

APEC기후센터

역학적 계절예측의 통계적 후처리의 성능향상을 위한 예측인자 선정 기법 개선

유진호 연구본부

APEC기후센터
연구보고서

역학적 계절예측의 통계적 후처리의 성능향상을 위한 예측인자 선정 기법 개선

Improving Predictor Selection for Better Performing Statistical
Post Process of Dynamical Seasonal Forecast

유진호 연구본부

연구보고서 2015-01

발간사

APEC 기후센터는 2005년 설립 이후 각국에서 생산되는 전지구 기후모형의 자료를 이용하여 계절예측을 생산하고 있습니다. 계절예측은 농업, 수자원, 보건 등 다양한 사회분야에서 한층 수준 높은 의사결정을 지원할 수 있는 높은 가능성을 가지고 있어 전세계 기후서비스의 근간이 되는 핵심 정보입니다.

그러나 계절예측 정보는 기후시스템에 내재한 복잡성으로 인해 불확실성이 클 뿐 아니라 예측 정보 생산의 주요 수단인 전지구 기후모형의 불완전함으로 인해 지역적인 의사결정에 활용되기 위해서는 많은 후처리 작업이 필요합니다. 이러한 후처리 작업은 과거에 생산된 기후모형의 정보와 해당기간의 실제 관측정보 사이의 상호관계를 분석하여 이루어지는데, 현존하는 기후모형의 과거 예측자료 특성상 우리나라가 속한 동아시아 지역 등에서는 통계적으로 유의미한 관계를 찾기 쉽지 않은 상황입니다.

본 연구에서는 우리나라 부근의 7월 강수량에 대한 관측 및 예측 자료의 대규모 특성을 분석하였습니다. 이를 바탕으로 APCC의 다중모델 앙상블(MME) 예측정보에 가장 영향을 크게 미치는 열대 해수면 온도 정보를 이용, 관측정보에서 나타나는 예측성이 낮은 성분을 소거하여 예측 정보의 후처리 과정에서 통계적 불확실성을 감소시킬 방법을 고안하였습니다. 고안된 방법을 통해 후처리 기법의 안정성이 증가하여 향후 계절 예측정보의 신뢰도가 향상되고 그에 따라 사회 각 분야에서 계절 예측 정보의 활용이 증대되기를 기대합니다.

본 연구를 맡아 수행해 주신 유진호 박사와 지원과 조언을 아끼지 않은 연구원 모두에게 감사드립니다.

2015. 3월
APEC 기후센터 소장
정진승

ABSTRACT

A new approach for improving stability and robustness of statistical post processing of dynamical seasonal prediction is proposed. It is to filter out unpredictable noise variability from predictand variable (observation) by assuming that predictable signal is only related with sea surface temperature (SST) variability in the tropics and sub-tropics. In reality, this assumption may not be true but in the context of statistical post-process aiming extraction of information from large scale field of Multi Model Ensemble seasonal prediction result, which will consist of climate feedbacks mostly associated with SST variability.

July mean precipitation over the east Asia was examined as a predictand. The filtering is performed by re-constructing precipitation data from leading SVD (Singular Value Decomposition) mode between east Asian precipitation and tropical and subtropical SST at Indo-Pacific basin. Filtered precipitation data is able to synchronize associated circulation patterns in the MME forecast and that of observation to some degree. A statistical post-process methods based on the stepwise pattern projection method using this filtered observation shows significant improvement of skill compared to that with original unfiltered observation. It is because of robust and organized selection predictor around tropical ocean in the filtered observation case, whereas predictor selection was not systematic and spreaded over the globe in the case of unfiltered observation in the post-process.

Although there are more issues to be resolved for implementation in the operational forecast, this approach suggest a possibility of improving statistical post-processing by re-constructing observation variable prior to correct forecast result. This is particularly effective in the situation of seasonal forecasts with short hindcast length to obtain robust statistical relationship for the conventional methods.

목차

역학적 계절예측의 통계적 후처리의 성능향상을 위한 예측인자 선정 기법 개선

발간사	i
ABSTRACT	iii
1. 서론	1
2. 연구 자료	6
3. 동아시아 강수 변동에 관련된 관측 및 모델의 순환장의 비교	7
3.1 관측의 순환장	7
3.2 예측의 순환장	8
3.3 관측과 예측의 상관성	12
4. 관측 자료의 재구성을 통한 예측성의 향상	15
4.1 관측 자료 재구성의 필요성	15
4.2 관측 자료의 필터	18
4.3 필터된 관측자료의 분석	21
4.4 통계적 후처리 기법의 적용 및 평가	22
5. 요약 및 토의	25
REFERENCES	28

1. 서론

일기예보와 다르게 계절예측은 대상이 되는 계절 내 매일 매일의 날씨 상태를 예측할 수 없고, 그 날씨의 통계적 특성을 예측하는 것으로 일반적으로 계절 평균 기온 및 강수량이 대표적인 예측 대상이 된다. 따라서 예측의 대상이 되는 현상은 매일 매일의 날씨를 좌우하는 이동성 고저기압 등의 단기적으로 변화하는 현상이 아니라 이 날씨 현상보다 느리게 변화하며 나타나는 대기의 저주파 변동(low frequency variability)이다. 이 대기의 저주파 변동은 날씨변동을 만들어내는 시스템보다 공간 규모 역시 크며 전지구 규모에 걸친 원격상관 (teleconnection) 형태로 나타나는 경우가 많아 열대 지방에서 벌어지는 현상이 중위도 지역의 계절 기후에 영향을 미치는 메커니즘이 된다.

이러한 저주파 변동은 해수면 온도의 변화와 같은 대기 하층의 경계조건과의 상호작용을 통해 “강제로(forced)” 만들어 지거나, 대기 내부의 비선형 역학에 의해 대기 “스스로” (free) 생성하게 되는데, 일반적으로 강제로 만들어 진 성분에 대해서는 예측이 가능하나 비선형 역학에 의해 만들어 진 부분은 계절예측이 어려운 것으로 알려져 있다. 따라서 외부에 의해 강제된 저주파 변동(external variability: signal)과 내부 비선형 역학에 의해 만들어진 변동(internal variability: noise)의 상대적인 비가 그 지역의 계절예측의 상한을 결정하게 된다. 현재까지 계절예측에서 열대지방의 경우 external variability, 즉 signal이, 중-고위도 지방의 경우 internal variability, noise가 큰 것으로 밝혀져 있다.

열대지방에서 signal의 큰 부분을 설명하는 현상은 열대태평양 해수면 온도의 경년 변동인 ENSO(El Niño Southern Oscillation)이며 엘니뇨나 라니냐의 발생에 따라 이와 관련된 무역풍 및 열대지방의 대규모 강수형태가 변화하면서 전지구적인 대기순환장의 변화를 초래하여 장기예측을 가능하게 하는 signal을 제공한다. ENSO 외에 열대 인도양이나 대서양의 해수면 온도의 변화 역시 지역적인 영향을 제공하나(Gadgil et al., 2004) ENSO에 비해 그 해수면 온도 예측의 성능 자체가 아직까지는 낮은 수준이어서 계절예측 성능의 많은 부분은 ENSO의 발생 및 강도와 밀접한 관련을 갖고 있다(Shukla et al. 2000).

현재 계절예측은 대기-해양-지면에서 나타나는 물리 과정들을 수치적으로 해석하는 전지구 기후모델(Global Climate Model)을 이용해 수행되고 있는데 이러한 역학적 계절예측

의 태동 역시 ENSO의 성공적인 예측에 기인한 바 크며 ENSO의 모의 성능은 모델 진단 및 예측성 평가의 중요한 기준이 된다. Fig. 1은 7월 달의 관측(a)과 예측(b)에서 나타나는 ENSO와 관련된 열대태평양 지역 강수량의 패턴이다. 관측 Nino3.4 지수와 각 지점 강수량 편차(anomaly) 값의 회귀식을 구하여 산출하였다. El Niño가 발생하면 적도지역에서 중태평양과 서태평양의 강수가 증가하며 Maritime continent 지역의 강수는 감소한다. 이러한 경향은 예측(b)에서도 잘 나타나고 있어 전반적인 모델의 ENSO관련 강수 패턴 예측이 우수한 수준임을 보여준다. 그러나 지역적으로 상세하게 살펴보면 남위 10도의 경우 위의 두가지 패턴의 강수는 관측에서는 El Niño때 증가, 예측에서는 감소하는 상반된 경향을 보이며(Fig. 1c) 동경 160도 근처에 위치하는 국가인 바누아투(Vanuatu) 지점에 해당하는 격자점의 강수 시계열을 보면 예측과 관측의 상관계수가 거의 0이어서 ENSO에 대한 전반적으로 우수하게 평가된 예측정보가 실제로 재해석 없이 활용될 때 열대태평양에 존재하는 어떤 국가에는 크게 유의미한 정보가 아닐 수도 있음을 보여준다.

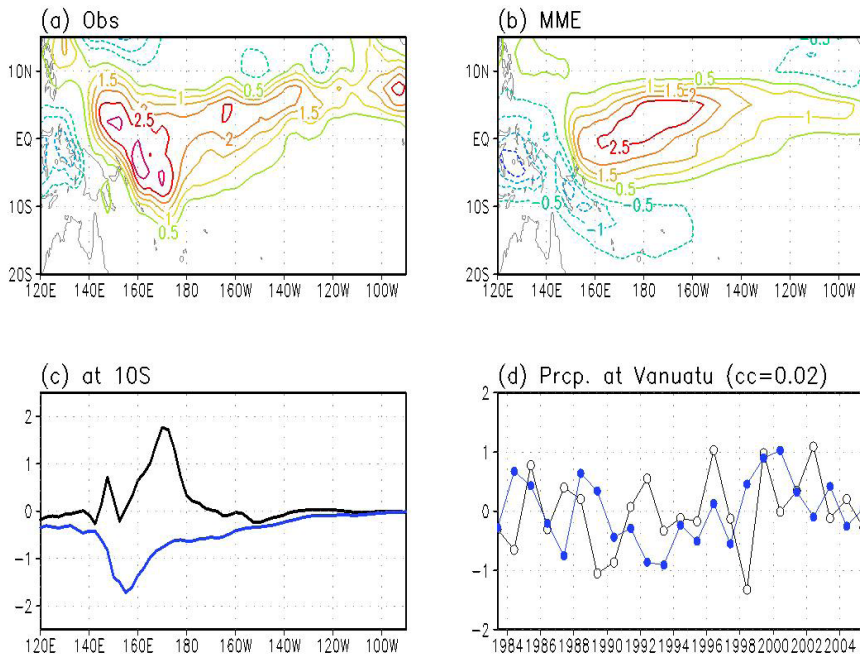


Figure 1. Precipitation anomaly distribution associated with ENSO in (a) observation and (b) MME forecast. (c) ENSO related precipitation anomaly at 10S in observation (black) and MME (blue) (d) Time series of precipitation anomaly at the location of Vanuatu in the observation (black) and MME forecast (blue)

Fig. 1의 경우에서 본 바와같이 다른 기후변동에 비해 비교적 예측성이 높은 ENSO 예측이라고 하더라도 다양한 오차가 발생하며 이러한 오차를 감소시키기 위해 기후모델의 개선이 꾸준히 진행되어 왔다. 모델의 개선을 통한 예측성능의 개선이 궁극적으로 옳은 방향이나 예측 정보의 시급한 수요에 대응하기 위해서는 현존하는 예측 자료 안에서 최선의 정보를 도출하는 통계적 후처리 기술의 개발 역시 병행되어야 하며 계절 예측 분야 초기에서부터 많은 시도들이 있어왔고 일부 성과를 거두었다(e.g. Feddersen 1994, Kang et al. 2004, Delsole and Shukla 2012).

통계적 후처리를 통한 예측성능의 개선은 항상 나타나는 것은 아니고 모델의 예측과 실제가 유의미한 시간적인 상관관계를 갖는 경우에 가능하다. ENSO를 예로 들면 예측된 ENSO정보가 공간적으로는 문제가 있을 수 있으나 ENSO라는 현상으로 표현되는 열대지방의 기후변동성의 경년변동이 실제와 유사하게 표현된다면 이러한 공간형태상의 문제를 예측 이후 통계적으로 해결할 수 있다. 열대 태평양에 강수에 대한 관측과 예측의 첫 번째 EOF(Empirical Orthogonal Function) 모드는 ENSO에 해당하는 모드인데(Not shown) 두 자료의 이 모드에 대한 PC(Principal Component) 시계열의 상관계수는 0.72로 모델의 ENSO 발생과 실제 ENSO 발생은 높은 시간적인 상관관계를 갖는다. 따라서 El Niño가 예측된 경우에 실제로 El Niño가 발생할 가능성이 매우 높으므로 대상 지역에 대한 최종 예보를 생산할 때 예측된 El Niño의 모델 강수 패턴(Fig.1b)이 아닌 El Niño시기에 실제 나타나는 강수 패턴(Fig.1a)을 고려하여 예측정보를 생산하는 것이 가능하다. 이러한 후처리에 필요한 조건은 예측과 관측의 높은 시간적 상관관계와 그에 해당하는 명확한 공간패턴이다.

이러한 장기예측의 통계적 후처리는 CCA(Canonincal Correlation Analysis) 등과 같은 서로 다른 변수 사이의 상관패턴을 찾아내는 통계적 기법을 이용해 주로 수행되었다(e.g. Feddersen 1994). CCA는 두 변수의 특정 공간 패턴의 시간변화의 상관계수를 최대화 하는 패턴과 시계열의 상관모드들을 찾는 방법이고 MCA(Maximum Covariance Analysis) 혹은 SVD(Singular Value Decomposition) 기법은 상관계수가 아니라 공분산을 최대화 하는 모드들을 찾는 기법이다. 상관계수 혹은 공분산을 크게 만드는 상관 모드들을 찾게되므로 이 기법들에 한 변수는 예측, 다른 한 변수를 관측으로 지정하여 적용하는 경우 자연스럽게 앞서 언급한 유의미한 시간적인 상관관계를 찾아내어 오차 보정 혹은 통계적 후처리에 사용할 수 있게 된다.

이러한 상관패턴을 찾는 기법의 활용은 비단 오차의 보정 뿐 아니라 대규모 순환장의 정보를 지역적 기후 패턴으로 전이하는 규모축소(downscaling)의 기법으로도 활용될 수 있는데 APCC에서는 Lee(2003)의 CPPM(Coupled Pattern Projection Method) 기법에 기반한 Kang et al.(2009)이 개발한 기법이 활용되었고 APCC의 다중모델 앙상블(Multi Model Ensemble) 기법 가운데 하나에 사용된 후처리 기법인 SPPM(Stepwise pattern projection model, Kug et al. 2008) 역시 특정 지점의 기후변동과 연관된 패턴을 찾고 이를 이용해 예측 정보를 보정·재생산하고 있다. 이 CPPM이나 SPPM은 특정 지점의 강수 변동 혹은 몬순 지수와 같은 관측된 단일 시계열에 관련된 예측장의 공간패턴을 찾아내는 것으로 관측 변동과 가장 유의미한 상관패턴을 찾아내어 이를 예측의 보정에 활용한다. 이 기법들은 예측장에서 상관패턴을 찾기 위해 패턴을 결정하는 창의 크기나 위치를 변화시키거나(CPPM), 상관계수가 높은 지점을 자유롭게 수집(SPPM)하여 관측 시계열과의 시간상관성을 극대화 하는 패턴을 찾고 있는데 SVD와 CCA 등의 기법을 이용한 방법에 비해 상관패턴을 찾는 방식에 있어 자유도가 큰 편이다. 최적의 상관패턴을 찾기에 유리하다는 장점을 갖는 한편 높은 자유도로 인해 찾아진 패턴이 주어진 자료(sample) 내에서만 최적이고 새로운 자료에 대해서는 낮은 적용성을 갖게 되는 샘플링 문제에 취약하다는 단점을 갖는다.

이러한 단점을 해결하기 위해 기존 APCC에서 운영되던 Kang et al.(2009)의 기법에 여러 가지 예측 변수를 활용하고 임의의 조건을 삽입하여 문제 발생을 억제하려는 시도가 있었다. 그럼에도 불구하고 통계적 규모축소 기법의 과거 예측성과 실제 예측 성능의 괴리가 다수 나타나게 되었다. Fig. 2는 해당 기법을 임의의 백색 잡음 자료에 1000회 적용해 얻은 예측 성능(시간상관계수)의 분포이다. 아무런 관련성이 없는 서로 다른 잡음 자료를 예측과 관측이라 가정하고 이용했으므로 상관계수의 기댓값은 0이 되어야 하며 그러한 경향이 실제 예측조건에서는 나타나나(청색) 과거 예측의 경우는 0.3-0.4 정도의 가상적인 성능(artificial skill)이 나타나고 있음을 알 수 있다. 이는 적은 샘플 자료를 이용해 최대한의 예측 성능을 얻어내고자 하는 가운데 발생한 과대적합(overfitting) 문제이며 이 과대적합 문제는 비단 이 기법만의 문제만이 아니라 과거 예측기간이 짧은 계절예측 분야에서 관측자료와 예측자료를 이용한 통계적 후처리 기법 개발에 모두 경험하는 문제이다.

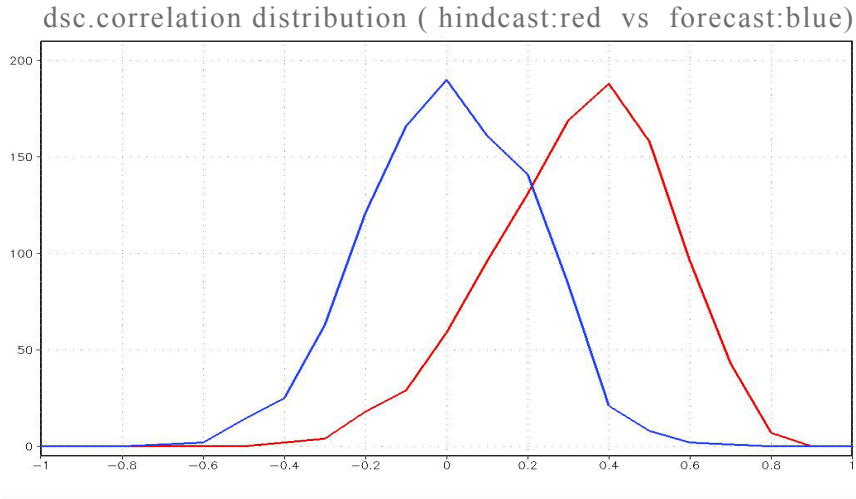


Figure 2. Distribution of correlation skills from 1000 cases using random noise data in hindcast (red) and forecast (blue)

역학적 계절예측의 통계적 후처리가 필요한 이상, 지속적인 후처리 기법의 개선은 필수적이다. 더불어 현재의 기법이 갖는 예측 인자 선정의 자유도나 유연성을 유지하면서 안정적인 예측정보를 도출하는 후처리 기법이 개발된다면 바람직할 것이다. 제한된 샘플로 인한 과대적합 문제 등의 해결을 위한 다양한 시도들이 진행되고 있는 가운데(e.g. Tippett et al. 2014) 본 연구에서는 현 APCC에서 활용 중인 통계적 후처리 기법의 샘플링 문제의 위험을 감소시켜 안정적인 예측정보를 생산하기 위한 방법을 제시하고자 한다.

본 연구에서 제시하고자 하는 방법은 일반적으로 많이 수행하는 예측 정보를 재가공하는 접근 방식과 다르게 잠재예측성(potential predictability)의 개념을 고려하여 관측 자료를 우선 재가공하고 이를 바탕으로 예측정보를 재가공하는 방식으로 이러한 방식의 적용을 통해 안정적인 예측 성능을 얻을 수 있음을 보이고자 한다. 2장에서 사용된 자료를 소개하고 3장에서는 이러한 접근 방법을 도입하기에 앞서 지역 강수와 대규모 순환장의 관계가 관측과 모델 예측에서 어떻게 나타나는지를 비교하여 재가공된 관측 정보에 필요한 정보가 무엇인지를 확인하고 4장에서는 관측정보를 재가공하여 가공하지 않은 관측에 비해 어떤 점들이 개선되었는지 제시할 것이다. 마지막 5장에서 결론과 토의사항을 정리하였다.

2. 연구 자료

NCEP에서 생산한 CMAP(CPC Merged Analysis Precipitation) 월평균 강수량 자료와 ERSST(Extended Reconstructed Sea Surface Temperature) 자료를 사용하였다. 대상 기간은 1982-2005년의 23년이며 이는 모델의 과거예측(hindcast) 기간에 해당하고 7월달의 자료만을 사용하였다. 7월 강수를 택한 것은 동아시아 몬순이 활발할 시기로 가장 사회적인 영향이 큰 변수 일 뿐 아니라 기온이나 다른 순환장에 비해 공간규모가 작고 지역적인 예측성이 낮아 과대적합의 문제를 다루는데 적합하다고 판단했기 때문이다.

APEC 기후센터 다중모델 앙상블(APCC MME)에 참여하는 10개 모델의 과거 예측자료(hindcast)를 활용하였다. 과거 예측자료의 기간은 1982-2005의 23년으로 6월 초기조건으로 예측한 7, 8, 9월에 대한 예측 가운데 선행시간이 1개월인 7월 강수자료를 활용하였다. MME는 10개 모델의 예측 편차값을 단순 평균한 SCM(Simple Composite Method) 자료가 활용되었으며 변수는 강수, 기온, 500hPa 지위고도, 850hPa 바람장 등이다. MME에 참여한 모델은 Table 1에 나타내었다.

Table 1. Participating models in MME of this study

Model Acronym	Institute / Country	System Configuration (ensemble member)
POAMA	Bureau of Meteorology / Australia	Coupled (10)
MSC_CANCM3	Meteorological Service of Canada / Canada	Coupled (10)
MSC_CANCM4	Meteorological Service of Canada / Canada	Coupled (10)
HMC	Hydrometeorological Center of Russia / Russia	2-tier (10)
GDAPS	Korea Meteorological Administration / Korea	2-tier (20)
JMA	Japan Meteorological Agency / Japan	Coupled (10)
NCEP	National Centers for Environmental Prediction / USA	Coupled (15)
CWB	Central Weather Bureau / Chinese Taipei	2-tier (10)
NASA	NASA / USA	Coupled (11)
PNU	Pusan National University / Korea	Coupled (5)

3. 동아시아 강수 변동에 관련된 관측 및 모델의 순환장의 비교

3.1 관측의 순환장

우선 통계적 후처리를 수행하기에 앞서 특정 지점의 강수 변동에 대해서 실제와 모델이 어떤 관련성을 갖는지 비교하여 어떤 변동이 모의되고 있는지 살펴보았다. Fig. 3은 관측 자료에서 나타나는 한반도 중부지방 부근 지점(127.5E, 37.5E)의 7월 평균 강수와 관련된 7월 대규모 순환장의 형태이다. 음영으로 표시된 것은 각 지점 강수와 한반도 강수와 상관계수, 그리고 한반도 강수에 선형회귀된 850hPa 바람 벡터(동서 및 남북 방향의 성분 가운데 한 가지라도 95% 이상의 신뢰도가 있는 경우 표시, 두가지 성분 모두 신뢰도가 유의한 경우는 보라색으로 표시함) 그리고 등치선은 500hPa 지위고도이며 역시 신뢰도가 유의한 경우 녹색으로 표시하였다.

이미 잘 알려진 대로 한반도 지역의 7월 강수는 동서로 이어진 강수밴드와 함께 나타나나 중국의 메이유나 일본의 바이유와는 시기 및 활성정도에서 약간의 차이를 보여 통계적으로 유의한 수준의 상관계수는 한반도 주변 지역에 국한되어 나타난다. 이러한 강수는 한반도 남쪽의 북태평양 고기압 및 그리고 북쪽 기압골의 발달과 관련되어 있으며 이러한 기압 배치에 따라서 남서풍이 한반도로 유입되는 형태를 보인다. 아울러 한반도 남쪽의 발달된 고기압 지역에서는 음의 강수편차가 나타나는 것을 볼 수 있다. 동아시아와 북서 태평양 지역을 제외하고는 다양한 변수가 조직적으로 연결된 패턴은 찾기 어려우며 한반도의 7월 강수를 지배하는 대규모 순환장은 동아시아와 북서태평양에 주로 존재하는 것으로 볼 수 있다.

Fig. 4는 기준이 되는 관측 지점을 2.5도 남쪽으로 내려 한반도 남부지방 부근으로 선정하고 Fig. 3과 같은 방식으로 표현한 그림이다. 단지 거리상으로 250km의 차이가 있을 뿐인데 흥미로운 차이점을 보인다. 우선 대상지역과 상관성이 높은 강수변동 지역이 한반도는 물론 일본열도를 포함하고 있고 한반도를 중심으로 뚜렷한 저기압성 회전이 나타나고 있다. 그 남쪽의 고기압은 남중국해에 한 개의 강한 중심을 가졌던 Fig. 3 과 다르게 두 개로 분리되어 보다 동쪽의 고기압이 더 강한 형태를 띄고 있어 한반도 지역의 저기압과 맞물려 일본 열도 쪽으로 남서기류를 유도하고 있다. 한편 이 고기압성 순환은 열대 태평양 지역에서 나타나는 강수 및 바람장과 보다 조직화 된 형태를 띄고 있는데, 마치 적도지역 중태평양에서 나타나는 양의 강수편차에 의해 유도되는 Gill-type 반응과 유사한 점이 있다. 한가지 주목할 점은

북미 서쪽의 저기압성 순환을 포함해 중위도 지역에서 나타나는 전지구적인 순환이 보다 뚜렷하게 나타난다는 점인데 마치 CGT(Circumglobal Global Teleconnection) 패턴과 유사한 형태를 띤다. 요컨대 남쪽으로 다소 이동한 지점의 강수 변동은 중부지방을 중심으로 한 강수 변동에 비해서 열대 지방은 물론 중위도 지역에도 많은 관련성을 가지며 변동하고 있음을 보여 이 지점에 대한 전지구적 대규모 순환장의 영향이 크고, 이는 예측가능성이 더욱 높을 수 있음을 시사한다.

3.2 예측의 순환장

그렇다면 예측 모델에서는 해당 지점의 강수 변동이 어떤 순환장과 관련이 있을까? Fig. 5는 Fig. 3와 같은 대상 지점에 대해서 MME에 참여한 모델 가운데 하나를 임의로 선택해 같은 방식으로 표현한 것이다. 참고로 모든 계산은 그 모델의 앙상블 평균값을 이용해 수행되었다. 이 모델에서는 한반도 지역의 강수는 관측에서 나타나는 바와 같이 동서 방향으로 연결된 강수밴드가 아닌 동아시아 지역 대륙 연안을 따르는 남서-북동 방향으로 기울어진 강수밴드와 연결되어 나타나고 있다. 이러한 강수밴드는 그 남쪽에서 적도지방에서부터 이어진 말굽 모양의 건조대와 인접해 있다. 이와 관련되어 강수대와 건조대 경계, 일본열도 지역에서 고기압이 발달하였으며 이 고기압은 남중국해에서 북태평양 지역에 까지 길게 이어져 있다. 이러한 기압 배치에 따라 한반도 지역으로 강한 남서기류가 유입되며 동시에 Maritime continent 지역에서부터 인도에 이르는 지역에 동풍이 나타나고 있다.

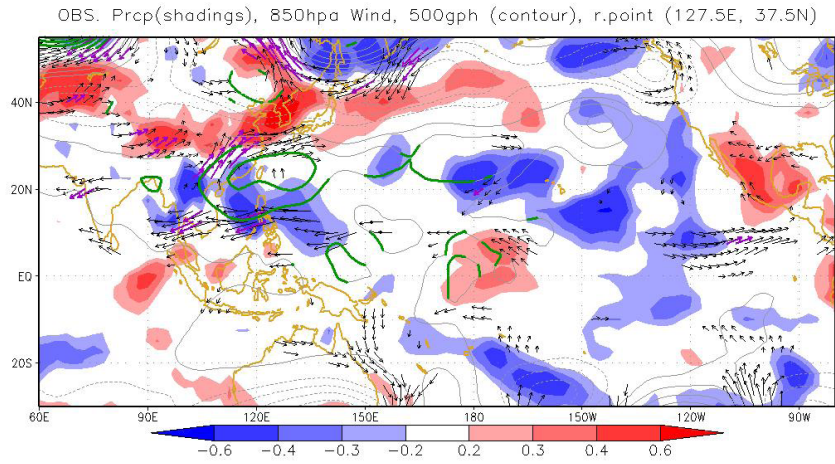


Figure 3. Observed distribution of precipitation (shading), 850hPa wind (vectors), 500hPa geopotential height associated with precipitation at reference point 1 (127.5E, 37.5N). Values shown are correlation (precipitation) and regression coefficients (wind, geopotential height). Statistically significant values (95%) were marked with green contour and violet vectors.

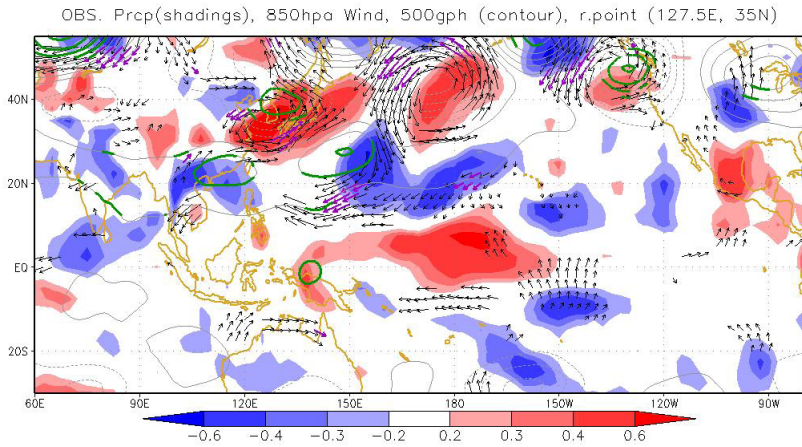


Figure 4. Same as Fig. 3 except for the reference point 2 (127.5E, 35N)

사용된 모델의 자료가 앙상블 평균장이므로 개별 앙상블 자료에서 나타나는 잡음이 소거되어 상관관계의 강도가 강하고 뚜렷하게 나타나는 경향이 있음을 감안하더라도 기본적으로 모델이 모의한 한반도 지역의 강수와 관련된 순환장은 그 구조면에서 큰 차이가 있음을 알 수 있다.

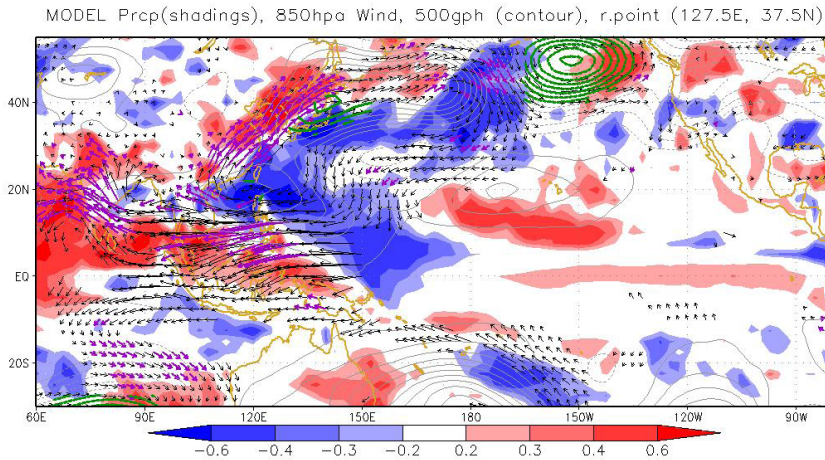


Figure 5. Same as Fig. 3 but for data from a model

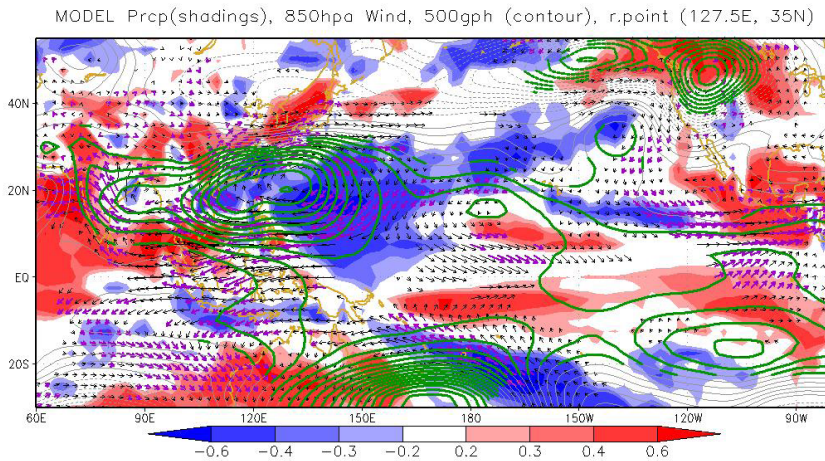


Figure 6. Same as Fig. 4 but for data from a model

한편 2.5도 남쪽의 지점 강수에 대한 모델 자료의 패턴은(Fig. 6) 중부지방 중심의 강수 변동과 관련된 패턴에 비해 남중국해 지역의 고기압의 발달이 두드러진다. 고기압의 중심은 관측의 한반도 중부지방 강수와 관련된 고기압의 중심부(Fig. 3)와 유사하고 통계적으로 유의한 값은 아니나 한반도 북쪽으로 기압이 낮은 곳이 위치하고 있어 그사이에서 남서기류가 한반도로 유입되는 구조를 보이고 있다. 동서방향 순환장의 폭은 관측에 비해 매우 크나 한반도를 중심으로 남북 고-저기압의 배치와 남서기류의 유입은 오히려 같은 지점에 대한 모델 결과(Fig. 5) 보다 관측과 유사한 형태를 보인다.

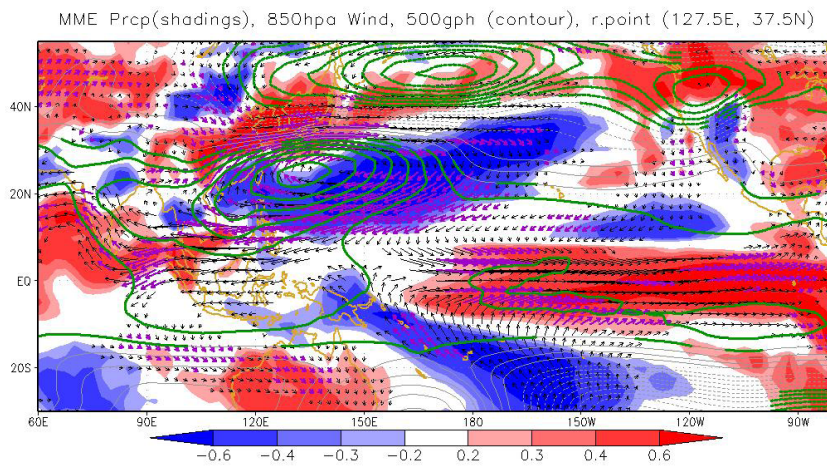


Figure 7. Same as Fig. 5 but for data from MME

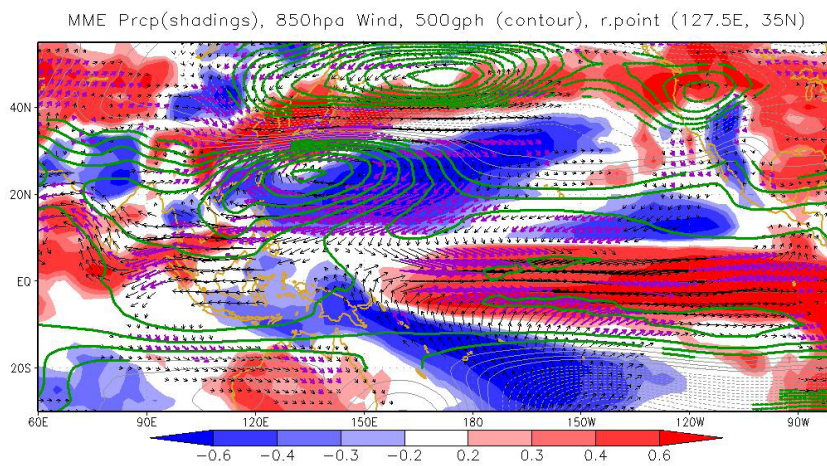


Figure 8. Same as Fig. 6 but for data from MME

그러나 이러한 결과가 모델은 한반도 강수와 관련된 역학적 구조를 남쪽에 모의하는 경향이 있다고 이야기 할 수 있는 수준은 아니다. 여기서 주목할 만 한 결과는 한반도 강수와 관련된 강수 패턴이 관측과 모델에서는 차이를 보이며 그 대상 지역이 다소 남쪽으로 치우쳤을 때 큰 범주에서의 유사성을 더 확인할 수 있다는 사실이다.

여러 개의 모델 결과가 평균된 다중모델앙상블(MME)의 결과가 Fig. 7에 제시되었다. Fig. 6에서 확인할 수 있었던 남고 북저의 기압배치와 그 사이의 강화된 남서풍이 매우 뚜렷하게 나타나고 있으며, 관련된 강수대는 중국에서부터 일본을 지나 알류우산 열도에 까지 이어지고 있다. 또한 그 남쪽의 건조대와 적도태평양 지역의 양의 강수 구역까지 나타나고 있음을 알 수 있다. 기압대의 동서방향 규모가 커졌으며 북쪽 저기압의 중심은 동아시아가 아닌 북태평양 지역에서 나타난다. MME가 모의하는 한반도 강수와 관련된 순환패턴은 대체로 Fig. 6, 즉 특정 모델의 한반도 남부 지역 기준 강수 관련 패턴, 에 나타난 패턴과의 유사성이 크다고 볼 수 있다. MME에서 나타나는 한반도 남부 지역 강수 기준 순환패턴은 중부 지역 강수 기준에 비해 큰 차이가 없다(Fig. 8). 즉, MME에서는 한반도 남부 지역과 중부 지역의 강수가 유의미하게 분리되어 해석하기 힘들다고 볼 수 있다.

단일 모델 앙상블, 다중모델 앙상블의 한반도 강수관련 패턴을 관측과 비교하면 많은 수의 예측 결과가 중첩될수록 상관패턴은 단순하고 큰 규모를 띄게 된다. 동아시아 지역에 나타나는 강수와 관련된 패턴은 관측에서는 한반도를 중심으로 남쪽에 고기압, 북쪽에 저기압이 나타나는 형태이며 그 사이에서 습윤한 남서기류가 유입되어 강수를 만들어 내는 구조이다. 이러한 패턴은 동아시아에 국한되어 비교적 작은 규모에서 나타나고 있다. 반면 MME에서는 대규모 스케일에서 남중국해에 중심을 둔 고기압의 발달과 북태평양에 중심을 둔 저기압의 발달 및 그 사이에서 강수대와 한반도로의 남서기류가 발달하는 형태를 띄어 큰 규모에서의 유사성을 보이고 있다. 이러한 유사성은 특정 모델에서 나타나는 구조적인 편차(Fig. 5) 등이 여러 모델들의 결과가 결합되면서 소거되어 모든 모델들이 모의할 수 있는, 열대 해양과의 상호작용에 주로 기인한 단순한 형태의 순환장만이 남아있기 때문으로 판단된다.

3.3 관측과 예측의 상관성

지금까지는 관측과 예측자료 각각에서 나타나는 특정 지점의 강수에 대한 순환장의 형태와 그 차이를 살펴보았다. 이러한 분석은 모델이 모의하는 역학적인 상호작용의 범위와 그

로인해 예측 정보에서 나타나는 구조적인 오차를 이해할 수 있게 한다. 그러나 실제 예측업무, 그리고 통계적 후처리는 관측자료와 예측자료를 연결짓는 작업으로 관측된 강수량과 모델의 예측자료를 비교할 필요가 있다. Fig. 9는 앞서 제시된 일련의 그림과 같은 방법으로 분석된 결과이나 관측된 한반도 지역 강수량자료와 연관된 MME 예측자료의 순환형태를 나타낸 것이다.

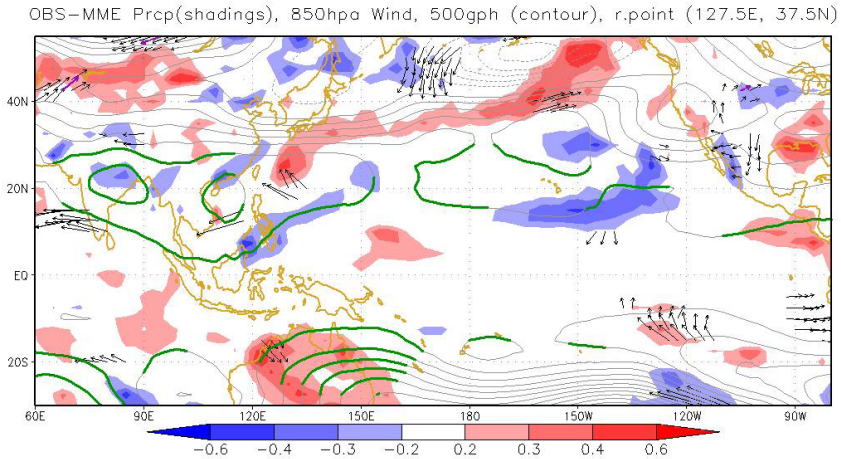


Figure 9. Same as Fig. 7 but the reference data is observed precipitation time series at the reference point

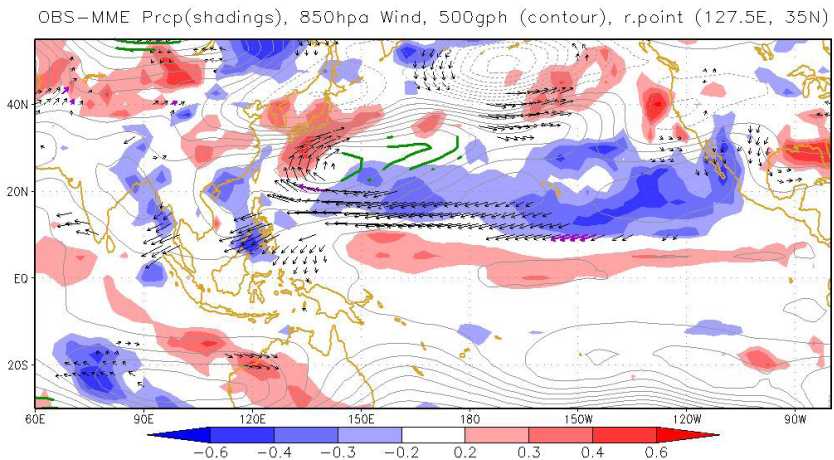


Figure 10. Same as Fig. 8 but for the reference point 2

가장 먼저 주목할 만한 것은 관측의 한반도 지역 강수와 MME 예측의 한반도 지역 강수는 유의미한 상관성을 보이지 않는다는 점이다. 대신 관련된 강수대가 다소 남동쪽에 치우쳐 나타나는 듯한 형태를 띠나 상관계수 값의 통계적 유의성은 높지 않다. 한편 남중국해에서 나타나는 고기압은 관측과 유사한 지역에서 나타나고 있어 일부 유사성을 보이고는 있으나 전반적으로 관측된 혹은 모델에서 나타나는 한반도 강수와 관련된 순환구조는 찾아볼 수 없다.

남부 지방을 중심으로 한 관측 강수에 관련된 MME 예측 패턴은(Fig. 10) 한반도와 일부를 중심으로 한 연관강수대의 존재, 그리고 그 남동쪽의 고기압 및 이와 관련된 남서기류의 존재 등 앞서 살펴본 관측 한반도 지역 강수와 관련된 순환패턴(Fig. 4)과 보다 유사한 형태를 갖는다. 또한 열대 태평양에 나타나는 강수대와 그 북쪽의 건조대 등 관측 남부 지방 강수와 관련된 특징 그리고 MME 결과에서 나타나던 특징이 함께 나타나고 있다. 이 경우 MME 자료와 관측자료는 한반도 중부지방에 관련된 상황보다 큰 스케일에서의 유사한 구조를 보인다고 볼 수 있다.

지금까지 한반도 지역의 강수 변동에 대한 관측 및 예측에 나타나는 순환패턴과 관측자료와 관련된 예측 순환패턴을 살펴본 결과를 정리하면 아래와 같다.

- 실제 한반도 중부지방을 중심으로한 7월 강수와 관련된 순환은 한반도 북쪽의 저기압과 한반도 남쪽의 고기압 발달과 그 사이에서 남서기류의 유입으로 정리되며, 관련 순환장은 동아시아 지역에 국한되어 있다.
- 특정 모델의 경우 이러한 순환장이 왜곡되어 나타나고 있으며 여러 개의 모델 자료가 합쳐진 MME에서는 큰 스케일에서의 고-저기압 분포 및 남서기류의 유입이라는 유사성은 있으나 이러한 순환패턴은 열대 태평양 등 더 큰 범위의 강수 편차와 높은 상관성을 보인다.
- 관측 강수변동과 관련된 MME의 순환패턴은 관측 혹은 예측에서 나타나는 강수와 관련된 순환패턴과 큰 유사성을 보이지 않는다.
- 한반도 남부 지방을 기준으로 한 분석 결과는 중부지방 강수를 기준으로 한 결과보다 모델 예측과 관측의 유사성이 더욱 높다.

서론에서 언급한대로 통계적 후처리가 성공적이기 위해서는 공간적인 패턴은 다르더라도

도 관련된 시간 변동의 상관성이 필요하다. 관측과 MME예측에서 한반도 지역 강수와 관련된 패턴은 일부 오차는 존재하나 대규모 스케일에선 기본적으로 유사성을 갖는다. 반면, 관측된 한반도 지역 강수량을 MME 예측에 연결지은 패턴이 관측 또는 예측장에서 나타났던 순환패턴과의 유사성이 떨어진다는 사실(Fig. 9)은 제시된 관측(Fig. 3)과 예측(Fig. 7)의 패턴의 시간 변동의 상관성이 떨어진다는 것이다. 즉, 통계적 후처리의 성공적이기 위한 조건을 만족하지 못한다고 볼 수 있다. 그러나 반대로 생각하면 Fig. 9에 해당하는 순환패턴과 Fig. 3 또는 Fig. 7과 유사성이 존재한다면 이는 통계적 후처리 적용이 가능함을 의미한다고 볼 수 있다. MME예측 결과는 많은 앙상블 예측이 평균되어 비교적 단순한 형태의 대규모 순환 정보만을 가지게 되므로 관측에서 나타나는 지역적으로 복잡한 형태의 순환정보를 기대하기는 어렵다. 따라서 최선은 Fig. 9와 Fig. 7 사이의 유사성을 높이도록 관측 강수량 시계열을 재가공하는 것이라 볼 수 있다.

4. 관측 자료의 재구성을 통한 예측성의 향상

4.1 관측 자료 재구성의 필요성

계절예측은 기후시스템의 비선형적인 역학과 초기조건 불확실성에 기인한 예측성능의 한계가 존재한다. 이 한계는 잠재적으로 예측가능한 성분(Signal)과 예측가능하지 않은 성분(Noise)의 크기에 의해 좌우된다. 어떤 변수(y)를 아래 식(1)과 같이 Signal(y_s)과 Noise(y_n)로 구분하여 설명하고 이를 통해 잠재예측성을 앙상블 예측자료에서 진단하는 시도는 오래전부터 있어왔다(e.g. Rowell 1998).

$$y = y_s + y_n \quad (1)$$

이 변수(e.g. 7월 평균 강수)에 대한 예측값을 x 라고 하고 이 역시 Signal과 Noise로 구분할 수 있다. 계산상의 편의를 위해 두 변수의 분산을 1이라고 가정하면, 두 변수의 상관계수는

$$y'(t) = a\Phi + b = a\left(\sum^N C(s)W(s)x(s,t)\right) \quad (2)$$

와 같다. 이때 일반적으로 Noise는 정규분포를 따르는 무작위 잡음으로 다른 변수와의 상관성은 없는 것으로 가정할 수 있고 충분히 많은 샘플을 갖는 경우 두 변수의 상관계수의 기댓값은 식(2)의 우변 첫 번째 항 $(\overline{x_s y_s})$ 과 같다. 그러나 샘플이 충분하지 않은 경우, 그리고 Noise를 무작위 잡음으로 가정할 수 없는 경우에는 다른 항들이 0이 아니게 되어 실제 상관계수는 Signal만으로 구한 상관계수와 같지 않게 된다. 모델 예측의 앙상블 평균은 잠재 예측성을 평가하는 데 있어 Signal로 간주되므로(Rowell 1998) 3장에서 다루어진 모델 예측결과 및 MME 예측 결과는 Noise가 소거된 즉, 모델의 입장에서 예측가능한 Signal만 있다고 보아야 한다. 따라서 상관계수는,

$$\overline{xy} = \overline{(y_s + y_n)x_s} = \overline{x_s y_s} + \overline{x_s y_n} \quad (3)$$

가 되며 예측 signal을 관측 signal의 선형회귀 형태로 아래와 같이 표현할 수 있는데,

$$x_s = ay_s + e \quad (4)$$

이 경우 $a = \overline{x_s y_s} / \overline{y_s^2}$ 이며 오차항 e 는 y_s 와 상관관계가 없다. 각 변수의 분산이 1로 같고 만일 잠재 예측성 $(V(y_s)/V(y))$ 이 예측과 관측 모두 같다고 정의하면 코시-슈바르츠 부등식에 의해 a 의 절대값은 1보다 작게 된다.¹⁾ 상관계수를 다시 정리하면 아래와 같다.

$$\overline{xy} = \overline{(y_s + y_n)x_s} = \overline{(y_s + y_n)(ay_s + e)} = \overline{ay_s^2} + \overline{y_n e} \quad (5)$$

만일 오차가 없는 완벽한 예측($y=x_s$)의 경우는 e 가 0이 되고 a 가 1이 될 것이므로 얻을 수 있는 최대 상관계수는 분산값이 1인 정규화된 변수(y)의 Signal 분산($\overline{y^2}$)과 같다. 그러나 실제 상황에서는 모델의 예측에는 항상 오차가 존재하므로 1보다 작은 크기의 a 값을 갖게 되며 특히 계절예측과 같이 과거 hindcast 샘플이 충분하지 않은 경우 식(5)의 우변에서 두 번째 항의 비중이 커질 수 있다. Fig. 9에서 나타난 관측 강수량의 변동과 관련된 MME 예측의 순환패턴은 상관계수와 회귀식을 통해 얻어졌으므로 결국 위에서 보인 식(5)의 결과로 볼 수 있고 만일 관측 강수량에서 차지하는 Noise의 비중이 크다면 두 번째 항의 영향이 크게 나타난 결과일 수 있다.

1) 코시 슈바르츠 부등식은 $(\overline{x_s y_s})^2 \leq \overline{x_s^2} \overline{y_s^2}$ 이며 총 분산과 잠재 예측성이 같다고 가정하면 Signal의 분산도 같다.

$(\overline{x_s^2} = \overline{y_s^2})$ 따라서 $a^2 = (\overline{x_s y_s})^2 / (\overline{x_s^2} \overline{y_s^2}) = (\overline{x_s y_s})^2 / (\overline{x_s^2} \overline{y_s^2}) \leq 1$ 이 성립한다.

APCC에서 활용되고 있는 후처리 기법과 같이 자유도가 커서 관측변수의 변동에 대해서 좁은 지역에서 높은 상관관계를 갖는 상관패턴의 정보가 활용될 가능성이 높은 기법의 경우 만일 예측 인자가 식 (5) 우변의 두 번째 항이 크게 작용하여 우연하게 높게 계산된 상관계수의 정보가 유입될 것이고 이는 이후 새로운 예측 상황에서 낮은 예측성능을 가져올 수 있다. 즉, 관측의 Noise 변동성에 최적화 된 통계적 후처리가 만들어 질 수 있다는 것이다. 이러한 기법의 안정성을 높이기 위해서 실무에서는 여러 가지 장치들을 도입해 왔으나 뚜렷한 이론적 근거는 없는 상황이며, 근본적인 해결책은 가능하다면 y_n , 즉 관측에서의 Noise를 소거하는 것이다.

그런데 관측자료는 앙상블 예측과 같이 Signal과 Noise를 구분할 수 있는 도구가 없으므로 Noise의 소거는 이론적으로 가능하지 않다. 최근에 관측자료만을 이용해 잠재 예측성을 평가하기 위한 시도들이 있어왔으나(Feng et al. 2014) 이러한 연구들은 시간 평균된 분산값의 관점에서 평가가 가능할 뿐 매 시점에서 Signal과 Noise의 구분은 가능하지 않다. 또한 이 연구들에서 Signal과 Noise를 구분하기 위한 여러 가지 가정들을 도입하고 있는데 Signal은 계절 규모를 넘어서는 time scale을 가진다거나 매일 매일의 변동성이 AR(Autoregressive) 모델을 만족한다는 등의 가정에 기반하고 있다(Delsole et al. 2013).

본 연구에서도 한가지 가정을 도입하여 관측자료에서 Signal과 Noise의 분리를 시도하고자 한다. 다만 본 연구에서 설정하는 Signal은 잠재적으로 예측가능한 성분으로 가정된 기존 연구와 약간의 차이를 보이는 데 이는 이 분리의 목적이 예측성을 평가하기 위함이 아니라 주어진 MME 예측자료에서 최적의 정보를 안정적으로 도출하는데 있기 때문에 MME 예측자료가 갖는 특성에 일부 영향을 받는다.

예측가능한 성분, 즉 Signal을 정의할 때 전지구 모델의 예측자료를 예측인자(predictor)로 활용할 것이므로 예측가능한 성분은 전지구 모델이 설명할 수 있는 혹은 전지구 모델 자료 내에 존재하는 역학적 상호작용과 관련이 있는 성분이어야 한다. 가령 어떤 지점 강수량의 실제 변동이 열대 태평양의 변동인 ENSO의 영향과 극지방의 변동인 AO의 영향을 받으나 모델의 한계로 인해 모델에서는 그 지점의 강수량과 AO에 관련된 역학적 관련성이 존재하지 않고 ENSO와만 관련된 경우, 모델 자료를 활용하기 위한 통계적 후처리에서 실제로는 AO와 관련되어 있다고 하더라도 모델에서 설명하지 못하는 부분이므로 이 조건에서 예측가능한 성분은 ENSO에 관련된 것에 국한되어야 한다. 즉 모델이 “이해” 할 수 있는 역학적

관계를 충실하게 재현하는 것이 주어진 예측자료를 적절하게 활용하는 것이라고 할 수 있다. 아직까지 기후모델들의 예측성은 열대 해수면 온도에 크게 의존하며 그 외 지역, 변수에 대한 역학적 피드백은 충실하게 예측하지 못하고 있는 점을 감안하여 예측가능한 성분을 열대 해수면 온도, 특히 인도-태평양 지역의 해수면 온도와와의 상관성을 크게 갖는 부분으로 정의하여 관측자료를 분리하였다.

4.2 관측 자료의 필터

관측 자료에서 강수량 ($Y(s,t)$)과 해수면온도 ($X(s',t)$)를 이용해 SVD 분석을 수행하고 관측 강수의 처음 3개 모드만을 이용해 원래 변수차원으로 환원하여 필터된 강수 $Y^*(s,t)$ 를 구성하였다. 이때 s 는 강수량의 영역이며 동아시아 지역(105E-150E, 27.5N-47.5N)을 선정하였고 s' 은 해수면 온도자료의 영역으로 인도양과 태평양을 포함하는 영역(60E-270E, 40S-40N)으로 선정하였다.

$$Y^*(s,t) = \sum_{i=1}^3 P_i(s) T_i(t) \quad (6)$$

여기서 P 는 강수에 대한 SVD 공간패턴, T 는 시계열, 그리고 i 는 모드를 나타낸다.

Fig. 11은 수행된 SVD의 첫 3개 모드의 공간 및 시간 분포이다. 첫 번째 모드(Fig. 11a)는 전체 공분산의 절반 이상을 설명하며 동아시아 강수대의 발달과 북서태평양 및 적도 동태평양의 해수면 온도 변동이 결합된 모드로 추출되었다. 정규화된 상관계수가 아닌 공분산을 최대로 하는 모드를 찾는 SVD는 한쪽 변수에 두드러진 변동모드가 있는 경우에 우선적으로 선택할 수 있으며 이 경우 첫 번째 모드가 동아시아 강수량의 대표적인 변동모드에 의해 우선적으로 선택된 것으로 판단된다. 북서태평양의 음의 해수면 온도는 동아시아 강수 강화에 따른 잠열 증가 및 단파복사 감소로 인한 해수면 온도 하강으로 보이며 이러한 동아시아 지역의 변동에 적도 동태평양의 해수면 온도의 변동이 일부 영향을 줄 수 있음을 시사한다. 두 번째 모드는 NINO3.4 지역(170W-120W 5S-5N)을 중심으로 강하게 나타나는 해수면 온도, 즉 ENSO와 관련된 동아시아 지역의 강수량 변동 모드이며 세 번째 모드는 동아시아 지역의 퀴드러플 형태의 강수패턴과 관련된 주변 지역 해수면 온도의 변동 및 북미 연안 부근의 해수면 온도 변동이 결합된 패턴이다.

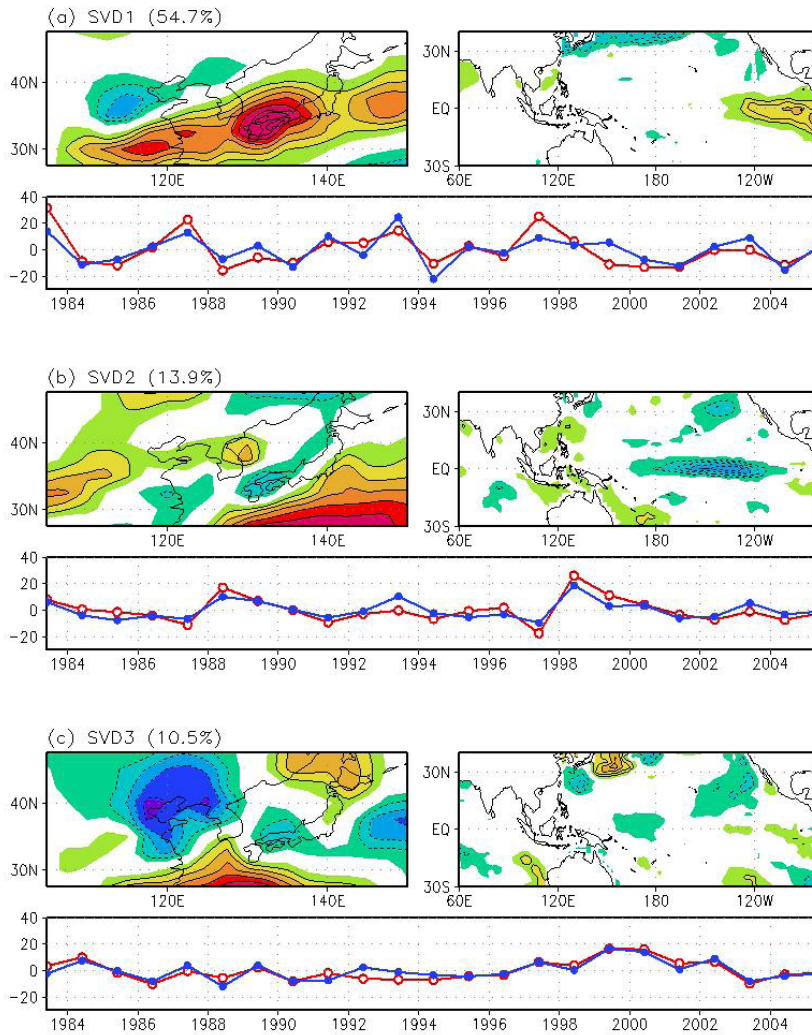


Figure 11. (a) Leading SVD modes between July precipitation over East Asia and Tropical Indo-Pacific SST. Precipitation(left) and SST(right) singular vectors are shown in top and associated time series are shown in the bottom. Contour intervals are arbitrary. (b),(c) same with (a) but for second and third modes.

식 (6)에 의해 동아시아 강수량 자료를 재구성하고 이 필터된 강수량이 실제 강수량의 어느 만큼을 설명하는 지 살펴보았다(Fig. 12). 전체 강수량 변동은 일본 열도 남쪽에서 최대 값이 나타나며 한반도 방향으로 일부 큰 값이 나타나고 있다. 반면 필터된 강수량의 분산은 전반적으로 값이 감소하였으며 일본 지역의 한반도 지역으로의 높은 분산이 많이 감소했음을 알 수 있다. 이러한 경향은 필터된 강수량의 분산이 전체 분산에서 차지하는 비에서 뚜렷하게 나타나는 데 대체로 일본 남쪽 태평양 지역에서의 분산비가 높고 연해주지역과 한반도 중부지역의 분산비가 0.2-0.3으로 낮은 편이다. 이러한 분산 비는 필터된 강수량이 예측가능한 Signal이라고 한다면 전체 분산가운데 Signal 분산의 비중인 잠재 예측성이 되고 따라서 실제 예측성능 역시 잠재예측성에 비례하는 모습을 보이게 될 것이다. 필터된 강수량의 한반도 지점에 대한 시계열을 Fig. 12d에 표시하였는데, 실제 관측 (검은색)의 일부 변동은 필터된 자료(푸른색)가 다른 모습을 보이고 있다. 즉, 그러한 변동은 해수면 온도의 변화와 무관한 지역적인 대기 순환의 영향으로 발생한 것이다. 두 시계열 간의 상관계수는 0.51이며 이 값은 별다른 의미를 갖지 않는다.

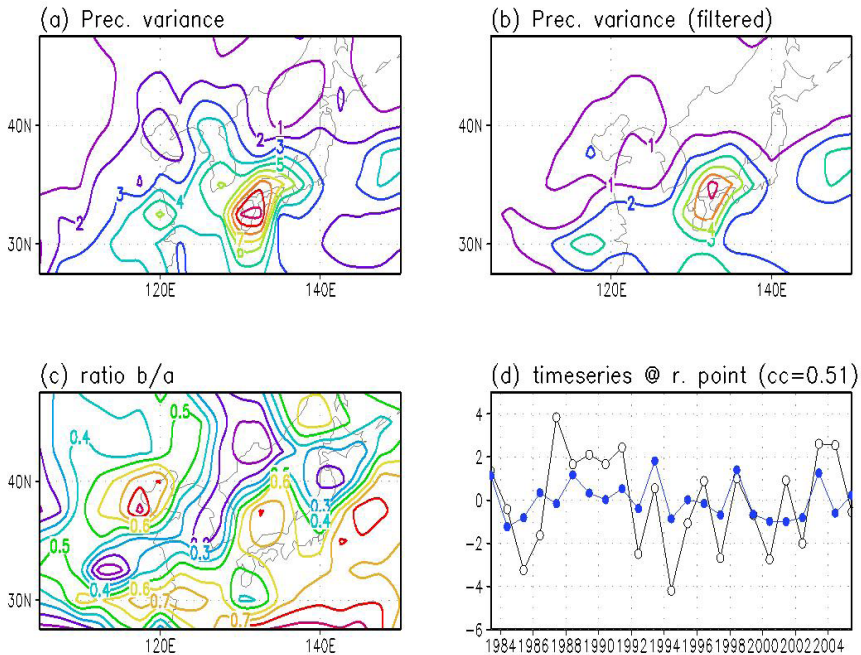


Figure 12. (a) Variance of monthly mean precipitation in July. (b) Same with (a) but for the filtered precipitation by using leading SVD modes shown in Fig. 12 (c) Ratio between b and a (b/a). (d) time series of precipitation at the reference point.

4.3 필터된 관측자료의 분석

앞 절에서 재구성한 관측 강수량자료를 이용해 앞서 분석한 바와 같은 한반도 지점의 강수 변동과 관련된 순환패턴을 재분석 하였다. Fig. 13은 한반도 중부지역에 해당하는 지점의 필터된 강수량의 변동과 관련된 순환 패턴이다. 순환 패턴을 구성하는 각 변수들은 필터하지 않았고 상관계수와 회귀계수를 구하는 강수 시계열만 필터된 값을 사용하였다. 한반도의 동쪽인 일본열도를 중심으로 저기압이 나타나며 그 남쪽에 남중국해와 서태평양을 포함하는 동서로 긴 고압대가 나타나고 있으며 고압대와 관련된 건조지역이 나타나고 있다. 이는 Fig. 3에서 보인 한반도 중부지역과 관련된 강수패턴과 그 보다 남쪽 지점의 강수변동과 관련된 패턴(Fig.4)이 혼합된 형태를 보인다. 그러나 관측자료의 다른 변수에는 여전히 Noise 성분이 남아 있으므로 강수량 시계열이 Signal이라고 하더라도 얻어진 결과는 작은 규모의 패턴들이 남아있을 것이다.

필터된 관측자료가 MME 예측의 특성을 고려한 가정에 의해 도출되었고 MME 예측의 통계적 후처리를 위해 활용될 것이므로 MME 예측에서 어떠한 패턴을 보이는 지가 더욱 중요하다. Fig. 14에서 필터된 관측자료와 MME예측의 상관패턴을 나타내었다. Fig. 9와 비교할 때 매우 조직적인 패턴을 보이고 있다. 이 패턴은 MME 예측 자체 자료만을 이용해 산출한 한반도 강수량과 관련된 순환패턴(Fig. 7)과 매우 유사하다. 즉 필터된 자료는 MME 예측이 물리적으로 설명할 수 있는 변동을 많이 포함하고 있다는 것이다. 물론 MME 예측 자료만을 사용한 경우와 비교해 적도태평양 강수의 변동성과의 상관성이 약하게 나타나는 등의 차이가 있으나 필터된 자료가 꼭 MME 자료에서 나타나는 모든 순환패턴을 도출할 필요는 없다. 가령 실제로 존재하지 않는 기후모드간의 물리적 관련성을 모델이 모의하는 경우라면 필터된 관측자료를 이용하더라도 모델이 임의로 모의한 관련성은 나타나지 않을 것이다.

즉 해수면 온도의 변동과 관련된 강수만을 도출하여 이를 MME 예측과 관련 짓는 경우 더욱 안정적이고 유의미한 관측-예측간 상관관계의 도출이 가능하며 이를 바탕으로 한 통계적 후처리는 그렇지 않은 경우에 비해 더 안정적인 예측 성능을 보일 것으로 기대할 수 있다.

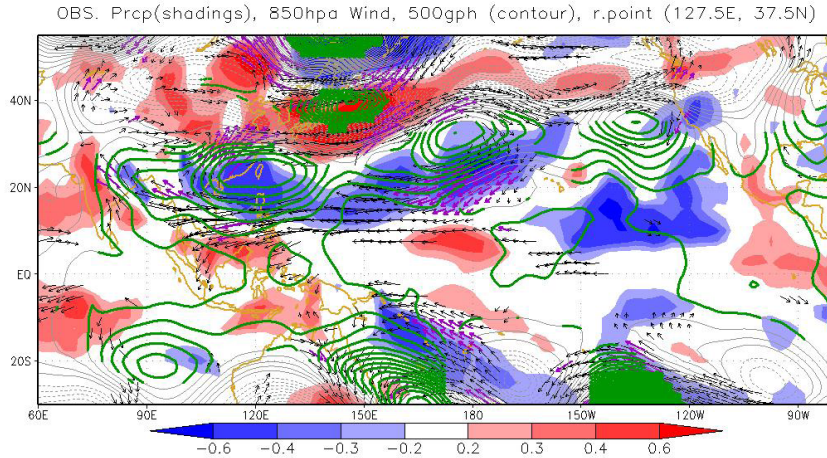


Figure 13. Same with Fig. 3 but for the filtered precipitation as a reference time series

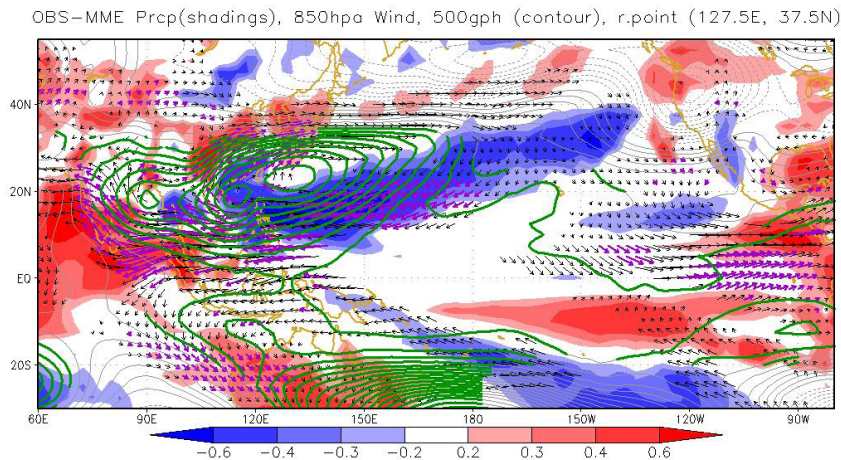


Figure 14. Same with Fig. 9 but for the filtered precipitation as a reference time series

4.4 통계적 후처리 기법의 적용 및 평가

필터된 동아시아 지역 강수량 자료를 이용해 통계적 후처리를 적용해 보았다. 사용한 후처리 기법은 Kug et al.(2008)의 SPPM기법에 기반한 간단한 선형회귀기법으로 아래의 식과 같이 표현된다.

$$y'(t) = a\Phi + b = a\left(\sum_{s=1}^N C(s)W(s)x(s,t)\right) + b \quad (7)$$

예측인자인 ϕ 는 대상 관측 지점의 강수 시계열(y)과 예측된 강수자료(x)의 각 지점별 과거 자료를 이용해 구한 공분산 $C(s)$ 을 새로운 예측값인 $x(s,t)$ 에 투영한 값이며 s 는 예측자료의 각 격자점을 나타낸다. 이때 어떤 격자점을 선택할 것인지를 결정하는 것이 가중치 함수인 $W(s)$ 이며 이 값은 0 또는 1로 특정 격자점을 포함시키거나 포함시키지 않을 수 있다. 이 포함여부를 결정하는 것은 과거 예측 및 관측자료의 상관계수로, 상관계수의 절대값이 0.3 이상인 격자점 가운데 상관계수가 높은 지점부터 일정 간격 (0.05)으로 등급을 나누어 공분산을 투영시킬 대상에 포함시키게 된다. 최대 200개 까지의 격자점을 활용하도록 설정하였다.

총 23년의 자료가 있으므로 1년씩 제외한 Cross-validation을 적용하여 후처리 기법을 적용하여 동아시아 지역의 모든 격자점에 대한 예측정보를 생산하였으며 그 결과 얻어진 각 해의 예측 결과들을 이용해 필터하지 않은 실제 관측값과의 상관계수를 구한 것이 Fig. 15이다. 필터된 관측 자료를 이용한 경우의 격자점 별 상관계수를 왼편(a)에, 그리고 필터하지 않은 관측 자료를 사용한 경우를 오른편에 나타내었다. 필터된 자료는 식(7)의 $C(s)$ 와 $W(s)$ 를 구하는데 사용되었다.

일본열도 남동쪽에 대해서는 실제 관측 자료를 이용한 경우와 필터된 관측자료를 이용한 경우가 큰 차이를 보이지 않으나 그 외 지역에 대해서는 기대했던 대로 필터된 자료에 기반한 후처리가 높은 성능을 보인다. 한반도 주변과 그 북서쪽 지역은 상대적으로 낮은 예측성을 보이는 것으로 나타나고 있는데 이러한 공간 패턴은 Fig. 12c에 보였던 필터된 자료의 분산이 전체 분산에서 설명하는 비율, 즉 잠재 예측성과 매우 유사한 형태이다. 즉, 잠재예측성이 높은 지역에서는 실제 관측자료에 Noise가 적게 포함되어 있으므로 필터된 자료와 실제 자료의 차이가 적어서 예측성능의 큰 차이가 없고 잠재예측성이 낮은 지역에서는 아무리 좋은 예측 기법을 활용하더라도 예측성능의 상한이 낮으므로 다른 지역에 비해 좋은 예측성능을 보일 수 없다. 이렇게 잠재 예측성이 낮은 지역은 관측에 포함된 Noise의 비중이 크므로 과대적합 문제에 취약해 통계적 후처리의 위험성이 더욱 커지게 된다. 따라서 실제 자료를 활용한 경우에 매우 큰 음의 예측성능을 보이게 된다.

통계적 후처리에 활용된 각 격자점의 분포를 Fig. 16에 제시하였다. 이 분포는 동아시아 지역 171개 격자점과 Cross-validation으로 인해 재 선택된 23회 모두를 누적해 산출한 것으로 동아시아 지역의 예측에 활용할 수 있는 MME 예측 강수정보의 분포를 나타낸다. 왼쪽은 필터된 강수자료를 이용한 경우, 오른쪽은 실제 강수자료를 활용한 경우이다. Fig. 15의 예측성능 비교에서 짐작할 수 있듯이 필터된 관측 강수자료를 활용한 경우에 예측인자로 활용된 지점의 분포가 조직적이고 같은 유사한 지점이 많이 선택되는데 반해 필터하지 않은 자료를 이용한 경우는 전구 모든 영역에 고르게 분포하고 있다. 이것은 Noise에 의해 얻어진 물리적 의미가 없는 상관관계에 의해 후처리 기법이 크게 영향을 받았으며 지역적 또는 시간적 일관성도 많이 떨어졌음을 의미한다. 필터된 자료를 이용한 경우의 예측인자 지점의 분포는 북서태평양과 중태평양, 인도양 등 해양에서 많이 선택되어 해수면 온도의 변동과 보다 많은 관련성이 있는 예측정보가 활용되고 있음을 알 수 있다.

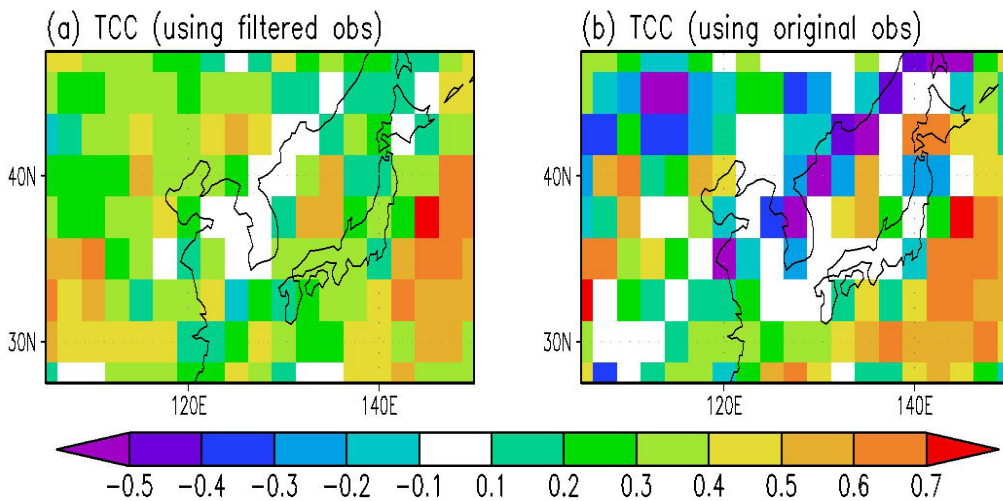


Figure 15. (a) Temporal correlation coefficient (TCC) at each grid point of corrected (downscaled) precipitation of GCM with observed precipitation. Downscaling was used as a filtered observation for training. (b) Same with (a) but with original observation for training.

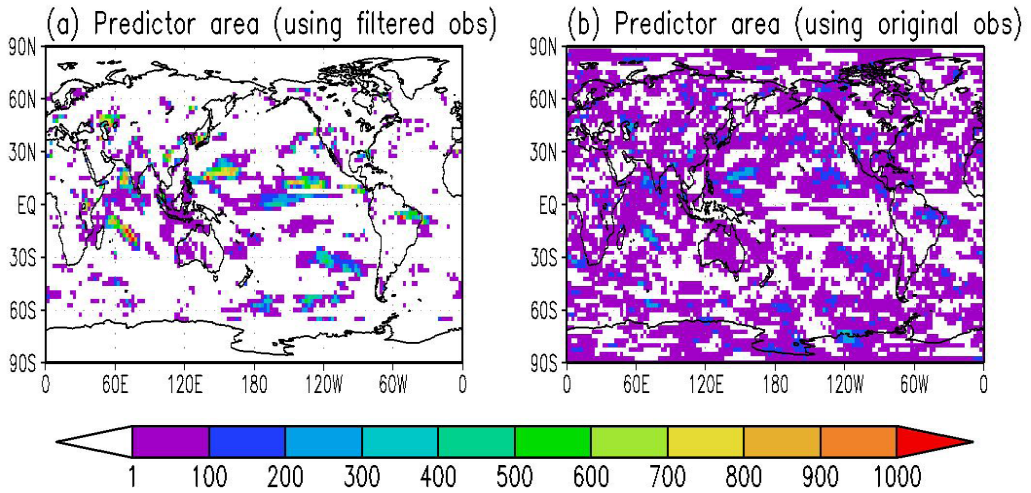


Figure 16. (a) Number of selection as a predictor for downscaling of precipitation over East Asia domain shown in Fig. 15 with filtered observation (b) Same with (b) but with original observation.

5. 요약 및 토의

역학적 계절예측에 있어 통계적 후처리 기법의 적용은 예측 모델이 가진 구조적 오차의 개선을 통한 지역 기후예측 정보의 품질 향상을 위해 필수적이다. 그러나 현재 계절예측을 산업적으로 생산하고 있는 기관 가운데 이러한 통계적 후처리를 거친 예측 정보가 제공되고 있는 곳은 매우 제한적인데 이는 기관의 성격이나 정보 제공에 있어 맡은 역할 등 과학기술 외적인 부분의 영향이나 상이한 후처리 기법을 적용해 생산된 다양한 변수에 대해 역학적으로 일관된 해석이 어려운 이유도 있지만 통계적 후처리 기법의 성능 자체가 만족스럽지 못한 것의 영향도 있다. 통계적 후처리 기법은 과거 예측 및 관측 자료에서 나타나는 상관관계를 이용해 새로운 예측의 오차를 감소시키는 방법으로 어떠한 상관관계에 기반하는가에 따라 그 성능이 좌우된다. 계절예측이 가진 가장 큰 문제 가운데 하나는 과거 예측의 기간이 길지 않아 이러한 과거 예측에 기반해 얻어진 상관관계의 안정성이 떨어진다는 점이다. 안정적이지 않은 상관관계에 기반한 후처리 기법은 새로운 예측정보에 대해 좋은 결과를 생산할 수 없

다. 따라서 안정적인 상관관계의 도출 및 이를 이용한 후처리 기법의 적용은 계절예측에 있어 통계적 후처리 기법의 활용성을 높이고 예측정보의 정확도 향상에 기여할 수 있다.

동아시아 지역 특정 지점의 7월 평균 강수의 경년 변동에 대해 관측과 모델이 어떠한 순환패턴을 보이는지에 대한 분석을 통해 관측에 비해 단일 모델예측 및 MME예측이 어떠한 한계를 갖는지 평가하고 이를 바탕으로 관측된 특정지점의 강수 자료를 재가공해 어떠한 정보를 설명하도록 해야하는 지에 대해 분석하였다.

이어 관측 강수량 변동 가운데 일부는 예측가능한 Signal, 그 나머지는 예측이 가능하지 않은 Noise로 구분하고 열대 해수면 온도의 변동에 관련된 역학적 관련성에 주로 의존하는 MME 예측의 특성을 감안하여 열대 해수면 온도와 SVD 선행모드를 이용한 동아시아 관측 7월 평균 강수량의 재구성을 통해 얻어진 강수자료를 이용해 결합 패턴을 투영하는 방식의 통계적 후처리 기법의 안정성을 높일 수 있음을 보였다. 즉, 관측에서 나타나는 열대 해수면 온도와 관련된 변동성을 추출해 통계적 후처리 기법의 개발에 활용할 수 있다는 점을 시사한다.

본 연구에서 제안한 방식은 예측자료의 가공을 목표로 하는 기존 연구와는 달리 예측대상이 되는 관측자료의 가공을 통해 안정성을 높이려 하였으며 실제로 안정성과 예측성이 높아지는 성과를 보였다. 그러나 동아시아 지역에 대한 적용성은 평가가 되었으나 다른 지역에 대한 일반적인 적용을 위해서는 해결해야 할 추가과제들이 있다. 우선 재가공에 사용된 3개의 선행 SVD 모드의 사용이 다른 지역에도 적절한 개수인지 확인할 필요가 있으며 3개라는 모드 개수 역시 각 모드가 설명하는 공분산의 크기를 보고 선정하였으나 다소 임의적인 선택일 수 있다. 또한 다른 계절이나 다른 변수에 대해서도 동일 조건이 적용가능한지에 대한 평가가 이루어질 필요가 있으며 궁극적으로는 이러한 모드 선택에서 비교적 자유로운 Coupled (Forced) Manifold (Navarra and Tribbia, 2005)와 같은 기법의 활용이 고려될 필요가 있다. 이 방법의 장점은 SVD는 공분산을 최대화 하는 여러 개의 모드들로 구분이 되어 몇 개의 모드를 선택하는 지에 따른 임의성이 존재하나 이 방법은 단 한 개의 값이 결과로 얻어진다는 점이다. 다만, 이 방법은 일종의 다변수를 이용한 선형회귀식으로 이 회귀식의 해 역시 샘플링 문제에서 자유롭지 않다는 단점이 있다.

궁극적으로는 기후 모델 및 예측 시스템의 꾸준한 개선을 통해 관측에서 나타나는 다양

한 기후 피드백을 실제에 가깝게 모의하도록 해야 할 것이며 추후 열대 해수면 온도의 변동 외에도 여러 기후모델들이 일관성을 갖고 모의, 예측하는 기후 피드백이 어떤 것인지에 찾아낼 수 있다면 이를 포함한 관측 자료의 재구성이 이루어질 수 있고 추가적인 예측인자의 도입을 통해 잠재 예측성 및 예측성의 향상을 기대할 수 있을 것이다.

REFERENCES

- DelSole, T., and J. Shukla 2012: Climate models produce skillful predictions of Indian summer monsoon rainfall, *Geophys. Res. Lett.*, 39, L09703, doi:10.1029/2012GL051279.
- DelSole, T., A. Kumar, and B. Jha, 2013: Potential seasonal predictability: Comparison between empirical and dynamical model estimates, *Geophys. Res. Lett.*, 40, 3200-3206
- Feng, X., T. DelSole, and P. Houser, 2014: Comparison of Seasonal Potential Predictability of Precipitation. *J. Climate*, 27, 4094-4110
- Feddersen, H., A. Navarra, and M. N. Ward, 1999: Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Climate*, 12, 1974-1989
- Gadgil, S., Vinayachandran, P. N., and Francis, P. A., and S. Gadgil, 2004: Extremes of the Indian summer monsoon rainfall, ENSO, and equatorial Indian Ocean oscillation, *Geophys. Res. Lett.*, 31, 1821, doi:10.1029/2004GL019733.
- 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. Climate*, 17, 834-844.
- Kang, H, C.-K. Park, S. N. Hameed, and K. Ashok, 2009: Statistical Downscaling of Precipitation in Korea Using Multimodel Output Variables as Predictors. *Mon. Wea. Rev.*, 137, 1928-1938.
- Kug, J.-S. J.-Y. Lee, and I.-S. Kang, 2008: Systematic Error Correction of Dynamical Seasonal Prediction of Sea Surface Temperature Using a Stepwise Pattern Project Method. *Mon. Wea. Rev.*, 136, 3501-3512.
- Lee, J.-Y., 2003: Assessment of potential seasonal predictability with a multi-model dynamical-statistical ensemble system. Ph.D. thesis, Seoul National University, Seoul, Korea, 164 pp.

- Rowell, D.P., 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations, *J. Climate*, 11, 109-120.
- Shukla, J., L. Marx, D. Paolino, D. Straus, J. Anderson, J. Ploshay, D. Baumhefner, J. Tribbia, C. Brankovic, T. Palmer, Y. Chang, S. Schubert, M. Suarez, and E. Kalnay, 2000: Dynamical Seasonal Prediction. *Bull. Amer. Meteor. Soc.*, 81, 2593-2606.
- Tippett, M. K., T. DelSole, and A. G. Barnston, 2014: Reliability of Regression-Corrected Climate Forecasts. *J. Climate*, 27, 3393-3404.

연구보고서 2015-01

역학적 계절예측의 통계적 후처리의 성능향상을 위한 예측인자 선정 기법 개선
Improving Predictor Selection for Better Performing Statistical Post Process
of Dynamical Seasonal Forecast

유진호 연구본부



APEC 기후센터

612-020 부산광역시 해운대구 센텀7로 12
Tel: 051-745-3900 Fax: 051-745-3949

www.apcc21.org

 www.facebook.com/apcc21

 www.youtube.com/APECClimateCenter21

 www.twitter.com/apcc21

 www.plus.google.com/+APECClimateCenter21

 www.flickr.com/apcc21