유사하게 나타나지만, 4주이상에서 동아시아의 예측 성능은 추운 겨울의 경우 약간 높게 나타난다.

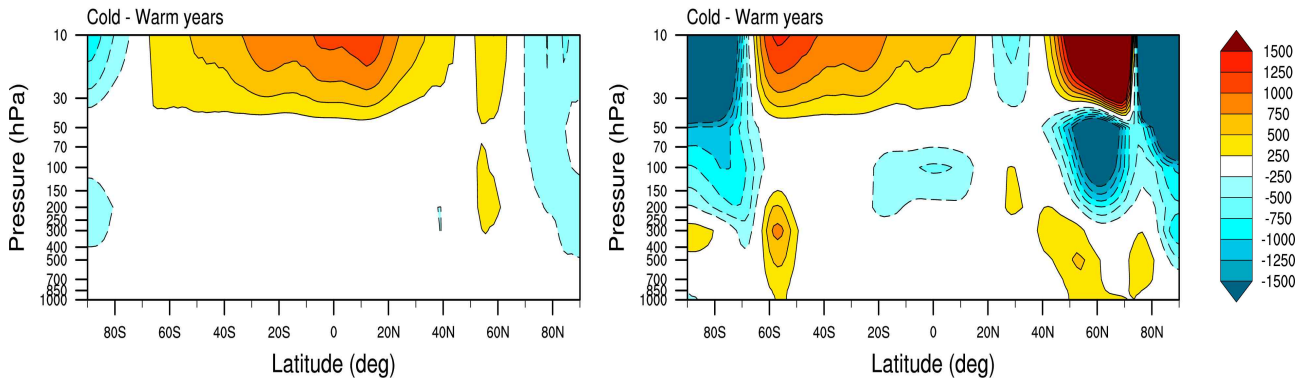


Figure 68. Difference of Mean Square Errors of zonal mean geopotential height between cold year and warm year cases in GloSea5 at lead weeks of 1-2 (left) and 3-4 (right) in DJF.

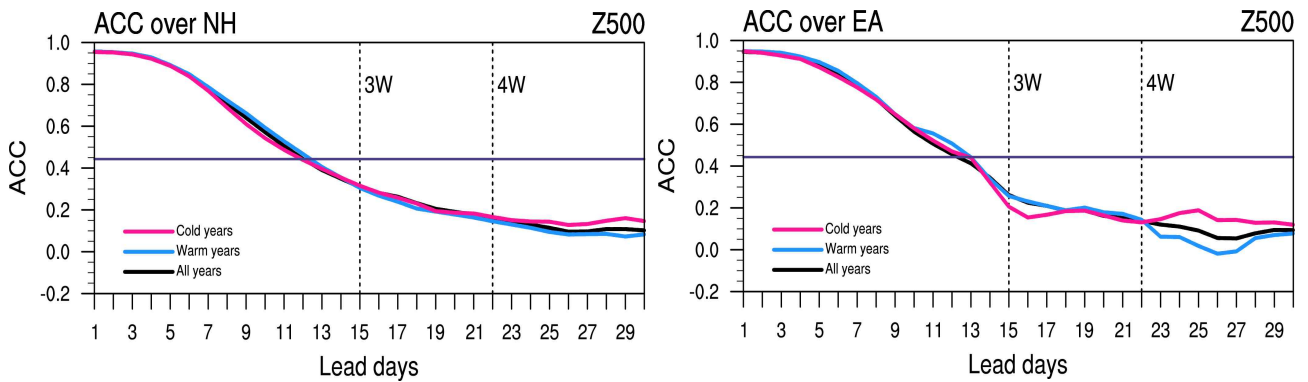


Figure 69. Anomaly Correlation Coefficient of 500hPa geopotential height over globe (left) and East Asia (right) Blue line is for warm year cases. Pink line is for cold year cases. Black line is for all year cases.

추운 겨울의 모든 사례의 경우 예측 성능이 떨어지지 않지만, 2017/18의 추운 겨울 사례에서 예측 성능이 과거재현기후보다 예측 성능이 낮게 나오는 원인 진단을 위해 추운 겨울 사례의 36개의 앙상블에 대해 해면기압에 대한 EOF 분석을 figure 70에 나타내었다. 50%의 변동율을 보이는 첫 번째 모드에서는 동아시아 겨울철의 기온에 영향을 주는 가장 큰 기후 모드 중 하나인 AO 예측 성능 및 북극해의 해빙 감소 등과 연관된 특성을 잘 모의하고 있다. 특히 선행주수 3-4주까지 관측이 나타낸 패턴을 잘 모의하고 있어 추운 겨울 사례의 예측 성능이 높은 원인을 이해할 수 있다. 20% 정도의 변동율을 보이는 두 번째 모드의 경우 유라시아 대륙 북서쪽, 우랄산맥, 동아시아로 연결되는 wave-train 기압패턴을 보인다. 이러한 패턴은 겨울철 한파와 크게 영향이 있으며, 특히 2017/18년 겨울철 사례의 가장 큰 특징인 우랄 블로킹 사례를 설명할 수 있는 패턴이다. 선행주수 1-2주의 경우 관측에서 보이는 패턴을 잘 모의하고 있으나, 선행주수 3-4주의 경우 이러한 wave-train 패턴을 거의 모의하지 못하고 있다. 즉, 2017/18 사례의 경우, 과거재현기후에서도 잘 모의하지 못하는 사례 중 하나이며 예측 성능을 떨어뜨릴 수 있는 조건을 가졌다고 할 수 있다. 다음 장에서는 이러한 성능이 낮은 사례에서 초기 조건이 어떠한 영향을 미치는지, 예측 성능을 높일 수 있는지를 알아보려고 한다.

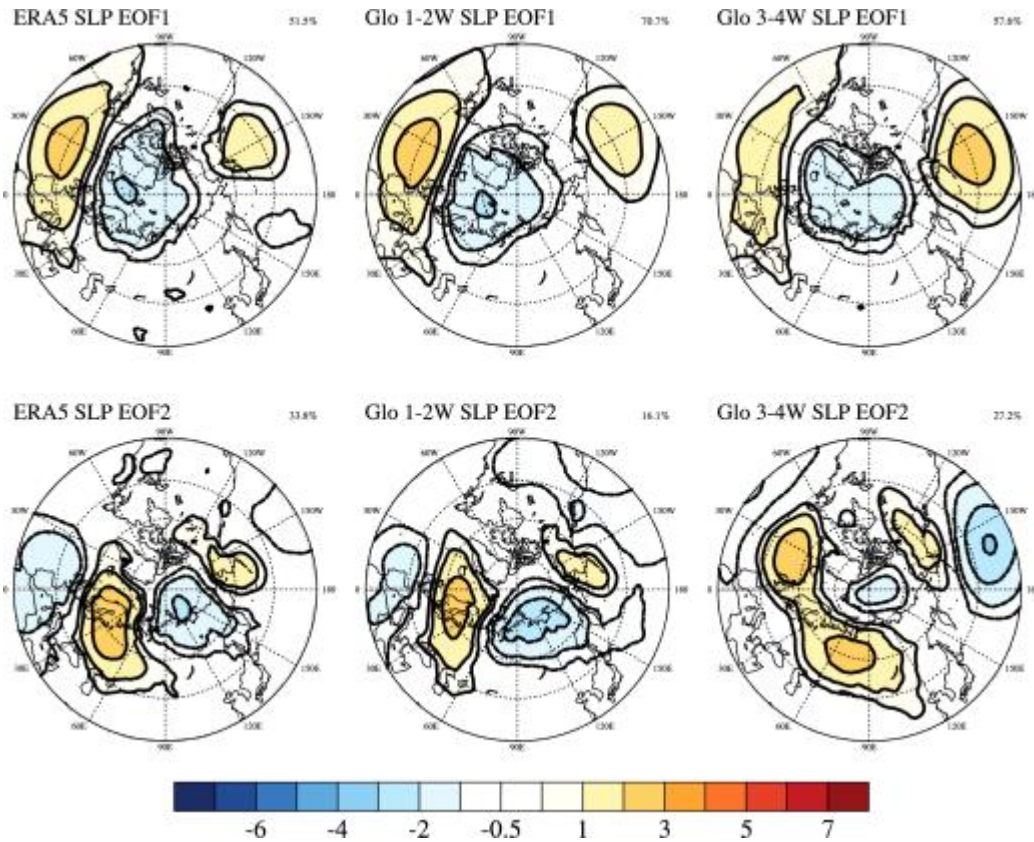


Figure 70. Sea level pressure distribution from 1st to 2nd mode of EOF analysis from ERA5 reanalysis, GloSea5 lead weeks 1-2, and lead weeks 3-4.

㊤ 초기장 차이에 따른 겨울철 예측 특성 진단

민감도 실험의 초기 조건의 차이를 비교해 보기 위해 규준 실험인 NWPiF와 ERA5의 차이를 통해 기존의 오차를 확인하고, 두 실험의 차이를 통해 오차의 차이에 대해 분석하였다. Figure 71은 선행시간 1일에서의 규준 실험의 해면 기압 오차 및 두 실험간의 차이를 나타낸 그림이다. NWPiF 실험의 해면기압 오차를 살펴보면, 극지역에서 동아시아까지 넓은 지역에 걸쳐 큰 오차를 보이고 있다. 그에 비해, ERAiF 실험의 경우 작긴 하지만 한반도를 비롯하여 동아시아 대륙까지 오차를 약하게 줄여주는 결과를 확인할 수 있다. 이러한 작은 오차의 개선이 계절내 예측에서는 어떠한 영향을 미치는지 figure 72에서 나타내었다. NWPiF의 경우 선행 주수 3-4주의 오차는 1일의 오차와 비교하여 보면, 패턴은 유사하지만 크기는 오히려 작아졌음을 확인할 수 있다. ERAiF 실험 결과 선행 주수 3-4주의 오차는 NWPiF의 오차를 상쇄시키고 있음을 알 수 있다. 특히, 고위도 지역의 해면기압의 오차를 효과적으로 감소시키고 있음을 알 수 있다.

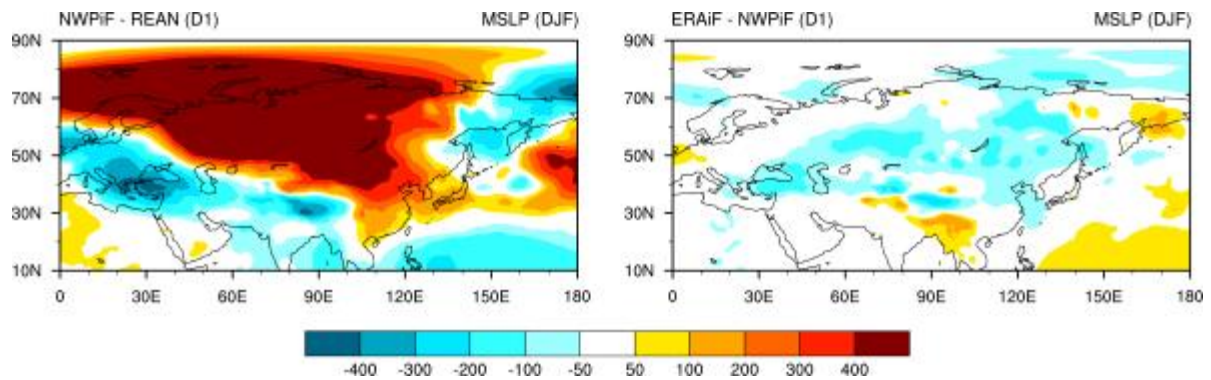


Figure 71. Difference distribution of sea level pressure at lead day 1 between NWPIF and ERA5 (left), ERAiF and NWPIF (right).

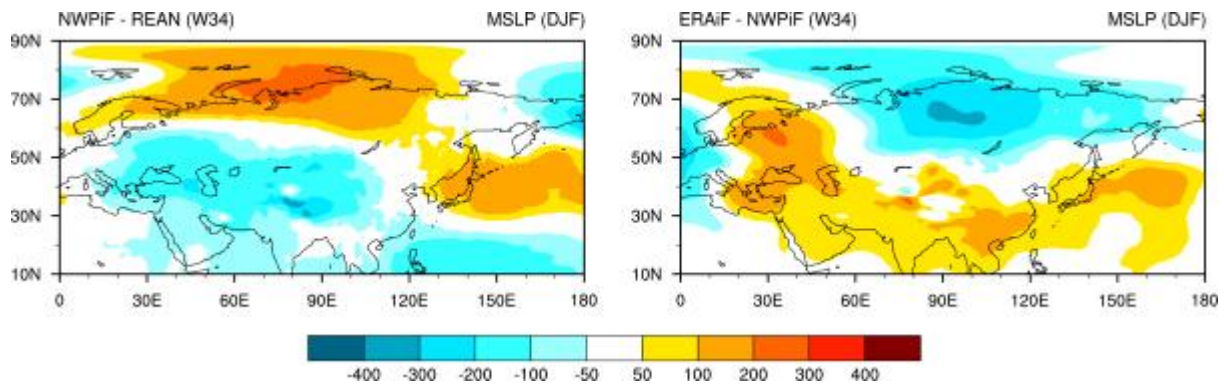


Figure 72. Difference distribution of sea level pressure at lead weeks 3-4 between NWPIF and ERA5 (left), ERAiF and NWPIF (right).

Figure 73에서 상층 순환장의 오차를 살펴보기 위해 선행시간 3-4주 500hPa 지위고도장의 결과를 비교하였다. 재분석자료의 결과를 살펴보면, 앞 장에서 설명했던 우랄 블로킹 현상의 특징을 잘 보여주고 있다. 우랄 산맥부근의 강한 고기압성 편차가 두드러지게 나타나고, 풍하측인 시베리아 및 동아시아에서의 기압골이 강화가 잘 나타나고 있다. 또한, 베링해 지역의 고기압성 편차가 지배적으로 나타나고 있다. NWPIF 실험의 결과 우랄 산맥 부근의 강한 고기압성 편차는 조금 약하지만 패턴은 잘 모의하고 있다. 하지만, 동아시아에서의 기압골은 아주 약하게 모의하며, 특히나 베링해 지역의 고기압성 편차는 잘 모의하지 못하고 있음을 확인할 수 있다. ERAiF 실험의 결과를 살펴보면, 우랄 산맥 부근의 강한 고기압을 더 강하게 모의하고 있으며, 동아시아의 기압골 강화 및 베링해 지역의 고기압성 편차를 더 관측과 유사하게 모의하고 있음을 알 수 있다. 즉, ERA-interim 재분석자료를 초기장으로 사용했을 경우 NWP 예측장을 초기장으로 사용했을 경우보다 상층 순환장을 관측과 더 유사하게 모의하고 있음을 알 수 있다. 즉, 과거재현기후에서 잘 모의하지 못하는 사례인 블로킹으로 인한 추운 겨울철 특성을 약간의 초기장의 오차 제거로 인해 더 잘 모의하고 있음을 확인할 수 있다. 다시 말해, NWP 예측장을 사용하는 현업 시스템에서 초기 오차로 인해 계절내 예측 성능이 영향을 받고 있음을 알 수 있다. 다음 장에서 겨울철을 비롯한 2018년 전체 예측 성능에 대해 평가하여 실제 영향 정도를 살펴보았다.

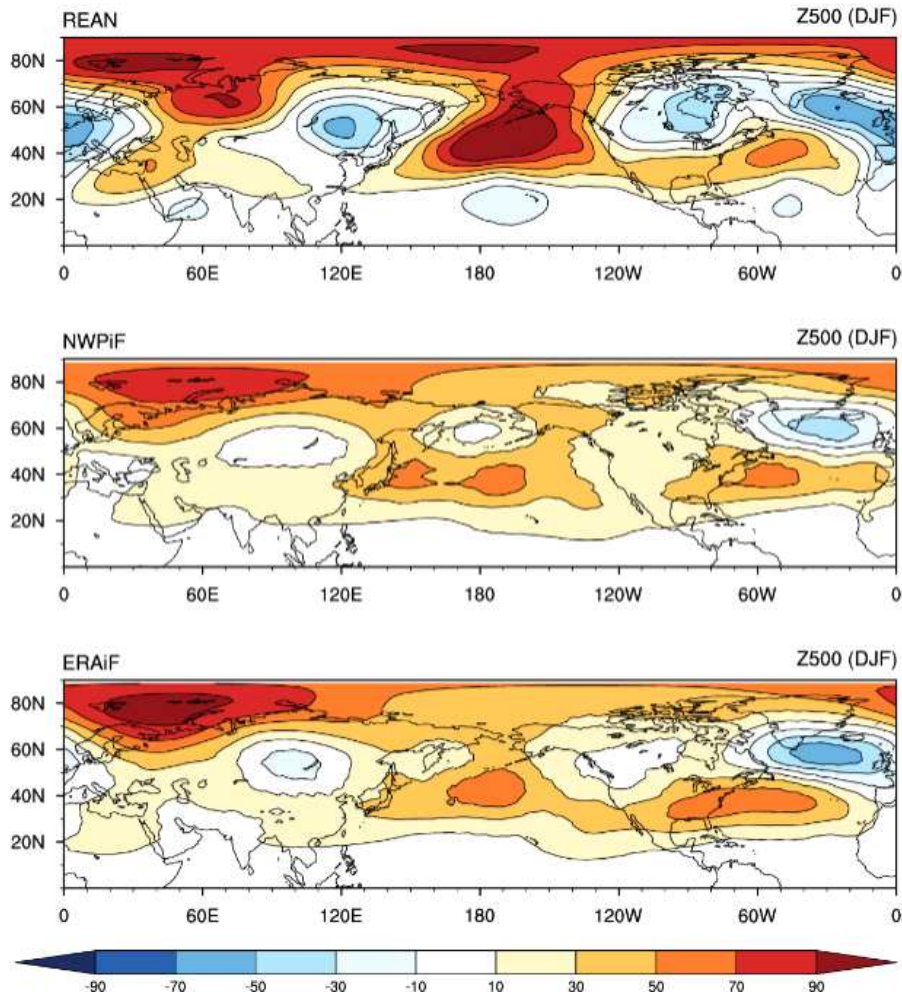


Figure 73. Horizontal distribution of 500hPa geopotential height at lead weeks 3-4 from ERA5 (top), NWPIF (middle), and ERAiF (bottom).

실제 지표 온도의 예측장은 어떠한 차이가 있는지 NWPIF 실험의 오차와 NWPIF와 ERAiF 실험들의 차이를 Figure 74에 나타내었다. 앞선 장에서 실시간 예측장의 온도 분포에서 양의 오차가 나타났던 것과 동일하게 NWPIF 실험에서는 극지역을 제외하고 양의 오차가 일반적으로 나타나고 있음을 알 수 있다. 특히 한반도를 중심으로 동아시아 지역 전반에 걸쳐 양의 오차가 두드러진다. 또한, 북미에서 알래스카에 이르는 지역에 양의 오차가 두드러지게 나타나고 있다. 극지역에서는 음의 오차가 나타나고 있다. 이와 비교했을 때, ERAiF 실험에서는 극지역에서의 음의 오차가 효과적으로 줄어들고 있다. 또한, 북미지역과 알래스카지역에 이르기까지 나타나고 있는 양의 오차를 효과적으로 줄여주고 있음을 알 수 있다. 한반도를 비롯한 동아시아에 나타나는 양의 오차 역시 다소 줄여주는 결과를 보여주고 있다. 즉, 과거재현기후와 실시간예측장에서의 초기 조건의 다른 처방이 결국 동아시아 지표 온도 편차장 예측에도 부정적으로 영향을 미치고 있음을 확인할 수 있다.

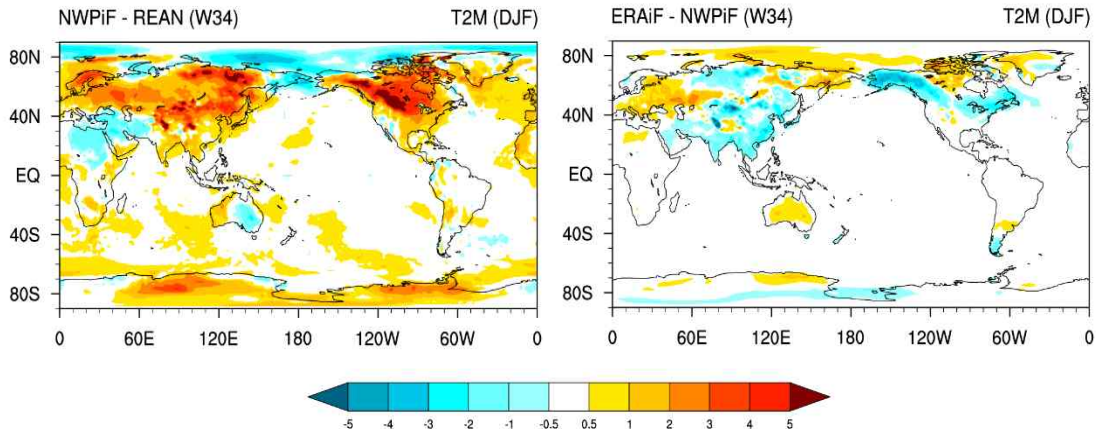


Figure 74. Difference distribution of surface temperature at lead weeks 3-4 between NWPIF and ERA5 (left), ERAiF and NWPIF (right).

㉔ 초기장 차이에 따른 계절내 예측 성능 기여도 평가

민감도 실험의 결과 비교를 통해 겨울철 예측 성능 이외에도 여름철, 연평균 등의 예측 성능에 어떤 차이가 있는지 알아보았다. Figure 75은 NWPIF 실험 결과로부터 계산된 2018년 여름 평균 지표온도의 오차 및 두 실험간 차이를 나타낸 그림이다. NWPIF 실험결과를 살펴보면, 겨울철 지표온도의 오차는 중위도 대륙을 중심으로 한랭 오차가 두드러졌지만, 여름철 지표온도의 오차는 고위도에서 온난 오차, 중위도 동아시아에서는 약한 한랭 오차가 나타나는 것이 특징이다. ERAiF 실험 결과의 경우, 북미지역 및 남미지역의 온난오차를 효과적으로 줄여주고 있다. 유라시아 대륙으로 온도를 더 높이고 있어 오히려 오차가 더 커지는 것을 알 수 있다.

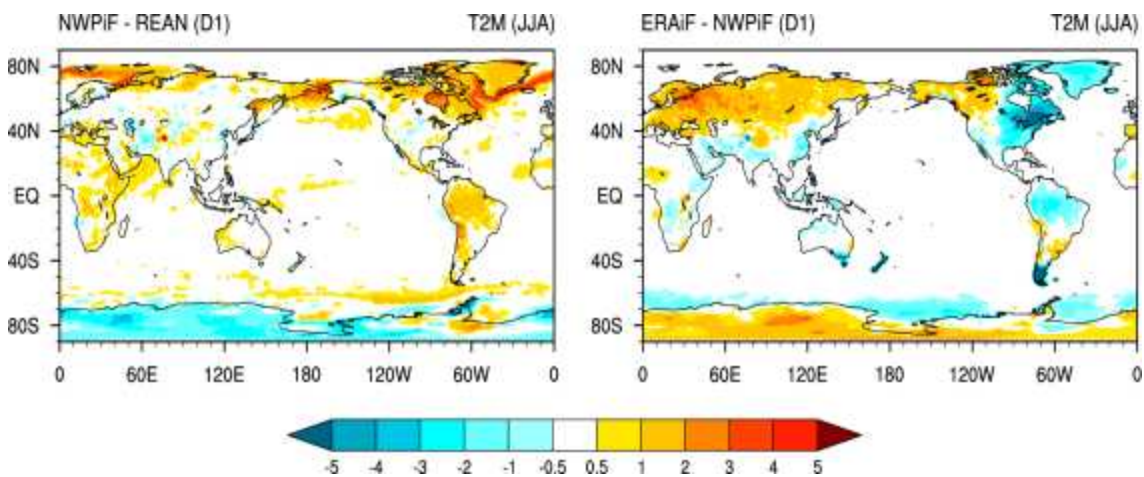


Figure 75. Difference distribution of surface temperature at lead day 1 between NWPIF and ERA5 (left), ERAiF and NWPIF (right).

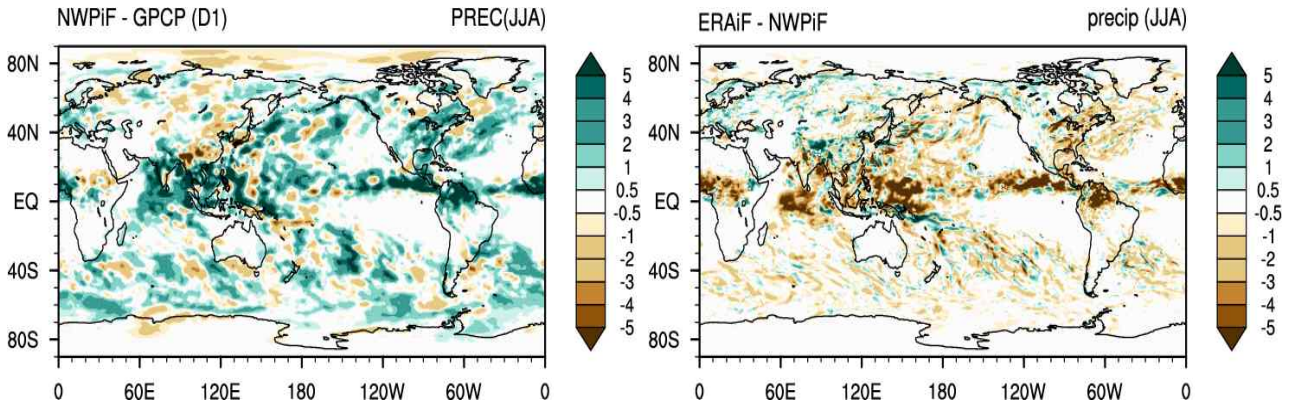


Figure 76. Difference distribution of surface precipitation at lead day 1 between NWPIF and ERA5 (left), ERAiF and NWPIF (right).

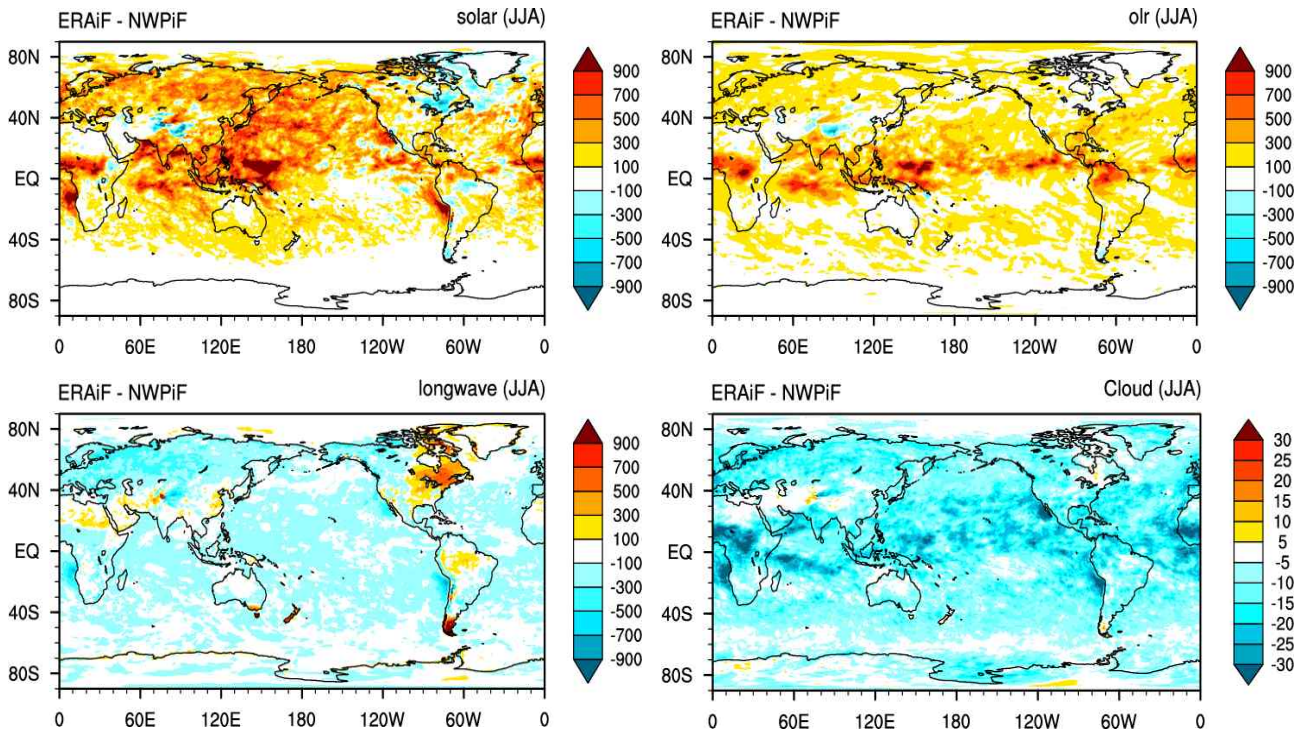


Figure 77. Difference distribution of surface solar radiation fluxes (top-right), outgoing longwave radiation fluxes (top-left), surface longwave radiation fluxes (bottom-right), and cloud amount (bottom-left) at lead day 1 between ERAiF and NWPIF.

Figure 76은 2018년 여름철 평균 강수의 오차를 나타내기 위해 NWPIF 실험의 강수 오차 및 ERAiF와 NWPIF 실험의 강수 차이를 나타낸 그림이다. NWPIF 실험의 초기 1일 강수 오차는 상당히 크게 나타나고 있다. 적도 지역을 중심으로 대부분의 지역에서 양의 오차가 지배적이다. 한반도를 중심으로 건조 오차가 나타나긴 하지만, 주요 강수 지역에서는 강한 강수가 나타나고 있다. 그에 비해 ERAiF 실험의 경우 NWPIF 실험에서 나타난 양의 오차를 효과적으로 줄여주고 있음을 확인할 수 있다. 특히 서태평양에서 인도양에 이르기까지 주요 대류 강수의

오차를 효과적으로 줄이고 있다. 강수의 경우 초기장에서 자료 동화되지 않는 예단 변수이기 때문에 특히나 오차의 경우 다른 변수들과 상호작용이 크게 일어나는 변수라고 할 수 있다.

강수와 크게 연관된 복사변수들과 운량의 실험간 차이를 figure 77에서 비교하였다. 지표 단과 복사 플럭스의 두 실험간 차이를 살펴보면, ERAiF 실험이 NWPiF 실험에 비해 더 많은 복사 플럭스를 많이 흡수하고 있음을 알 수 있다. 또한 대기 상층에서도 ERAiF 실험의 결과 더 많은 장파 복사 플럭스를 내보내고 있다. 특히 두 실험 간의 강수 차이의 패턴과 매우 유사하게 적도 지역을 중심으로 많은 차이를 보이고 있음을 알 수 있다. 온도 변화와 매우 연관이 큰 지표 장파 복사 플럭스의 경우, ERAiF 실험의 플럭스가 NWPiF 실험의 플럭스에 비해 작은 양을 보이고 있다. 복사 플럭스 값에 가장 영향을 크게 주는 총 운량에 차이를 살펴보면, ERAiF의 결과가 NWPiF의 결과에 비해 적은 양을 나타내고 있다. 즉, 기존 실험에 비해 적게 모의된 운량은 지표로 도달하는 태양 복사 플럭스를 증가시켰으며, 상층에서 방출되는 장파 복사 플럭스 역시 증가시켰음을 알 수 있다. 또한, 이러한 현상은 최종적으로 지표의 온도의 증가를 야기시킨다.

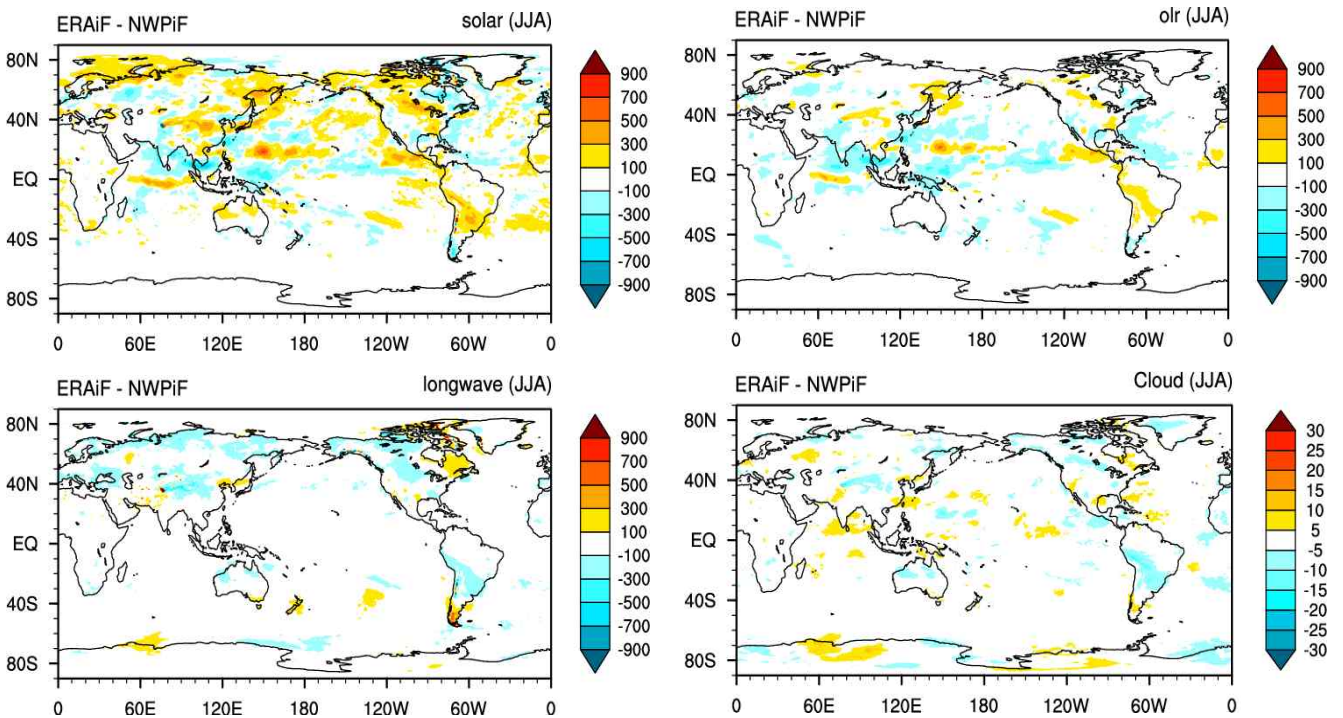


Figure 78. Difference distribution of surface solar radiation fluxes (top-right), outgoing longwave radiation fluxes (top-left), surface longwave radiation fluxes (bottom-right), and cloud amount (bottom-left) at lead weeks 3-4 between ERAiF and NWPiF.

선행시간이 증가에 따라 영향이 어떻게 바뀌는지를 살펴보기 위해 선행 주수 3-4주의 변수별 차이를 figure 78에 나타내었다. Figure 76에서 보였던 초기 차이와 비교하였을 때, 선행 주수 3-4주의 결과는 다소 줄어든 것을 확인할 수 있다. ERAiF의 지표 단과복사의 경우 서태평양 강수대 및 적도 지역 일부에서 NWPiF에 비해 적게 입사되고 있으며, 한반도를 비롯한 동

아시아 일부, 적도 태평양 일부에서 많은 지표 단파복사가 입사되고 있음을 알 수 있다. 상층 장파복사의 경우, 서태평양 및 인도양 지역에서 ERAiF이 NWPiF에 비해 적게 방출되고 있음을 알 수 있다. 지표 장파복사의 경우 중위도 대륙 일부에서 ERAiF가 NWPiF보다 다소 적게 방출되고 있다. 총운량의 경우 두 실험의 차이가 아주 크지 않게 나타나고 있다. 이러한 차이가 온도와 강수에 어떠한 영향을 미치는지 다음과 같이 살펴보았다.

Figure 79에서 선행주수 3-4주의 NWPiF의 강수 오차 및 두 실험 사이의 강수 차이를 나타내었다. NWPiF의 결과를 보면, 서태평양을 중심으로 강수의 음의 오차를 보이고 있으며, 동태평양 및 북서 태평양을 중심으로 양의 오차가 나타나고 있다. NWPiF실험에서 나타나는 이러한 오차는 ERAiF 실험에서 효과적으로 상쇄되고 있음을 확인할 수 있다. 특히 서태평양 및 인도양에서의 강수 오차는 효과적으로 줄어들고 있음을 알 수 있다. 이러한 강수 및 복사 플럭스의 차이는 지표 온도 예측에도 영향을 미친다. 특히 복사 플럭스의 변화는 열대 대류 활동과도 큰 관련이 있기 때문에, 열대 계절내 진동에서의 영향에 대해서는 다음 장에서 다시 논하기로 한다.

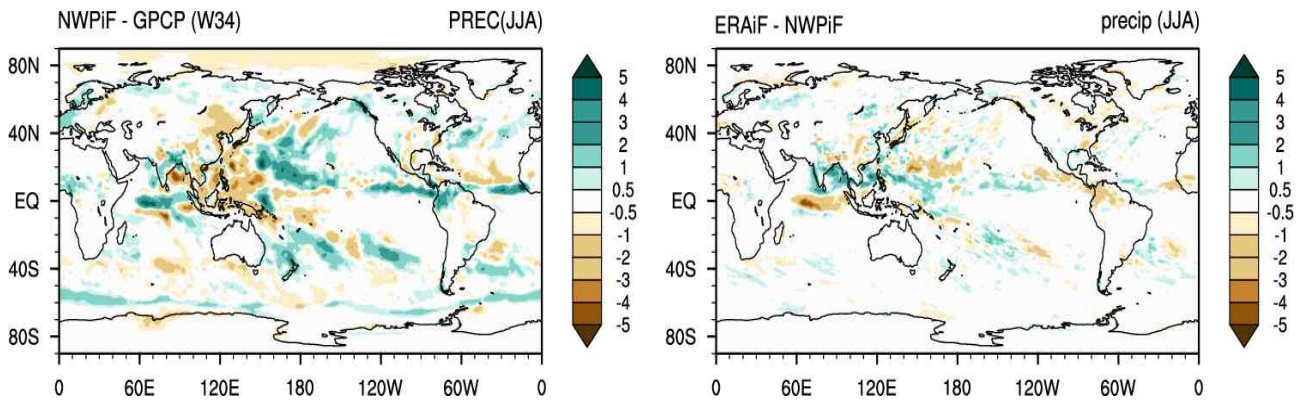


Figure 79. Difference distribution of surface precipitation at lead weeks 3-4 between NWPiF and ERA5 (left), ERAiF and NWPiF (right).

Figure 80에서 선행주수 3-4주의 NWPiF의 지표 온도 오차 및 두 실험 사이의 온도 차이를 나타내었다. NWPiF 실험에서의 온도의 오차는 앞서 살펴본 실시간 예측장에서의 오차와 마찬가지로 온난 오차가 두드러지게 나타난다. 특이할만한 점은 겨울철, 여름철에 따라 다르게 나타나지 않고 동일하게 온난 오차가 나타나는 점이다. 이러한 점은 GloSea5 시스템에서 나타나는 계통적인 오차의 특성이라고 할 수 있으며, 이러한 특성을 고려하여 예측장을 해석할 때 참고할 수 있을 것이다. ERAiF 실험에서는 NWPiF 실험과 비교하여 볼 때, 전반적으로 온난 오차를 다소 줄여주고 있는 역할을 하고 있다. 즉, NWP 예측장을 대기 초기장으로 사용할 경우 계절내 예측장에서의 온난 오차를 더욱 크게 가져올 수 있음을 알 수 있다.

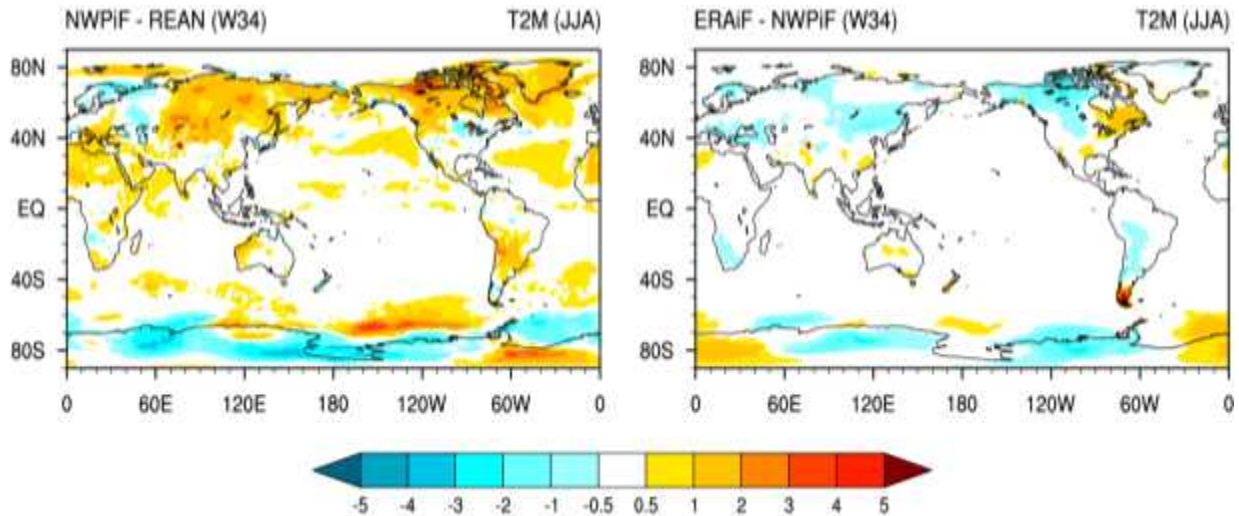


Figure 80. Difference distribution of surface temperature at lead weeks 3-4 between NWPIF and ERA5 (left), ERAiF and NWPIF (right).

초기조건 변화로 인한 개선 정도를 나타내기 위해 두 실험 간의 RMSE를 비교하여 개선율을 figure 81에 나타내었다. 두 실험의 RMSE의 변화율을 나타낸 그림으로 빨간색으로 나타난 부분은 ERAiF 실험이 NWPIF 실험에 비해 RMSE가 줄어들었음을 나타내며, 파란색으로 나타난 부분은 RMSE가 증가했음을 나타낸다. 사용된 결과는 2018년 1월-12월 전체 결과를 사용하여 평균하였다. 예를 들어, 0.5 이상 크기로 빨간색이 나타난 지역은 NWPIF 실험보다 ERAiF 실험이 50% 이상 RMSE가 감소, 개선되었음을 나타낸다. 선행시간 1일의 결과를 살펴보면, 중위도 대륙 지역 일부를 제외하고 대부분의 지역에서 개선이 되고 있음을 알 수 있다. 특히 저위도 및 고위도에서 효과적인 개선을 보이고 있다. 선행시간 3-4주의 결과를 살펴보면, 북반구 고위도 및 중위도 대륙 일부에서 RMSE가 다소 감소되는 것이 보이지만, 대부분의 지역 특히 중위도에서 큰 개선을 보이고 있음을 알 수 있다.

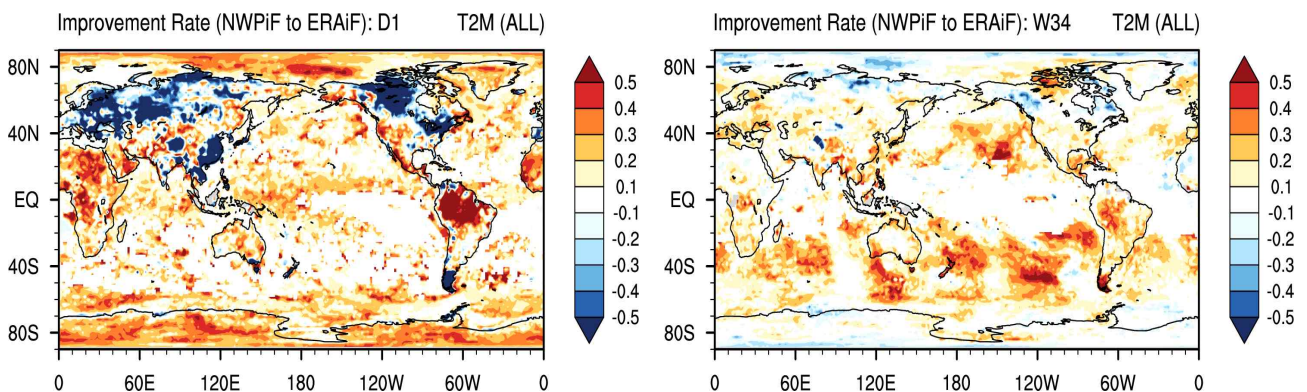


Figure 81. Improvement rate (*100%) of the ERAiF compare to the NWPIF using RMSE of surface temperature at lead day 1 (left), and lead weeks 3-4 (right).

동아시아 지역을 대상으로 계절별 예측 성능의 차이를 살펴보기 위해, 동아시아 계절 기후에 가장 연관이 있는 지표 온도 및 500hPa 지위고도의 ACC를 figure 82와 83에 나타내었다. 선행시간 1일에서의 지표 온도의 동아시아 지역 평균 ACC를 살펴보면, 전 계절에서 모두 0.7 이상의 높은 성능을 보이고 있다. 다만 NWPIF 실험의 경우 1-2월에 다소 낮은 성능을 보이고 있다. 그와 비교하였을 때, ERAiF 실험의 경우 모든 계절에 대해 높아진 ACC를 볼 수 있다. 특히 NWPIF 실험에서 낮게 나타났던 1-2월의 경우 더욱 개선된 ACC를 가져온다. 500hPa 지위고도의 경우 NWPIF 실험과 ERAiF 실험의 ACC는 거의 동일하게 나타나고 있다.

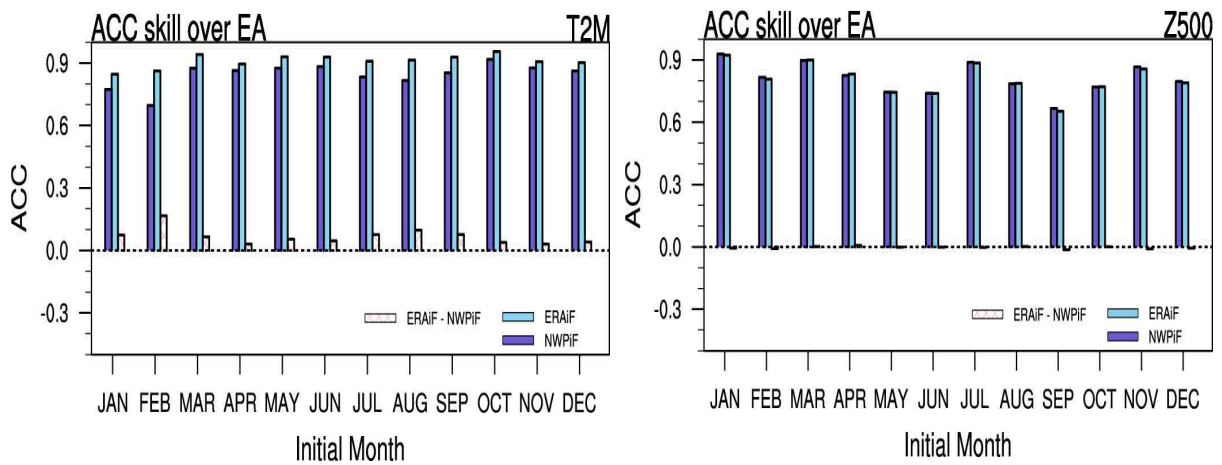


Figure 82. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) at lead day 1.

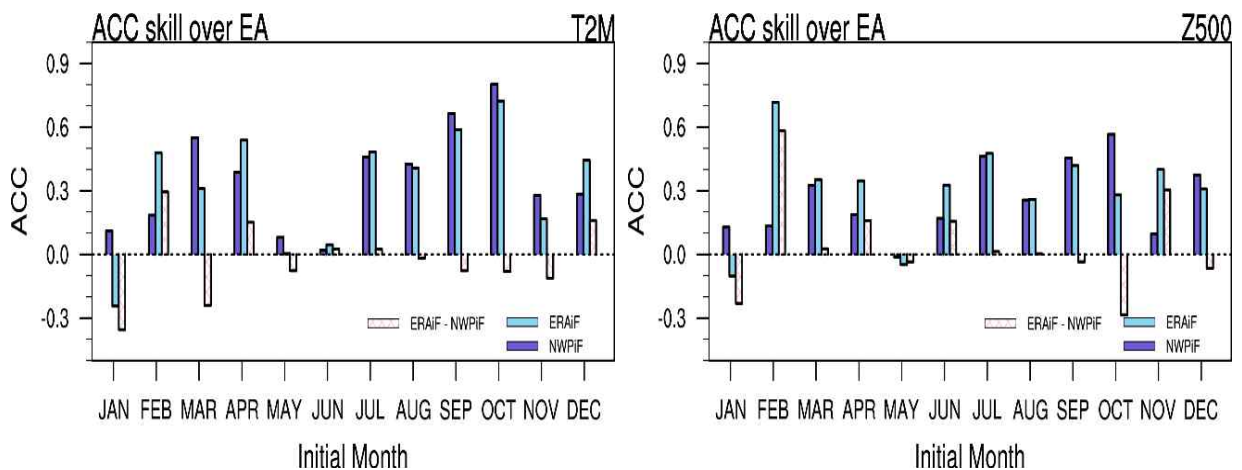


Figure 83. Monthly averaged Anomaly Correlation Coefficient of surface temperature over globe (left) and East Asia (right) at lead weeks 3-4.

선행시간 3-4주 결과의 동아시아 지역 예측 성능을 나타낸 그림을 살펴보면, 선행시간 1일과 비교할 때 그 효과가 랜덤으로 나타나고 있다. 지표 온도의 ACC의 경우 선행시간 1일에서

모든 계절 개선이 되었던 반면, 선행시간 3-4주의 결과는 오히려 안좋아지는 경우도 많다. 특히 1월의 경우 선행시간 1일에서 많은 개선을 보였지만, 3-4주의 결과는 ACC가 0보다 작아지고 있음을 알 수 있다. 하지만, 500hPa 지위고도의 경우, 1월, 10월을 제외하고 ERAiF의 실험이 NWPIF실험보다 대체적으로 ACC가 높게 나타나고 있음을 알 수 있다.

계절별로 어느 정도 차이가 있지만 전체적인 예측 성능의 차이를 살펴보기 위해 선행주수 1-8주까지 전지구 및 동아시아 지역의 전계절 평균된 ACC를 figure 84에 나타내었다. 계절별 차이가 있긴 하지만, 전계절 평균 ACC를 살펴보면 전지구 및 동아시아 모두 NWPIF보다 ERAiF의 예측 성능이 다소 높게 나타나고 있음을 알 수 있다. 동아시아 지역의 지표 온도의 경우 선행주수 1주부터 5주에 이르기까지 모두 향상된 ACC를 보이고 있다. 동아시아 지역의 500hPa 지위고도 역시 선행주수 1주에서 7주에 이르기까지 모두 향상된 ACC를 보이고 있다. 즉, 대기 초기 조건이 향상된 ERAiF 실험의 경우 어느정도의 오차를 포함하고 있는 NWP 예측장 자료를 초기장으로 사용하는 NWPIF 실험의 경우에 비해 동아시아 계절내 예측 성능을 효과적으로 향상시킬 수 있음을 확인할 수 있다. 즉, 현재 현업시스템에서의 초기장의 제한적 현실에 의해 NWP 예측장을 대기 초기 조건으로 사용하고 있기 때문에, 계절내 예측 성능에서 그로 인한 부정적 영향을 받고 있음을 알 수 있다. 다만, 이러한 효과는 월별로 다르게 나타나고 있기 때문에 쉽게 오차 제거를 통한 방법으로 해결할 수 없음을 알 수 있다.

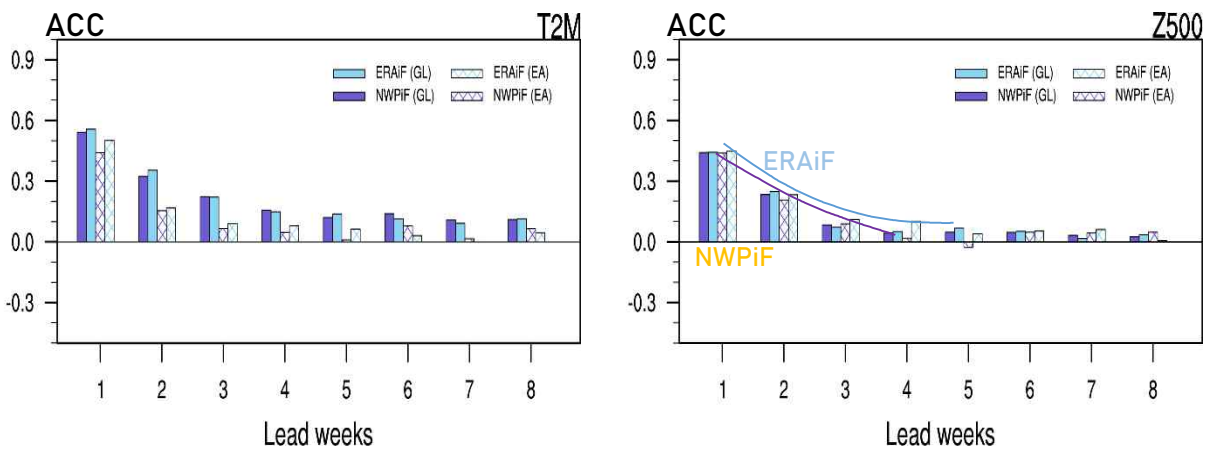


Figure 84. Anomaly Correlation Coefficient of surface temperature (left) and 500hPa geopotential height (right) over globe (shaded bar) and East Asia (dashed bar).

예측 성능을 살펴보면, 선행시간 3-4주이후에도 그 영향이 유지되는 것을 알 수 있다. 따라서 계절예측에는 어떠한 영향이 있는지 해수면 온도의 월별 평균의 분포를 figure 85에서 살펴보았다. 북반구 여름철의 경우 재분석 자료를 초기장으로 사용한 ERAiF 실험의 경우에서 한반도 주변 및 일부 지역에서의 온도 변화를 볼 수 있다. 선행 시간이 4-6개월로 길어질수록 중위도 로컬 지역의 SST의 변화가 무시할 수 없을 정도로 나타나고 있다. 해양 및 해빙 초기장의 변화가 없기 때문에, 해양 연직 온도 및 혼합층 깊이의 영향은 아주 미미하게 나타난다 (보이지 않음). 다만, 해수면 온도의 차이 및 해수면 염분 농도의 차이가 북극 지역에서 다소 나

타나기 때문에 북극-중위도 역학 관계에 영향이 있을 수 있다. 계절 예측 성능의 척도로 흔히 사용되는 적도 지역 SST의 예측 성능에는 큰 영향을 미치지 못하지만 (보이지 않음), 로컬한 지역에서 SST 변화가 나타나기 때문에 지역적 영향장을 살펴보는 연구에서는 주의할 필요가 있다.

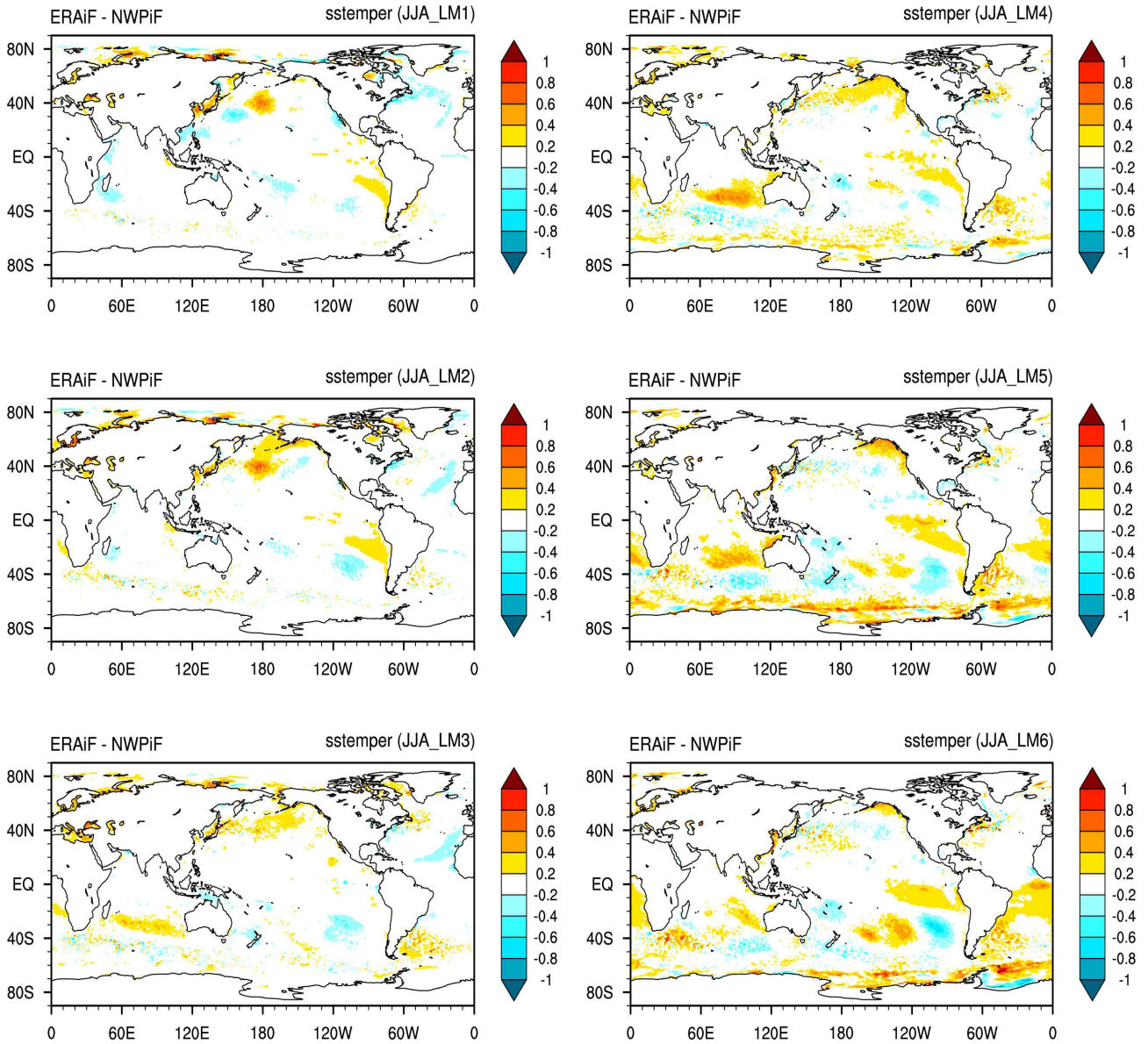


Figure 85. Difference distribution of sea surface temperature at lead month 1-6 between ERAiF and NWPiF.

(2) MJO 오차진단 툴 개발

(가) 과거재현기후에서의 계절내 변동성 특성

GloSea5 과거재현기후의 계절내 변동성 및 MJO 예측성능을 분석하기 위해 1991년부터

2010년까지의 계절내 예측 결과를 분석하였다. 선행 시간에 따른 모델 예측성능을 비교하기 위해 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34)을 분석 기준으로 설정하였다. 주요 분석 변수는 하층 동서 바람장(U850)과 상층 동서 바람장(U200), 상층 발산장(VP200)이다.

① 선행 시간에 따른 MJO 예측 성능 및 오차 분석

Figure 86는 관측(OBS)과 기후예측모델(GL5) 과거재현기후의 하층 동서 바람장(U850)과 상층 동서 바람장(U200), 상층 발산장(VP200)의 겨울철(11월~4월) 평균 공간분포(shading)와 변동성 공간분포(contour)를 선행 시간(W12:선행 1-2주, W34: 선행 3-4주)에 따라 제시한 결과이다. 계절내 변동성의 공간 분포 특성을 분석하기 위해 20~70일 band pass filter를 적용하였다. 선행시간 1-2주에서는 주요 변수의 평균 공간분포와 변동장 공간분포가 관측과 유사하게 확인되지만, 선행시간 3-4주에서는 공간 분포 특징의 예측성능이 낮게 나타났다. 기후예측모델 과거재현기후의 계절내 변동성의 두드러진 특징을 보이는 적도 인도양 지역의 겨울철(11~4월)에 대한 관측 및 모델의 하층 동서 바람장(U850)과 상층 동서 바람장(U200), 상층 발산장(VP200) 파워 스펙트럼 분석 결과를 Figure 87에 제시하였다. 모델 결과는 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34)을 분석 기준으로 나타났다. 선행시간 1-2주에서는 관측의 계절내 변동성 주파수의 기간에 따른 빈도를 유사하게 예측하였으나, 선행시간 3-4주에서는 스펙트럼 분포의 예측성능이 낮아지는 것을 확인하였다. Figure 88은 기후예측모델 과거재현기후의 관측 및 모델의 적도지역 하층 동서 바람장(U850)과 상층 동서 바람장(U200), 상층 발산장(VP200)에 대한 겨울철(11월 - 4월) 파수 진동수(zonal wavenumber frequency) 분석 결과를 제시하였다. 모델은 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34)을 분석 기준으로 설정하였다. 적도 지역 파수는 계절내 변동성 특징이 나타나는 30~80일의 빈도에서 두드러진 특징을 보였고 선행주수가 증가할수록 모델의 예측성능이 낮아짐을 확인하였다.

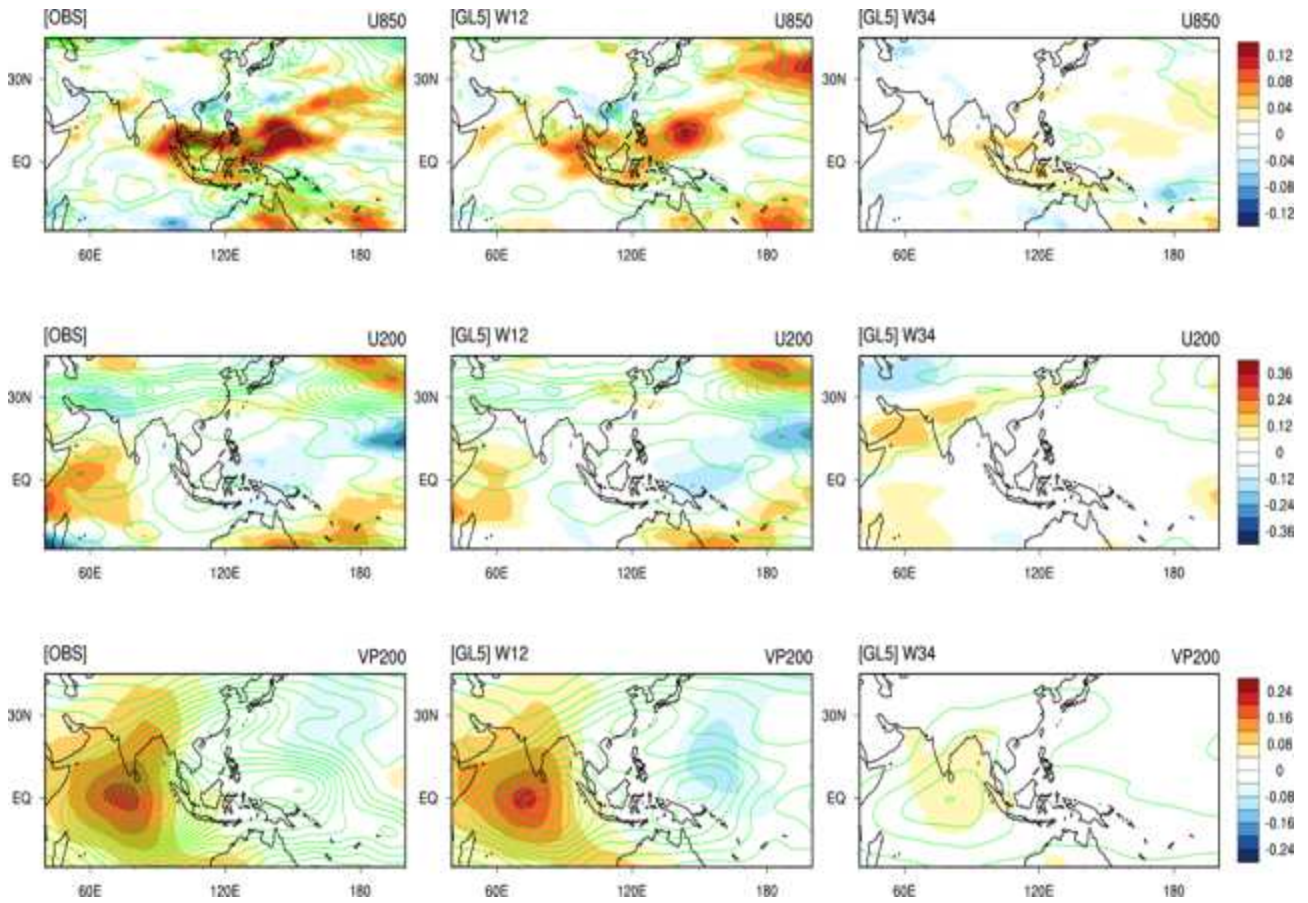


Figure 86. Spatial distributions of mean field (shading) and variance field (contour) for 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), and 200 hPa divergence (VP200) in observations (OBS) and GloSea5 model (GL5). GL5 is classified by the forecasting 1-2 weeks (W12) and the forecasting 3-4 weeks (W34).

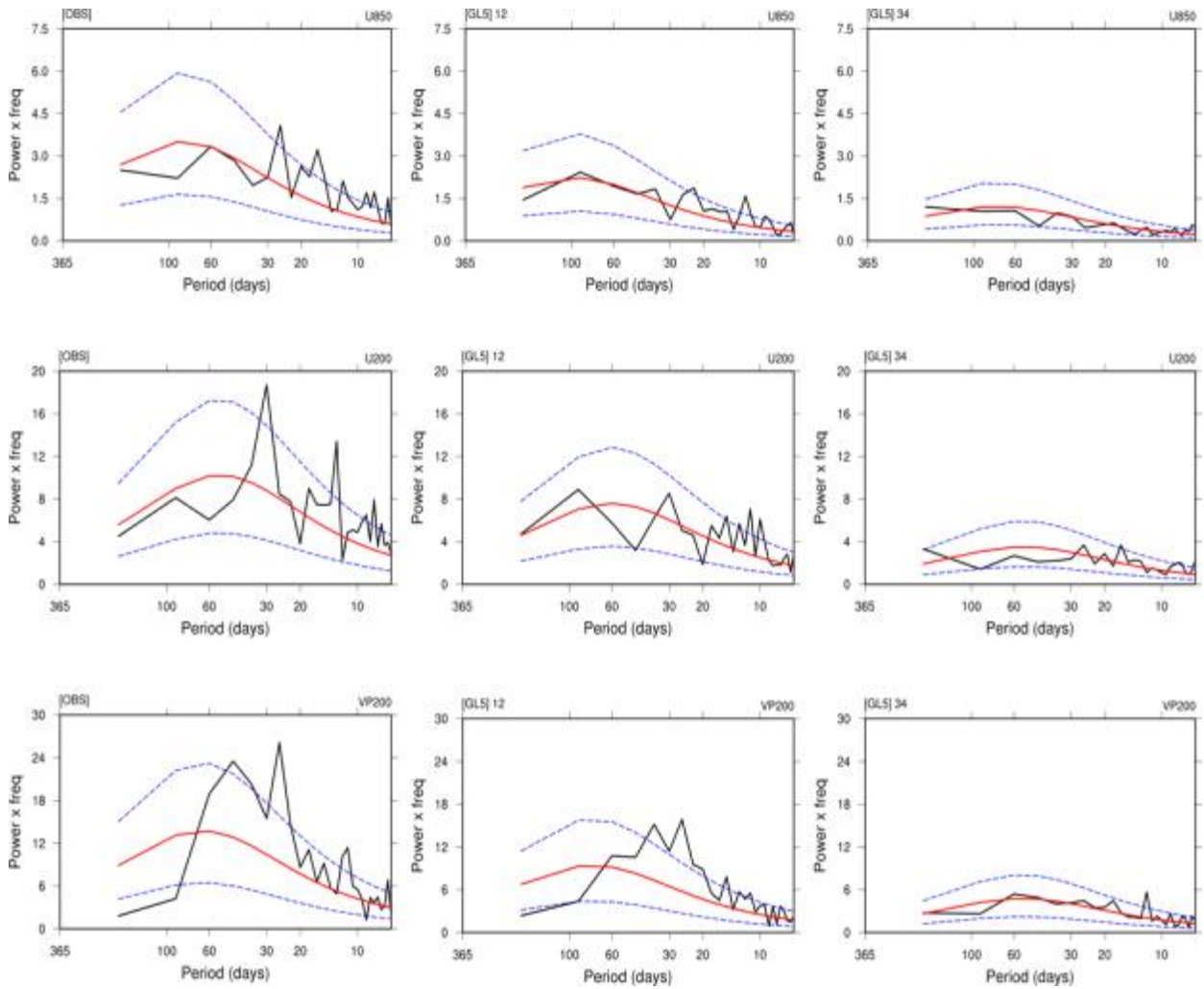


Figure 87. Power spectrum results of 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), and 200 hPa divergence (VP200) in observations (OBS) and GloSea5 model (GL5). GL5 is classified by the forecasting 1-2 weeks (W12) and the forecasting 3-4 weeks (W34).

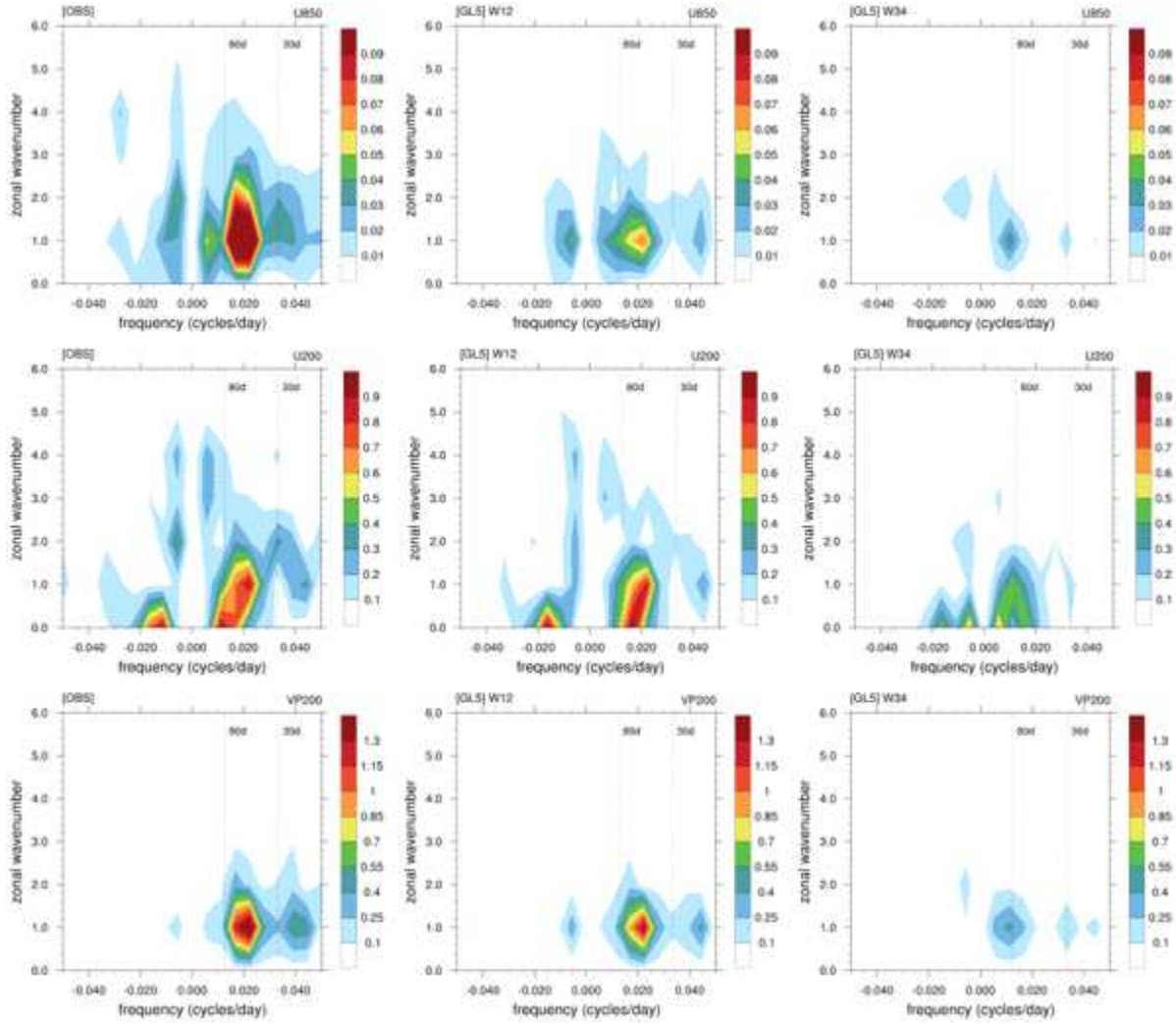


Figure 88. Wavenumber frequency results of 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), and 200 hPa divergence (VP200) in observations (OBS) and GloSea5 model (GL5). GL5 is classified by the forecasting 1-2 weeks (W12) and the forecasting 3-4 weeks (W34).

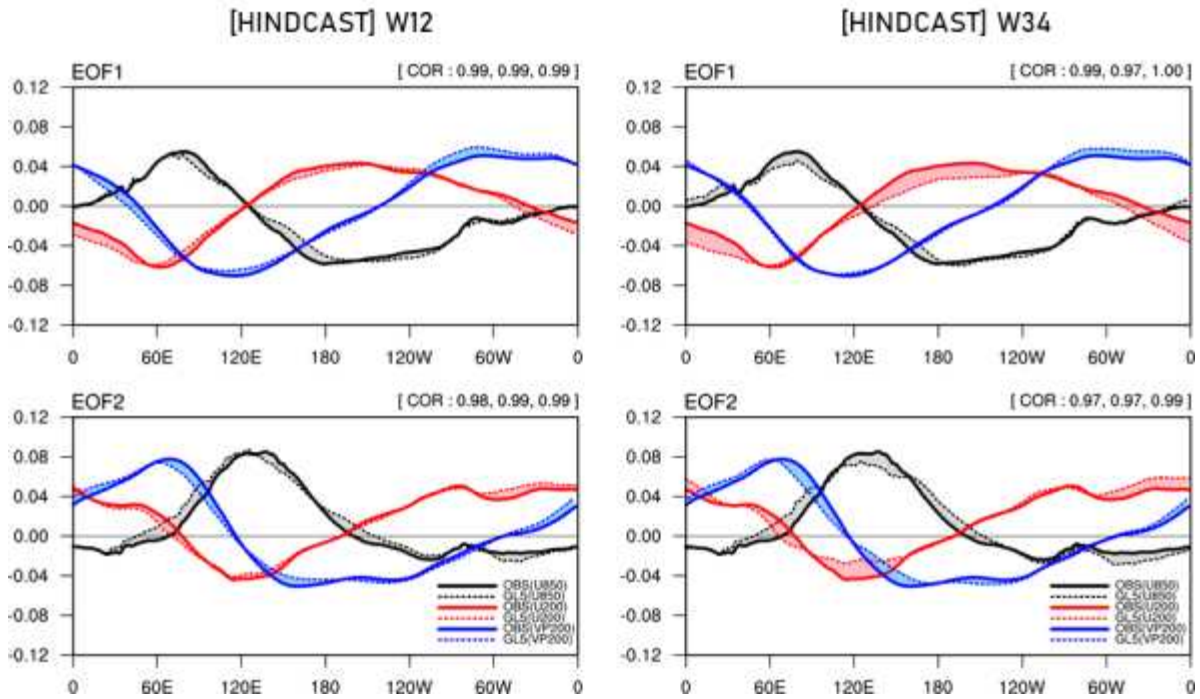


Figure 89. MEOF (Multivariate Empirical Orthogonal Function) results of 850 hPa zonal wind (U850-black line), 200 hPa zonal wind (U200-red line), and 200 hPa divergence (VP200-blue line). (OBS-line, GL5-dash line).

기후예측모델 과거재현기후의 적도 변동성을 살펴보기 위해 MJO 지수(Velocity Potential MJO Index, VPM) 분석에 사용되는 하층 동서 바람장(U850), 상층 동서 바람장(U200) 및 상층 발산장(VP200)의 다변량 직교함수 분석(Multivariate Empirical Orthogonal Function, MEOF) 첫 번째, 두 번째 모드 결과를 Figure 89에 제시하였다. 관측 및 모델의 결과는 실선과 점선으로 각 분석 변수를 나타냈다. 모델은 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34) 기준으로 제시하였고, 각 분석 변수의 관측과의 상관계수를 오른쪽 상단에 표시했고, 관측과의 차이를 면적(shading)으로 나타냈다. 예측 선행 주수가 증가 할수록 면적이 다소 넓게 나타나 오차가 증가하는 것을 확인할 수 있지만, 선행시간 3-4주까지 MJO 변동성의 설명 변수는 높은 예측성능을 유지되는 것으로 확인된다.

Figure 90은 기후예측모델 과거재현실험의 MJO 지수(Velocity Potential MJO Index, VPM) 예측성능(Bivariate Anomaly Correlation Coefficient, COR)을 예측 일수에 따라 제시한 그림이고, Figure 91는 오차(Bivariate Root Mean Square Error, RMSE)를 예측 일수에 따라 제시한 그림이다. 기후예측모델 과거재현실험 MJO 지수의 예측 성능은 예측 일수가 증가할수록 감소하는 것으로 나타났고, 오차는 예측 일수가 증가할수록 증가하는 것을 확인할 수 있다.

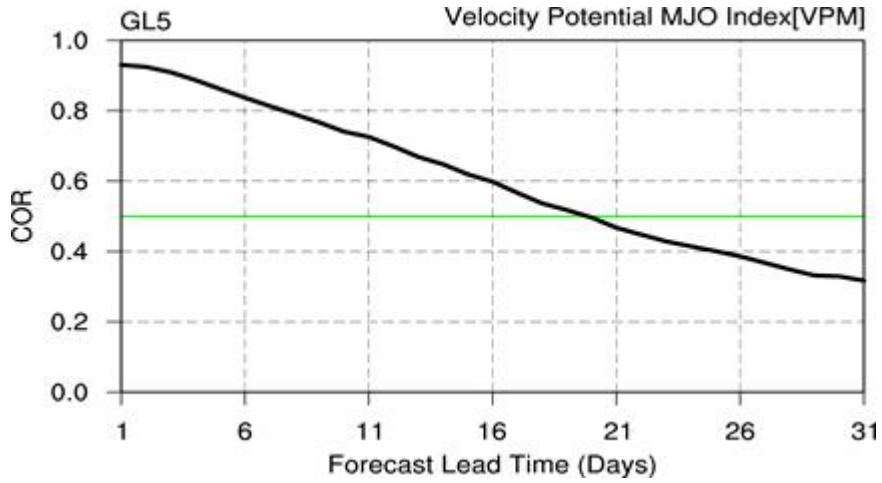


Figure 90. COR (Bivariate Anomaly Correlation Coefficient) for Velocity Potential MJO indices in GloSea5 HINDCAST (green line : 0.5).

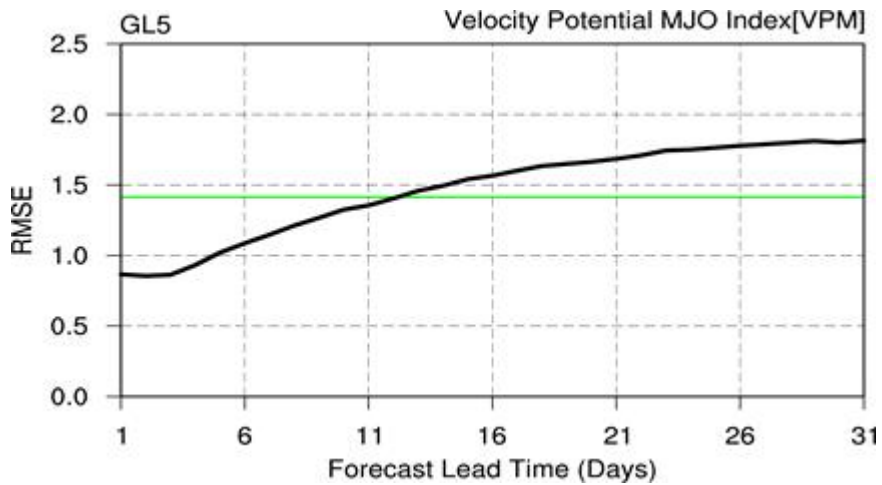


Figure 91. RMSE (Bivariate Root Mean Square Error) for Velocity Potential MJO indices in GloSea5 HINDCAST (green line : $\sqrt{2}$).

(나) 실시간 예측장에서 초기장 차이에 따른 MJO 예측 성능 및 오차 분석

초기 오차에 의한 적도 변동성의 공간 분포 특징을 분석하기 위해 재분석 자료인 ERA5를 처방한 초기장 민감도실험(ERAiF)과 NWP 초기장을 사용하는 현업 실시간예측장(FORECAST) 결과인 규준실험(NWPiF)을 선행 주수에 따라 비교하였다.

① 초기장 차이에 의한 적도 변동성 오차 특징

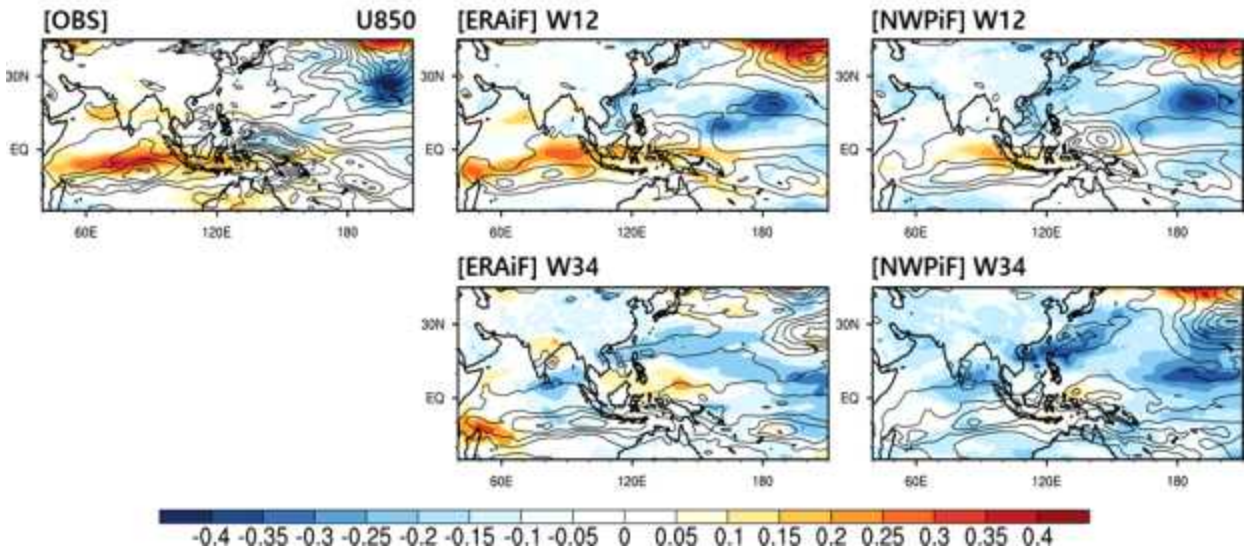


Figure 92. Spatial distributions of mean field (shading) and variance field(contour) for 850 hPa zonal wind (U850) in observations (OBS) and initial sensitivity experiments (ERAiF) and controlled experiment (NWPiF).

Figure 92는 관측(OBS)과 재분석자료 초기장 실험(ERAiF), NWP 초기장 실험(NWPiF)의 하층 동서 바람장(U850)의 겨울철(2017년 11월~2018년 4월) 평균 공간분포(shading)와 변동성 공간분포(contour)를 선행 시간(W12:선행 1-2주, W34: 선행 3-4주)에 따라 제시하였다. 계절내 변동성의 공간 분포 특성을 분석하기 위해 20~70일 band pass filter를 적용하였다. 겨울철 U850 변동성은 MJO 활동 영역인 인도양과 서태평양 부근에서 크게 확인되었다. W12의 변동장 공간상관계수(Patten Correlation Coefficient, PCC)는 ERAiF실험과 NWPiF실험에서 각각 0.91, 0.92로 나타났고, W34에서는 각각 0.73, 0.67로 재분석자료 초기장을 처방한 실험의 공간 변동장 예측성능이 NWP 초기장 실험 대비 약 9% 높은 예측성능을 보였다.

Figure 93는 적도(10S-10N) 인도양-서태평양 영역(60-160E)의 시간에 대한 하층 동서 바람장(U850)의 분포를 분석하기 위해 관측(OBS)과 재분석자료 초기장 실험(ERAiF), NWP 초기장 실험(NWPiF)의 겨울철(2017년 11월~2018년 4월) 시간에 따른 U850의 변화를 Hovmoller diagram을 사용하여 선행 주수(W12:선행 1-2주, W34: 선행 3-4주)에 따라 제시한 결과이다.

초기장 처방에 따른 선행 시간 1-2주의 U850 분포 차이는 크지 않았으나, 선행 시간 3-4주 결과에서는 NWP 초기장 결과 대비 재분석자료를 초기장으로 처방한 실험결과의 U850 분포가 관측과 유사하게 확인되었다.

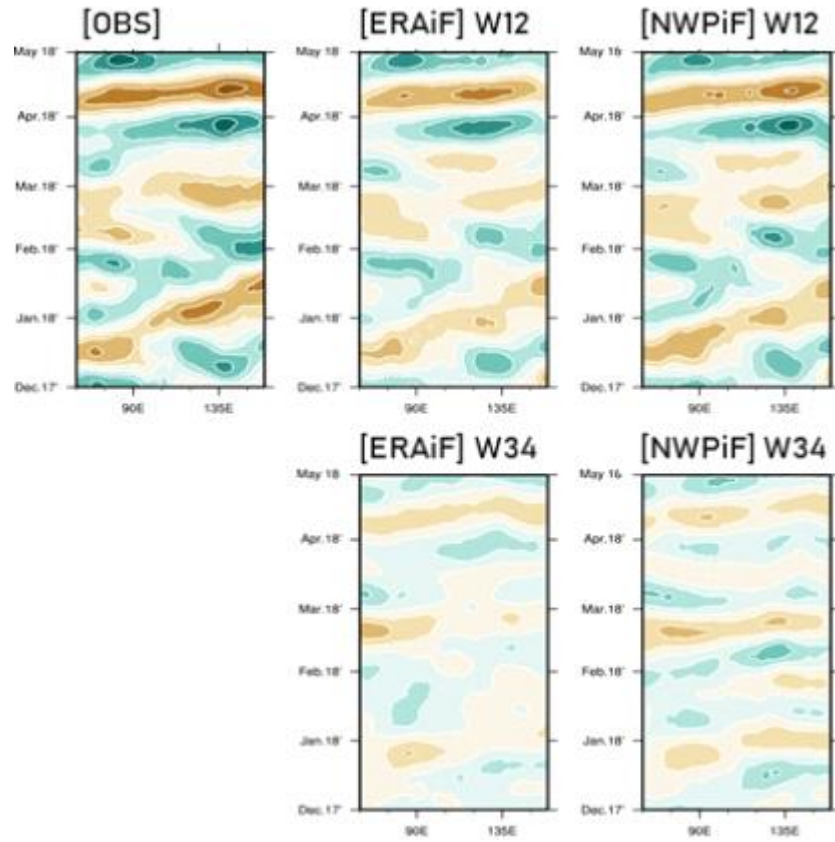


Figure 93. Hovmoller diagram of 850 hPa zonal wind(U850) in observations (OBS) and initial sensitivity experiments (ERAiF) and controlled experiment (NWPIF) during the winter (Dec 2017~Apr 2018).

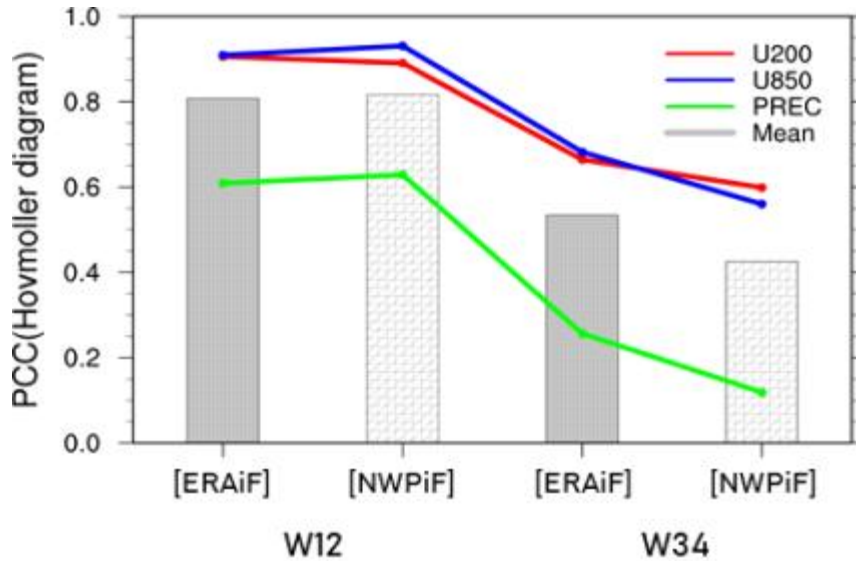


Figure 94. PCC skill of hovmoller diagram for 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), precipitation (PREC), and the mean of all variables in initial sensitivity experiments (ERAI f - filled bar) and controlled experiment (NWPI f - patterned bar).

Hovmoller diagram 분석을 MJO 변동성의 주요 변수인 상층 동서 바람장(U200), 하층 동서 바람장(U850), 강수(PREC)에 대해 적용하였다. Figure 94은 재분석자료 초기장 실험(ERAI f - filled bar)과 NWP 초기장 실험(NWPI f - patterned bar)의 MJO 변동성을 설명하는 개별 주요 변수(U200, U850, PREC)와 주요 변수 평균(Mean)의 Hovmoller diagram의 공간 상관 계수(PCC)를 선행주수(W12/W34)에 따라 제시한 결과이다. 주요 변수의 예측성 평균 비교를 통해 재분석자료를 초기장으로 처방한 실험과 NWP 초기장 실험의 선행시간 1-2주의 예측성은 유사하게 나타났지만, 선행시간 3-4주의 예측성은 NWP 초기장 실험 대비 재분석자료를 처방한 실험 결과에서 25.4% (Improvement Rate) 높게 확인되었다.

Figure 95은 관측(OBS)과 재분석자료 초기장 실험(ERAI f), NWP 초기장 실험(NWPI f)의 겨울철(2017년 11월~2018년 4월) 하층 동서 바람장(U850)의 지연 시간 경도 상관분석(lag time-longitude correlation) 결과로, 선행1-2주(W12) 및 3-4주(W34)로 제시하였다. 각 실험 결과에서 동일하게 선행시간이 1-2주에서 3-4주로 증가할수록 MJO 전파(propagation) 모의 능력이 낮아지는 것이 확인된다. 재분석자료 초기장 실험(ERAI f - filled bar)과 NWP 초기장 실험(NWPI f - patterned bar)의 상층 발산장(VP200), 하층 동서 바람장(U850), 강수(PREC) 그리고 각 변수 평균(Mean)의 MJO propagation diagram(60-180E, 10S-10N, 20~70 days filtered)의 공간 상관 계수(PCC)를 선행주수(W12/W34)에 따라 Figure 96에 제시하였다. MJO 주요 변수의 MJO propagation diagram 예측성 평균 비교를 통해 재분석자료로 초기장을 처방한 실험과 NWP 초기장을 처방한 실험의 선행시간 1-2주의 예측 성능은 유사하나, 선행시간 3-4주의 예측 성능은 NWP 초기장 실험 대비 재분석자료를 처방한 실험에서 33.4% (Improvement Rate) 높게 확인된다.

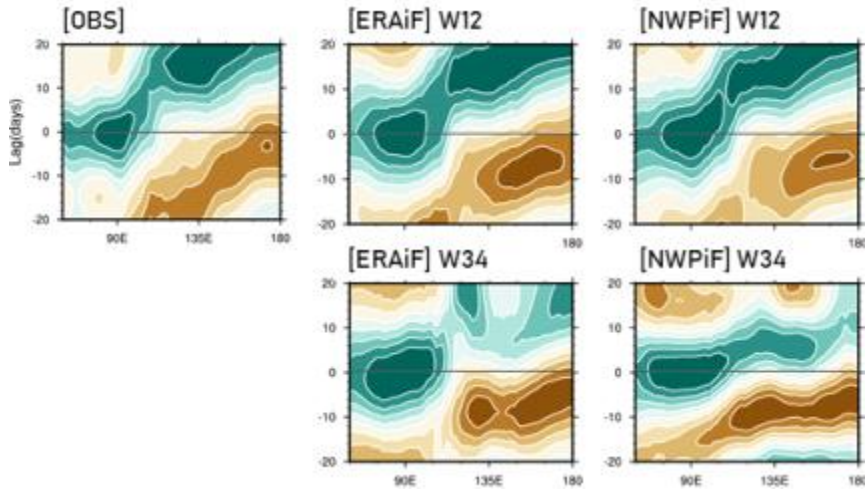


Figure 95. Lag time-longitude correlation (i.e. MJO propagation diagram) of 850 hPa zonal wind (U850) in observations (OBS) and initial sensitivity experiments (ERAiF) and controlled experiment (NWPiF) during the winter (Dec 2017~Apr 2018) (60-180E, 10S-10N, 20~70 days filtered).

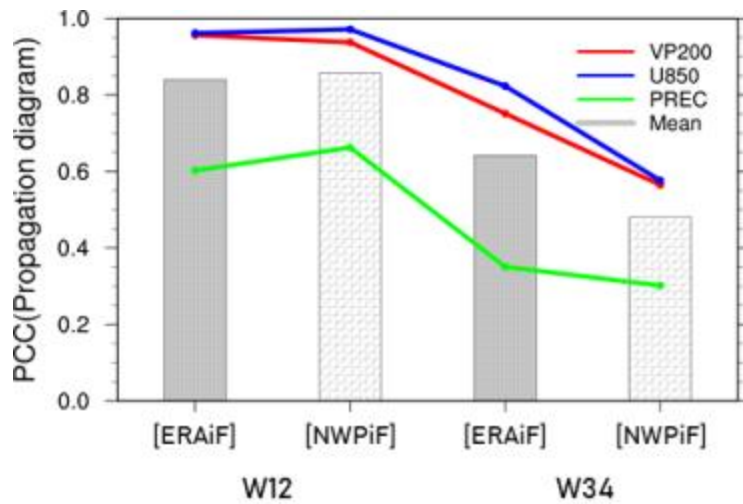


Figure 96. PCC skill of MJO propagation diagram (60-180E, 10S-10N, 20~70 days filtered) for 850 hPa zonal wind (U850), 200 hPa zonal wind (U200), 200 hPa divergence (VP200), and the mean of the variable in initial sensitivity experiments (ERAiF - filled bar) and controlled experiment (NWPiF - patterned bar).

② 초기장 차이에 의한 수증기 연직 오차 구조 특징

재분석자료를 초기장으로 처방한 민감도 실험 결과와 NWP 초기장 결과의 중층대기 수증기 연직구조 비교를 위해 GloSea5 연직 자료와 비교 가능하도록 관측(ERA5) 연직자료를 추출하였으며, 모델 및 관측처리의 1991-2010년 기준으로 기후값을 산출하여 아노말리값을 분석에 사용하였다.

Figure 97과 Figure 98은 재분석자료를 초기장으로 처방한 민감도 실험 결과(ERAiF)와 NWP 실험 결과(NWPiF)의 적도지역의 MJO 기간 중층대기 비습과 상당온위 결과를 선행주수(W12, W34)에 따라 제시하였다. Figure 97의 점선으로 표시된 영역은 MJO 수증기 변동성이 가장 두드러지는 Maritime Continent 지역이다. 850 hPa 부터 500 hPa 까지의 중층 대기 수증기량을 관측과 비교했을 때, 모델의 두 실험 결과에서 선행시간이 1-2주에서 3-4주로 증가하면서 오차가 상대적으로 크게 확인된다. 특히 선행시간 3-4주에서는 재분석자료를 초기장으로 처방한 실험에서 수증기량을 일괄되게 과소모의하는 것으로 나타났고, NWP 초기장을 처방한 실험에서는 오차의 범위가 더 크게 확인되었다. 중층 대기의 상당 온위를 관측과 비교한 Figure 98 결과에서는, 수증기량의 재현성능과 유사한 패턴을 확인하였다. 재분석자료 초기장을 처방한 실험과 NWP 초기장을 처방한 실험 결과의 중층 대기 수증기의 계절내 변동성 예측성능은 선행시간이 늘어날수록 오차가 크게 증가하였으며 다른 변수에 비해 초기장 처방에 따른 예측력 향상이 두드러지지 않음을 확인하였다. 수증기의 계절내 변동성의 예측성을 확보하기 위해 강수 및 구름 모수화와 관련된 추가적인 연구가 필요할 것으로 생각된다.

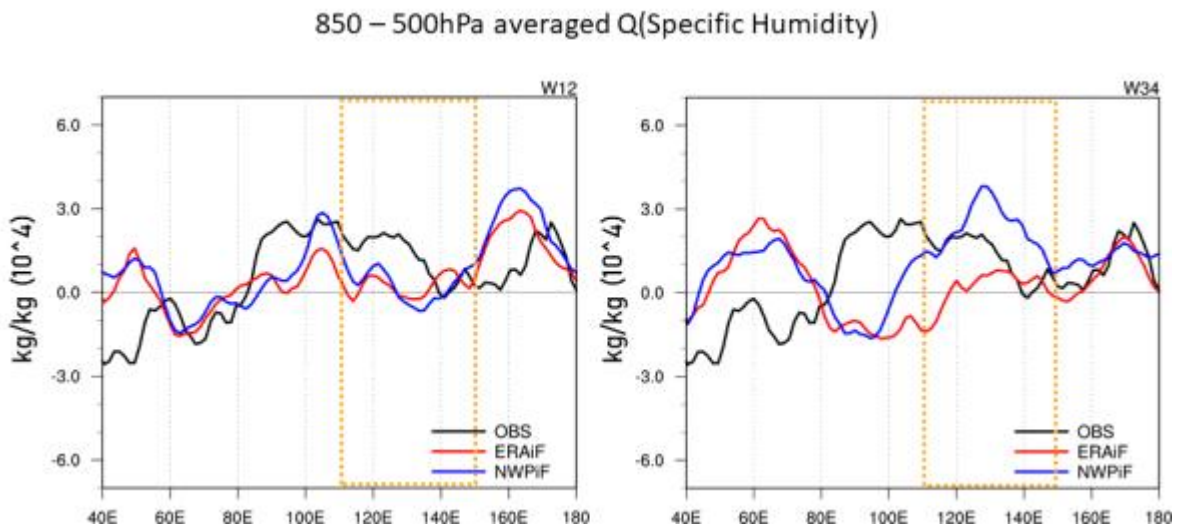


Figure 97. Distribution of specific humidity on middle atmosphere (850-500hPa) in observation (black) and ERAiF (red) and NWPiF (blue) experiments by the forecasting 1-2 weeks (left) and forecasting 3-4 weeks (right) (10S-10N, 20-70 days filtered).

850 – 500hPa averaged Q(Equivalent Potential Temperature)

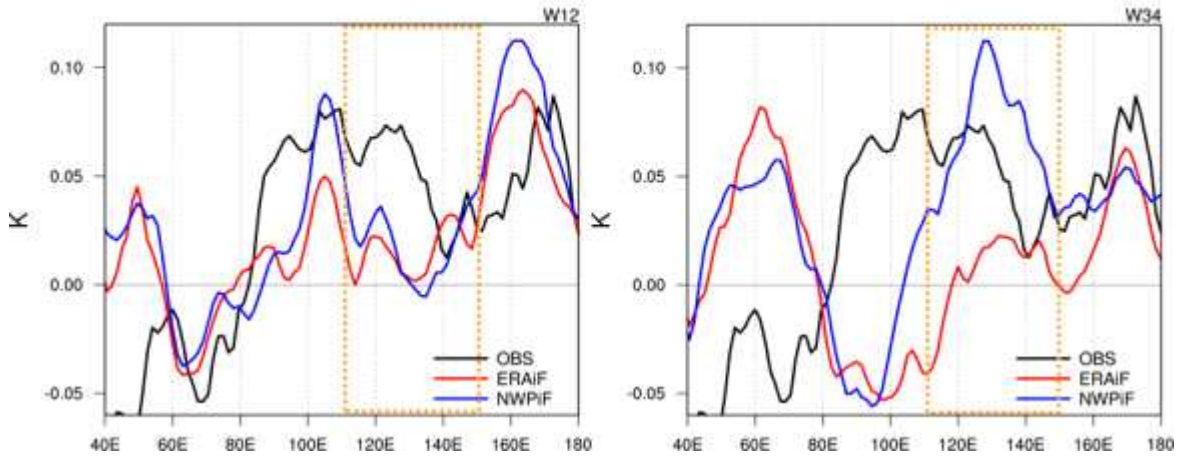


Figure 98. Distribution of equivalent potential temperature on middle atmosphere (850-500hPa) in observation (black) and ERAiF (red) and NWPiF (blue) experiments by the forecasting 1-2 weeks (left) and forecasting 3-4 weeks (right) (10S-10N, 20~70 days filtered).

③ 초기장 차이에 따른 MJO 예측 성능 및 오차 비교

이 연구에 사용한 MJO 지수는 속도 포텐셜 MJO 지수(Velocity Potential MJO Index, VPM)이며, MJO 변동성을 설명하는 주요 변수인 하층 수평 바람장(U850), 상층 수평 바람장(U200), 상층 발산장(VP00)의 MEOF를 통해 산출된다(Ventrice et al. 2013). 하층 수평 바람장(U850), 상층 수평 바람장(U200), 장파 복사(OLR)로 산출하는 RMM 지수에 비해 예측성능이 약 5~10일 짧게 확보되는 것으로 알려져 있다. GloSea5의 계절내예측 실험은 60일 기준으로 예측자료를 생산한다. MJO 지수 분석을 위해서는 60일 예측자료의 연간 변동성을 제거하는 과정(harmonics 전처리)과 ENSO 영향을 제거하는 과정(과측자료를 활용한 120일 이동평균)을 선행하였다. 초기장 처방에 따른 실험결과의 MJO 지수는 2017년 10월부터 2018년 12월 기간을 기준으로 산출되었고, GloSea5 과거모의 실험결과(HCST)의 MJO 지수는 1991년 1월부터 2010년 12월 기준으로 산출되었다.

MJO 지수 분석에 활용되는 주요 변수는 하층 동서 바람장(U850), 상층 동서 바람장(U200) 및 상층 발산장(VP200)이며, 해당 변수의 전체 실험 기간(2017년 11월~2018년 12월)의 주요 변동성을 살펴보기 위해 MEOF(Multivariate Empirical Orthogonal Function) 분석을 수행하였다. Figure 99는 재분석자료를 초기장으로 처방한 실험(ERAiF)과 NWP 초기장을 처방한 실험(NWPiF)의 선행시간 3-4주(W34)의 MEOF 결과이다. 각 변수에 대해 관측(실선)과 모델(점선)의 차이를 면적으로 제시하였고, 각 변수(U850, U200, VP200)의 상관 계수는 Figure 99의 오른쪽 상단에 표시하였다. 재분석자료를 초기장으로 처방한 실험의 각 변수의 상관계수가 NWP를 초기장으로 처방한 실험에 비해 높고 오차(면적)가 효과적으로 감소하는 것을 확인하였다. 이러한 차이는 MJO 변동성이 활발한 지역인 인도양과 서태평양 지역에서 두드러지게 나타났고, 재분석자료 초기장 처방에 따른 MJO 주요 변수의 예측 성능의 향상을 확인하였다.

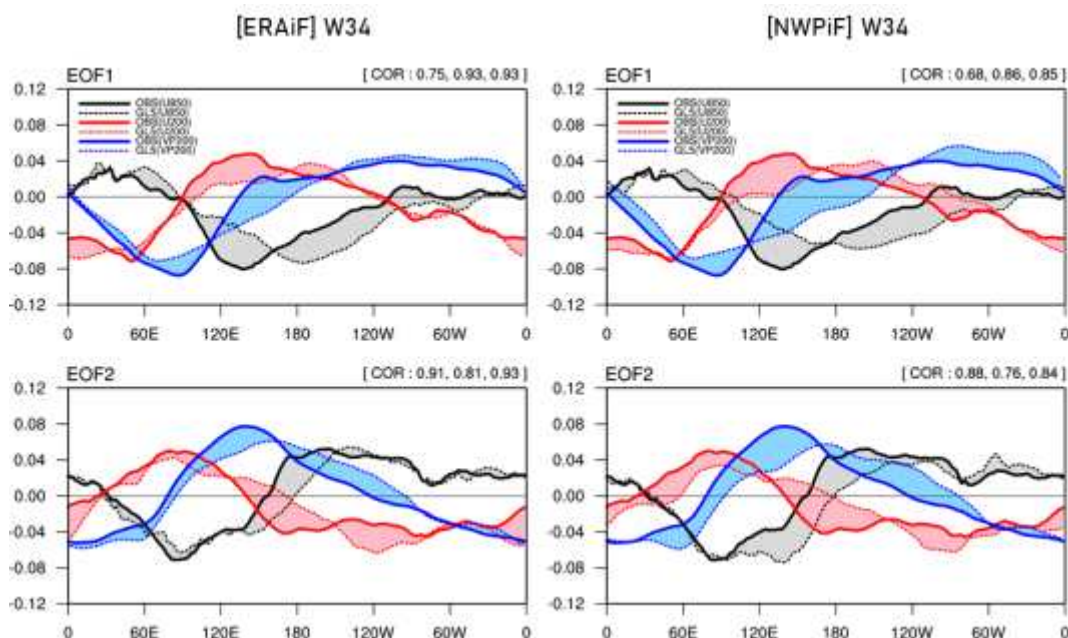


Figure 99. MEOF (Multivariate Empirical Orthogonal Function) results of 850 hPa zonal wind (U850-black), 200 hPa zonal wind (U200-red), and 200 hPa divergence (VP200) in initial sensitivity experiments (ERAiF) and controlled experiment (NWPiF) for the forecast 3-4 weeks.

Figure 100에 GloSea5 과거모의실험(HCST) 및 재분석자료 초기장 실험(ERAiF)과 NWP 초기장 실험(NWPiF)의 MJO 지수(Velocity Potential MJO Index, VPM)의 예측성능(Bivariate anomaly correlation coefficient, BCOR)을 예측 일수(Lead Time)와 분석 기준이 되는 예측 주수(Lead Week)인 1-2주와 3-4주의 평균으로 제시하였다. 예측 일수가 증가함에 따라 재분석자료 초기장을 사용한 실험의 예측성능이 NWP 초기장 실험 대비 높게 유지되는 것으로 나타났다. 선행 주수 3-4주(W34)에서의 예측성능은 재분석자료 초기장을 처방한 실험결과에서 NWP 초기장 실험 대비 12% (Improvement Rate) 높게 확인되었다.

Figure 101은 GloSea5 과거모의실험(HCST) 및 재분석자료 초기장 실험(ERAiF)과 NWP 초기장 실험(NWPiF)의 MJO 지수(Velocity Potential MJO Index, VPM)의 오차(Bivariate Root Mean Square Error, BRMSE)를 예측 일수(Lead Time)와 분석 기준이 되는 예측 주수(Lead Week)인 1-2주와 3-4주의 평균으로 제시한 Figure이다. 예측 일수가 증가함에 따라 과거모의 실험과 민감도 실험, 규준실험의 오차가 증가하는 것으로 확인되었으나 선행 주수에 따른 차이는 예측성능(BCOR)에 대비 두드러지지 차이가 나타나지 않았다.

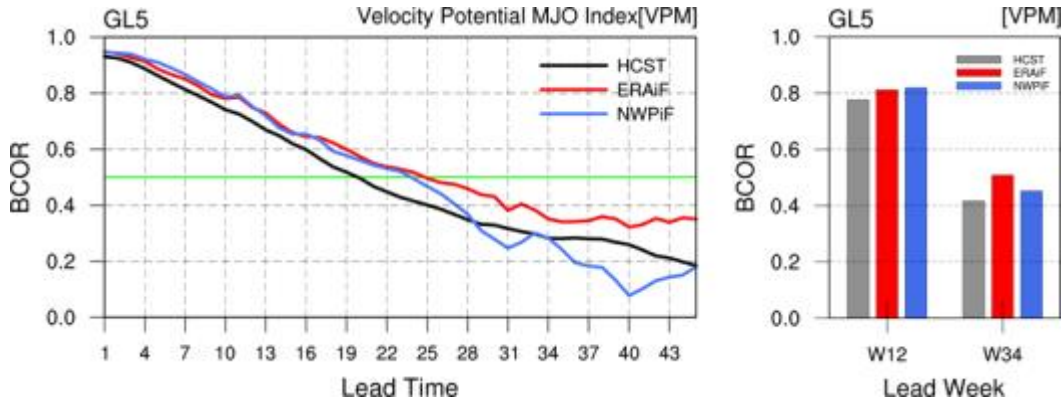


Figure 100. Bivariate Correlation coefficient (BCOR) for Velocity Potential MJO indices in GloSea5 HINDCAST (grey), initial sensitivity experiments (ERAiF-red), and controlled experiment (NWPiF-blue).

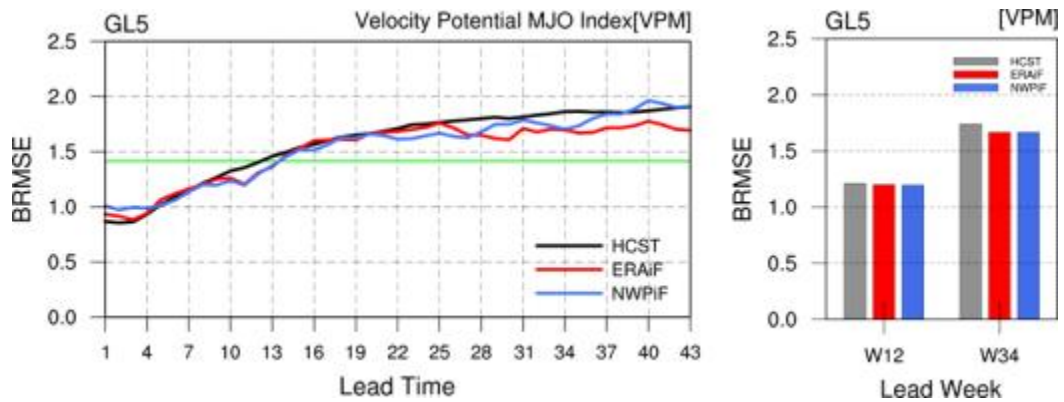


Figure 101. Bivariate Root Mean Square Error (BRMSE) for Velocity Potential MJO indices in GloSea5 HINDCAST (grey), initial sensitivity experiments (ERAiF-red), and controlled experiment (NWPiF-blue).

다. 앙상블 생산 방안에 따른 계절내 기온 예측 민감도 평가

(1) 계절내 규모에서의 실시간 예측장 앙상블 특성진단

현 시스템에서의 예측 개선을 위하여 새로운 혹은 최적의 앙상블 방안을 찾기 위해서는 현재 운영되고 있는 GloSea5 시스템에서의 앙상블의 기본 특성을 먼저 파악하는 것이 중요하다. 유의미한 결과 도출을 위해 샘플수가 많은 GloSea5 현업 후보실험을 사용하여 앙상블 특성을 파악하고자 하였다. 앙상블 특성으로는 앙상블의 유형, 신뢰도 평가, 불확실성 추정 이렇게 세 가지 관점을 선택하였다.

(가) 앙상블 유형

앙상블의 유형을 진단하기 위해 ranked histogram 분석을 하였다. GloSea5 현업 후보실험의 앙상블은 3개이므로 관측의 결과를 포함하여 1위부터 4위까지의 순위를 매길 수 있었다. 리드

1일부터 49일까지의 ranked histogram을 파란색 막대로 Figure 102에 나타내었다. 리드 초반에는 강한 U자형 모습을 보여 앙상블이 underdispersion하다면 리드 3주차부터는 평편하거나 약한 음의 바이어스를 가진다.

객관적으로 앙상블 유형을 진단하기 위하여 3차 다항식 추세선을 fitting시켜 히스토그램의 곡률을 계산하였다. 추세선은 빨간색 선으로 히스토그램과 함께 Figure 102에서 제시하였다. 곡선의 기울기와 곡선의 시종값 등의 관계를 사용하여 네 가지 앙상블 유형을 정의하였다. 이렇게 구해진 GloSea5의 앙상블 유형은 Figure 103에서 각 계절과 리드타임에 대해 잘 정리되었다.

Figure 103은 ranked histogram을 각 계절별로 제시한 그림이다. 짧은 리드타임에서의 작은 앙상블 스프레드는 모든 계절에서 동일하게 나타나는 현상이고 리드 전반에 걸쳐 겨울철 예보는 음의 바이어스를 가지고 있다는 점이 특징적이다.

대체로 GloSea5 후보실험의 기온 예측에 있어 앙상블은 리드 초반에는 그 스프레드가 관측을 모의할 만큼 충분히 크지 않으나 2주 이후부터는 예보 불확실성을 표현하기에 충분한 분포를 가지고 있다고 보여진다. 단, 겨울철 앙상블은 예보 시작 약 2주 후부터는 음의 바이어스를 가지는 경향이 있다.

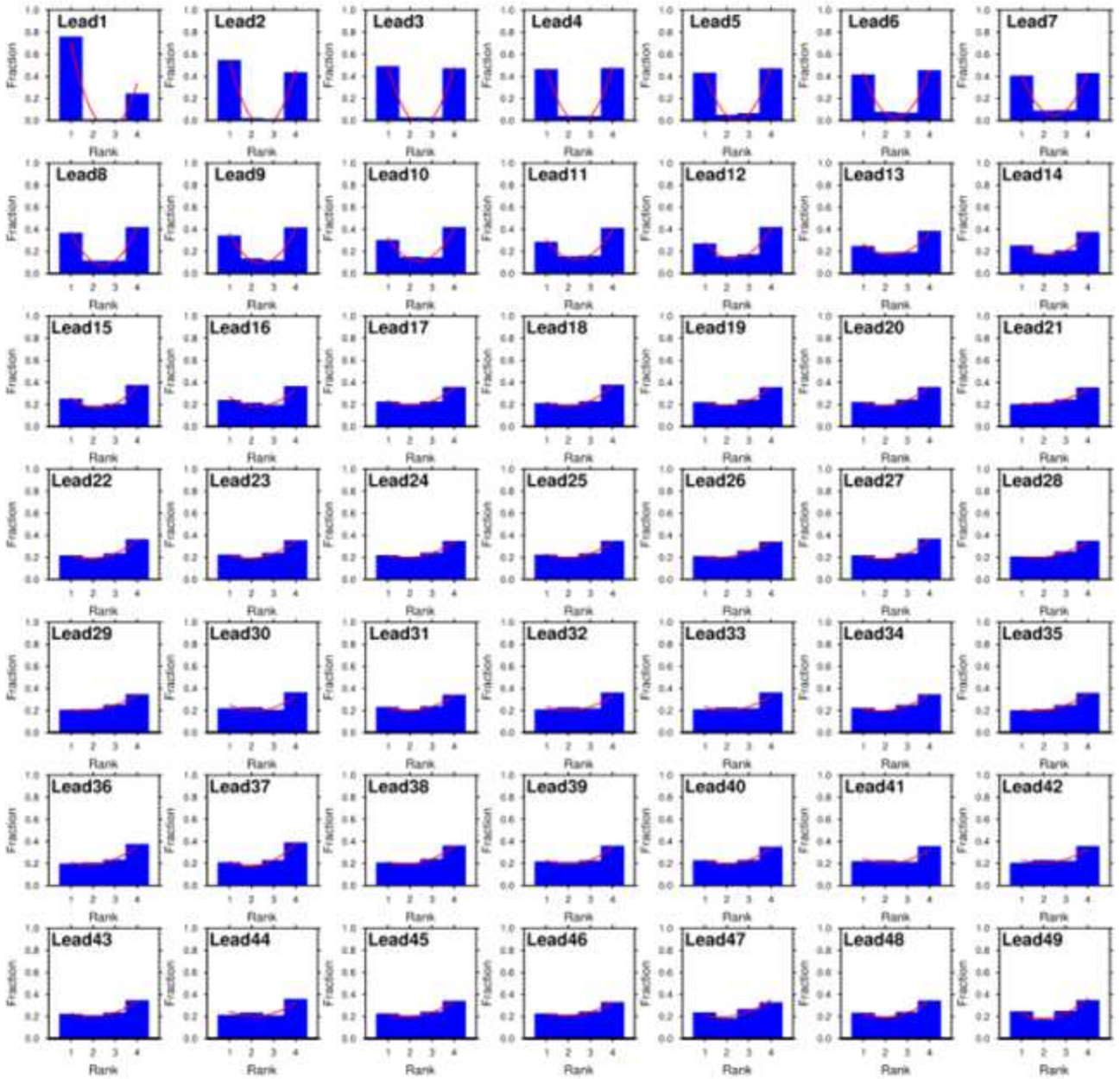


Figure 102. Ranked histogram with 3 ensemble members of GloSea5 hindcasts and observed 2m temperature over the East Asia.

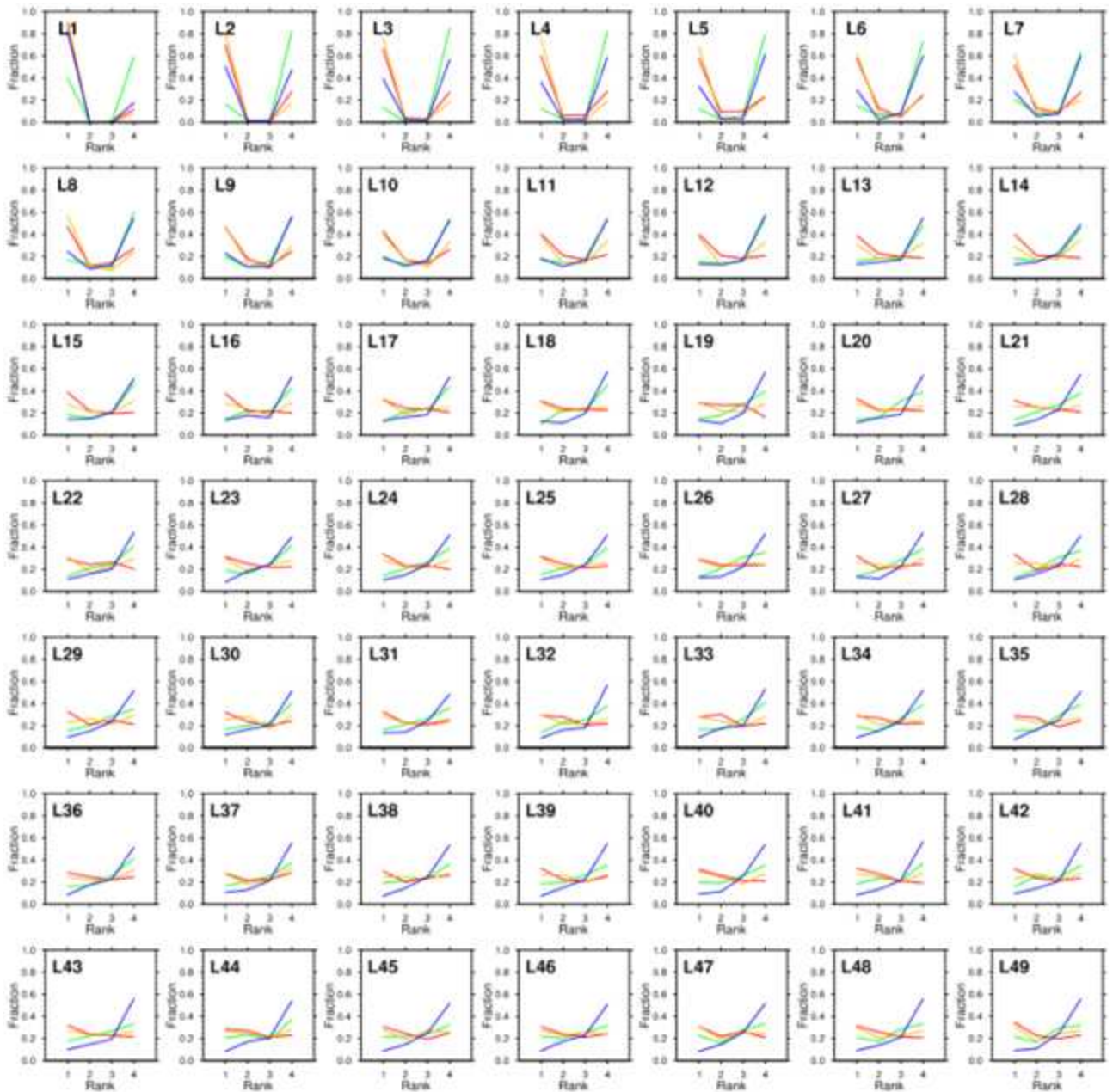


Figure 103. Same as Figure 102, but for each season, spring (green), summer (red), autumn (orange), winter (blue).

(나) 앙상블 신뢰도

각 계절과 리드타임마다 신뢰도 지수를 계산하였고 앙상블 유형과 함께 Figure 104에 타일 맵으로 정리하였다. 가장 이상적인 형태의 앙상블 유형인 평편한 히스토그램은 초록색으로 표시되고 GloSea5는 2주 이후 예보부터는 앙상블이 consistent함을 알 수 있다 (겨울철 제외). 앙상블이 골고루 잘 분포되어 있을 뿐만 아니라 신뢰도 지수 역시 20 이하의 좋은 값을 가지고 있었다. 리드 2주 이내의 기온은 앙상블이 underdispersion 하고 신뢰도 지수가 크게 도출 되어 그 분포가 관측을 모의하는데 한계가 있다고 보여 진다. 초기 쇼크의 영향으로 판단되므로 추후 충격을 최소화 할 수 있는 초기장의 개선이 요구되어지는 포인트이다.

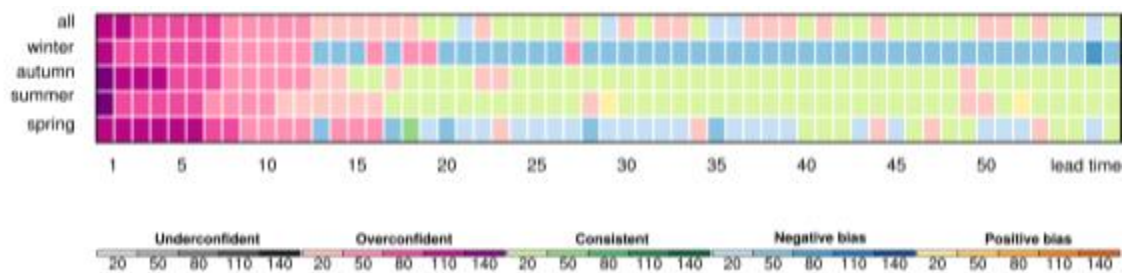


Figure 104. Tile map for ensemble type and reliability index in regard to 2m temperature over East Asia using GloSea5 hindcast. The color of tile indicates ensemble type and the depth of color means reliability index. The more reliable ensemble distribution has the lighter color.

동아시아 계절내 기온을 모의함에 있어 GloSea5의 현업 후보실험은 리드 초반에는 앙상블이 underdispersion하여 그 스프레드가 관측을 예측할 만큼 충분히 크지 않다. 그러나 2주 후부터는 앙상블 스프레드가 예보 불확실성을 표현하기에 충분한 consistent 유형이다. 단, 겨울철 앙상블은 예보 시작 약 2주 후부터는 Negative bias를 가지는 경향이 있다. 리드타임 2주 이후에는 약 20에 해당하는 낮은 신뢰도 지수 값을 나타내고 있어 GloSea5 앙상블은 계절내 규모에서 대체로 신뢰할만한 앙상블을 보유하고 있다고 보여진다.

따라서 GloSea5 앙상블에 대한 일관성과 신뢰도 진단 결과 예보 2주 이후 앙상블 자료들은 동아시아 기온 예보에 있어서 믿을 만한 수준으로 평가되므로 활용하기 좋은 앙상블 분포를 가지고 있다.

(다) 에러-스프레드 관계

Figure 105에서는 에러와 앙상블 스프레드의 관계를 나타내었다. 일반적으로 앙상블 스프레드가 에러보다 작으면 앙상블이 underdispersion 하다고 본다. Figure 104의 앙상블 유형에서 나타난 특징들이 Figure 105에서도 두드러진다. 모든 계절을 통틀어 분석한 결과를 보면 짧은 리드 타임에서는 에러가 스프레드보다 큰 값을 가지고 있어 앙상블이 관측을 표현하기에 작은 스프레드를 가지고 있다. 2주 이후 예보에서는 스프레드가 에러보다 커지게 된다. 이러한 경향

은 봄, 여름 보다는 가을, 겨울 후보실험에서 크게 나타난다. 봄과 여름 후보실험에서는 전체 리드타임에 걸쳐 스프레드가 에러보다 작은 경향이 있다. 이러한 계절적인 경향이 합쳐져 전체 에러-스프레드 관계에 영향을 주어 종합적으로는 앙상블이 리드 20일 정도까지는 underdispersion, 이후는 overdispersion 하는 것으로 나타나고 있다.

에러와 스프레드의 관계가 좋은 선형관계를 가지고 있으면 통계적으로 신뢰도가 높은 앙상블이라고 평가한다. GloSea5는 후보실험의 계절내 기온 앙상블은 스프레드가 커짐에 따라 에러가 증가하는 좋은 선형관계를 가지고 있다 (Figure 106). 일부 스프레드가 에러보다 작아 underdispersion 하는 구간이 있으나 이는 2주 이내의 짧은 리드타임의 영향이 반영되었기 때문으로 보여진다. 전체적으로 평균값이 대각선에서 크게 멀리 떨어져있지 않고 에러와 스프레드가 함께 증가하여 좋은 에러-스프레드 관계를 보유하고 있다. 이러한 이상적인 에러-스프레드 관련성은 스프레드가 에러를 추정할 수 있는 좋은 지표가 됨을 의미한다.

또한 스프레드의 공간적인 특징(Figure 107)을 살펴보면 리드가 길어질수록 고위도 대륙에서부터 점차 남쪽으로 스프레드가 증가하여 에러 역시 유사한 전파 경향을 가지고 있을 거라 추정할 수 있고 고위도 대륙에서 에러의 근원을 찾을 수 있을 것이라 추정된다.

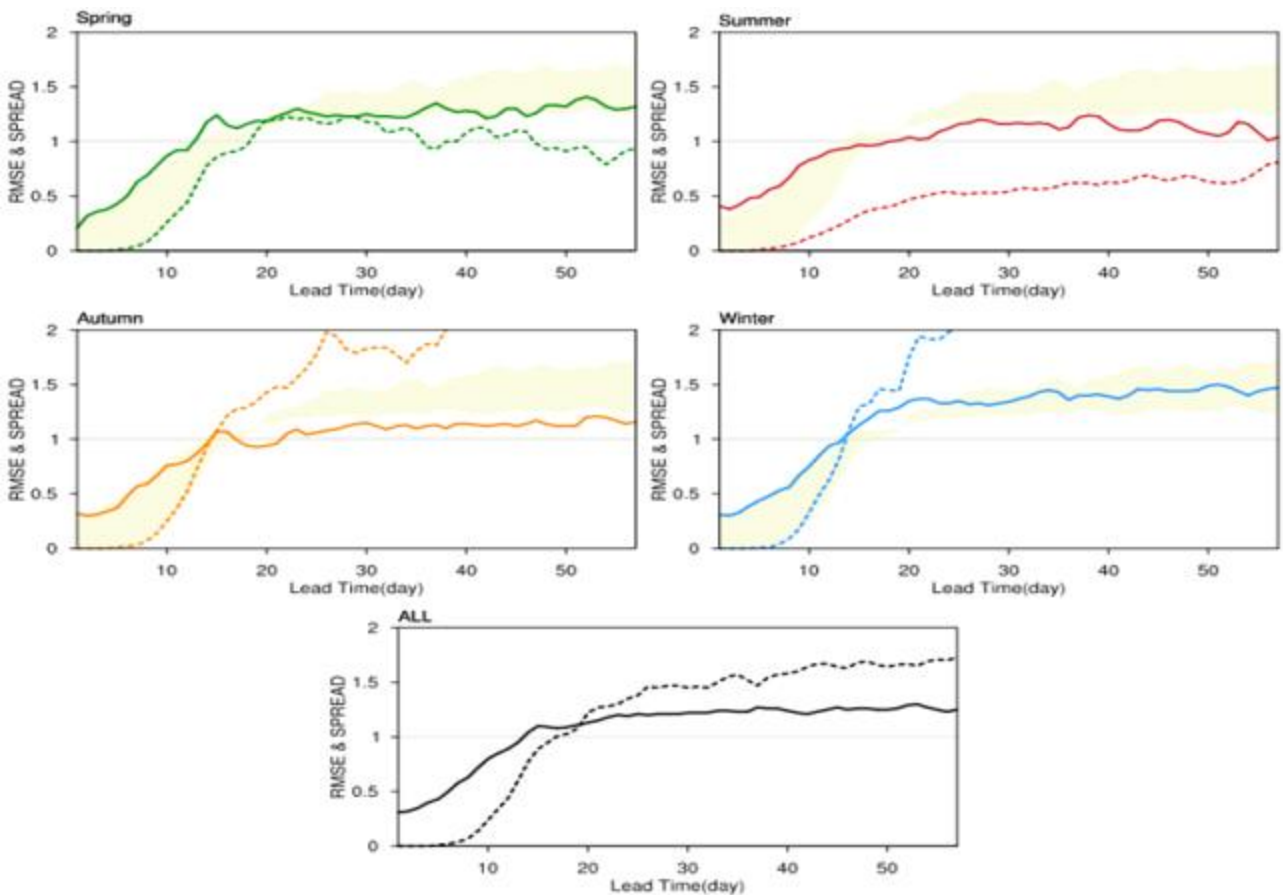


Figure 105. Root mean square error (solid line) and spread (dashed line) of ensemble for 2m temperature in GloSea5 hindcast.

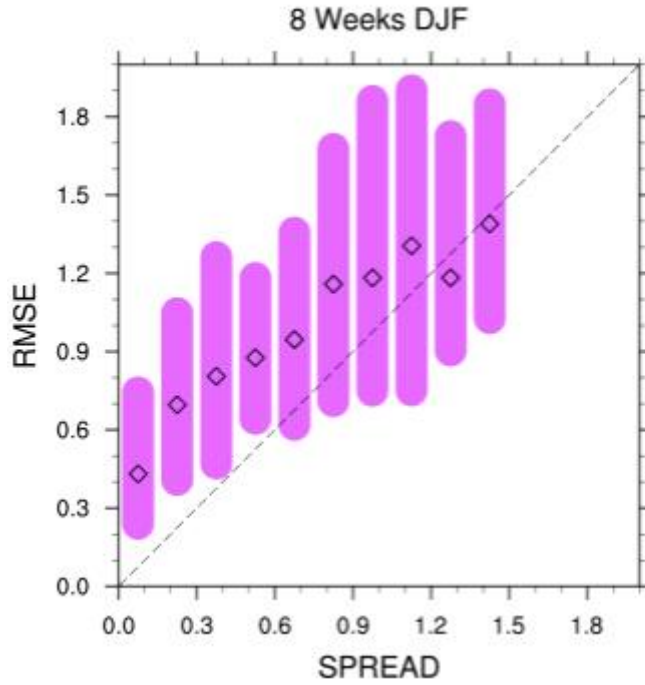


Figure 106. Error-spread relationship of winter(DJF) temperature hindcast with 1 to 8 weeks lead time in GloSea5.

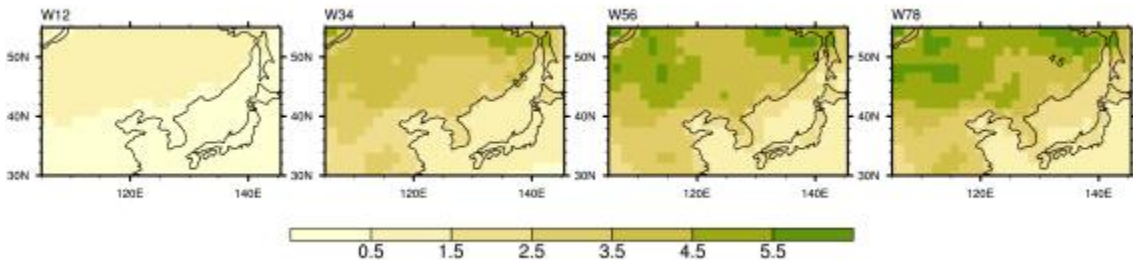


Figure 107. Spatial pattern of ensemble spread for winter (DJF) temperature hindcast with 1 to 8 weeks lead time in GloSea5.

(라) 불확실성 추정

앙상블 에러와 스프레드의 관계를 통해 좀 더 예측 가능한(불가능한) 상태를 만드는 대기 흐름(불확실성)을 표현할 수 있다. Figure 108은 flow-dependent spread 분석을 통해 나타난 GloSea5의 8주 이내 겨울철 기온 예측의 불확실성을 보여준다. 불확실성이 큰 구간(스프레드가 1.0이상인 구간)의 패턴을 자세히 보기 위해 Figure 109에서 따로 도식화하였다. GloSea5에서 동아시아 겨울철 계절내 기온 앙상블의 큰 스프레드는 동아시아를 포함한 유라시아 대륙의 저온 아노말리와 관련되어 있고 이 저온 아노말리는 우랄산맥 인근 상층 고기압성과 북동아시아 저기압성 아노말리와 연관된다. 몇몇 연구들에서 계절내 기온 예측에 있어서 모델에서의 눈-기

온 커플링 과정의 개선(Diro and Lin, 2020)과 현실적인 초기 토양 수분 및 눈 상태 처방(Koster et al. 2010; Jeong et al. 2013; Hirsch et al. 2014.; Orsolini et al. 2013; Thomas et al. 2016)이 중요한 역할을 할 수 있음을 시사한 바 있다. 따라서 GloSea5의 동아시아 기온 예보에 대한 불확실성을 줄이기 위해서 GloSea5의 고위도 지면/해빙 처방에서 개선을 하나의 방안으로 제안한다.

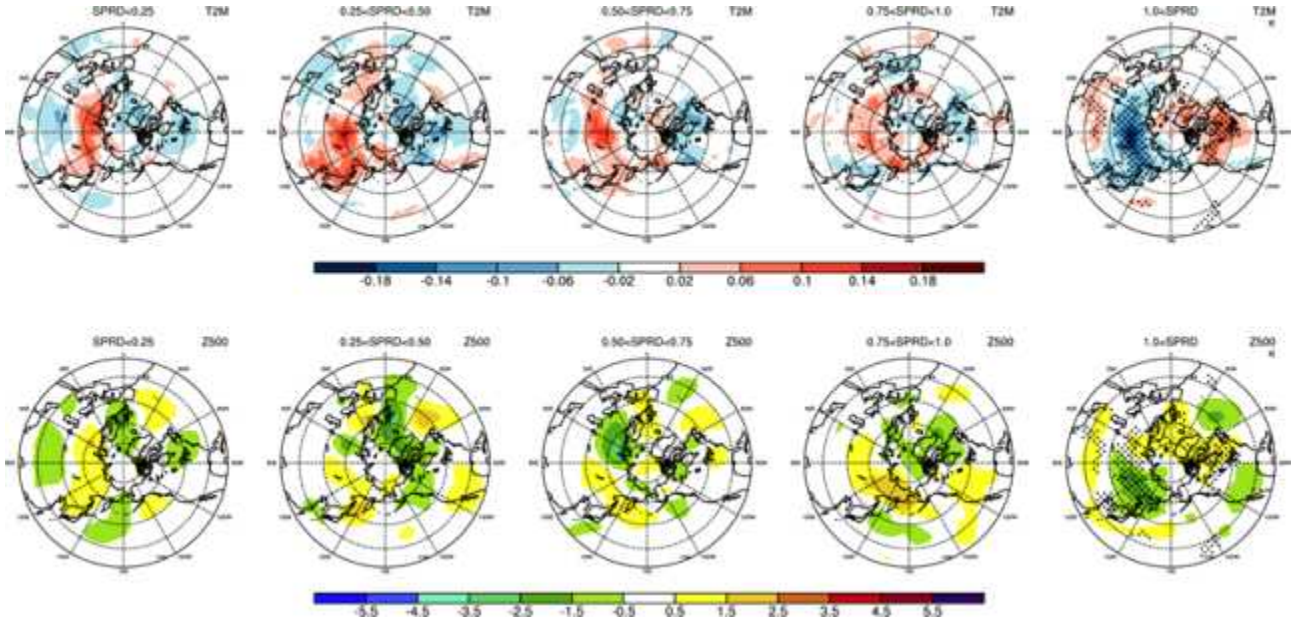


Figure 108. Uncertainty of winter temperature (upper panel) and geopotential height at 500hPa (below panel) with 8 weeks lead time in GloSea5. Uncertainty with 5 spread interval comes from flow-dependent spread analysis and spread increases from left to right.

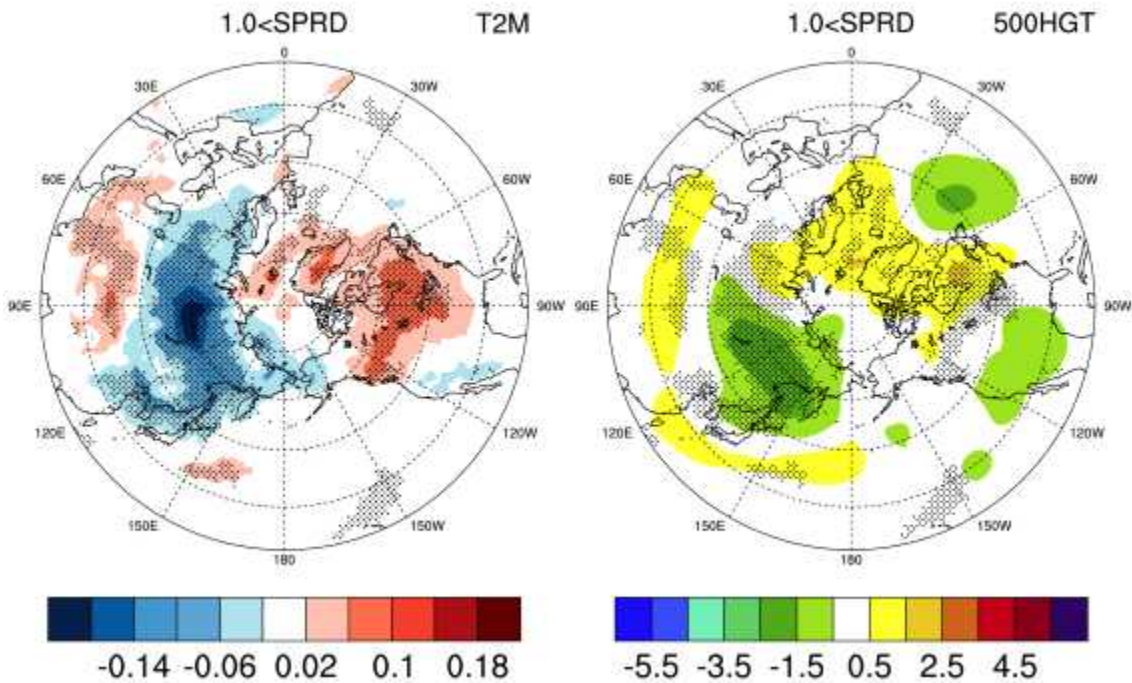


Figure 109. Uncertainty of winter temperature (left panel) and geopotential height at 500hPa (right panel) in the largest spread section.

(2) 앙상블 생산방안별 기온 예측 민감도 평가

이 연구에서는 동아시아 계절내 기온 예측에 있어서 앙상블 개수, 생산 일자, 생성 기법에 변화를 준 민감도 실험들의 정확도, 유용성, 활용도를 기존 현업 예보와 비교함으로써 기존 현업예보의 예측성을 향상시켜 줄 수 있는 앙상블 방안을 찾고자 하였다. 결정론적 예보에 대한 민감도 평가는 선행 연구에서 이미 수행된 바 있어 이 연구에서는 확률 예보의 예측 성능을 주로 비교하였고 그를 위해 평년 보다 낮음(below normal), 평년(normal), 평년보다 높음(above normal) 세 가지 카테고리로 확률을 계산하였다.

(가) 정확도 측면 평가

신뢰도 다이어그램은 확률 예보와 관측에서 해당 사건이 일어나는 빈도를 매칭 시킨 그림이다. Figure 110의 신뢰도 다이어그램은 현업 예측 실험에서의 동아시아 기온 예보의 정확도를 볼 수 있는데 8주 이내 예보에 대한 결과로 각 주의 결과는 색깔로 구분하였다. 예보가 정확할수록 검은 실선으로 표시된 대각선에 예측 신뢰도 선이 가까이 접근하고 신뢰도 선이 가로축과 평행하게 기울어질수록 resolution이 없는 예보로 간주된다. Figure 110의 색깔로 표시된 숫자는 각 리드타임의 brier score이다.

동아시아 기온 확률 예보는 적분 1주후에는 brier score가 0.13으로 낮고 신뢰도 선이 다이어그램의 대각선에 가까우나 전 확률 구간 안에서 예보가 관측에 비해 과대모의 하는 경향이 있다. brier score의 증가와 신뢰도 선의 기울기로 확인할 수 있듯이 리드가 길어질수록 예보의 정확도가 떨어지는 경향이 있다. 특히 3,4주부터 resolution이 크게 낮아지고 5,6주에 brier score나 resolution이 유지되다가 7,8주에 소폭 감소하는 경향이 있다. 3,4주 예보 개선에 대한 필요성을 암시하는 대목이다.

Figure 111은 각 카테고리에 대한 brier score이다. 공통적으로 리드 5일경 스코어가 크게 증가한뒤 이후 즉시 감소하였다가 리드 14일경 다시 증가하여 보름정도 유지된 뒤 리드 30일부터 리드 42일까지 낮은 값을 나타낸다. 평년 이하와 평년 이상 예보에 비해 평년 수준의 예보는 리드 타임의 변화에 따라 스코어의 변화가 그리 크지 않고 대략 0.2~0.25 사이에서 분포한다. 평년 이하와 평년 이상 예보에서는 적분 초기 1~3일은 상당히 좋은 예측을 모델이 생산

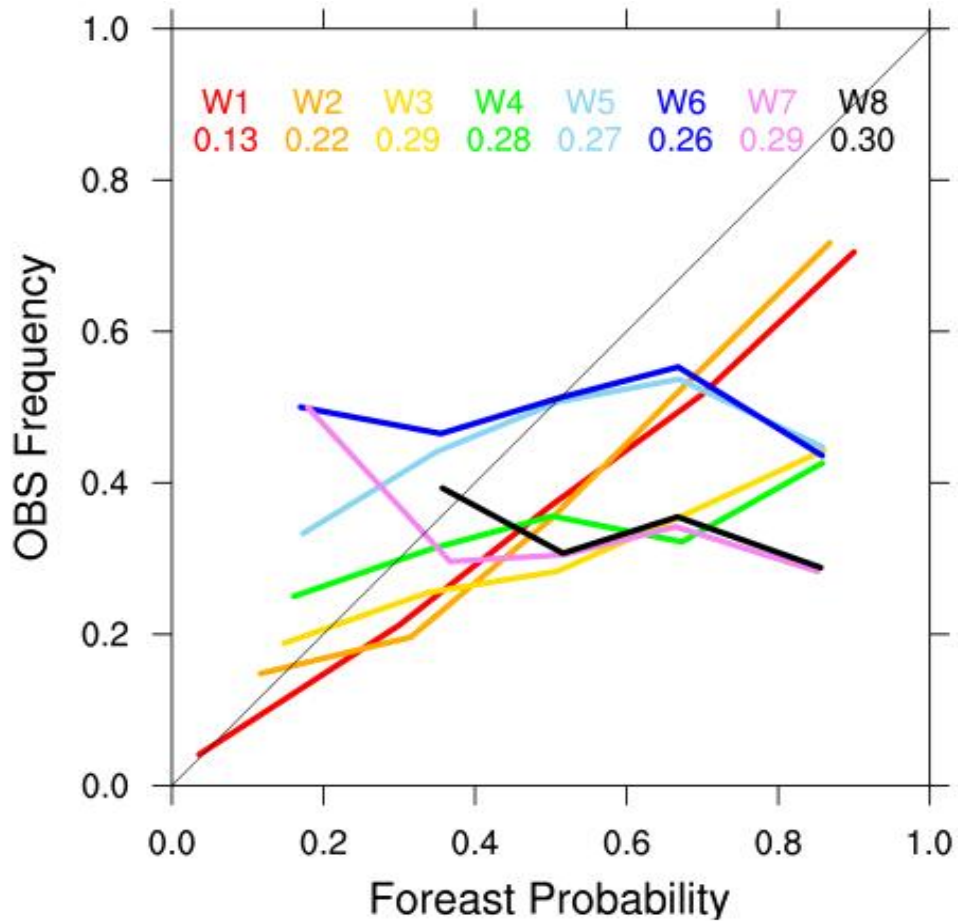


Figure 110. Reliability diagram for weekly probability forecast of 2m temperature over East Asia.

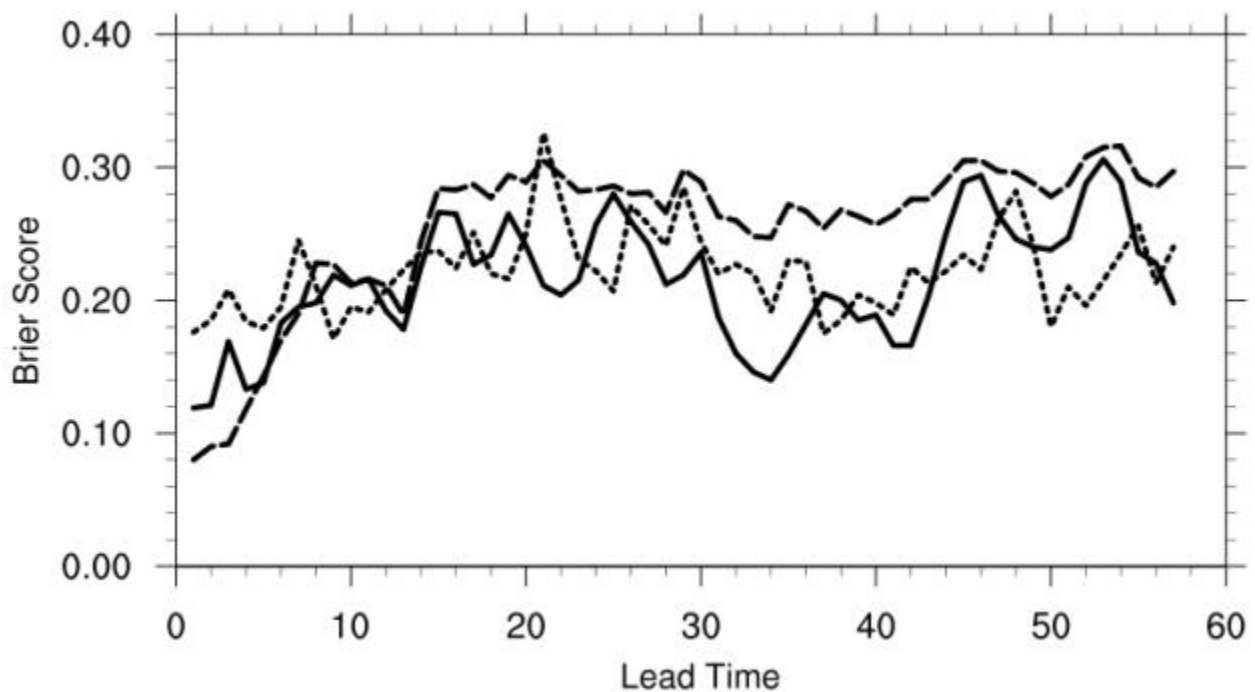


Figure 111. Brier score for daily probability forecast of EA 2m temperature in each category (below normal (short dashed line), normal (solid line), above normal (long dashed line)).

하나 이후 급격히 예측 정확도가 떨어지는 경향이 있다. 초기 5일 이내에서는 다른 카테고리 비해 평년보다 따뜻한 기온 예보가 정확도가 높지만 리드 3,4주에서는 오히려 평년보다 낮거나 비슷한 정도의 기온 예보가 더 정확도가 높은 것을 볼 수 있다.

예측 스킬은 주로 스킬 스코어로 표현이 되는데, 기준 예보 대비 개선된 비율로 해석한다. 본 연구에서 기준 예보는 현업 예보로 설정하였다.

$$SS_{ref} = \frac{(A - A_{ref})}{A_{perf} - A_{ref}} \times 100$$

brier score의 경우 완벽한 예보는 스코어가 0이 되므로 이 연구에서의 brier skill score는 아래식으로 정리될 수 있다.

$$BSS = \frac{(BS - BS_{ref})}{0 - BS_{ref}} = 1 - \frac{BS}{BS_{ref}} = 1 - \frac{BS_{EXP}}{BS_{OPER}}$$

Figure 112~115는 각 실험마다의 brier skill score를 비교한 그림이다. Figure 112에서는 현업 예보 대비 앙상블 멤버 수 증가와 생산 일자 변경의 효과를 볼 수 있다. 각 그림마다 막대 그래프가 기준선 대비 오른쪽으로 치우치면 민감도 실험의 효과가 있다고 해석할 수 있고 좌측으로 치우치면 현업 예보가 더 예측의 정확도가 높다고 해석할 수 있다. 왼쪽 그림은 앙상블 멤버 수 증가 실험과 현업 예보의 스킬을 비교하였는데 매일 6개로 앙상블 수를 증가시켰을 때 전체 리드타임의 72%에서 정확도가 개선이 되었다. 앙상블 생산 일자를 주중 1회 적분으로 변경했을 때(중간)는 전체 리드타임의 36%만이 개선이 되므로 원래 현업 적분 방식이 더 효과적이다. 주말 1회로 적분 일자를 변경(오른쪽)한다면 전체 리드타임의 절반 정도는 개선 효과를 볼 수 있고 절반 정도는 효과가 없으나 개선율의 크기로 본다면 주말런을 통해 현업 예보의 정확도를 누적 613% 개선할 수 있다. 주중런과 주말런을 통해서는 초기 10일은 개선 효과를 보기 힘들다는 특징이 있다. 개선폭이 가장 큰 실험은 주말런의 10일~17일 정도로 최대 30%의 개선율을 가져다줄 수 있다. 정확도가 개선된 일 수 측면에서는 멤버수 증가(72%) > 주말런(49%) > 주중런(36%) 순으로 효과를 기대할 수 있고, 개선율 측면에서도 멤버수 증가(238) > 주말런(226) > 주중런(-958) 순으로 효과를 기대할 수 있다. 예보 카테고리를 보자면 동아시아 기온이 평년 이상보다는 평년/평년 이하일 때 개선 효과가 있다.

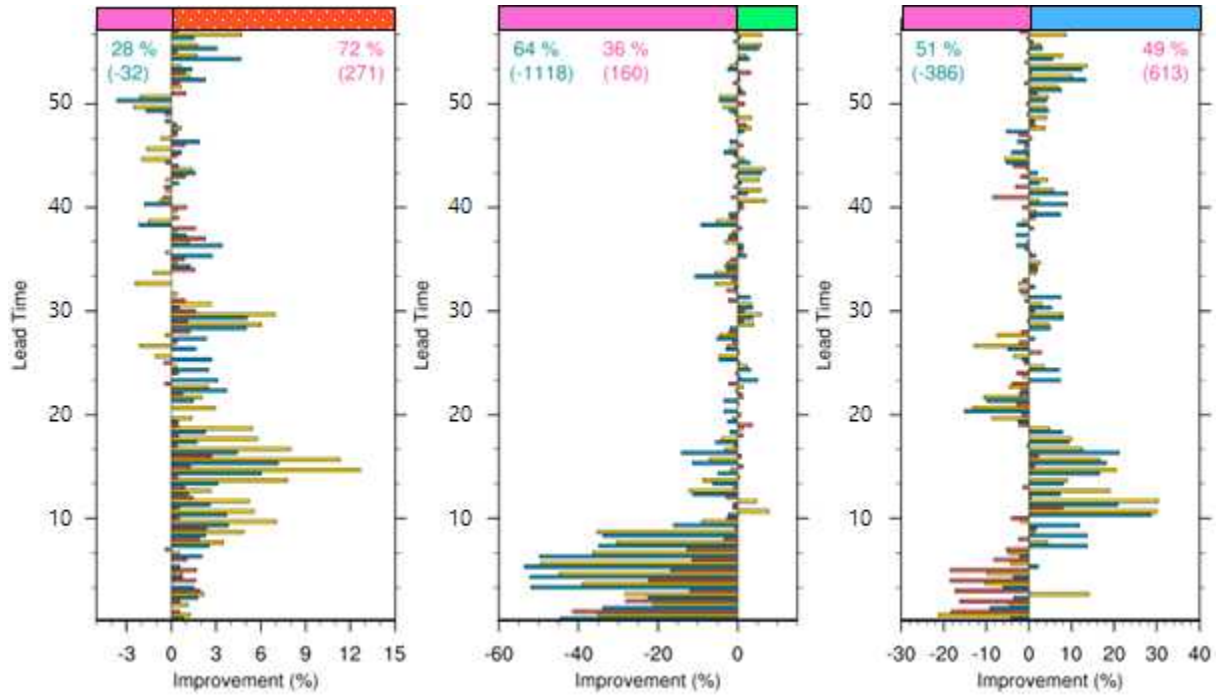


Figure 112. Brier skill score of 2m temperature probability forecast in each category (below normal (blue bar), normal (yellow bar), above normal (red bar)). The effect of increasing ensemble size (left panel), changing issuing date in weekdays (middle panel) and weekend (right panel) are shown.

양상블 수를 28개에서 42개로 증가시켰을 때 생산 일자별로 효과를 살펴보았다 (Figure 113). 왼쪽은 매일 생산, 중간은 주중 1회 생산, 오른쪽은 주말 1회 생산 실험에서 양상블 수를 증가시켰을 때의 결과이다. 예보 정확도 측면에서 매일 생산 시 양상블 증가 효과는 전 리드타임의 72%에서 개선이 되었으므로 상당히 크다고 볼 수 있다. 주중 1회 생산 실험에서 양상블 수의 증가도 효과가 있다. 그러나 주말 1회 생산 실험에서는 리드타임마다 그 크기가 달라서 8주 예보에 있어 양상블 수 증가에 대한 효과는 미미한 편이다. 양상블 수가 늘어나면 대체로 개선이 되고 최대 10%까지 개선 효과를 볼 수 있다. 정확도가 개선된 일 수와 개선을 두 측면 모두 매일런 > 주중런 > 주말런 순으로 효과를 기대할 수 있다. 매일 런하거나 주중에 런하면 양상블 수가 많을수록 유리하고 여기서도 평년보다 따뜻할 때보다 평년 혹은 평년보다 추울 때 양상블 증가 효과 있다.

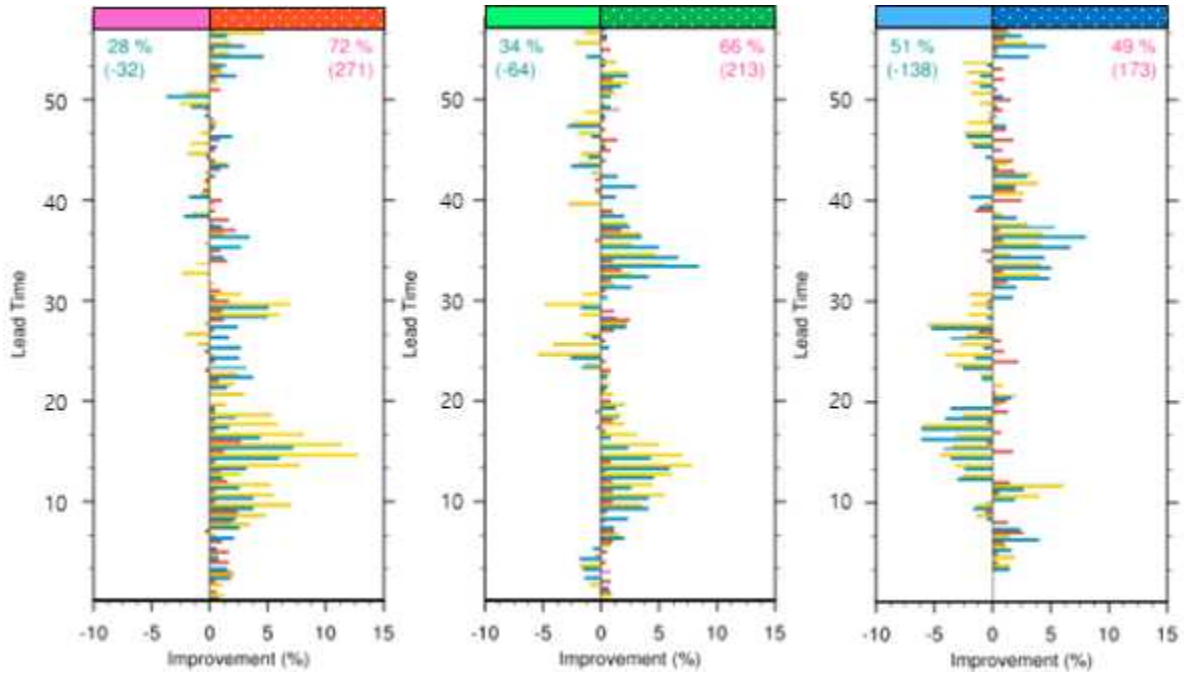


Figure 113. Improvement ratio of brier score for 2m temperature probability forecast in each category (below normal (blue bar), normal (yellow bar), above normal (red bar)). The effect of increasing ensemble size for the experiment issuing every day (left panel), weekdays (middle panel) and weekend (right panel) are shown.

그러나 현업 대비 개선율이 아닌 brier score 자체에 대한 민감도 실험 간 비교를 해보면 (Figure 114~115) 값의 차이가 겨우 0.1이 될 정도로 매우 작은 차이라는 점을 참고할 필요가 있다. 앙상블 수와 무관하게 리드타임 초반에서의 스코어 차이가 큰 편이고 리드 20일 이후에서는 실험간 스코어 차이가 거의 없다. LAF와 burst 기법을 병행하여 적분한 기온 예보는 burst기법 만으로 목요일에 런을 한 예보보다 좋은 정확도를 나타내고 월요일에 런을 한 예보와는 비슷한 정확도를 가진다 (Figure 114). burst기법 만으로 일주일에 한 번 적분을 한 실험인 주중런과 주말런을 상호 비교하자면 이 또한 앙상블 수와 무관하게 주말에 1회 적분을 한 예보가 더 좋은 정확도를 가진다 (Figure 115).

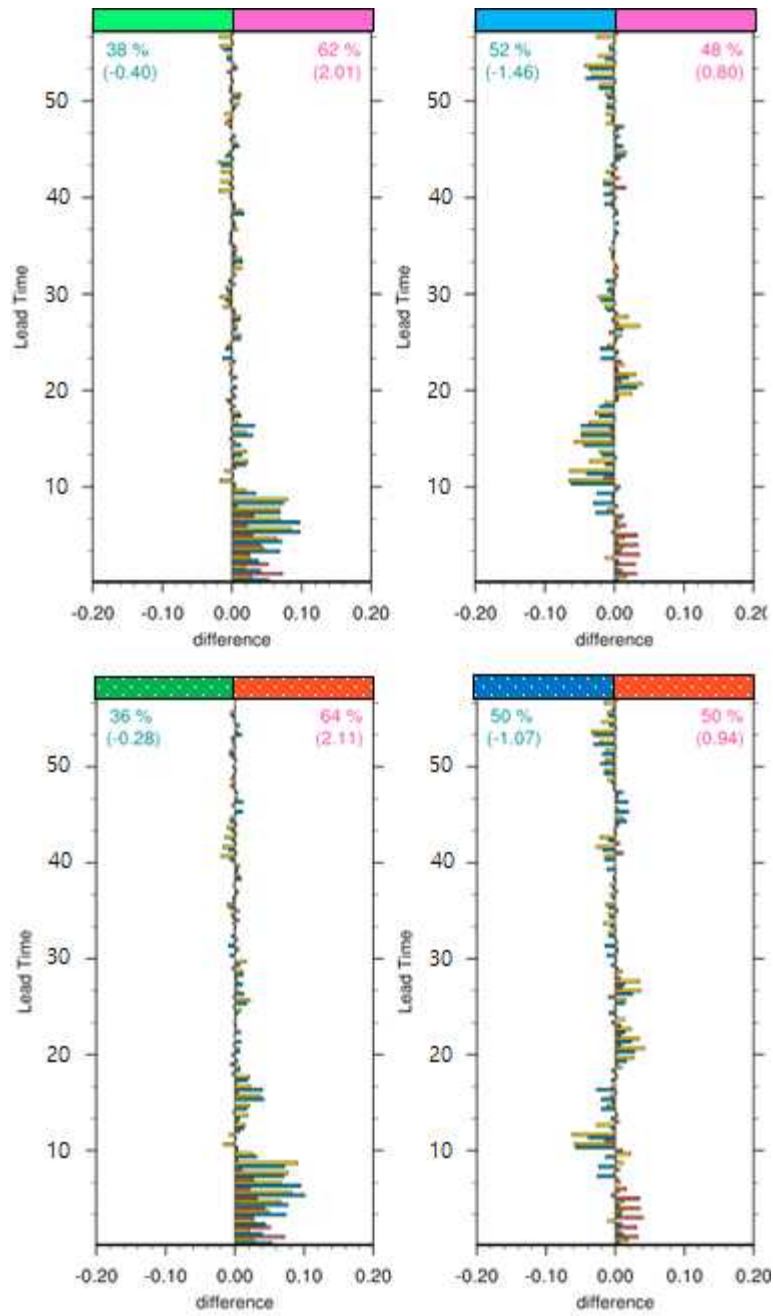


Figure 114. Difference of brier score between experiments regarding ensemble technique, LAF or LAF+burst. Upper panel is for 28 ensemble members and below panel is for 42 ensemble members.

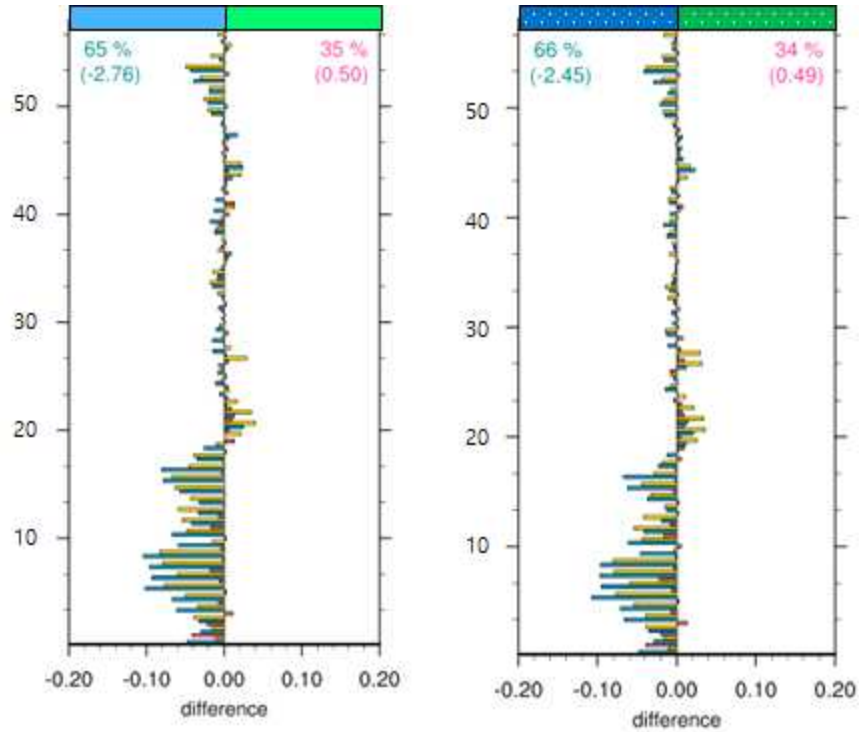


Figure 115. Difference of brier score between experiments regarding issuing date, Thursday or Monday.

확률 예보 정확도를 예보 카테고리 전체에 대해서 종합적으로 살펴보기 위해 ranked probability score (RPS)를 통해서도 살펴보았다. 이 경우 역시 완벽한 예보의 RPS는 0이므로 현업 예보 대비 RPS 개선율은 아래와 같이 정의하고 Figure 116에 도식화 하였다.

$$RPS_{improvement} = 1 - \frac{RPS_{EXP}}{RPS_{OPER}}$$

현업의 RPS는 0.2 정도의 값으로 적분이 시작되었다가 이후 5일 경 급격히 증가하여 0.4~0.5 수준의 값을 가진다. 리드가 길어질수록 정확도가 떨어진다. brier score의 결과와 마찬가지로 현업 시스템 대비 앙상블 수를 50% 증가시킬 경우 전체 리드타임의 77%에서 RPS 개선을 도모할 수 있었다. 현업 대비 burst 기법만을 쓴다면 주말런의 경우 현업 예보의 정확도가 향상되었고 주중런은 개선 효과를 볼 수 없었다. 특히 주중런의 경우 리드타임 초반 10일까지는 현업 예보에 비해 40% 정도의 정확도를 감소시키므로 단기 예보 전략 수립 시 고려해야할 점이다. 개선율의 크기 측면에서는 주말런이 멤버수 증가보다 더 효과적이고 개선된 리드타임 수 측면에서는 멤버 증가가 주말런보다 더 효과적이다.

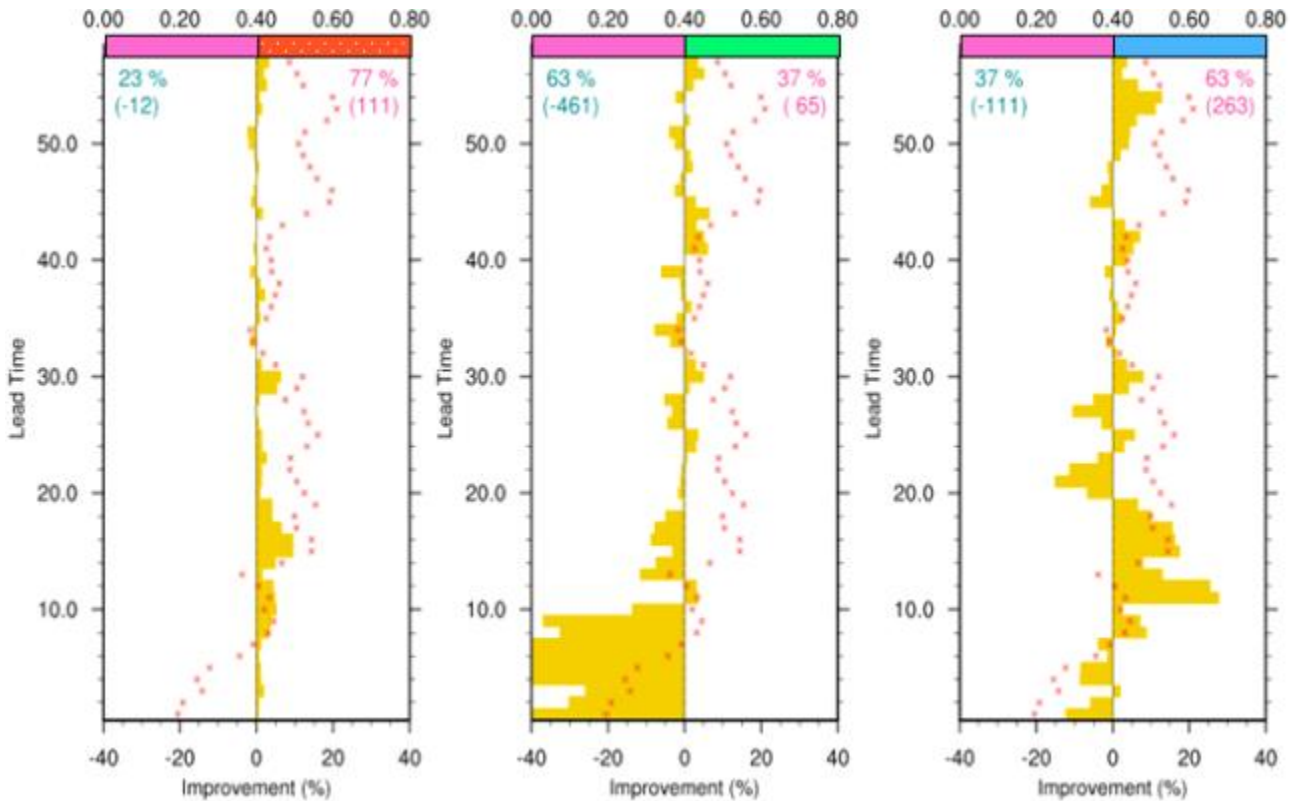


Figure 116. Same as Figure 112, but for ranked probability skill score (yellow bar). Red asterisk indicates RPS itself for OPER.

(나) 예보 유용성

Receiver Operating Characteristic Curve와 곡선 아래 면적(AUC)를 통해 예보의 분포가 사건과 무사건 사이에서 잘 분류되어 있는지를 평가함으로써 예보의 유용성을 살펴보았다. 이 경우 역시 기준 예보는 현업 예보로 두고 민감도 실험의 예보 유용성 개선율을 비교/평가 하였다.

$$Improvement\ Rate = \frac{(AUC - AUC_{ref})}{1 - AUC_{ref}} \times 100 = \frac{(AUC_{EXP} - AUC_{OPER})}{1 - AUC_{OPER}} \times 100$$

민감도 실험들의 ROC 다이어그램을 Figure 117에 제시하였고 리드타임에 따른 예측성의 변화 경향을 살펴보았다. 첫 번째 리드타임에서 민감도 실험들의 기온 예보 AUC는 0.88로 높은 예측성을 볼 수 있고 실험 간 예측 능력의 차이가 거의 없다. 리드타임 전반에 걸쳐 실험 간 커브의 모양과 AUC 값의 차이가 크지 않다. 또한 커브는 유용한 예보라 의미부여를 할 수 있는 마지노선인 다이어그램의 대각선과 리드가 길어질수록 점차 가까워지지만 대각선의 오른쪽 쪽 아래로는 치우치지 않고 AUC 또한 0.5 아래로 크게 벗어나지 않아 7주차 예보까지는 쓸 만 하다고 판단된다.

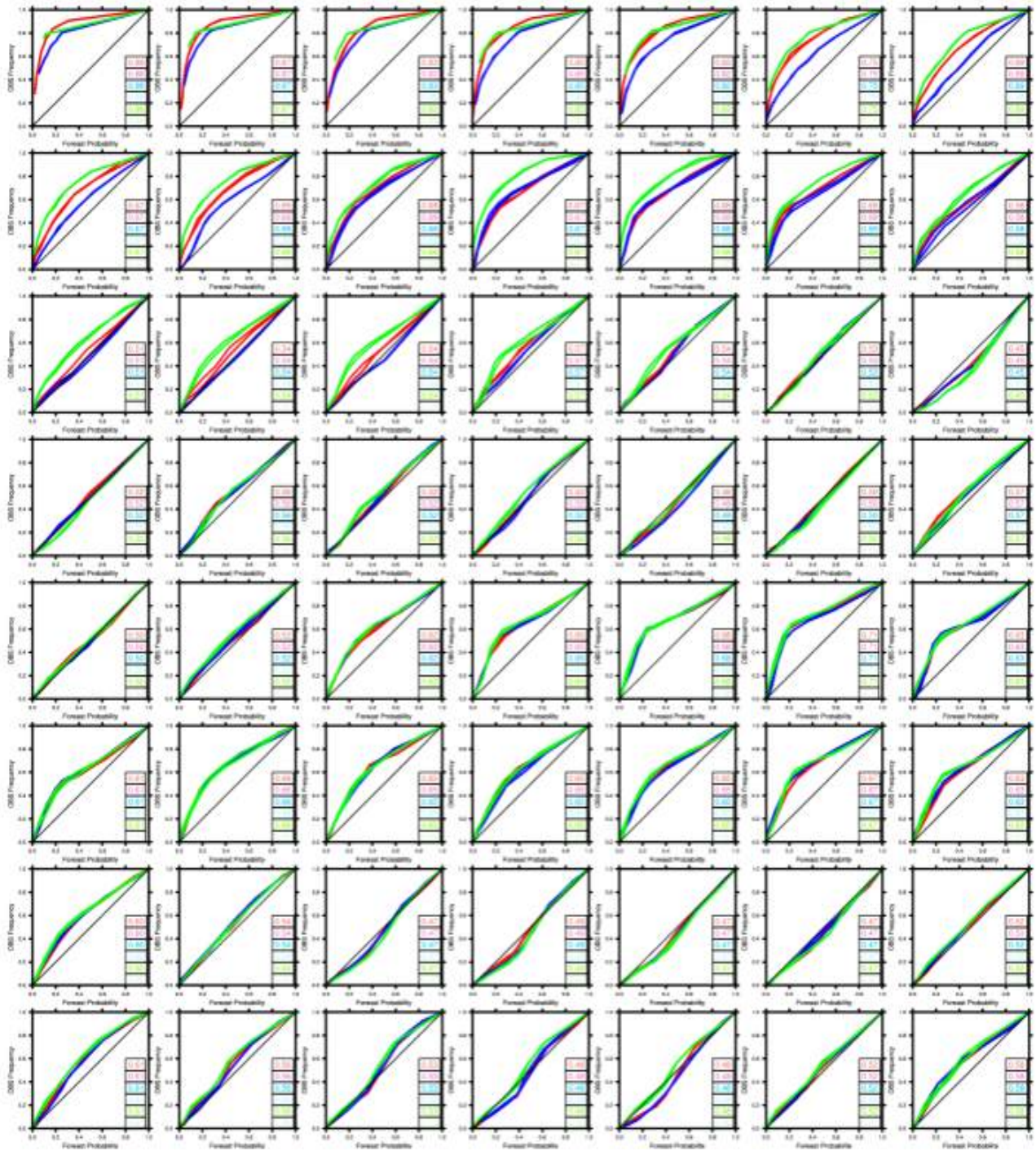


Figure 117. ROC diagram of 2m temperature probability forecast for from lead 1 day to 49 day (sequence from left to right and top to down). Red, green, blue colors indicate experiments initialized in everyday, Thursday, Monday, respectively. Solid and dashed lines are for 28 and 42 ensembles. Values in boxes are AUC for sensitivity experiments.

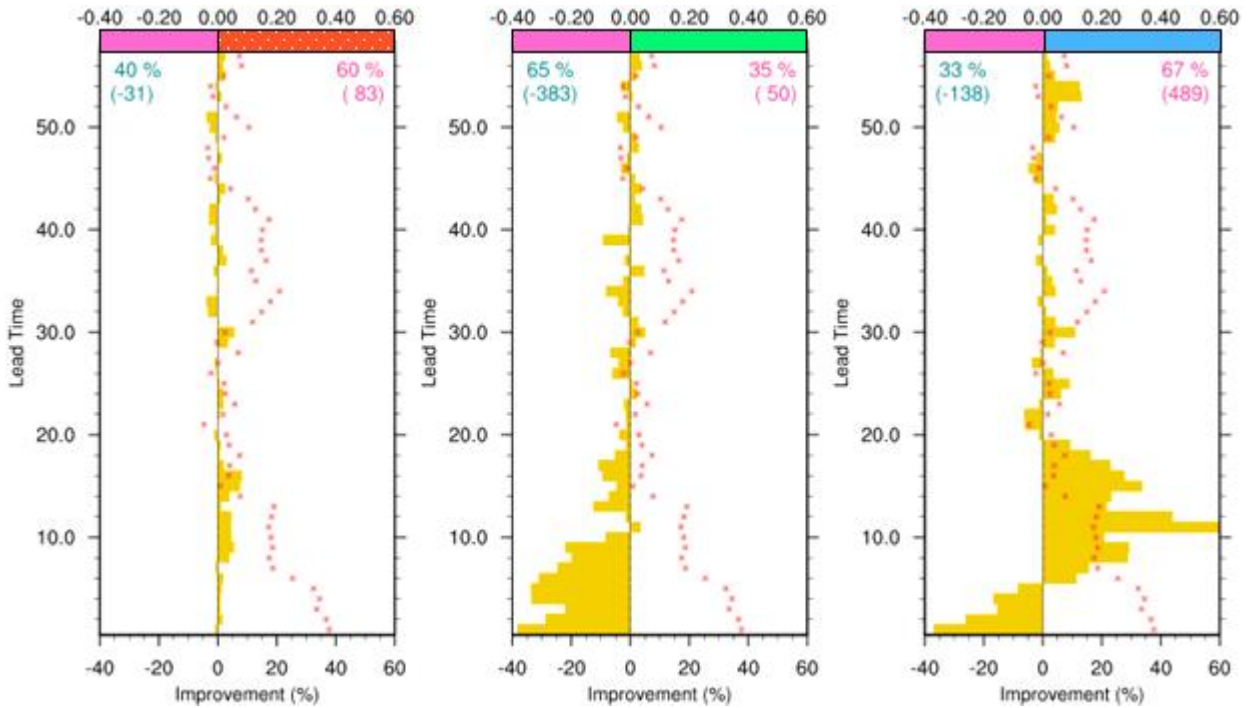


Figure 118. Improvement ratio for AUC of 2m temperature probability forecast (yellow bar) and the value of 0.5 subtracted from AUC itself for OPER (red asterisk). The effect of increasing ensemble size (left panel), changing initial date for weekdays (middle panel) and weekend (right panel) are shown.

Figure 118에서는 현업 예보 대비 앙상블 방안 변화에 따른 AUC의 개선 정도를 나타내었다. 현업 예보의 AUC는 0.4로 매우 유용하나 시간의 흐름에 따라 0에 가깝게 떨어지다가 30일 이후 다시 회복하는 경향이 있다. 적분 초기에는 현업 예보의 유용성 자체가 원래 상당히 좋기 때문에 앙상블 수를 증가하거나 생산 일자를 변경하여도 쉽게 개선이 되지 않는다. 그러나 예측성 개선이 필요한 2주 이후 예보에 대해서는 주말런을 통해서 약 10%~20% 유용성 개선을 도모할 수 있다. 앙상블 수 증가효과는 유용성 개선 관점에서는 5% 미만의 개선을 가져다줄 수 있고 주중런은 유용성 개선에 긍정적인 효과를 가져다주지 못한다.

Figure 119는 민감도 실험들의 예보 정확도와 유용성 그리고 현업 예보 대비 개선율을 종합적으로 보여준다. 기본적으로 현업 예보의 경우 0.5이상의 AUC 값을 보이므로 대체로 8주까지 유용한 예보를 생산해내고 있다. 유용성과 정확도 평가에서 4주(28일) 이후의 예보에 대해서는 민감도 실험간 유용성의 차이가 크지 않다. 초기 2,3주에서의 현업 예보의 스킬 저하는 주말런을 통해 향상시킬 수 있다. 현업 예보 대비 민감도 실험의 개선정도를 살펴보면 주말런을 통해 적게는 5%에서 많게는 45%(예보 2주차)의 유용성 개선을 도모할 수 있다.

기본적으로 현업 실험의 경우 대체로 8주까지 유용한 예보를 생산해내고 있고 그 중 3,4주 예보는 가장 개선이 필요한 기간으로 평가되었다. 현업 예보의 계절내 예측 스킬을 주말런(정확도 ~15%, 유용성 ~45%)을 통해 향상시킬 수 있었고 기존 앙상블 생산 방법을 유지할 경우 앙상블 수를 증가시킴으로써 예측성 개선에 일부 효과(정확도와 유용성 5% 개선)를 볼 수 있다. 그러나 주 1회 적분에서는 앙상블 크기 증가 효과가 미미하므로 예보 유용성과 정확도 개선을 위해서는 주중런은 권장하지 않는다.

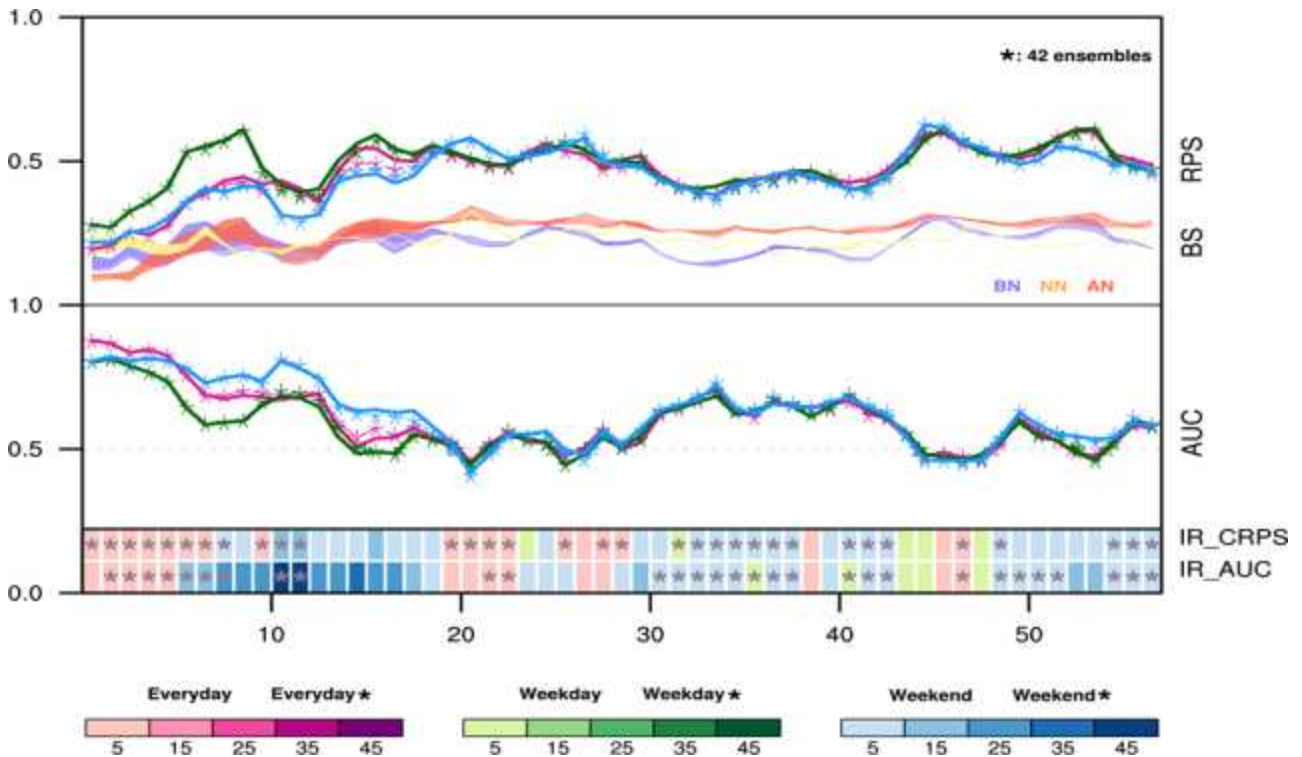


Figure 119. Forecast accuracy (brier score, ranked probability score, improvement ratio of continuous ranked probability score) and useless (area under ROC curve, improvement ratio of area under curve) for 2m temperature probability forecast from lead 1 day to 57 day. Experiments for issuing date are distinguishable by color. Results on experiments with 42 ensemble members has asterisk. Each box in the below tile map shows an experiment with the maximum improvement ratio among 6 experiments along with an amplitude of the ratio.

계절내 기간 중 예측 성능의 개선이 요구되는 예보 3,4주차에 대하여 좀 더 면밀히 살펴 보기위해 동아시아 3,4주 기온 예측에 대한 민감도 실험 예보 개선 정도를 Figure 120에서 비교하였다. 앙상블 멤버수가 증가하면 리드 타임 전반(정확도 기준 93%, 유용성 기준 64%)에 걸쳐 스킬이 증가한다. 개선율은 정확도 기준으로 3.2% 증가하였고 유용성을 기준으로 1.5% 증가하였다. 주중 1회 런 중 주말에 런을 하면 전체 14일의 리드타임 중 절반 정도(정확도 기준 50%, 유용성 기준 57%)의 리드타임에서 스킬이 증가하였고 이는 앙상블 수 증가 실험에 비해 개선율의 폭이 넓다. 개선율은 정확도 기준 1.4% 증가하였고 유용성을 기준으로 5.8% 증가하였다. 주중에 런을 하는 것보다 현재의 현업 실험을 통해 전체 리드타임 중 약 80%의 리드타임에서 스킬 개선을 도모할 수 있다. 개선율 역시 현업 대비 약 3%(정확도 -2.5%, 유용성 -3.8%) 감소되었다. 예보 카테고리 별로 살펴보자면 대체로 평년 이하와 평년 이상에서 개선 효과가 있었다. 단, 좀 더 안정적인 결과 도출을 위해 추후 다수의 예측장을 생산하고 검증하여 통계적 유의성을 확보할 필요가 있다. 이 연구에서는 분석 영역의 모든 공간 격자점을 사용하는 방식을 취함으로써 예측장의 수가 작은 문제를 극복하고자 하였고 이는 선행 연구들에서 샘플의 크기가 작을 때 종종 사용되는 접근법으로 (e.g. MacLeod et al. 2018, Christensen et al. 2015, Haiden et al. 2016) 결과를 신뢰하기에 충분한 방식이다.

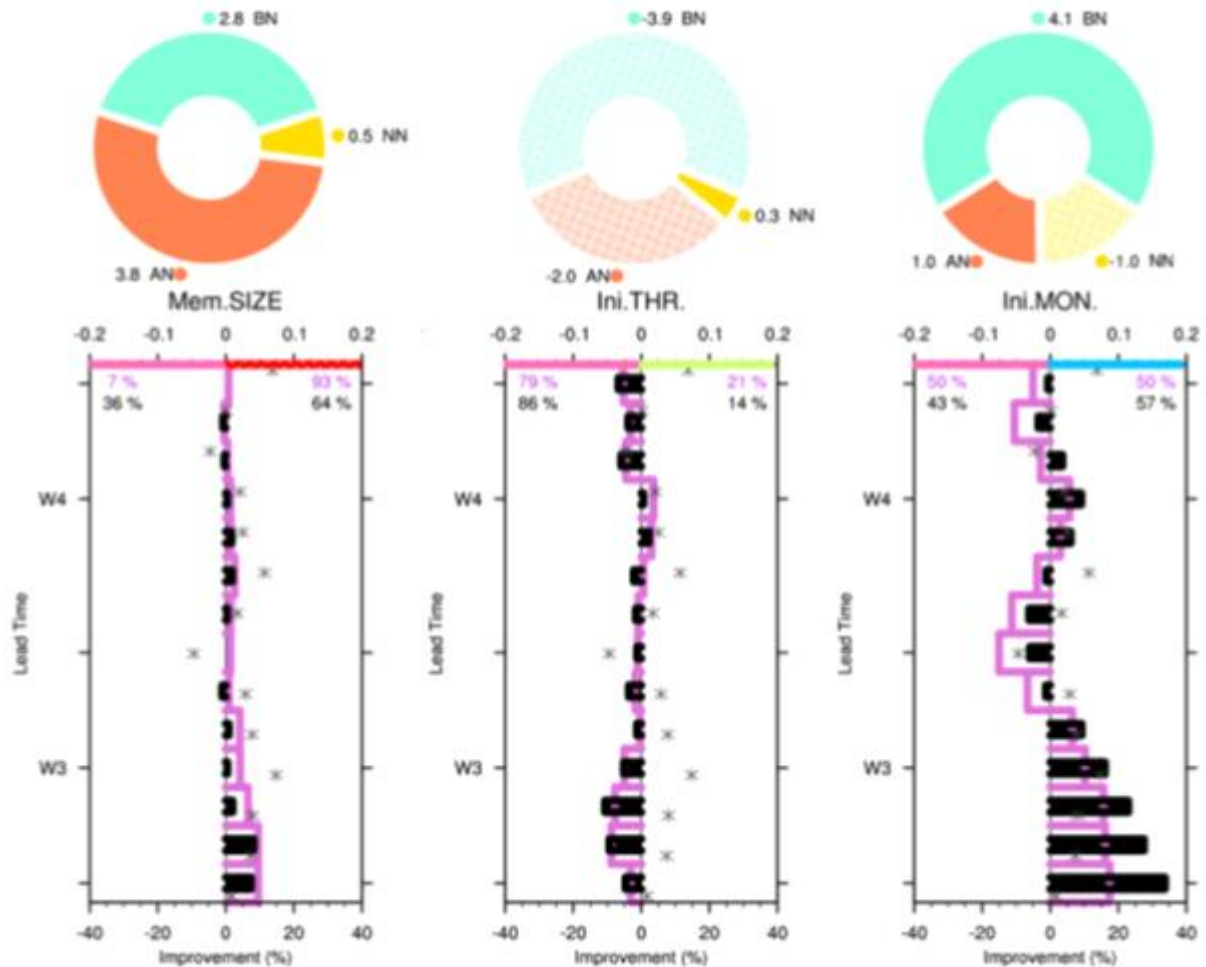


Figure 120. Improvement ratio for 2m temperature probability forecast in week3 and 4. Purple and black bar indicates improvement of RPS and AUC, respectively. Gray asterisk is the value of AUC minus 0.5 and the value over 0 means the forecast is useful. The improvement of brier score for each forecast category (below normal, normal and above normal) is shown in the upper panel as a pie chart. The hatched area in the pie chart means there is nothing to improve in the category comparing to OPER.

(다) 활용도 측면 평가

이 절에서는 Economic Value(EV)와 Rate of Return(ROR)를 통해 예보 3,4주 차의 동아시아 기온 예측에 대한 활용도를 살펴보았다. EV는 비용과 손실에 근간하여 예보의 가치를 진단하는 방법이고, ROR은 주어진 확률로 예측에 투자할 경우 가질 수 있는 수익률을 나타낸 지표로서 두 진단 방법을 통해 예보를 활용하는 사용자 중심에서의 이득을 가늠할 수 있다.

Figure 121에서는 8주 예보에 대해서 비용과 손실의 비율에 따른 각 민감도 실험별 EV를 도식화하였다. 현업 예보(붉은색 음영)를 사용할 시 0.3의 C/L에 해당하는 예보 사용자는 완벽한 예보가 가지는 가치의 약 55%에 해당하는 이득을 얻을 수 있다. 그러나 0.1 이하 그리고 0.8 이상의 C/L에 해당하는 사용자는 현업 예보로부터 이득을 얻을 수 없다. 초기 시간을 주 1회 월요일로 두었을 경우 대부분의 사용자에게 있어서 현업 예보에 비해 대략 5%의 이득을 줄

수 있는 반면 적분 시점이 목요일인 경우에는 약 10%의 이득이 감소하게 된다. 앙상블 수 증가는 EV 증감에 큰 영향을 주지 않는 것으로 보인다. 리드타임별로 6개의 민감도 실험 중 사용자에게 가장 큰 이득을 가져다주는 실험을 확인하기 위하여 타일맵을 활용하였다 (Figure 122). 리드가 증가할수록 사용자 혜택 또한 줄어드는 경향이 있다. C/L이 0.3~0.5에 해당하는 사용자가 계절내 기온 예보 활용 시 많은 혜택을 얻는다고 볼 수 있는데 그 가치의 정도가 주말런에서 리드 2주 차까지는 40~50%에 육박한다. 이후 10% 미만으로 EV가 감소하다가 리드 5주 차가 되면 약 10% 더 증가한다. 리드타임과 사용자 구간에 따라 일부 차이는 있으나 대체로 월요일에 예보를 시작한 경우 많은 사용자에게 이득이 주어진다고 볼 수 있다.

Figure 123을 통해 4주 예보에 있어서 현업 예보에 비해 EV가 개선된 정도를 정리해보았는데 월요일에 예보를 시작할 때 사용자(C/L이 0.3인 경우)들은 현업 예보 보다 더 나은 기온 예보를 받아보게 되고 종전 대비 약 8%의 경제적 이득을 가질 수 있다. 반면 목요일에 예보를 시작할 경우 현업 예보를 활용할 때 보다 4%~8% 정도의 경제적 손실을 가지게 된다.

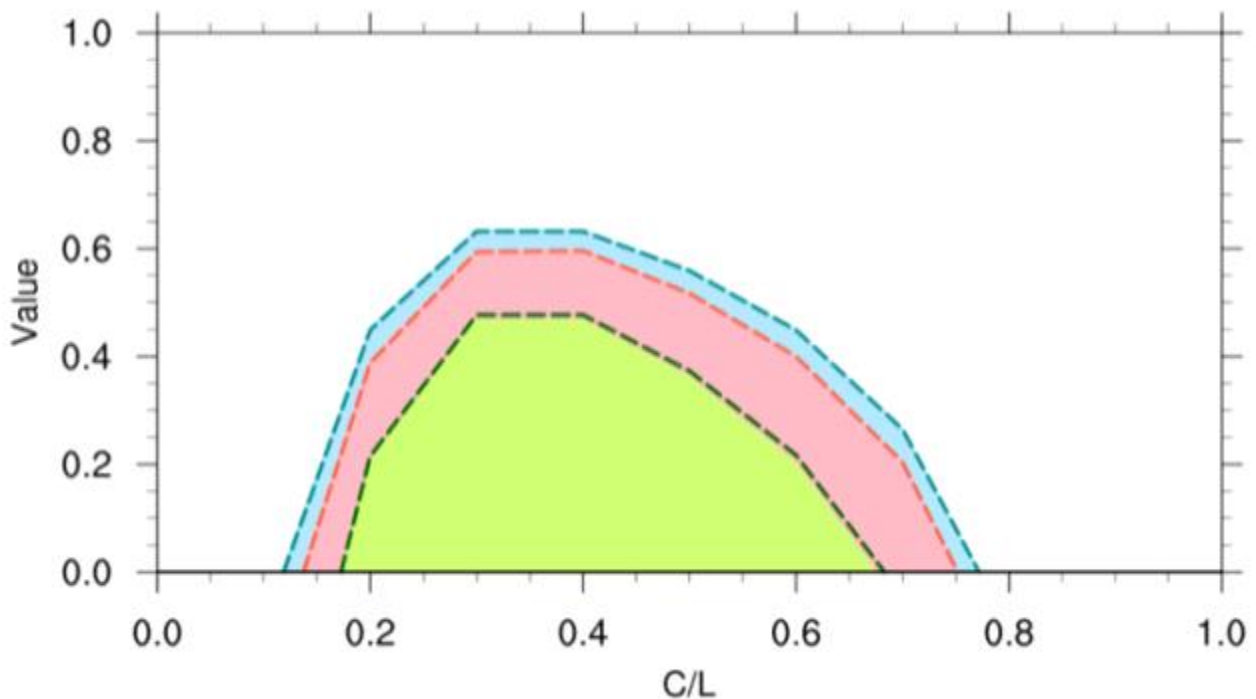


Figure 121. Economic value for 2m temperature probability forecast in 6 sensitivity experiment. Red, green, blue colors indicate experiments initialized in everyday, Thursday, Monday, respectively. Results on experiments with 42 ensemble members has dashed line.

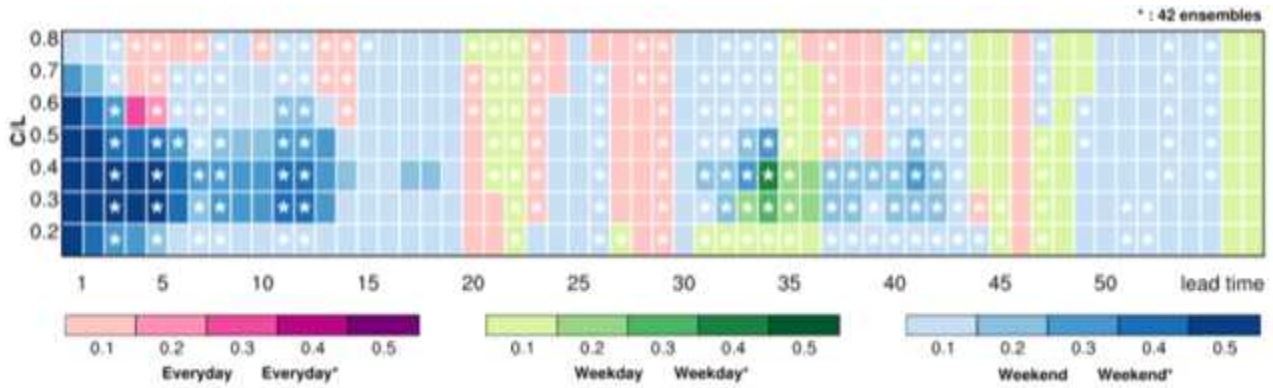


Figure 122. Experiment with the largest economic value for 2m temperature probability forecast from lead 1 day to 57 day and the value in that case. Experiments for issuing time are distinguishable by color. Boxes with asterisk are for the experiments with 42 ensemble members.

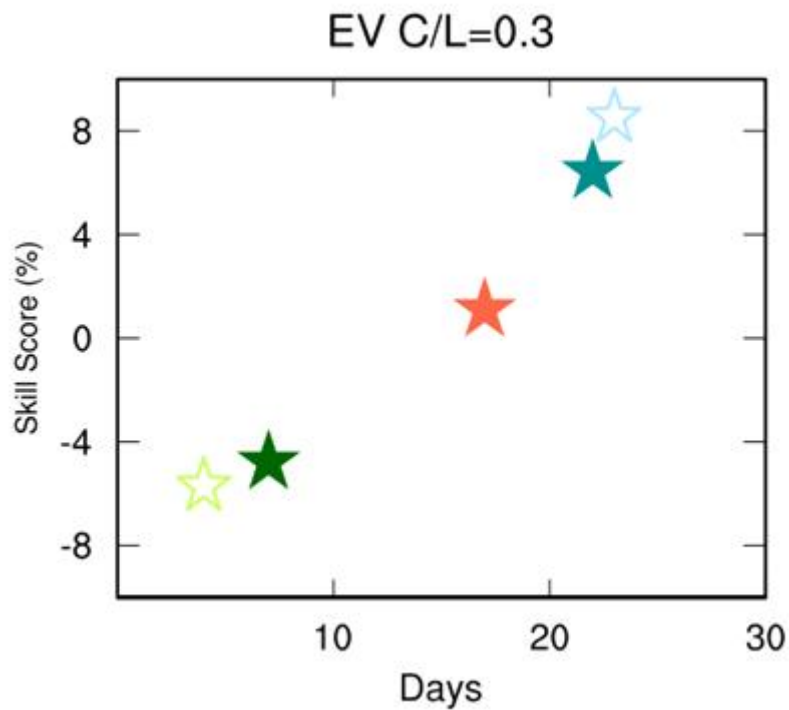


Figure 123. Skill improvement and the number of days with improved skill of economic value within eight weeks forecast comparing to OPER. Red is for E3. Green and Blue is for Thursday and Monday run. Open (closed) stars indicates the experiment with 28 (42) ensemble members.

예측 성능의 개선이 요구되는 예보 3,4주 차에 대한 기온 예측 활용성을 Figure 124를 통해 또 살펴보았다. 예보 3주차에서 현업 시스템의 동아시아 기온은 0.3의 비용과 손실의 비(C/L)를 가진 사용자에게 가장 가치가 가장 높다. 이 사용자들에게 있어서 현업시스템은 완벽한 예보의 약 20%의 가치에 달하는 예보를 생산한다. 이 3주 차 현업 예보는 주말에 런을 한다면 5~9% 더 많은 가치를 얻을 수 있고 3주에 해당하는 7일 중 절반의 리드타임에서 더 나은 이득을 가져올 수 있다. 또한 예보 4주 차의 동아시아 기온에 대해 우리가 현재 현업 시스템을 가지고 예측을 한다면 기후학적 분포로 판단하는 것보다는 10~20% 더 많은 예측 가능성을 가지고 있다. 그런데 주말런을 활용한다면 이 현업 시스템 보다 약 4% 더 많은 수익률을 가질 수 있고 리드 타임 7일 중 거의 대부분에서 더 나은 예측이 가능하다.

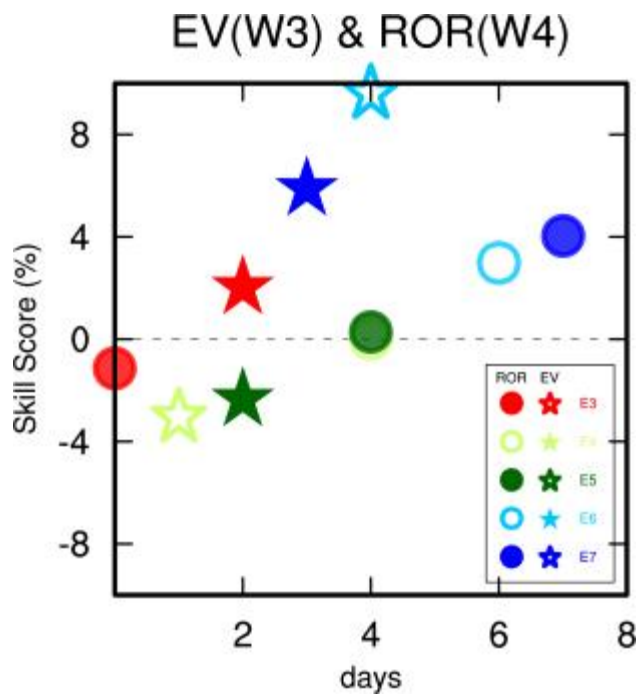


Figure 124. Skill improvement and the number of days with improved skill of economic value for W3 and rate of return for W4 comparing to OPER.

(라) 사례 분석

이 절에서는 극한 기상에 대해서도 예보를 개선시킬 수 있는 앙상블 방안이 존재하는지 살펴보고자 2017년 12월 중순 한파에 대한 사례분석을 하였고 Extreme Forecast Index(EFI)를 통해 진단하였다. 2017년 12월은 11월 중반에 시작된 음의 북극진동이 12월 중반까지 지속되면서 북극의 찬 공기가 중위도로 남하하기 좋은 조건이었으며, 우랄산맥-카라 해 부근에 형성된 상층 고기압이 정체하여 상층의 찬 공기가 우리나라 부근으로 지속적으로 유입되어 기온이 평년보다 낮은 날이 많았고 특히 11일~14일에는 찬 대륙고기압이 일시적으로 강하게 발달하여 기온이 큰 폭으로 떨어졌다(Figure 125). 이 시기는 2017년 11월 20일을 초기장으로 사용한 예

보의 리드 4주차에 해당한다. EFI로 살펴본 결과 이 한파는 기후학적으로 하위 10%~20%에 해당하는 추위였으나 현업시스템에서는 -10%의 비정상성만 나타내었다. 그러나 주말런은 -30%까지의 비정상성을 보였다.

EFI는 주로 Hit Rate(HR)와 False Alarm Rate(FAR)로 검증을 한다. 주말런을 통해 한파에 대한 4주 예보는 현업 대비 HR은 0.48에서 0.67로 증가하였고, FAR은 0.46에서 0.38로 감소하여 예보관 입장에서 주 1회 주말런은 극한 기상 발생에 활용하기 좋은 앙상블 방안으로 보인다 (Figure 126).

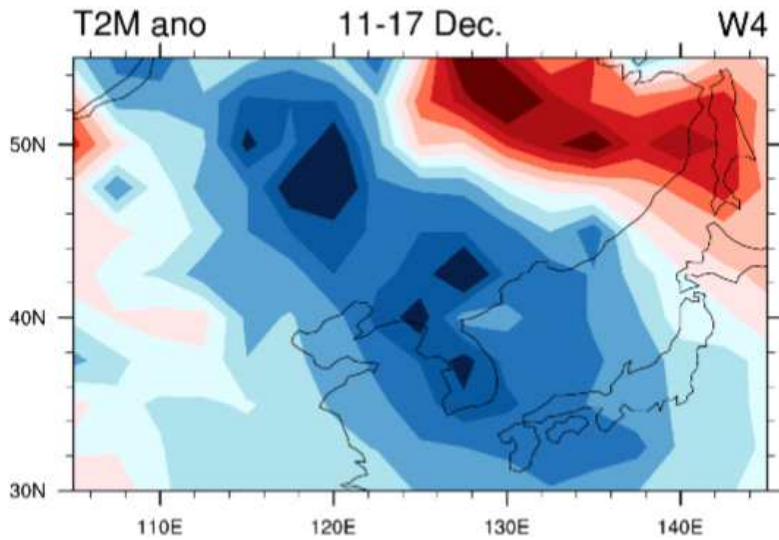


Figure 125. Spatial pattern of temperature anomaly for a cold surge on the third week of December in 2017.

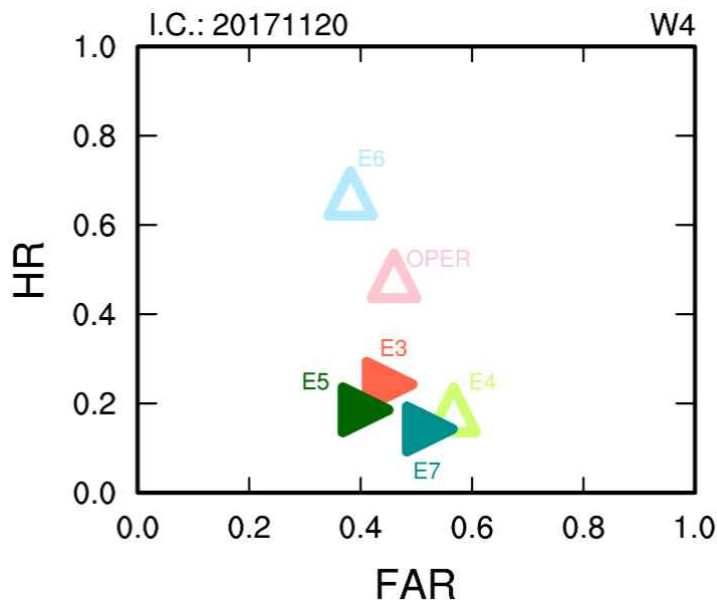


Figure 126. Scatter plot for hit rate and false alarm rate of temperature for the cold surge forecasts in each sensitivity experiment.

동아시아 3,4주 기온 예측에 대한 민감도 실험 간 예보 개선 정도를 종합적으로 비교해보면 Figure 127로 나타낼 수 있는데, 다양한 평가 기법을 통해 산출된 예보 스킬의 개선율과 개선된 예보일 수를 앙상블 크기 증가와 생산 일자 변화의 측면에서 정리하였다. 각 box의 오른쪽의 막대에는 사용된 예보 스킬들이 나타나고 그중 짧은 막대들은 주중런의 스킬이다. 종합 분석 결과 예측성 개선이 필요한 3,4주 기온 예보에 대해서 앙상블 크기를 증가시킬 경우 현업 예보 대비 최대 약 3%의 스킬 개선과 최대 13일, 평균 6일의 예보일 개선을 기대할 수 있다. 매일보다는 일주일에 한 번 예보를 생산하고 그것이 월요일이라면 최대 6%의 스킬 개선과 4~11일의 리드 타임 개선을 가져올 수 있다. 결국 같은 시간적 물적 자원을 사용할 때 선행시간이 가까울수록 더욱 신뢰할만한 예측 정보를 생산할 수 있고 비용면에서 투자가 가능한 상황이라면 앙상블 수를 늘리는 것도 더 나은 동아시아 기온 예보를 얻는 데 도움을 줄 수 있다.

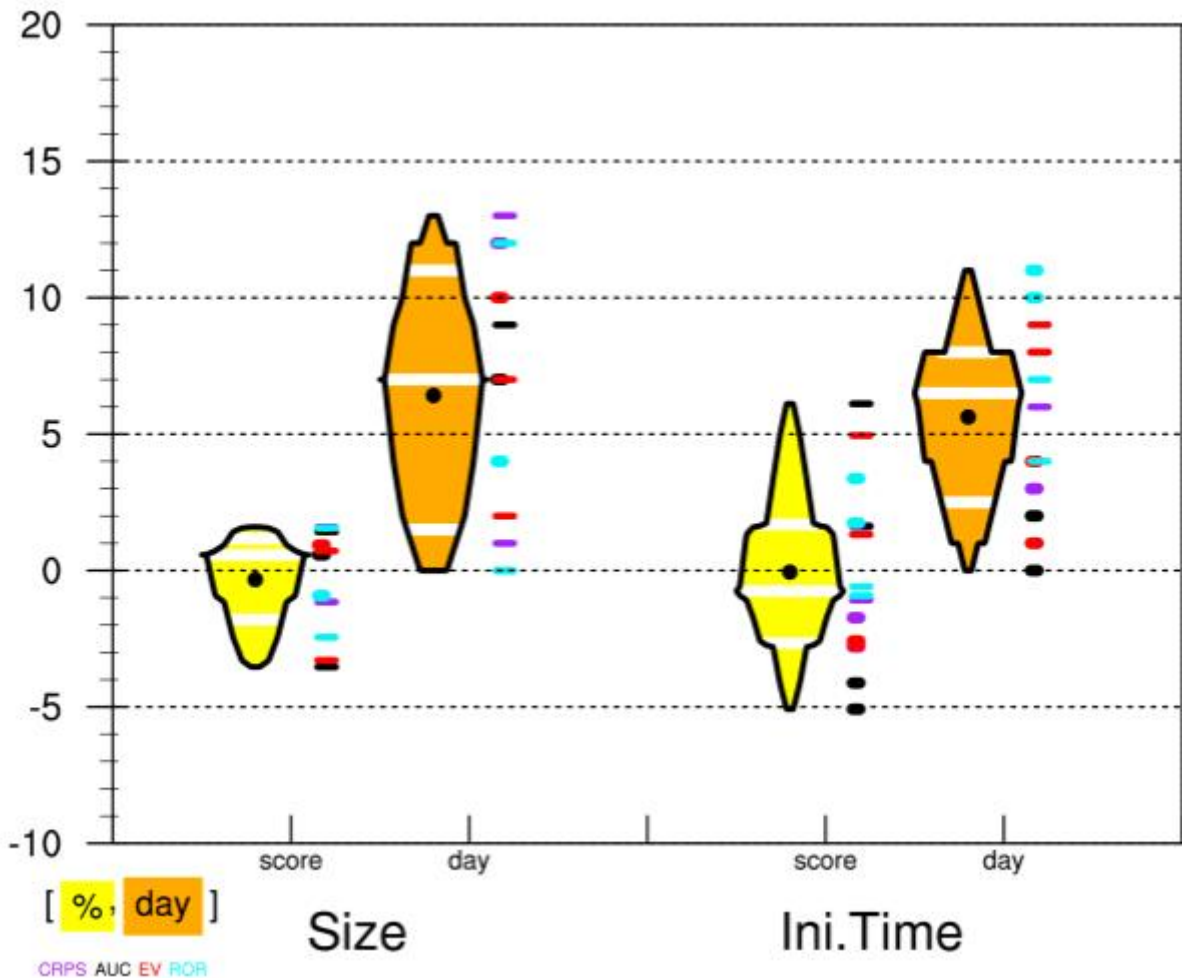


Figure 127. The violin plot for comprehensive intercomparison of all skill improvement on EA temperature forecast 3 and 4 weeks in advance. The skill improvement is the improvement rate of skill and the number of improved days for changing ensemble size and initial time. The bars on the right in each violin indicate the values of improved forecast skill or days and the shorter bars among them are for Thursday run.

4. 요약 및 결론

A. 기상청 계절내 예측 시스템 (GloSea5) 예측성 평가

첫째, 기상청 계절내 현업 예보에 대한 객관적이고 일관되며 여러 측면을 검토하되 그것들을 종합적으로 아우를 수 있는 평가의 기틀을 마련하고자, 주평균 예측 통합 평가프레임 (CREDYT: Comprehensive and Relative Evaluating weekly mean DYNAmical forecast)을 개발하고 양질의 평가 정보 생산을 위해 효율적이고 새로운 형태의 검증·진단 방법론을 고안하였다. 이를 통해, 세계 우수 기관 대비 KMA 현업 예측의 상대적 비교 우위를 판단하고, 향후 예보 개선 도모는 물론 그 방향성을 도출할 수 있을 것으로 기대된다.

- 네 부문(기후값, 실시간 예측력, 기후 리짐, 역학적 커플링)에 대한 필수 평가 콘텐츠를 발굴하고, 이를 구조화하여 평가 프레임(CREDYT)을 정립하였다.
- 실시간 예측 검증 및 모형 성능 진단에 대한 새로운 방법론(순위 기반 검증 지수 합산 평가 기법(RVMM), 지구시스템 간 연계성 진단법, 초기 기후 모드 상태에 대한 예측성 의존도 진단법 등)을 고안하였다.

둘째, 2019년 52주 KMA-GloSea5 주평균 예측에 대한 평가 정보를 CREDYT 프레임에 기반해 생산하였다. 특히, 세계 우수 기관(ECCC, ECMWF, NCEP, UKMO) 예측 시스템 대비 상대적인 성능 평가에 중점을 두었다. 실시간 예보에 대한 결정론적/확률론적 검증에 더해 2019년 이전 1999~2010년 reforecast에 대한 평년값 오차, 계절성, 기후 리짐, 역학적 커플링 재현에 대한 구체적인 분석이 이루어졌다.

- KMA, UKMO 등 GloSea 계열 모형의 겨울철 남극 기온에는 warm bias가 존재하는데, 이는 모형 초기장인 ERA-Interim과 모델 검증자료인 ERA5의 차이로부터 기인하는 것으로 판단된다. KMA 평년값에서는 아시아 대륙과 북미 일부에 온난편차가 큰데, 같은 계열 모형임에도 불구하고 UKMO 평년값에서는 이러한 편차가 나타나지 않아 그 원인을 파악할 필요가 있다. KMA-GloSea5는 여름 계절 길이는 길게, ITCZ/SPCZ 계절 스윙은 다소 강하게 모의하는 특성이 있다.
- KMA 모델의 2019년 주별 기후값 재현성능과 실시간 예측성능은 모두 ECMWF에 비해 뒤처지나, ECCC, NCEP, UKMO와는 전반적으로 비슷하거나 약간 우위에 있다. 특히 동아시아 가을 기온 재현성능은 ECCC, NCEP, UKMO에 비해 우수함을 확인하였다.
- ACC를 이용한 KMA의 2019년 일별 예측성능 역시 ECMWF에 비해 뒤떨어지나, 다른 모델에 비해서는 더 좋은 예측가능성을 보여주어 중간정도의 일별 예측성능을 기록하였다. 6개 검증 지수값을 RVMM 기반으로 합산했을 때, KMA-GloSea5의 상대적 성능은 변수, 계절에 따라 다르지만 지역별로 ECMWF에 이어 2위에 오르는 등 양호한 성능을 보인다.
- KMA-GloSea5를 활용한 2019년 주별 MJO 예측은 선행시간 1주에서는 VPM1과 VPM2 모

두 신뢰수준 95% 이상 유의한 상관계수를 보여주나, 3주에서는 성능이 확연히 감소한다. 그러나 겨울철(NDJFMA) 선행시간 3주의 VPM2 예측성능은 0.70으로 95% 이상 유의한 상관계수를 보임을 확인하였다.

- 10개 기후 리짐의 기본 변수 예측력 기여도를 평가한 결과, 가을철 동아시아 지위고도 예측 성능은 선행 1~2주에서 MJO, ART1 초기 조건 등에 민감한 것을 알 수 있다. Chord 다이어그램이 제시하는 수권-빙권-대류권-성층권 접합강도는 선행 3주 예보에서 겨울철에 과대모의, 가을철에 과소모의가 나타났다. 또한, 전 계절에서는 수권 내 연결성 과소모의와 대류권 내부 연결성 과대 모의가 뚜렷하다.
- 대기-해양, 열대-중위도, 성층권-대류권 기본 역학적 커플링 진단 결과, 강수-SST 상관성의 극지역 과대 모의, 동아시아 기온 상승에 선행하는 열대 대류활동 부재, 성층권 극소용돌이에 대한 지표 기온 반응 지연 등 KMA-GloSea5 역학과정 모의 오차를 확인할 수 있었다.

셋째, KMA-GloSea5의 재현/예측성능을 향상시키기 위해 현재 기상청에서 현업으로 사용하는 기후값 생산방안을 개선한 3가지 기후값 생산방안을 제안하고 재현/예측성능을 살펴보았다. 3가지 기후값 생산방안은 기본적으로 KMA 기후값 생산방안을 따르도록 설계하였다. EXP1은 거리가중치만을 제거한 방법, EXP2는 거리가중치를 그대로 두고 reforecast 샘플 수를 증가시킨 방법이다. EXP3는 EXP1과 EXP2의 효과를 동시에 볼 수 있는 방법으로, 거리가중치를 제거하고 reforecast 샘플 수를 증가시켰다.

- KMA 기후값에서 나타나는 겨울철 남극의 극단적인 warm bias와 대륙 전반의 cold bias가 기후값 실험을 통해 일정 수준 보정되었고, EXP3의 보정 효과가 가장 크다. 같은 GloSea5 모델인 UKMO에서도 겨울철 남극 warm bias가 보정되는 등의 유사한 결과가 나타났다.
- KMA와 각 기후값 처방 실험의 상대적인 재현/예측성능을 분석한 결과, 기후값 재현성능이 좋을수록 실시간 예측성능 역시 좋아지는 경향이 있으나, 변수별/계절적 차이가 있었다.
- 전반적으로 EXP3 방법의 과거 재현성능 및 실시간 예측성능이 KMA나 다른 실험에 비해 우세한 것으로 보인다. 특히 가을철 동아시아 예측성능은 앙상블 증가에 의한 효과가 더 커지게 되어 예측성능이 매우 향상되었다. 향후 실시간 예측성능을 판단하기 위한 샘플이 증가되면 더 안정적인 결과를 얻을 수 있을 것으로 기대된다.
- 기후값 생산을 위해 reforecast 앙상블 멤버간의 날짜에 따른 거리가중치를 주는 것 보다 예측시점에서 다소 멀어지더라도 사용하는 reforecast 날짜 수를 확대하는 방향으로 리소스를 집중하는 것이 예측성능 향상에 도움이 될 것으로 판단된다.

넷째, MME 기반으로 개발된 동아시아 겨울 3주차 주평균 기온 예측성 추정 알고리즘을 KMA-GloSea5 모형에 맞게 최적화 하고 2019/2020 겨울 12주 예보에 대하여 그 적용성을 평가하였다.

- KMA-GloSea5 reforecast 자료 분석을 통해, 예측 성능에 민감한 기후 모드 교차 조건, 선행 1주 성능 지속성 조건, 선행 1~4주 예보 안정도 조건을 확인하였다. 그 중, 성능 지속성 조건은 만족되지 않아 추정 인자에서 탈락시켰다. 나머지 인자를 활용하여 KMA-GloSea5에 최적화된 겨울철 동아시아 주평균 기온 선행 3주 예측 스킴 추정을 위한 의사결정트리를 제시하였다 (부록 A).
- MME 기반으로 작성한 알고리즘에 배해, 1991~2010(20년)으로 확장된 과거재현 기간을 적용하였고 선행 1~2주 교차 조건을 추가하여 그 성능 추정 인자의 안정도를 높이고자 하였다.
- 성능 추정 결정트리를 2019/2020 겨울철 12개 주평균 예보에 대해 적용성 평가 결과, 기후 리짐 교차 조건을 적용했을 때 50%의 삼분위 성능 카테고리 일치성을 보였다. 하지만, 기온 안정도 인자를 적용했을 때는 17% 정도에 불과했다. low와 moderate 성능 카테고리에는 기후 리짐 교차 조건 적용이, 반대로, high 성능 카테고리에는 기온 안정도 조건 적용이 상대적으로 성능 추정에 유리함을 확인할 수 있었다.
- KMA-GloSea5에 맞춰 수정된 결정트리를 적용하여 예보 시점에서 실시간으로 성능을 추정하고 불확실성을 줄일 수 있다면, 매뉴얼 예보에서 모형 정보 가중치 선정의 객관적 정보로 활용될 수 있을 것이다.

B. 초기장에 따른 오차 성장 특성 진단

첫째, 초기 오차 성장 특성을 살펴보기 위해 먼저 기후값으로 사용되는 과거재현기후의 예측 특성에 대해 살펴보았으며 다음과 같이 요약하였다.

- 평균적으로 10일을 전후로 예측 성능(ACC)이 0.5 아래로 떨어지며 선행시간 3~4주를 전후로 0.1~0.2 정도에서 수렴하고 있음을 볼 수 있다.
- 선행시간 1일의 지표온도 및 강수의 오차의 패턴은 선행시간 3~4주까지 거의 동일한 패턴으로 유지되고 있으며 그 강도는 더 커지는 계통적 오차가 지배적이다.
- 적도 지역을 중심으로 과도한 증발이 발생하고, 증발된 수증기는 강수 물리 과정을 통해 많은 강수로 전환되어 오히려 건조한 대기를 만들고 있다.
- 중위도 지역에서는 대류 활동에 의한 강수보다 미세물리과정에 의한 강수 과정이 중요하다 할 수 있는데, 모델에서는 이러한 현상을 잘 설명하지 못하고 대부분의 강수를 대류 과정에 의한 강수로 모의하고 있다.
- 그 결과 약한 강수는 과소 모의하고 있으며, 중간 강도의 강수를 과대 모의하고 있는 것

을 알 수 있다. 특히, 한반도 지역은 구름의 양도 작게 모의가 되며, 대기 중의 수증기도 관측보다 더 많이 남아있는 것을 알 수 있다.

둘째, 초기장 차이에 따른 계절내 예측 성능에 미치는 영향에 대해 살펴보기 위해, 초기 조건을 NWP 예측장과 재분석 자료를 사용한 민감도 실험을 구성하여 초기 조건에 대한 예측장의 차이에 대해 분석하였다.

- NWP 예측장을 초기장으로 사용할 때 보다, 재분석자료를 초기장으로 사용할 때, 동아시아 겨울철 기온 및 순환장의 계절내 예측 성능이 개선된다.
- NWP 예측장을 초기장으로 사용할 때 보다 재분석자료를 초기장으로 사용할 때, MJO 예측 성능이 향상된다.
- 선행시간 1일에서 윤량 및 복사 값들의 오차가 매우 크게 나타나 초기값 처방에서의 이슈가 있을 것으로 판단된다.
- 계절내 규모에서의 오차의 감소를 위해 지면오차 및 구름/복사와 연관된 초기조건들의 개선이 필요하다.

결론적으로, GloSea5의 계통적 오차를 개선하기 위해서는 적도 지역에서 중요한 강수 과정인 강수모수화의 특성과 더불어 중위도 지역의 강수와 큰 연관이 있는 미세물리과정 역시 자세한 분석이 필요하다 할 수 있다. 또한, 과거재현기후와 실시간 예측장 사이의 초기 조건의 차이는 특히 편차장에서 계통적 오차의 제거에 영향을 줄 수 있다. 이러한 오차를 줄이기 위해, 윤량 및 복사 값 처방에 대한 자세한 분석 및 강수 모듈의 적절한 개선이 필요할 것이다. 또한, 이 연구에서 사용된 MJO 분석 및 진단 방법은 기후예측모델의 개발/개선시 과거재현기후 및 실시간 예측장, 그리고 민감도 실험 등에 적용이 용이하도록 관측과 모델의 자료 전처리 방법과 분석 방법을 준자동화가 가능한 진단 툴로 구성하였으며, 추후 지속적이고 일관성 있는 기후예측모델 진단에 활용할 수 있을 것으로 기대된다.

C. 앙상블 생산 방안에 따른 계절내 기온 예측 민감도 평가

첫째, 기상청 현업 장기예측시스템(GloSea5)의 기온 예측성 개선을 위해 계절내 예보에 대한 GloSea5의 앙상블 특성과 개선 방향에 대한 제언하였다.

- GloSea5 현업 hindcast의 1주~8주에 해당하는 동아시아 기온 예측에 대한 앙상블의 기본 특성을 진단한 결과 예보 2주 이후 일관되고 신뢰도가 높은 앙상블을 생산하고 있다고 평가되므로 활용하기 좋은 앙상블이다.
- GloSea5 앙상블에 대한 불확실성 진단을 통해 살펴본 결과 유라시아 대륙 고위도에서의 저온 아노말리와 그에 수반되는 상층 고/저기압의 동서 배치가 8주 이내 동아시아 기온 예보의 에러가 커지게 하는 요인으로 작용할 수 있어 고위도 지면/해빙 처방에서의 개선을 제안한다.

둘째, 동아시아 계절내 기온 예측에 있어서 GloSea5의 효율적인 앙상블 예측 시스템을 구축을 위해 앙상블 개수, 생산 일자, 생성 기법에 변화를 준 민감도 실험들을 분석하였고 정확도, 유용성, 활용도 측면에서 기존 현업 예보와 비교하여 최적의 구축 방향에 대한 제언하였다.

- 기본적으로 현업 실험의 경우 대체로 8주까지 유용한 예보를 생산해내고 있고 그 중 3, 4주 예보는 가장 개선이 필요한 기간으로 판단된다. 계절내 예측에 대한 현업 예보의 정확도와 유용성을 월요일 적분으로 생산 일자 변경을 통해 각각 최대 15%, 45% 향상시킬 수 있다. 기존 앙상블 생산 방법을 유지할 경우 앙상블 수를 42개로 증가시키는 방안도 정확도와 유용성 개선에 일부(5% 정도) 효과를 볼 수 있다. 그러나 burst 방법에서는 앙상블 크기 증가 효과가 미미하고 주중런은 권장하지 않는다.
- 예측 성능의 개선이 요구되는 예보 3~4주차의 동아시아 현업 기온 예보는 기본적으로 유용하다. 이 기존 현업시스템에서 앙상블 멤버수가 증가하면 리드 타임 전반에 걸쳐 스킵을 증가시킬 수 있다. 기존 현업 예보 대비 민감도 실험 간 개선 정도를 상호 비교하자면 개선율의 크기에 있어서는 앙상블 멤버수 증가와 주말런이 주중런보다는 효과를 가져다 줄 수 있고, 개선된 리드타임의 수 측면에서도 앙상블 크기 증가가 가장 개선 효과가 크고 그 다음은 주말런 순서이다. 대체로 평년 이상과 평년 이하 카테고리에서 개선 효과가 있다.
- 동아시아 기온의 3~4주 예보에 있어서, 예보를 활용하는 사용자 중심에서 얻을 수 있는 이득은 현업 예보 대비 4% 이상의 수익률과 예보 가치를 5~9% 더 얻어낼 수 있는 주말런에서 가장 크다. 또한 예보관 입장에서도 주말런으로 생산 일자를 변경하는 것이 극한 기상 발생에 가장 활용하기 좋은 예보를 생산하는 것으로 보인다.
- 결국 같은 시간적 물적 자원을 사용할 때 선행시간이 가까울수록 더욱 신뢰할만한 예측 정보를 생산할 수 있고 비용면에서 투자가 가능한 상황이라면 앙상블 수를 늘리는 것도 더 나은 동아시아 기온 예보를 얻는 데 도움을 줄 수 있다.

사사

본 연구는 APCC의 지원을 통해 수행되었습니다.

References

- 김가영, 김해정, 오지현, 유태우, 이윤영, 정여민, 정유란, 정유림, 함수련, 윤순조, 2019: 다중모델 기반 계절내 예측 활용 가이드라인 개발, *APEC기후센터 연구보고서*, APEC기후센터.
- 김상욱, 2019: *Optimal ensemble size for Sub-seasonal to Seasonal (S2S) prediction system*. 석사학위논문, 지구환경과학부, 서울대학교, 41 pp.
- 서은교, 이명인, 정지훈, 강현석, 원덕진, 2016: 전지구 계절 예측 시스템의 토양수분 초기화 방법 개선, *대기*, **26**(1), 35-45.
- 이우섭, 윤순조, 김가영, 김미아, 신지현, 이현록, 2017: WMO 장기예보 선도센터 운영(7차년도), *APEC기후센터 연구보고서 2017-20*, APEC기후센터.
- 정영운, 문일주, 장필훈, 2016: 기상청 계절예측시스템(GloSea5)의 해양성층 강화시기 단기 해양 예측 정확도 및 대기-해양 접합효과, *대기*, **26**(4), 599-615.
- Baggett, C.F., Barnes, E.A., Maloney, E.D. and Mundhenk, B.D., 2017: Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophysical Research Letters*, **44**(14), 7528-7536, <https://doi.org/10.1002/2017GL074434>
- Baldwin, M.P., L.J. Gray, T.J. Dunkerton, K. Hamilton, P.H. Haynes, W.J. Randel, J.R. Holton, M.J. Alexander, I. Hirota, T. Horinouchi, D.B.A. Jones, J.S. Kinnersley, C. Marquardt, K. Sato, and M. Takahashi, 2001: The Quasi-Biennial Oscillation. *Reviews of Geophys.*, **39**, 179-229, <https://doi.org/10.1029/1999RG000073>
- Buizza, R., D. S. Richardson and T. N Palmer, 2003: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Quart. J. Roy. Met. Soc.*, **129**, 1269-1288, <https://doi.org/10.1256/qj.02.92>
- Choi, J., S. W. Son, Y. G. Ham, J. Y. Lee, and H. M. Kim, 2016: Seasonal-to-interannual prediction skills of near-surface air temperature in the CMIP5 decadal hindcast experiments. *J. Clim.*, **29**, 1511-1527, <https://doi.org/10.1175/JCLI-D-15-0182.1>.
- Christensen, H. M., I. M. Moroz, and T. N. Palmer, 2015: Evaluation of ensemble forecast uncertainty using a new proper score: application to medium-range and seasonal forecasts. *Quart. J. Roy. Met. Soc.*, **141**, 538-549, <https://doi.org/10.1002/qj.2375>.
- DelSole, T., Trenary, L., Tippet, M.K. and Pegion, K., 2017: Predictability of week-3-4 average temperature and precipitation over the contiguous United States. *Journal of Climate*, **30**(10), 3499-3512, <https://doi.org/10.1175/JCLI-D-16-0567.1>
- Diro, G. T., and H. Lin, 2020: Subseasonal Forecast Skill of Snow Water Equivalent and Its

- Link with Temperature in Selected SubX Models. *Wea. Forecasting*, **35**, 273–284, <https://doi.org/10.1175/WAF-D-19-0074.1>
- Ferranti, L., C. Susanna, and J. Martin, 2015: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quart. J. Roy. Met. Soc.*, **141**, 916–924, <https://doi.org/10.1002/qj.2411>.
- Good I. J., 1952: Rational decisions. *Journal of the Royal Statistical Society, Series B*, **14**, 107–114.
- Gossart, A., S. Helsen, J. T. M. Lenaerts, S. Vanden Broucke, N. P. M. van Lipzig, and N. Souverijns, 2019: An evaluation of surface climatology in state-of-the-art reanalyses over the Antarctic Ice Sheet. *J. Clim.*, **32**, 6899–6915, <https://doi.org/10.1175/JCLI-D-19-0030.1>.
- Green, B. W., S. Sun, R. Bleck, S. G. Benjamin, and G. A. Grell, 2017: Evaluation of MJO Predictive Skill in Multiphysics and Multimodel Global Ensembles, *Monthly Weather Review*, **145**(7), 2555–2574, <https://doi.org/10.1175/MWR-D-16-0419.1>
- Hagedorn, R., and L. A. Smith, 2009: Communicating the value of probabilistic forecasts with weather roulette. *Meteorol. Appl.*, **16**, 143–155, <https://doi.org/10.1002/met.92>.
- Haiden, T., M. Janousek, J. Bidlot, L. Ferranti, F. Prates, F. Vitart, P. Bauer and D. S. Richardson, 2016: Evaluation of ECMWF forecasts. including the 2016 resolution upgrade (ECMWF Tech. Memo 792). Retrieved from <http://www.ecmwf.int/sites/default/files/elibrary/2015/15275-evaluation-ecmwf-forecasts-including-2014-2015-upgrades.pdf>
- Ham, H., D. Won, and Y. Lee, 2017: Performance Assessment of Weekly Ensemble Prediction Data at Seasonal Forecast System with High Resolution. *Atmosphere (Basel)*, **27**, 261–276, <https://doi.org/10.14191/Atmos.2017.27.3.261>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hirsch, A. L., J. Kala, A. J. Pitman, C. Carouge, J. P. Evans, V. Haverd, and D. Mocko, 2014: Impact of land surface initialization approach on subseasonal forecast skill: A regional analysis in the southern hemisphere. *J. Hydrometeorol.*, **15**, 300–319, <https://doi.org/10.1175/JHM-D-13-05.1>.
- Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.*, **2**, 36–50, [https://doi.org/10.1175/1525-7541\(2001\)002<0036:GPAODD>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2).

- Jeong, J.-H., H. W. Linderholm, S.-H. Woo, C. Folland, B.-M. Kim, S.-J. Kim and D. Chen, 2013: Impacts of snow initialization on subseasonal forecasts of surface air temperature for the cold season. *J. Climate*, **26**, 1956–1972, <https://doi.org/10.1175/JCLI-D-12-00159.1>.
- Jung, M., S.-W. Son, Y. Lim, K. Song, D. Won, and H.-S. Kang, 2016: Assessment of Stratospheric Prediction Skill of the GloSea5 Hindcast Experiment. *Atmosphere (Basel)*, **26**, 203–214, <https://doi.org/10.14191/atmos.2016.26.1.203>.
- Kang, I.-S., J.-Y. Lee, and C.-K. Park, 2004: Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *J. Clim.*, **17**, 834–844, [https://doi.org/10.1175/1520-0442\(2004\)017<0834:PPOSMP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0834:PPOSMP>2.0.CO;2).
- Kim, S.-W., H. Kim, K. Song, S.-W. Son, Y. Lim, H.-S. Kang, and Y.-K. Hyun, 2018: Subseasonal-to-Seasonal (S2S) Prediction Skills of GloSea5 Model: Part 1. Geopotential Height in the Northern Hemisphere Extratropics Atmosphere. *Korean Meteorological Society*, **28(3)**, 233–245, <https://doi.org/10.14191/Atmos.2018.28.3.233>.
- Koster, R. D., P. A. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C. T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C.-H. Lu, S. Malyshev, B. McAvaney, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y. C. Sud, C. M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, <https://doi.org/10.1126/science.1100217>.
- Koster, R.D., S. P. P. Mahanama, T. J. Yamada, G. Balsamo, A. A. Berg, M. Boissarie, P. A. Dirmeyer, F. J. Doblas-Reyes, G. Drewitt, C. T. Gordon, Z. Guo, J.-H. Jeong, W.-S. Lee, Z. Li, L. Luo, S. Malyshev, W. J. Merryfield, S. I. Seneviratne, T. Stanelle, B. J. J. M. Van Den Hurk, F. Vitart, and E. F. Wood, 2011: The Second phase of the global land-atmosphere coupling experiment: Soil moisture contributions to subseasonal forecast skill. *J. hydrometeorology*, **12**, 805–822, <https://doi.org/10.1175/2011JHM1365.1>
- Kug, J.S., Jeong, J.H., Jang, Y.S., Kim, B.M., Folland, C.K., Min, S.K. and Son, S.W., 2015: Two distinct influences of Arctic warming on cold winters over North America and East Asia. *Nature Geoscience*, **8(10)**, 759–762, <http://dx.doi.org/10.1038/ngeo2517>
- Lee, S. M., H.-S. Kang, Y.-H. Kim, Y.-H. 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(1)**, 59–72, <https://doi.org/10.14191/Atmos.2016.26.1.059>
- Li, S. and Robertson, A.W., 2015: Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Monthly Weather Review*, **143(7)**, 2871–2889.
- Li, Y., and S. Yang, 2010: A dynamical index for the East Asian winter monsoon. *J. Clim.*, **23(15)**, 4255–4262, <https://doi.org/10.1175/2010JCLI3375.1>.

- Lim, Y., S.-W. Son, and D. Kim, 2018: MJO prediction skill of the subseasonal-to-seasonal prediction models. *J. Clim.*, **31**(10), 4075–4094, <https://doi.org/10.1175/JCLI-D-17-0545.1>.
- Lin, H., G. Brunet, and J. Derome, 2008: Forecast skill of the Madden–Julian oscillation in two canadian atmospheric models. *Mon. Weather Rev.*, **136**, 4130–4149, <https://doi.org/10.1175/2008MWR2459.1>.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0).
- MacLachlan, C., A. Arribas K. A. Peterson A. Maidens D. Fereday A. A. Scaife M. Gordon M. Vellinga A. Williams R. E. Comer J. Camp P. Xavier and G. Madec, 2015: Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**(689), 1072–1084, <https://doi.org/10.1002/qj.2396>.
- MacLeod, D., C. O’ Reilly, T. Palmer, and A. Weisheimer, 2018: Flow dependent ensemble spread in seasonal forecasts of the boreal winter extratropics. *Atmos Sci Lett.*, **19**(5), <https://doi.org/10.1002/asl.815>.
- Madec, G., 2008: NEMO ocean engine. *IPSL Tech. Rep.* **27**, 401 pp
- Monache, L. D., J. P. Hacker, Y. Zhou., and X. Deng., 2006: Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *J. Geophys. Res.*, **111**, D24307, <https://doi.org/10.1029/2005JD006917>.
- Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Weather Rev.*, **105**, 803–816, [https://doi.org/10.1175/1520-0493\(1977\)105<0803:TVOCCA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1977)105<0803:TVOCCA>2.0.CO;2)
- Naujokat, B., 1986: An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. *J. Atmos. Sci.*, **43**, 1873–1877, [https://doi.org/10.1175/1520-0469\(1986\)043<1873:AUOTOQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1986)043<1873:AUOTOQ>2.0.CO;2)
- Orsolini, Y., R. Senan, G. Balsamo, F. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco and R. Benestad, 2013: Impact of snow initialization on sub-seasonal forecasts. *Climate Dyn.*, **41**, 1969–1982, <https://doi.org/10.1007/s00382-013-1782-0>.
- Panagiotopoulos, F., M. Shahgedanova, A. Hannachi, and D. B. Stephenson, 2005: Observed trends and teleconnections of the Siberian high: A recently declining center of action. *J. Clim.*, **18**, 1411–1422, <https://doi.org/10.1175/JCLI3352.1>
- Rae, J. G. L., H. T. Hewitt, A. B. Keen, J. K. Ridley, A. E. West, C. M. Harris, E. C. Hunke, and D. N. Walters, 2015: Development of the Global Sea Ice 6.0 CICE configuration for the Met Office Global Coupled model. *Geosci. Model Dev.*, **8**, 2221–2230,

<https://doi.org/10.5194/gmd8-2221-2015>.

- Rashid, H. A., H. H. Hendon, M. C. Wheeler, and O. Alves, 2011: Prediction of the Madden-Julian oscillation with the POAMA dynamical prediction system. *Clim. Dyn.*, **36**, 649–661, <https://doi.org/10.1007/s00382-010-0754-x>.
- Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.*, **20**, 5473–5496, <https://doi.org/10.1175/2007JCLI1824.1>.
- Rodney, M., Lin, H. and Derome, J., 2013: Subseasonal prediction of wintertime North American surface air temperature during strong MJO events. *Monthly weather review*, **141**(8), 2897–2909, <https://doi.org/10.1175/MWR-D-12-00221.1>
- Roebber, P. J. and L. F. Bosart, 1996: The complex relationship between forecast skill and forecast value: A real-world analysis. *Wea. Forecasting*, **11**, 544–559, [https://doi.org/10.1175/1520-0434\(1996\)011<0544:TCRBFS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0544:TCRBFS>2.0.CO;2)
- Son, S. W., H. Kim, K. Song, S. W. Kim, P. Martineau, Y. K. Hyun, and Y. Kim, 2020: Extratropical Prediction Skill of the Subseasonal-to-Seasonal (S2S) Prediction Models. *J. Geophys. Res. Atmos.*, **125**, 1–14, <https://doi.org/10.1029/2019JD031273>.
- Thomas, J. A., A. A. Berg and W. J. Merryfield, 2016: Influence of snow and soil moisture initialization on sub-seasonal predictability and forecast skill in boreal spring. *Climate Dyn.*, **47**, 49–65, <https://doi.org/10.1007/s00382-015-2821-9>.
- Thompson, D.W. and Wallace, J.M., 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical research letters*, **25**(9), 1297–1300, <https://doi.org/10.1029/98GL00950>
- Ventrice, M. J., M. C. Wheeler, H. H. Hendon, C. J. Schreck, C. D. Thorncroft, and G. N. Kiladis, 2013: A Modified Multivariate Madden-Julian Oscillation Index Using Velocity Potentia, *Monthly Weather Review*, **141**(12), 4197–4210, <https://doi.org/10.1175/MWR-D-12-00327.1>
- Vitart, F., A. W. Robertson and D. L. T. Anderson, 2012: Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, **61**(2), 23–28.
- Vitart, F., C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, H. Hendon, J. Hodgson, H.-S. Kang, A. Kumar, H. Lin, G. Liu, X. Liu, P. Malguzzi, I. Mallas, M. Manoussakis, D. Mastrangelo, C. MacLachlan, P. McLean, A. Minami, R. Mladek, T. Nakazawa, S. Najm, Y. Nie, M. Rixen, A. W. Robertson, P. Ruti, C. Sun, Y. Takaya, M. Tolstykh, F. Venuti, D. Waliser, S. Woolnough, T. Wu,

D.-J. Won, H. Xiao, R. Zaripov, and L. Zhang, 2017: The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bull. Amer. Meteor. Soc.*, **98**(1), 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>

Wang, B., J.-Y. Lee, I.-S. Kang, J. Shukla, C.-K. Park, A. Kumar, J. Schemm, S. Cocke, J.-S. Kug, J.-J. Luo, T. Zhou, B. Wang, X. Fu, W.-T. Yun, O. Alves, E. K. Jin, J. Kinter, B. Kirtman, T. Krishnamurti, N. C. Lau, W. Lau, P. Liu, P. Pegion, T. Rosati, S. Schubert, W. Stern, M. Suarez, and T. Yamagata, 2009: Advance and prospectus of seasonal prediction: Assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Clim. Dyn.*, **33**, 93–117, <https://doi.org/10.1007/s00382-008-0460-0>.

Walters, D., I. Boutle, M. Brooks, T. Melvin, R. I. Stratton, S. Vosper, H. Wells, K. Williams, N. Wood, T. Allen, A. Bushell, D. Copsey, P. Earnshaw, J. Edwards, M. Gross, S. Hardiman, C. Harris, J. Heming, N. Klingaman, R. Levine, J. Manners, G. Martin, S. Milton, M. Mittermaier, C. I. Morcrette, T. Riddick, M. Roberts, C. Sanchez, P. Selwood, A. Stirling, C. Smith, D. Suri, W. Tennant, P. Li Vidale, J. Wilkinson, M. Willett, S. Woolnough, and P. Xavier, 2017: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geosci. Model Dev.*, **10**, 1487–1520, <https://doi.org/10.5194/gmd-10-1487-2017>.

Zhu, Y., 2005: Ensemble Forecast: A new approach to uncertainty and predictability. *Adv. Atmos. Sci.*, **22**, 781–788, <https://doi.org/10.1007/BF02918678>

부록 A. 동아시아 겨울 기온의 GloSea5 선행3주 주평균예보 스킬추정을 위한 지침서

V2020 by Yun-Young Lee

동아시아 겨울 기온의 GloSea5 선행3주 주평균예보 스킬추정을 위한 지침서
A guide to estimating lead week3 GloSea5 forecast skill of boreal winter weekly mean East Asia temperature

수주의 계절내 규모에서 역학 예측은 그 성능이 매우 낮고 불확실성이 큰 것이 잘 알려져 있음. 하지만, GloSea5 20년 (1991-2010) 과거 예측 분석에 따르면, 동아시아 겨울 주평균 기온 예측의 성능이 뚜렷이 구분되는 여러 조건들이 존재함. 본 지침서는 이러한 정보를 활용하여 선행 3주 (예측 1주) 스킬을 미리 가늠할 수 있는 방법을 제시함. 이렇게 추정된 스킬을 바탕으로 다른 정보 원천 활용 및 그에 따른 GloSea5 예측 hedging 여부를 결정할 수 있고, 궁극적으로 객관적이고 최적화된 consolidated 예보 생산에 일조할 것으로 기대함.

Target	Input information transformed
겨울철 동아시아 지역 (100.5°-135.5°E, 30°-40°N) 선행3주 주평균 기온 (T2m) 예측의 anomaly correlation coefficient (ACC) 삼분위 카테고리	- EATpcl_Stability: 1/abs(std/avg) during 1-4 lead weeks - Simulated weekly mean Climate_Regimes: (Table 1,2) 동아시아 기온 변동성분 (EATa, EATpcl, EATpc2) 기후지수 in 수문, 남빙권동(NINO3.4), 대서양 적도 변동(AEM), 평권, 북극기온 변동성분(ART1 & ART2), 성층권: 극소용돌이(PV), 대류권: 동아시아 겨울한순(EAWM), 시베리아고기압(SH), 극전동(AOI)

Decision tree based on tercile category information

① prepare categorized inputs of PCI_stability and climate regimes (Tbl.1&2)

② identify the ACC degree for each input

AN or red? YES → +1 to "HIGH", No → +1 to "LOW"
BN or blue? YES → +1 to "LOW", No → +1 to "MODERATE"
(NN or nothing) → +1 to "MODERATE"

③ summarize counts

④ judge the reliability of W3 prediction and incorporate it into the manual forecast (Tbl.3)

V2020 by Yun-Young Lee

Table 1 East Asia temperature variability

Abbr.	EATa	EATpcl	EATpc2
Definition	T2m ^a average over (30°-65°N, 105.5°-145.5°E)	PC timeseries of 1 st EOF of T2m ^a over (30°-65°N, 105.5°-145.5°E)	PC timeseries 2 nd EOF of T2m ^a over (30°-65°N, 105.5°-145.5°E)
note	T2m refers to air temperature at 2 meters above the surface.	(a) EOF1	(b) EOF2

^a indicates normalized value

Table 2 Climate indices of four earth systems simulated in GloSea5

	Index	Definition	Reference
Hydrosphere	NINO3.4	Sea Surface Temperature (SST) ^a average (5°S-5°N, 170°W-120°W)	
	Atlantic Equatorial Mode (AEM)	SST ^a average (3°S-3°N, 30°W-0°W)	
Cryosphere	Arctic Temp. index 1 (ART1)	T2m ^a average (70°N-80°N, 30°E-70°E)	Kug et al. (2015)
	Arctic Temp. index 2 (ART2)	T2m ^a average (65°N-80°N, 160°E-160°W)	Kug et al. (2015)
Stratosphere	Polar Vortex (PV)	50hPa zonal wind (U50°) average within 60°N-80°N	
Troposphere	East Asia Winter Monsoon (EAWM)	200hPa zonal wind (U200) gradient $\{[(30^{\circ}\text{-}35^{\circ}\text{N}, 90^{\circ}\text{-}160^{\circ}\text{E}) - (50^{\circ}\text{-}60^{\circ}\text{N}, 70^{\circ}\text{-}170^{\circ}\text{E})] - [(30^{\circ}\text{-}35^{\circ}\text{N}, 90^{\circ}\text{-}160^{\circ}\text{E}) - (5^{\circ}\text{S-}10^{\circ}\text{N}, 90^{\circ}\text{-}160^{\circ}\text{E})]\} / 2$	Li and Yang (2010)
	Siberia High (SH)	1000hPa geopotential height (Z1000) ^a average (40°-65°N, 80°-120°E)	Panagiotopoulos et al. (2005)
	Arctic Oscillation (AO)	PCI of Z1000 ^a poleward of 20°N, its loading pattern as below	Thompson and Wallace (1998)

^a indicates normalized value (anomalies)

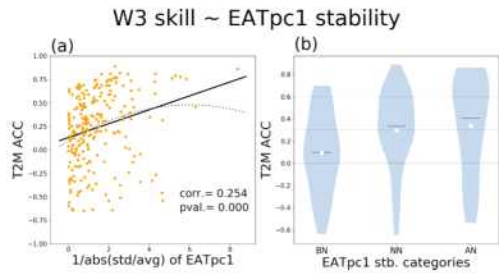


Figure 1 (a) scatter plot between simulated PC1 stability and EA T2m ACC and (b) violin plot of EA T2m ACC for three PC1 stability categories for lead time of 3 weeks from 240 samples (12 weeks for 20 winter seasons). X-axis displays the stability which is defined as an inverse of PC1 standard deviation for 4 lead weeks divided by absolute value of their average. The first and second order least square fitted lines are delineated (black solid and gray dash) in (a). The three categories in (b) are based on Upper (0.43) and lower (-0.43) tercile of standardized PC1 stability. Horizontal lines and white circles in (b) indicate median and mean values of samples for three stability categories.

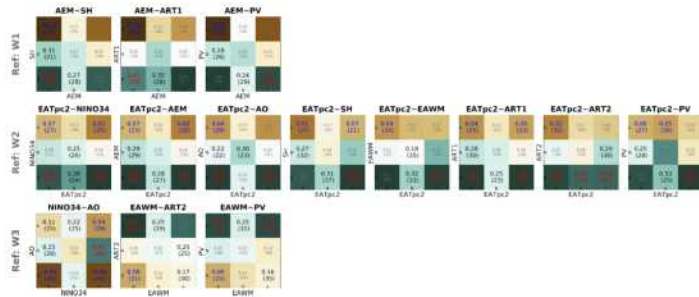


Figure 2 W3 mean ACC of 9 categories based on two different indices' tercile criteria of (top) W1, (middle) W2, and (bottom) W3. For the tercile categorization, anomaly of each index is calculated then it is normalized by its standard deviation. Upper and lower tercile is -0.43 σ and 0.43 σ , respectively. ACC average value and sample size within parenthesis are displayed in each box. Bold blue (red) color denotes the category of distinguished large (small) ACC average value greater (less) than 0.43 σ (-0.43 σ) of all ACCs, having the sample size 20 and over and the sample spread less than 1 σ of all ACCs. Bold black color denotes the category of ACC average between -0.43 σ and 0.43 σ , having the sample size 20 and over and the sample spread less than 1 σ of all ACCs. Here, only the cross-conditions whose 9 categories include at least one of all three blue, black, and red categories are displayed.

Table 3 Verification of estimated ACC categories for winter season 12 weeks in 2019 utilizing three metrics (Proportion Correct, Hit Rate and Post Agreement). Numbers in brackets indicate the sample size (denominator) considered in the score calculation.

Ingredient	Proportion Correct (PC)	Hit Rate (HR)			Post Agreement (PA)		
		ACC: Low	ACC: Moderate	ACC: High	ACC: Low	ACC: Moderate	ACC: High
Clim. Reg. based	0.50 [12]	0.60 [5]	0.50 [4]	0.33 [3]	0.75 [4]	0.50 [4]	0.25 [4]
EATpc1 stability	0.17 [12]	0.00 [5]	0.00 [4]	0.67 [3]	0.00 [4]	0.00 [2]	0.33 [6]

Supplementary

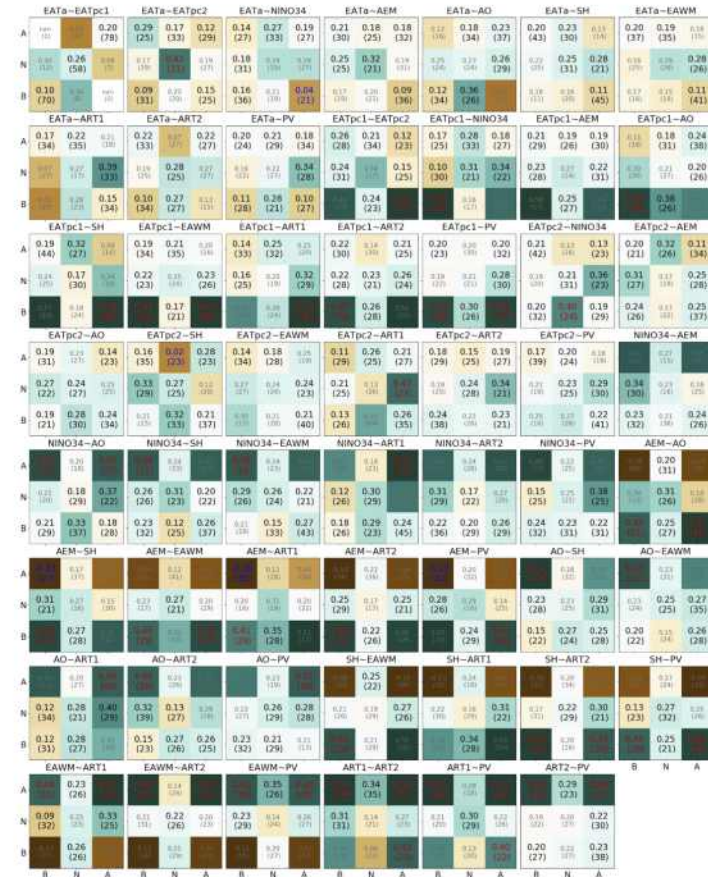


Figure 3 W3 mean ACC of 9 categories based on all available two different indices' tercile criteria of W1. Details are the same as in Figure 2.

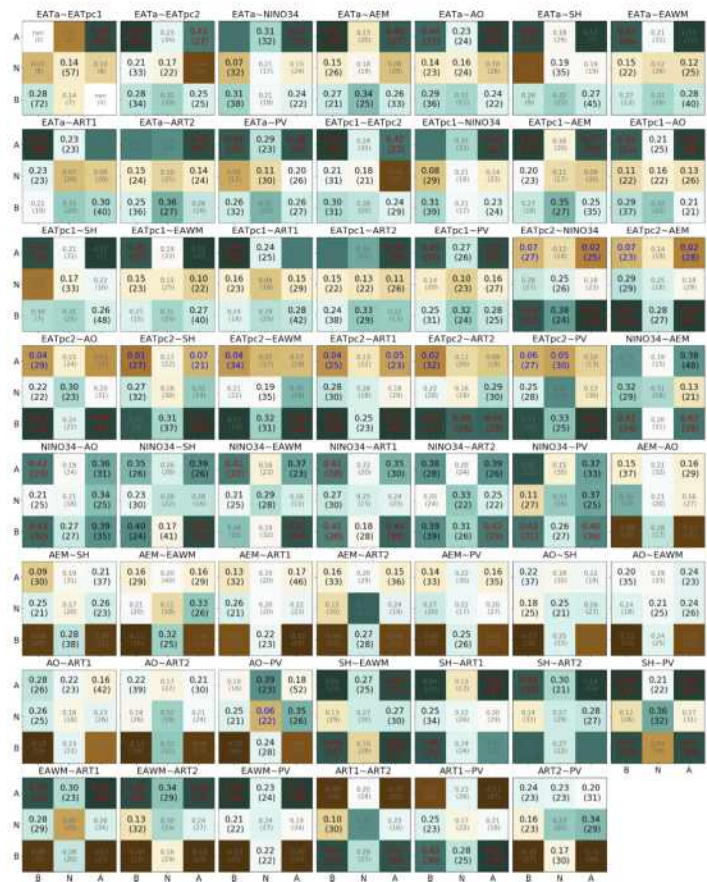


Figure 4 Same as Figure 3, but for 9 categories based on all available two different indices' tercile criteria of W2.

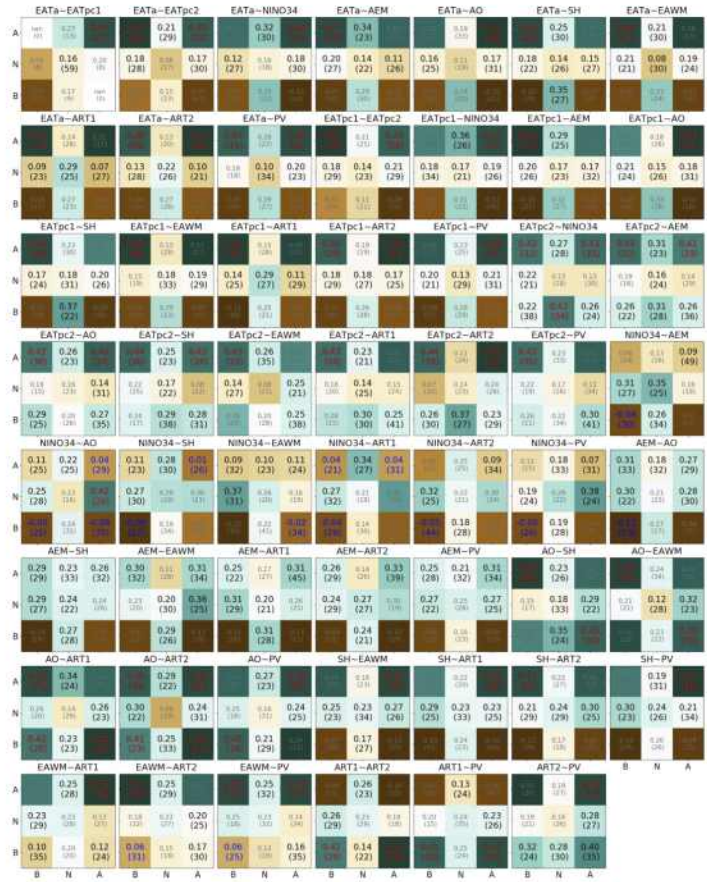


Figure 5 Same as Figure 3, but for 9 categories based on all available two different indices' tercile criteria of W3.

부록 B. KMA-GloSea5 기후값 생산 방안 가이드라인

KMA-GloSea5 기후값 생산 방안 가이드라인

김가영, 이윤영

APEC기후센터 예측기술과

Nov. 2020

- 본 가이드라인에서는 KMA-GloSea5(이하 현업) 계절내 규모 기후값 개선을 위해 검토한 3가지 기후값 생산방안의 민감도 평가 결과를 제시함.
- 세 기후값 실험 모두 기존 현업 기후값에 비해 선행시간 3주에서 에러(RMSE)가 줄어들어 기후값 재현성능이 향상되었고, 실시간 예측성능(ACC) 또한 현업 기후값을 적용한 경우 보다 대체로 향상되었음.
- 전반적으로 EXP3(거리가중치 미적용, 기후장 날짜 수 증가)가 가장 향상된 재현/예측성능을 보여주었음.
- 현업 기후값 개선을 위해 거리가중치를 주는 대신 기후장 날짜 수를 47개로 늘리는데 리소스를 투입할 것을 제언함.

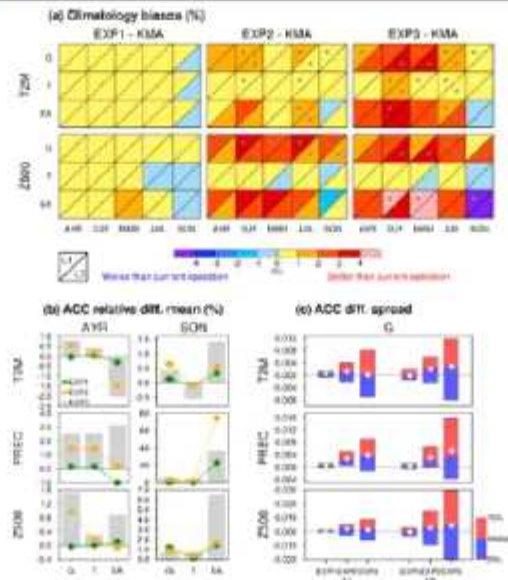


그림 1 KMA 기온과 500hPa 지위고도의 계절별/지역별 '19년 기후값 대비 각 기후값 실험의 상대적 재현성능(RMSE) 차이 (단위: %). 파상단과 우하단 삼각형은 각각 선행 1주와 3주를 나타냄. (b) KMA 기온, 강수, 500hPa 지위고도의 전계절/가을철 '19년 실시간 예측성능 대비 각 기후값 실험을 적용한 실시간 예측성능(ACC)의 상대적 차이 (단위: %). (c) KMA 전구 기온, 강수, 500hPa 지위고도의 '19년 예측성능과 각 기후값 실험과의 '19년 예측성능 차이의 상/하위 25% 분포, 중앙값, 평균값.

1. 기후값 생산방안

1) 기상청 현업 기후값 생산방안

- 기상청에서는 예측장을 매일 생산하나 기후장은 특정일(1, 9, 17, 25일)에만 생산하기 때문에 예측장에 준하는 매일의 기후장의 결과가 필요함
- 따라서 계절내 예측의 경우 예측장 날짜를 기준으로 전후의 가장 가까운 기후장 날짜를 선택하여 해당하는 기후장의 자료를 날짜 거리(d1, d2)에 따른 가중평균을 하여 예측장의 날짜에 준하는 기후장을 생산함
- 기상청 주변 앙상블 예측은 최근 7일의 예측자료인 28개의 앙상블멤버(4개*7일)를 수집하여 시간차이에 따른 앙상블을 구성하며, 기후장 역시 예측장에 준하는 날짜의 가중평균 기후장을 이용하여 주변 앙상블 예측자료를 생산함 (한현준 등, 2017)

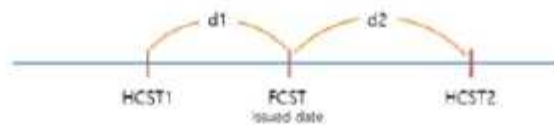


그림 2 선택된 기후장 날짜에 대한 예측장의 날짜와의 거리

$$\omega = e^{-d^2/100}$$

$$\text{normalized weight} = \omega_n / \sum_{n=1}^k \omega_n$$

2) 3가지 기후값 생산방안

- 현업 기후값 생산방안과 비교할 수 있도록 전반적인 생산과정은 기본적으로 현업 기후값 방안을 따르되, 거리가중치를 적용하지 않거나 기후장 날짜 수를 증가하는 등 변형을 주어 민감도 실험을 수행함
- 기후장 날짜의 경우, 앞/뒤로 총 2개의 멤버를 추가함

예) 20190502 기후값에 사용된 기후장 날짜:

(기준) 0501, 0509 → (추가) 0425, 0501, 0509, 0517

EXP	거리가중치	기후장 샘플 수	비고
KMA	적용	(3개*2일)*7일=42	기상청 현업 기후값
EXP1	미적용	(3개*2일)*7일=42	거리가중치 미적용
EXP2	적용	(3개*4일)*7일=84	기후장 날짜 수 증가
EXP3	미적용	(3개*4일)*7일=84	거리가중치 미적용, 기후장 날짜 수 증가

2. 분석기간

- 기후값 기간: 2019년 reforecast의 1999-2010년
- 실시간 예측 기간: 2019년 매주 목요일(0103-1226, 총 52개 주)

3. 관측자료

변수	관측자료	참고문헌
T2M, Z500	ERA5	Molard et al. (2015)
PREC	GPCP	Huffman et al. (2001)

4. 결과

1) 2019년 기후값 재현

[가] 겨울철 기온 Mean bias

- 기존 KMA 현업 기후값은 겨울철 남극 기온에서 극단적인 warm bias가 나타나고 대륙에서는 전반적으로 cold bias가 지속적으로 나타난다. 이는 UKMO에서도 동일하고 GloSea5계열 모델의 공통적인 특징으로 보임 (그림 3a,c)
- 각 기후값 실험을 통해 KMA와 UKMO 모두 겨울철 남극 warm bias와 북반구 대륙의 cold bias가 다소 보정됨. 보정이 가장 많이 된 실험은 EXP3으로 판단됨 (그림 3b,d)

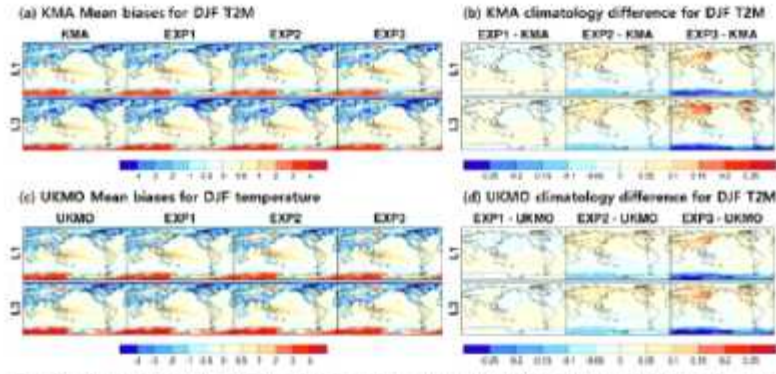


그림 3 선행시간 1주와 3주의 겨울철 기온에 대한 (a) KMA 현업 및 각 기후값 실험의 mean bias와 (b) 각 기후값 실험과 KMA 기후값 간의 차이. 동일한 방법을 적용한 (c, d) UKMO 결과

[L] KMA 현업 기후값 대비 상대적 RMSE 차이

$$\text{Relative difference(\%)} = -\frac{RMSE_{EXP} - RMSE_{KMA}}{RMSE_{KMA}} \times 100$$

- 세 가지 기후값 처방방안 모두 전반적으로 현업 기후값에 비해 나은 기후값 재현 성능을 보여줌
- 특히 EXP3은 선행시간 3주 동아시아 강수의 경우 DJF를 제외하면 현업에 비해 재현성능이 2~4%정도 향상되고, 선행시간 3주 동아시아 기온의 경우에도 역시 JJA를 제외하면 재현성능이 향상되어 다른 실험에 비해 더 좋은 재현성능을 보여줌 (그림 4).

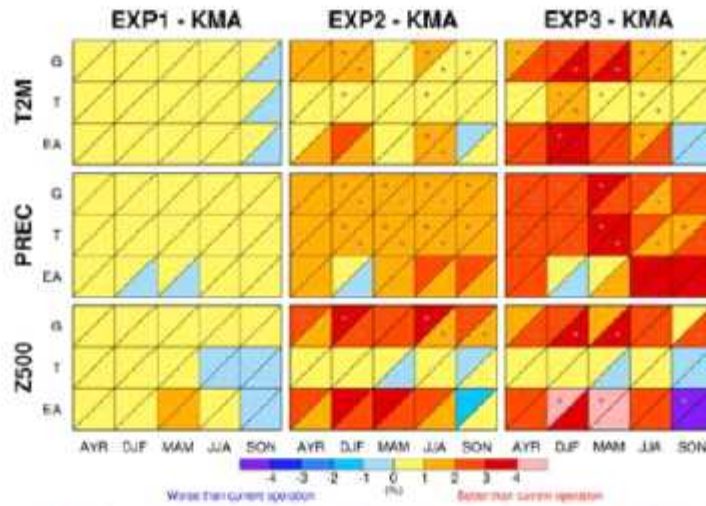


그림 4 기온, 강수, 500hPa 적위고도의 계절별/지역별 '19년 KMA 현업 기후값 대비 각 기후값 실험의 상대적 제원성능(RMSE) 차이 (단위: %). 확장단과 우하단 삼각형은 각각 선행 1주와 3주를 나타냄. 회색 점은 통계적으로 유의미한 차이를 나타냄

2) 2019년 실시간 예측성능

[가] 현업 실시간 예측 대비 상대적 ACC 차이

$$\text{Relative difference(\%)} = \frac{ACC_{EXP} - ACC_{KMA}}{ACC_{KMA}} \times 100$$

- 변수별/계절별 차이는 있으나 전반적으로 각 실험의 예측성능은 현업에 비해 향상되었고 이 중 EXP3의 예측성능이 가장 크게 향상됨. 특히 EXP3를 사용한 동아시아 예측성능은 가을철에 향상되는 경향이 있음 (그림 5).

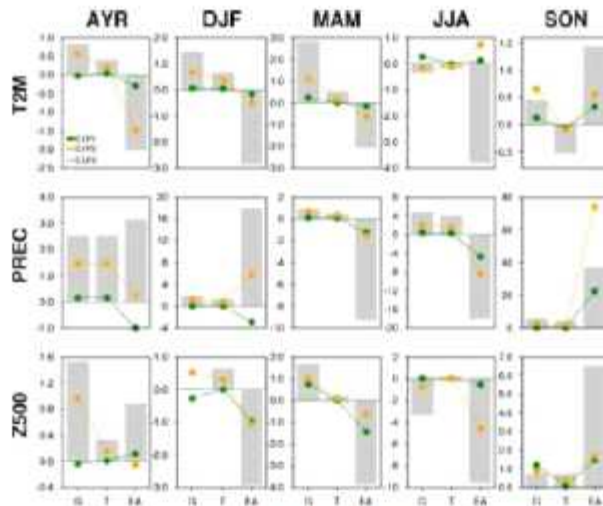


그림 5 선행시간 3주의 지역별/계절별 기온, 강수, 500hPa 지위고도의 '19년 KMA 현업 실시간 예측성능(ACC) 대비 각 기후값 실험용 적용한 예측성능의 상대적 차이 (단위: %).

[나] KMA 현업 실시간 예측 대비 ACC 차이의 분포

$$ACC\ difference = ACC_{EXP} - ACC_{KMA}$$

- 변수별/지역별 차이가 있으나 전반적으로 EXP3의 예측성능이 다른 실험에 비해 향상됨 (그림 6).

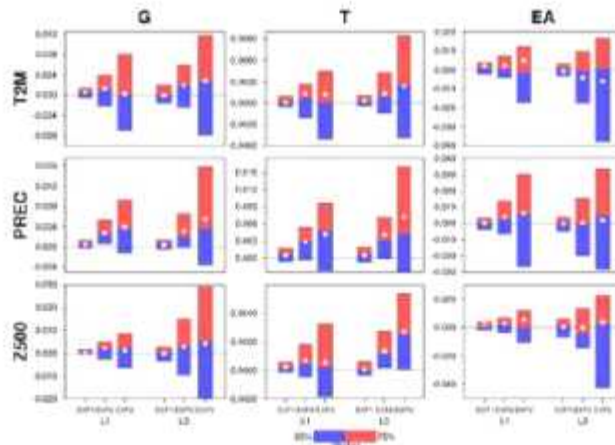


그림 6 전계절 지역별 기온, 강수, 500hPa 지위고도에 대한 '19년 예측 성능: 현업과 각 기후값 실험 간 성능 차이의 상/하위 25% 분포 중앙값, 평균값.

5. 결론 및 제언

- KMA와 UKMO의 기후값에서 공통적으로 나타나는 겨울철 남극의 극단적인 warm bias와 대륙 전반적인 cold bias가 기후값 실험을 통해 일정 수준 보정되었고, EXP3의 보정효과가 가장 큼.
- 현업과 각 기후값 실험의 상대적인 재현/예측성능을 분석한 결과, 기후값 재현성능이 좋을수록 실시간 예측성능 역시 좋아지는 경향이 있으나, 변수별/계절적 차이가 있음.
- 전반적으로 EXP3 방법의 과거재현성능과 실시간 예측성능이 기존 현업에 비해 좋은 것으로 나타남. 특히 동아시아 예측성능의 경우는 가을월에 앙상블 증가의 효과가 더 커짐. 현재 사용한 샘플이 2019년에 한정되어 있기 때문에 향후 샘플이 확대되면 안정적인 결과를 얻을 수 있을 것으로 기대됨.
- 최종적으로, 기후값 계산에 hindcast 앙상블 멤버간 거리가중치를 주는 것 보다 예측시점에서 다소 멀어지더라도 사용하는 기후값 날짜 수를 늘리는 방향으로 리소스를 집중하는 것이 효과적일 것으로 판단됨.

Reference

- 함현준, 원덕진, 이예숙, 2017: 고해상도 장기예측시스템의 주변 앙상블 예측자료 성능 평가. 대기 27(1), 599-615.
- Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B. McGavock, J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution from Multi-Satellite Observations. *J. Hydrometeor.*, 2(1), 36-50.
- Malardel, S., N. Wedi, W. Deconinck, M. Diamantakis, C. Kuhnlein, O. Mozdzyński, M. Hamrud, and P. Smolarkiewicz, 2015: A new grid for the IFS. Newsletter No. 146 - Winter 2015/16, ECMWF, 6 pp.

부록 C. GloSea5 MJO 오차 진단 툴 매뉴얼

GloSea5 현상기반(MJO) 오차 진단 툴 매뉴얼

2020년 12월

APEC 기후센터 기후사업본부
예측기술과 정여민, 함수련

차 례

I. 서론	1
II. 주요 내용	2
1. 모델 및 재분석자료 처리 과정	2
1) GloSea5 자료 전처리	2
2) 재분석 자료 전처리	2
3) GloSea5 및 재분석 자료 후처리	2
2. GloSea5 계절내 변동성 MJO 분석	3
1) 과거재현실험(HINDCAST)	3
2) 미래예측실험(FORECAST)	3
3) 사례연구(CASE STUDY)	3
III. MJO 오차 진단 툴	4
1. 디렉토리 구조 및 코드 설명	4
1) 디렉토리 구조	4
2) 코드 설명	5
2. 흐름도	8
3. 수행 방법	9
1) GloSea5 MJO 오차 진단 툴 설치	9
2) 파라미터 설정	9
3) GloSea5 MJO 오차 진단 툴 수행	10
4) GloSea5 MJO 오차 진단 툴 최종 결과물 예시	11
참고문헌	14

I 서론

APEC 기후센터(APEC Climate Center, APCC) 기후사업본부에서는 2019년부터 기상청 기후예측시스템(KMA-Met Office Joint Seasonal Forecasting System)인 GloSea5(Global Seasonal Forecast System version5) 모델의 계절내(Subseasonal to seasonal) 예측결과의 예측성능 평가 및 오차 분석을 통한 개선 지원 업무를 진행 중이다. 더불어 학계와 유관기관의 GloSea5의 개선을 위한 다양한 연구도 진행되고 있으며, 이를 반영할 수 있는 일관된 오차 진단 시스템의 필요성에 따라 GloSea5 계절내 현상기반 오차진단 틀을 개발하였다. 특히, 계절내 시간 규모의 대표적인 열대 대류 현상으로 알려져 있는 매든-줄리안 진동(Madden-Julian Oscillation, MJO) 현상을 대상으로 오차 진단 틀을 구상하여, 선행시간에 따른 GloSea5 모델의 계절내 변동성 예측성 및 오차 분석을 용이하게 하였다.

GloSea5 계절내 현상기반(MJO) 오차 진단 틀은 1991년부터 2010년을 기준으로 산출되는 과거재현실험(HINDCAST)과 2016년부터 현업 예측 자료로 사용되는 예측실험(FORECAST), 그리고 GloSea5 모델과 기타 S2S 모델, 또는 개선/개발된 버전의 GloSea5 모델의 비교 분석이 가능한 사례연구(CASE STUDY)로 분류하여, 사용자의 필요에 따른 오차 진단이 가능하도록 설계하였다. 해당 오차 진단 틀은 HINDCAST, FORECAST, CASE STUDY 분석에 필요한 자료 처리 과정과 MJO 계절내 변동성 분석에 필수적인 그림을 제공하여 GloSea5 모델의 개선/개발 결과를 반영한 계절내 예측성 및 오차 진단에 지속적으로 활용하고자 한다.

II 주요내용

1. 모델 및 재분석자료 처리 과정

1) GloSea5 자료 전처리

GloSea5 계절내 예측 결과(S2S)는 매달 1, 9, 17, 25일 예측일 기준으로 $1.5^{\circ} \times 1.5^{\circ}$ 공간해상도의 60일 예측 자료로 제공되며, 해당 자료는 분석을 위해 HINDCAST 기간(1991-2010)의 기후값(climatology) 기준으로 편차값(anomaly)을 산출하는 전처리 과정을 수행한다.

- 결과 예시 : GL5_{VAR}_HCST_1991-2010_{MMDD}_ANO.nc
- 변수 정보 : VAR [year(20), time(60), lat(121), lon(240)]

2) 재분석 자료 전처리

GloSea5 계절내 예측 결과의 오차 진단을 위해 사용된 재분석 및 관측 자료는 ERA5(ECMWF)와 GPCP(NASA)이며 모델과의 비교를 위한 공간해상도와 시간규모 변환 과정을 수행한다. 재분석 자료를 $1.5^{\circ} \times 1.5^{\circ}$ 공간해상도로 재격자화된 일 단위 자료로 변환하여, GloSea5 계절내 예측 결과와 동일한 예측일 기준 60일 자료를 생산하고, 분석을 위해 HINDCAST 기간(1991-2010)의 기후값(climatology) 기준으로 편차값(anomaly)을 산출하는 전처리 과정을 수행한다.

- 결과 예시 : ERA5_{VAR}_HCST_1991-2010_{MMDD}_ANO.nc
- 변수 정보 : VAR [year(20), time(60), lat(121), lon(240)]

3) GloSea5 및 재분석 자료 후처리

GloSea5 선행 주수(Lead week)에 따른 계절내 변동성 분석을 위해 앞서 생산한 예측일 기준 60일 편차값 자료를 변환하는 후처리 과정을 수행한다. 선행 주수(Lead week 1-4)에 따른 분석 자료 생산은 표 1에 제시한 것과 같이 선행 1주(W1)는 예측일(Initial time) 기준으로 1-8일 자료를, 선행 2주(W2)는 예측일 기준 9-16일 자료를 각각 수집하여 분석기간(예: HINDCAST의 경우 20년)의 365일 연간자료를 생산하여 분석에 활용한다.

- 결과 예시 : GL5_{VAR}_HCST_1991-2010_POST_W1.nc
- 변수 정보 : VAR [time(20*365), lat(121), lon(240)]

<표 1> GloSea5 계절내 변동성 분석에 활용되는 후처리 분석 자료의 초기장 정보 예시

Lead week	Initial time											
	01-01	01-09	01-17	01-25	02-01	02-09	02-17	02-25	-	12-09	12-17	12-25
W1												
W2												
W3												
W4												

II 주요내용

2. GloSea5 계절내 변동성 MJO 분석

1) 과거재현실험(HINDCAST)

GloSea5 HINDCAST 결과의 계절내 변동성의 오차 진단을 위해 후처리 과정을 통해 생산한 선행 주수(W1-4) 분석 자료를 사용하여 관측과 GloSea5 모델의 선행 1~2주(W12)와 선행 3~4주(W34) 결과를 비교한다. 분석에 사용되는 주요 변수는 U850(하층 수평 바람장), U200(상층 수평 바람장), VP200(상층 속도 포텐셜), PREC(강수량)이며, 계절내 변동성 특징 비교를 위해 20~70일 band pass filter를 적용한다. 평균장 및 변동장의 공간 분포 비교, 파워 스펙트럼(Power spectra) 분석, 파수 진동수(Wavenumber frequency) 분석, Multivariate EOF 분석은 W12, W34 기준으로 관측과 제시하며, MJO 지수(Velocity Potential MJO Index, VPM)의 Bivariate ACC 및 Bivariate RMSE 통계값은 선행 일수 기준으로 관측과 제시한다.

2) 미래예측실험(FORECAST)

GloSea5 FORECAST 결과의 계절내 변동성의 오차 진단을 위해 후처리 과정을 통해 생산한 선행 주수(W1-4) 분석 자료를 사용하여 관측과 GloSea5 모델의 선행 1~2주(W12)와 선행 3~4주(W34) 결과를 비교한다. 분석에 사용되는 주요 변수는 U850(하층 수평 바람장), U200(상층 수평 바람장), VP200(상층 속도 포텐셜), PREC(강수량)이며, 계절내 변동성 특징 비교를 위해 20~70일 band pass filter를 적용한다. 평균장 및 변동장의 공간 분포 비교, 지연 시간-경도 교차 상관(Lag time-longitude cross correlation) 분석, 파수 진동수(Wavenumber frequency) 분석, Multivariate EOF 분석은 W12, W34 기준으로 관측과 제시하며, MJO 지수(Velocity Potential MJO Index, VPM)의 Bivariate ACC 및 Bivariate RMSE 통계값은 선행 일수 기준으로 관측과 제시한다.

3) 사례연구(CASE STUDY)

GloSea5 CASE STUDY 결과의 계절내 변동성의 오차 진단을 위해 후처리 과정을 통해 생산한 선행 주수(W1-4) 분석 자료를 사용하여 관측과 GloSea5 모델, 비교 대상 모델의 선행 1~2주(W12)와 선행 3~4주(W34) 결과를 비교한다. 분석에 사용되는 주요 변수는 U850(하층 수평 바람장), U200(상층 수평 바람장), VP200(상층 속도 포텐셜), PREC(강수량)이며, 계절내 변동성 특징 비교를 위해 20~70일 band pass filter를 적용한다. 평균장 및 변동장의 공간 분포 비교, 시간-경도 Hovmoller diagram 분석, 지연 시간-경도 교차 상관(Lag time-longitude cross correlation) 분석, Multivariate EOF 분석은 W12, W34 기준으로 관측과 제시하며, MJO 지수(Velocity Potential MJO Index, VPM)의 Bivariate ACC 및 Bivariate RMSE 통계값은 선행 일수 기준으로 관측과 제시한다.

III MJO 오차 진단 툴

1. 디렉토리 구조 및 코드 설명

1) 디렉토리 구조

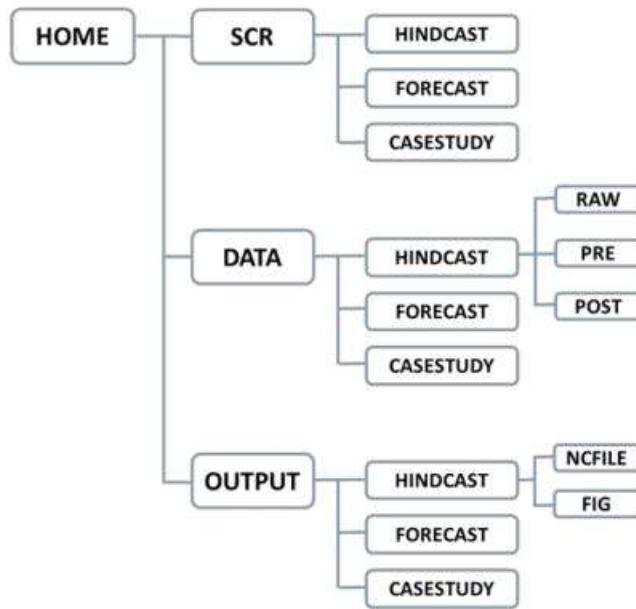


그림 1. GloSea5 계절내예측 현상기반 오차진단 툴 디렉토리 구조

GloSea5 계절내예측 현상기반(MJO) 오차 진단 툴은 과거재현실험(HINDCAST)과 예측실험(FORECAST), 그리고 GloSea5 모델과 기타 S2S 모델, 또는 개선/개발된 버전의 GloSea5 모델의 비교 분석이 가능한 사례연구(CASE STUDY)로 분류하여, 사용자의 필요에 따른 실험 결과의 오차 진단이 가능하다. [SCR] 경로에 모델 및 관측의 자료처리 코드와 계절내 변동성 분석 코드가 실험 분류에 따라 위치한다. 자료처리 코드의 결과물은 [DATA] 경로로 저장되고, 계절내 변동성 분석 코드의 결과물은 [OUTPUT] 경로로 저장된다.

III MJO 오차 진단 틀

2) 코드 설명

• HINDCAST

- `$(HOME)/SCR/HINDCAST`

1-0.Pre_process_data.csh

: HINDCAST 실험 분석시 모델 및 재분석자료 일괄 처리 쉘 코드

- 결과 `$(HOME)/DATA/HINDCAST/PRE(POST)`

1-0-1.HCST_GL5_ANO.ncl

1-0-2.HCST_GL5_POST_by_LW.ncl

1-0-3.HCST_ERA5_RAW2DAILY.ncl

1-0-4.HCST_ERA5_DAILY2ANO.ncl

1-0-5.HCST_ERA5_FORMATTING.ncl

1-0-6.HCST_ERA5_POST_by_LW.ncl

※ 주의 | 단일 쉘 작업시 HPC CPU 소모량이 크기 때문에 PBS 사용 권장

1-1.Cal_mean.variance.field.ncl

: 관측과 모델(W12,W34)의 평균·변동 공간분포 가시화 코드

- 결과 `$(HOME)/OUTPUT/HINDCAST/FIG`

1-2.Cal_power.spectrum.ncl

: 관측과 모델(W12,W34)의 주파수 분석 가시화

- 결과 `$(HOME)/OUTPUT/HINDCAST/FIG`

1-3.Cal_wavenumber.freq.ncl

: 관측과 모델(W12,W34)의 파수 진동수 분석 가시화

- 결과 `$(HOME)/OUTPUT/HINDCAST/FIG`

1-4.Cal_MEOF.ncl

: 관측과 모델(W12,W34)의 MJO 변수 다변량 경험적고함수 가시화 코드

- 결과 `$(HOME)/OUTPUT/HINDCAST/FIG`

1-5.Cal_ACC.RMSE.ncl

: 모델의 예측 일수에 따른 MJO지수 예측성능 및 오차 분석 가시화

- 결과 `$(HOME)/OUTPUT/HINDCAST/FIG`

III MJO 오차 진단 틀

2) 코드 설명

• FORECAST

- \${HOME}/SCR/FORECAST

2-0.Pre_process_data.csh

: FORECAST 실험 분석시 모델 및 재분석자료 일괄 처리 쉘 코드

- 결과 \${HOME}/DATA/FORECAST/PRE(POST)

2-0-1.FCST_GL5_ANO.ncl

2-0-2.FCST_GL5_POST_by_LW.ncl

2-0-3.FCST_ERA5_RAW2DAILY.ncl

2-0-4.FCST_ERA5_DAILY2ANO.ncl

2-0-5.FCST_ERA5_FORMATTING.ncl

2-0-6.FCST_ERA5_POST_by_LW.ncl

※ 주의 | 단일 쉘 작업시 HPC CPU 소모량이 크기 때문에 PBS 사용 권장

2-1.Cal_mean.variance.field.ncl

: 관측과 모델(W12,W34)의 평균·변동 공간분포 가시화 코드

- 결과 \${HOME}/OUTPUT/FORECAST/FIG

2-2.Cal_time-lon.correlation.ncl

: 관측과 모델(W12,W34)의 지연 시간-경도 교차 상관 분석 가시화 코드

- 결과 \${HOME}/OUTPUT/FORECAST/FIG

2-3.Cal_wavenumber.freq.ncl

: 관측과 모델(W12,W34)의 파수 진동수 분석 가시화 코드

- 결과 \${HOME}/OUTPUT/FORECAST/FIG

2-4.Cal_MEOF.ncl

: 관측과 모델(W12,W34)의 MJO 변수 다변량 경험적교함수 가시화 코드

- 결과 \${HOME}/OUTPUT/FORECAST/FIG

2-5.Cal_ACC.RMSE.ncl

: 모델의 예측 일수에 따른 MJO지수 예측성능 및 오차 분석 가시화 코드

- 결과 \${HOME}/OUTPUT/FORECAST/FIG

III MJO 오차 진단 틀

• CASE STUDY

- \${HOME}/SCR/CASESTUDY

3-0.Pre_process_data.csh

: CASE STUDY 실험 분석시 모델 및 재분석자료 일괄 처리 쉘 코드

- 결과 \${HOME}/DATA/CASESTUDY/PRE(POST)

※ 주의 | 단일 쉘 작업시 HPC CPU 소모량이 크기 때문에 PBS 사용 권장

3-1.Cal_mean.variance.field.ncl

: 관측과 GL5, 비교 모델의 평균·변동 공간분포 가시화 코드

- 결과 \${HOME}/OUTPUT/CASESTUDY/FIG

3-2.Cal_time-lon.hovmoller.ncl

: 관측과 GL5, 비교 모델(W12,W34)의 시간-경도 hovmoller 가시화 코드

- 결과 \${HOME}/OUTPUT/CASESTUDY/FIG

3-3.Cal_time-lon.correlation.ncl

: 관측과 GL5, 비교 모델(W12,W34)의 시간-경도 상관 분석 가시화 코드

- 결과 \${HOME}/OUTPUT/CASESTUDY/FIG

3-4.Cal_MEOF.ncl

: 관측과 GL5, 비교 모델의 다변량 경험적함수 가시화 코드

- 결과 \${HOME}/OUTPUT/CASESTUDY/FIG

3-5.Cal_ACC.RMSE.ncl

: GL5과 비교 모델의 MJO지수 예측성능 및 오차 분석 가시화

- 결과 \${HOME}/OUTPUT/CASESTUDY/FIG 코드

[관측 및 재분석자료 일괄 처리 쉘 코드]

- \${HOME}/SCR/HINDCAST/1-0.Pre_process_data.csh

- \${HOME}/SCR/FCSTCAST/2-0.Pre_process_data.csh

- \${HOME}/SCR/CASESTUDY/3-0.Pre_process_data.csh

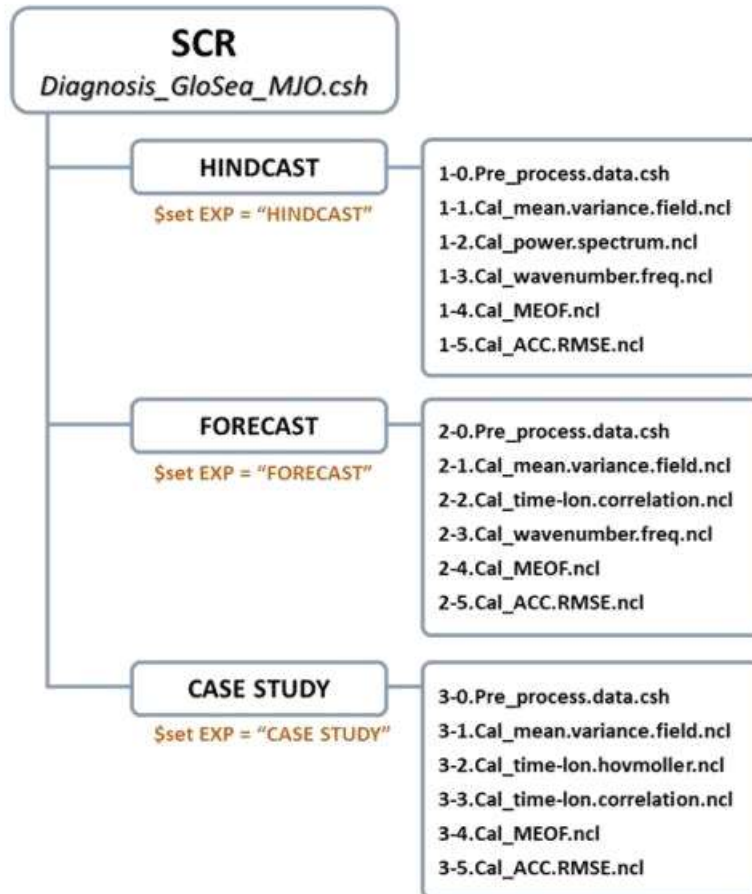
: 관측 및 재분석자료 일괄처리 쉘 코드

※ 주의 | 단일 쉘 작업시 HPC CPU 소모량이 크기 때문에 PBS 사용 권장
(HINDCAST 기준 약 30시간 소요)

III MJO 오차 진단 틀

2. 흐름도

- `$(HOME)/SCR/Diagnosis_GloSea_MJO.csh`



[자동화 코드]

- `$(HOME)/SCR`

`Diagnosis_GloSea_MJO.csh`

: 오차 진단 총괄 스크립트 (분석 실험[EXP] 선택)

※ 주의 | 작업시 HPC CPU 소모량이 크기 때문에 PBS 사용 권장

III MJO 오차 진단 툴

3. 수행방법

1) GloSea5 MJO 오차 진단 툴 설치

```
> cp Diagnosis_GloSea_MJO.tar ${HOME}
> cd ${HOME}
> tar -xvf Diagnosis_GloSea_MJO.tar
> ls
DATA Diagnosis_GloSea_MJO.tar OUTPUT SCR
```

2) 파라미터 설정

```
> cd ${HOME}/SCR
> vi 0.Parameter.Users.ncl
```

```
begin
*****
; Set Parameter
*****
; "Date information for HINDCAST"
sHYear = 1991
eHYear = 2010
; "Date information for FORECAST"
sFYear = 2016
eFYear = 2019
; "Date information for CASE STUDY"
sCYear = 2017
eCYear = 2018
sCMon = 10
eCMon = 12
*****
; Set Directories
*****
DirSCR      = "/"
DirDATA     = "./DATA/"
DirOUTPUT   = "./OUTPUT/"
*****
; Set Subseasonal data | No need to change
*****
. . .
```

III MJO 오차 진단 툴

3) GloSea5 MJO 오차 진단 툴 수행

```
> cd ${HOME}/SCR
> csh Diagnosis_GloSea_MJO.csh
```

```
### 1. User Option #####
set EXP = "HINDCAST" # ("FORECAST"/"CASESTUDY")
set SYear = 1991
set EYear = 2010

### 2. Set Directories #####
set DirSCR = ./${EXP}/
set DirDATA = ../DATA/${EXP}/
set DirOUTPUT = ../OUTPUT/${EXP}/
set DirFIG = ../OUTPUT/${EXP}/FIG/

### 3. Run script by EXP #####

if (${EXP} = "HINDCAST" ) then
echo "[step 1-0] Pre-processing for HINDCAST"
csh ${DirSCR}/1-0.Pre_process_data.csh

echo "[step 1-1] mean & variance field for HINDCAST"
ncl - Q ${DirSCR}/1-1.Cal_mean.variance.field.ncl

echo "[step 1-1] power.spectrum for HINDCAST"
ncl - Q ${DirSCR}/1-2.Cal_power.spectrum.ncl
. . .

else if (${EXP} = "FORECAST" ) then
echo "[step 2-0] Pre-processing for FORECAST"
csh ${DirSCR}/2-0.Pre_process_data.csh

echo "[step 2-1] mean & variance field for FORECAST"
ncl - Q ${DirSCR}/2-1.Cal_mean.variance.field.ncl

echo "[step 1-1] power.spectrum for HINDCAST"
ncl - Q ${DirSCR}/2-2.Cal_time-lon.correlation.ncl
. . .

end if
echo "--- END ---"
```

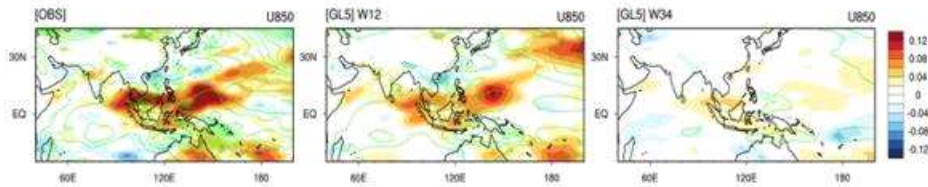
III MJO 오차 진단 틀

4) GloSea5 MJO 오차 진단 틀 최종 결과물 예시

- 예시 : HINDCAST 결과물

① FIG.1-1.mean.variance.field

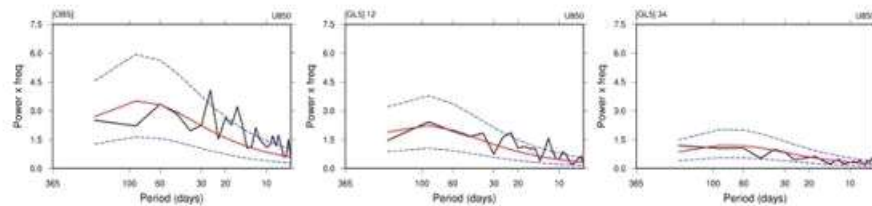
GloSea5 과거재현실험의 적도 변동성의 공간 분포 특징을 분석하기 위해 관측 및 모델의 평균 공간분포(shading)와 변동성 공간분포(contour)를 제시하였다. 모델은 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34)을 분석 기준으로 설정하였다. 계절내 변동성 분석을 위해 20~70일 band pass filter를 적용하였다. 제시된 그림은 하층 동서 바람장(U850)의 겨울철(11월 - 4월) 평균 및 변동성 공간분포이다.



< GloSea5 MJO 오차 진단 틀 결과물 예시 FIG.1-1.mean.variance.field >

② FIG.1-2.power.spectrum

GloSea5 과거재현실험의 계절내 변동성이 크게 나타나는 적도 인도양 지역의 관측 및 모델의 파워 스펙트럼 분석 결과를 제시하였다. 모델은 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34)을 분석 기준으로 설정하였다. 제시된 그림은 하층 동서 바람장(U850)의 겨울철(11월 - 4월) 파워 스펙트럼 결과이다.



< GloSea5 MJO 오차 진단 틀 결과물 예시 FIG.1-2.power.spectrum >

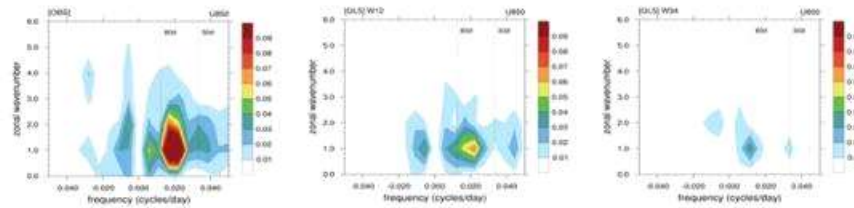
III MJO 오차 진단 틀

4) GloSea5 MJO 오차 진단 틀 최종 결과물 예시

- 예시 : HINDCAST 결과물

③ FIG.1-3.wavenumber.freq

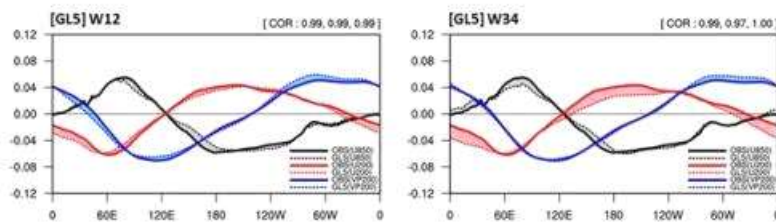
GloSea5 과거재현실험의 관측 및 모델의 적도지역 파수 진동수 분석 결과를 제시하였다. 모델은 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34)을 분석 기준으로 설정하였다. 제시된 그림은 하층 동서 바람장(U850)의 겨울철(11월 - 4월) 파수 진동수 결과이다.



< GloSea5 MJO 오차 진단 틀 결과물 예시 FIG.1-3.wavenumber.freq >

④ FIG.1-4.MEOF

GloSea5 과거재현실험 기간의 적도 변동성을 살펴보기 위해 MJO 지수 (Velocity Potential MJO Index, VPM) 분석에 사용되는 하층 동서 바람장(U850), 상층 동서 바람장(U200) 및 상층 발산장(VP200)의 다변량 직교함수 분석 (Multivariate Empirical Orthogonal Function, MEOF) 첫 번째 모드 결과물 제시하였다. 관측 및 모델의 결과는 실선과 점선으로 각 분석 변수를 나타냈다. 모델은 선행 1-2주 평균(W12)과 선행 3-4주 평균(W34) 기준으로 제시하였고, 관측과의 차이를 면적(shading)으로 나타냈다.



< GloSea5 MJO 오차 진단 틀 결과물 예시 FIG.1-4.MEOF >

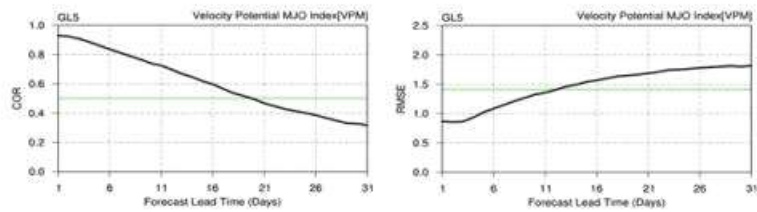
III MJO 오차 진단 툴

4) GloSea5 MJO 오차 진단 툴 최종 결과물 예시

- 예시 : HINDCAST 결과물

⑤ FIG.1-5.ACC.RMSE

GloSea5 과거재현실험의 MJO 지수(Velocity Potential MJO Index, VPM) 예측성능(Bivariate Anomaly Correlation Coefficient, COR)과 오차(Bivariate Root Mean Square Error, RMSE)를 예측 일수에 따라 제시하였다.

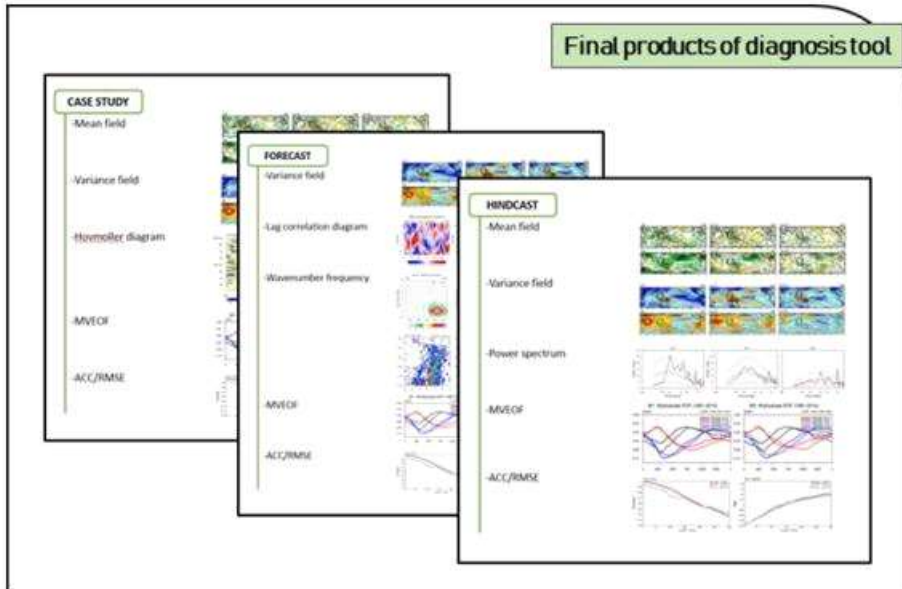


< GloSea5 MJO 오차 진단 툴 결과물 예시 FIG.1-5.ACC.RMSE >

III MJO 오차 진단 툴

4) GloSea5 MJO 오차 진단 툴 최종 결과물 예시

- 예시 : 최종 결과물



※ 참고문헌

신현희, 김가영, 2018, 동아시아 겨울몬순지수 하이브리드 예측시스템 운영매뉴얼, APEC 기후센터

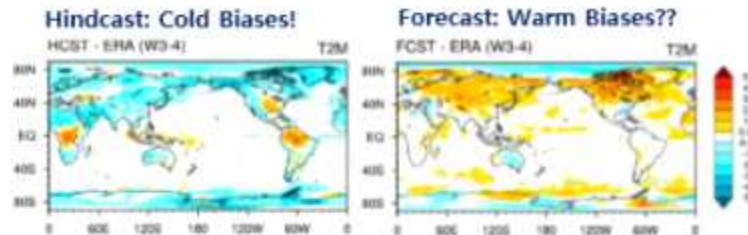
정여민, 2019, GloSea5 자료처리 매뉴얼, APEC 기후센터

함수련, 정여민, 2020, GloSea5 기술보고서: 초기 오차가 계절내 예측 성능에 미치는 영향, APEC 기후센터

부록 D. Met Office-KMA Wiki science page 결과 공유

How important is it to have consistency between the initial condition of the hindcast and real-time forecast?

Purpose and current status



- Why the forecast (18-19) temperature biases show to different with hindcast (92-10) temperature biases?
 - How much can be different initial condition impact the sub-seasonal forecast?
- * Atmosphere initial condition – Hindcast: ERA-interim, Forecast: KMA-NWP forecast

Experimental Design

* For Forecast case (Oct2017-Dec2018 initial with 4-ensemble), N216L85/ORCA025L75 GC2 version:

- NWPiF: KMA-NWP initial data (Current operation system)
 - ERAiF: ERA-interim initial data (Same as Hindcast system)
- ⇒ Only different to atmospheric forcing

Results for Sub-seasonal prediction (Weeks 3-4)

- NWP initial data can lead to the increased temperature biases in subseasonal forecast, especially Arctic, East Asia, and North America. (Fig. 1) [DJF dataset]

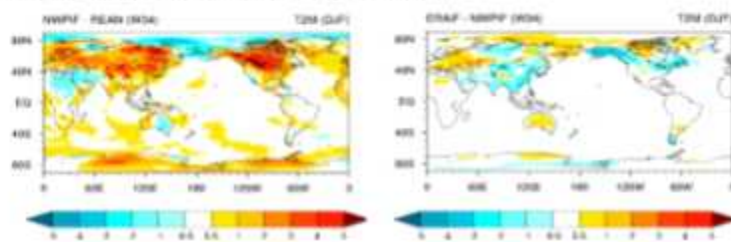


Figure 1 Difference distribution of surface temperature of lead week 3-4 between NWPiF and ERA5 (left), and between ERAiF and NWPiF (right)

- Anomaly correlation coefficients of surface temperature and 500 hPa geopotential height from ERAiF are generally higher than those from NWPiF. (Fig. 2) [All dataset]

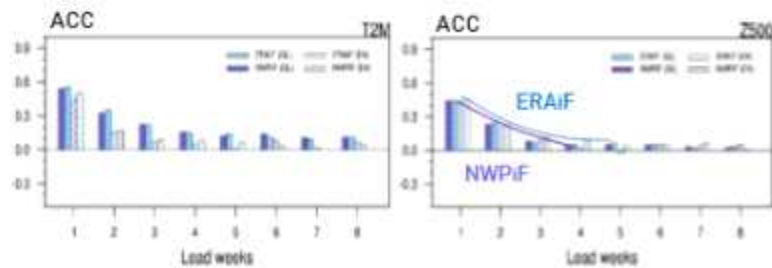


Figure 2 Anomaly correlation coefficient of surface temperature (left) and 500 hPa geopotential height (right) on lead weeks 1 to 8 for global and East Asia domain.

- Forecast using NWP initial data can leads to poor MJO prediction skill from bivariate anomaly correlation coefficient compared to the forecast using reanalysis initial data. (Fig. 3) [NDJFMA dataset]

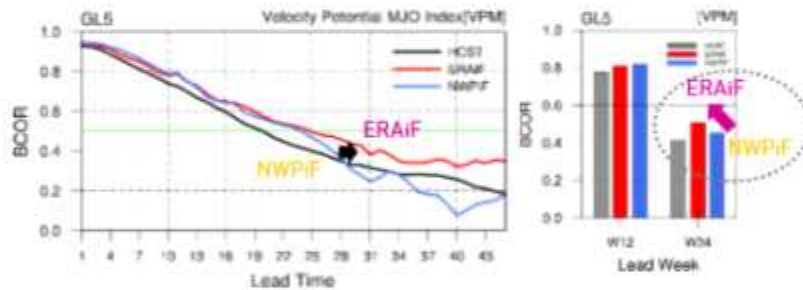


Figure 3 Bivariate anomaly correlation coefficient as a function of forecast days (left) and weeks (right)

Concluding remarks

- In realtime forecast, initial data using NWP prediction (not reanalysis as hindcast) leads to decrease in subseasonal prediction skill due to the fact that includes increased initial errors of soil temperature, cloud amount, radiation forcing, etc.
- The discrepancy of the initial data between hindcast and forecast can affect the information of the anomaly field in realtime forecast due to the difference in systematic error.

From Suryun Ham and Yeomin Jeong at APEC Climate Center (with NIMS/KMA), Korea

(suryun01@gmail.com)