



Improvement of precipitation forecasting skill of ECMWF data using multi-layer perceptron technique

Lee, Seungsoo^a · Kim Gayoung^b · Yoon, Soonjo^c · An, Hyunuk^{d*}

^aResearch Fellow, APEC Climate Center, Busan, Korea

^bResearcher, APEC Climate Center, Busan, Korea

^cResearcher, APEC Climate Center, Busan, Korea

^dProfessor, Department of Rural and Environmental Engineering, Chungnam National University, Daejeon, Korea

Paper number: 19-014

Received: 11 March 2019; Revised: 19 May 2019 / 20 June 2019; Accepted: 20 June 2019

Abstract

Subseasonal-to-Seasonal (S2S) prediction information which have 2 weeks to 2 months lead time are expected to be used through many parts of industry fields, but utilizability is not reached to expectation because of lower predictability than weather forecast and mid- / long-term forecast. In this study, we used multi-layer perceptron (MLP) which is one of machine learning technique that was built for regression training in order to improve predictability of S2S precipitation data at South Korea through post-processing. Hindcast information of ECMWF was used for MLP training and the original data were compared with trained outputs based on dichotomous forecast technique. As a result, Bias score, accuracy, and Critical Success Index (CSI) of trained output were improved on average by 59.7%, 124.3% and 88.5%, respectively. Probability of detection (POD) score was decreased on average by 9.5% and the reason was analyzed that ECMWF's model excessively predicted precipitation days. In this study, we confirmed that predictability of ECMWF's S2S information can be improved by post-processing using MLP even the predictability of original data was low. The results of this study can be used to increase the capability of S2S information in water resource and agricultural fields.

Keywords: S2S, ECMWF, Multi-layer perceptron

다층퍼셉트론 기법을 이용한 ECMWF 예측자료의 강수예측 정확도 향상

이승수^a · 김가영^b · 윤순조^c · 안현욱^{d*}

^aAPEC 기후센터 선임연구원, ^bAPEC 기후센터 연구원, ^cAPEC 기후센터 연구원, ^d충남대학교 지역환경토목학과 조교수

요 지

2주에서 2개월까지 선행기간을 가지는 계절내-계절(Subseasonal-to-Seasonal, S2S) 예측결과는 산업전반에 걸쳐 다양한 분야에 활용이 가능할 것으로 기대되고 있으나, 일기예보나 중장기 예보대비 낮은 예측성으로 인하여 현재까지 활용성이 매우 낮은 실정이다. 본 연구에서는 기계학습 기법 중 비선형회귀 분야에서 좋은 결과를 보여주는 다층퍼셉트론 기법을 이용하여 S2S 예측자료의 후처리를 통한 국내 영역에서의 강수예측성 향상에 관한 연구를 수행하였다. 후처리 모형의 학습을 위한 입력자료로는 ECMWF의 S2S 과거예측(Hindcast) 정보를 이용하였으며 양분예보기법에 기반하여 학습된 다층퍼셉트론 모델을 이용한 후처리 결과와의 비교 분석이 수행되었다. 비교분석 결과 편차도(Bias score)는 평균 59.7% 감소하였고, 정확도(Accuracy)는 124.3% 증가하였으며, 임계성공지수(Critical Success Index)는 88.5% 향상된 것으로 분석되었다. 탐지확률(Probability of detection)의 경우 원자료 대비 평균 9.5% 감소하였으나 이는 ECMWF의 예측모델이 강수의 발생일을 과도하게 예측하였기 때문인 것으로 분석되었다. 본 연구 수행 결과 비록 ECMWF의 S2S 예측자료의 예측성이 낮더라도 후처리를 통해 예측성을 향상시킬 수 있음을 확인하였으며, 본 연구 결과는 향후 수자원과 농업 분야에서 S2S 자료의 활용성을 높이는데 도움이 될 수 있을 것으로 판단된다.

핵심용어: 계절내-계절 예측, 기계학습, 다층퍼셉트론

*Corresponding Author. Tel: +82-42-821-5797

E-mail: hanlight20@hotmail.com (H. An)

1. 서론

2주에서 2개월까지의 시간규모를 가지는 계절내-계절(Subseasonal-to-Seasonal; S2S) 예측은 대기의 초기조건에 영향을 받기에는 시간 규모가 길고, 해양 경계조건에 영향을 받기에는 상대적으로 짧은 시간규모로써 수치모형을 이용해 정확도 높은 예측을 수행하기에는 현실적으로 어려운 실정이다. 그러나 날씨예보와 기후예측을 이어주는 이음새 없는 예보에 대한 높은 관심과 요구(Vitart *et al.*, 2017)에 힘입어 세계기상기구(World Meteorological Organization, WMO)의 세계기상연구프로그램(World Weather Research Programme, WWRP)과 세계기후연구프로그램(World Climate Research Program, WCRP)이 공동으로 추진하는 S2S 프로젝트가 2013년부터 시작되었다(WMO Publication, 2015).

S2S 프로젝트에서는 고영향(High-impact) 기상 이벤트를 특별히 강조하고 계절내에서 계절 규모의 시간규모에 대한 이해를 바탕으로 모델의 예측력을 향상시키고, 응용(적용)분야에서의 활용성 증진 및 운영 센터에 의한 이니셔티브 도입을 독려함과 동시에, 기상 및 기후 연구 커뮤니티의 전문성을 활용하여 기후 서비스를 위한 글로벌 프레임 체계의 중요성을 제기하기 위한 목적으로 수립되었다. 이와 관련된 S2S 프로젝트의 연구 우선순위는 1) 예측 스킬을 증가시키기 위한 기회의 창을 정의하는 것을 포함하여 계절내 이벤트의 잠재적 예측성을 평가하고, 2) 계절내에서 계절 예측 범위의 시스템적 에러와 편의(bias)에 대한 이해, 3) 다중 모델 결합을 통한 계절내 예측 자료의 비교, 검증, 테스트와 불확실성의 정량화, 4) 특정한 극한 이벤트의 연구 등이다(WMO, 2015). 현재 전 세계적으로 11개 현업기관이 S2S 프로젝트에 참여하고 있으며, 각 기관에서 생산된 S2S 예측자료는 유럽중기기상예보센터(European Centre for Medium-Range Weather Forecasts, ECMWF)와 중국기상청(China Meteorological Administration, CMA) 데이터 포털을 통해 사용자에게 제공된다. 그러나 다양한 분야에서의 높은 활용성이 기대되는 S2S 자료는 일기예보나 중장기 예보대비 예측성이 낮은 것으로 인식되어(Brunet *et al.*, 2010; WMO, 2012; Robertson *et al.*, 2015) 현재 활용성이 매우 낮다.

기상현상 중 강수는 일반 시민들의 일상생활뿐만 아니라 농업과 수자원을 포함하여 산업 전반에 걸쳐 가장 큰 영향을 미치는 현상 중 하나이다(Vitart *et al.*, 2012). 그러나 강수량뿐만 아니라 강수의 발생 유무조차 기상예보에서 가장 예측하기 어려운 현상으로 인식되고 있다(Seo *et al.*, 2012). 일상생활과 밀접한 관련이 있는 강수 예측을 위해 그동안 다양한 연구가 수행되었는데 최근에는 인공지능망과 기계학습을 이용한

강수예측의 정확도 향상을 위한 연구가 활발히 수행되고 있다. Hall *et al.* (1999)은 기계학습 기법을 이용하여 텍사스 달라스 지역에 위치하고 있는 36개의 강우 관측소에 대해서 확률강수예측(Probability of Precipitation, PoP)과 정량적 강수예측(Quantitative precipitation forecast)을 수행한 바 있다. Kim *et al.* (2001)은 구간 연산 신경망을 이용하여 한반도 영역에 대하여 계절별로 강수량 예측을 수행한 바 있으며, Cha and Ahn (2005)은 전국-지역 모델에서 생산된 예측인자들을 다중 선형회귀(Multiple Linear Regression, MLR) 기법과 인공신경망(Artificial Neural Network, ANN)의 입력자료로 활용하여 남한지역의 여름철 강수를 보정하는 연구를 수행하였다. Kang and Lee (2011)는 한반도 영역에서의 정량적 강수추정을 위하여 중규모 수치예보자료, 상층기상관측소와 자동기상관측소(Automatic Weather Station, AWS) 관측자료를 이용하여 인공신경망 모델을 구성하여 중규모 수치예보모형의 3시간 누적 강수에 대한 추정을 수행하였다. Ha *et al.* (2016)은 내일 강수 예측을 위하여 예측인자로 기온, 전일-전주 강수일, 태양과 달 궤도 관련 자료 등을 이용하여 비지도 사전 학습을 통해 초기 가중치를 설정할 수 있는 딥빌리퍼네트워크(Deep belief network, DBN)를 이용하여 새로운 강수예측 방법을 제안하였으며, 다층퍼셉트론(Multi-layer Perceptron, MLP) 기법과 비교한 결과 학습에 소요되는 시간이 짧고 양분 예보기법에 기반한 정확도가 향상됨을 보였다. Xingjian *et al.* (2015)은 강수의 현재 예보를 시공간적(spatiotemporal)으로 연속하는 문제로 인식하여 레이더 관측 이미지를 이용하여 미래 12시간까지 강수 예측을 수행하였다. 예측 모델을 구축하기 위하여 일반적인 Long Short-term Memory (LSTM) 기법의 내부에 Convolution 기법을 적용한 ConvLSTM을 제안하였으며, 다른 기법들에 비해 우수한 성능을 보여주는 결과를 도출하였다. Nair *et al.* (2018)은 ANN 기법을 이용하여 Global Climate Models (GCMs) 자료의 예측성을 향상시키는 연구를 수행한 바 있으며 인도지역에 대해 몬순(monsoon) 기간의 강수량 연 변동(year to year variations)의 예측성을 향상시킬 수 있음을 보인 바 있다.

이렇듯 강수의 예측 정확성을 향상시키기 위해 많은 연구가 수행되었으나 대부분의 연구가 단기예보 및 계절규모 예보 관점에서 이루어졌고, S2S 시간 규모에서의 강수 예측성 향상에 관한 연구는 매우 미진한 실정이다.

따라서 본 연구에서는 기계학습을 이용하여 S2S 예측자료의 후처리를 통한 남한 영역의 예측성 향상 가능성을 알아보기 위한 연구를 수행하였다. 연구를 위한 S2S 예측자료로는 예측성이 가장 높다고 알려진 European Centre for Medium-Range Weather Forecasts (ECMWF) 자료를 선정하였으며,

기계학습 방법으로는 비선형 회귀분석에 가장 많이 이용되는 MLP기법을 이용하였다. 학습 모델을 구축하기 위하여 ECMWF (2014)에서 밝힌 바와 같이 강수량 산정에 이용되는 지표면 온도(T2M), 850 hPa 기압면에서의 동서, 남북방향 바람(U850, V850), 해면기압(MSLP), 강수(PREC) 등 5개의 변수를 이용하여 입력자료를 구성하고 재분석자료인 ERA-Interim 자료를 관측값으로 하여 MLP 모델을 이용하여 학습 모델을 구축하였다. 전체 자료 기간은 ECMWF 모델의 과거 예측자료(hindcast) 기간인 1999년 1월 4일부터 2011년 1월 3일까지 12년 동안의 일 단위(Daily) 자료를 이용하였으며 이중 80%에 해당하는 기간을 학습에 이용하고 이후 20%에 해당하는 기간에 대해서 양분예보기법에 기반한 검증을 수행하여 MLP 기법을 이용한 후처리 자료의 예측성 향상 정도를 분석하였다. 본 논문의 구성은 2장에서 ECMWF자료와 MLP기법에 대한 간략한 소개와 학습모델 구축 방법 및 자료의 예측성 평가 방법을 기술하였다. 3장에서는 검증기간에 대해서 ECMWF의 예측자료와 학습된 모델의 출력 결과를 비교 분석하여 예측성 향상 정도에 대해 평가하였다. 마지막으로 4장에서는 결론 및 향후 연구 과제를 기술하였다.

2. 자료 및 방법

2.1 ECMWF 자료

S2S 예측을 위해 현재 11개 기관에서 생산하는 S2S 자료는 예측선행시간, 수평 및 연직 해상도, 앙상블 수, 과거예측자료

생산방법(고정 또는 on the fly), 과거예측자료 기간, 예측자료 생산주기 등이 모두 다르다. 본 연구에서는 현재까지 예측성이 가장 우수하다고 평가되고 있는 ECMWF의 S2S 자료(de Andrade *et al.*, 2019; Vitart and Robertson 2018)를 사용하였으며, ECMWF S2S 자료의 특징은 Table 1과 같다.

Table 1에 표시된 바와 같이 ECMWF모델은 일주일에 2번씩 향후 46일에 대해 전 지구 예보를 생산하고 있다. 모델의 예측성능을 파악하기 위해 과거 20년 기간에 대해서 11개의 앙상블 멤버를 이용하여 과거예측을 수행하고 51개의 앙상블 멤버를 구성하여 향후 46일에 대해서 예측을 수행하고 있다.

2.2 MLP 기법

본 연구에서는 인공신경망 기법 중 하나인 MLP 기법을 이용하여 ECMWF 예측자료의 후처리를 통해 예측성능을 향상시키기 위한 연구를 수행하였다. 단층퍼셉트론(Single-layer Perceptron, SLP)과 MLP는 입력층(Input layer), 은닉층(Hidden layer), 출력층(Output layer), 가중치(Weight), 편이(Bias)으로 구성되어 있다는 부분에서는 동일하나 Fig. 1과 같이 은닉층의 개수가 2개 이상인 경우를 MLP라고 한다.

학습 모델을 구축하기 위하여 입력층에는 학습을 시키고자 하는 데이터를 배치하고, 은닉층에 위치하고 있는 가중치와 편이의 값을 조정하여 최종적으로 출력하고자 하는 결과에 가장 적합한 값을 출력하도록 반복학습을 수행하게 된다.

퍼셉트론의 특징은 은닉층의 활성화 함수가 -1에서 1사이의 값을 반환한다는 것인데 그로 인하여 비선형적 함수식을 표현하기 위해서는 입력 데이터를 정규화 해 줄 필요가 있다.

Table 1. Information summary of ECMWF S2S data

Center	Forecast Frequency	Forecast Time range	Forecast Ens. Size	Hindcast Frequency	Hindcast Ens. Size	Hindcast length
ECMWF	2/week (Mon,Thu)	0-46 days	51	2/week (Mon, Thu)	11	past 20 years

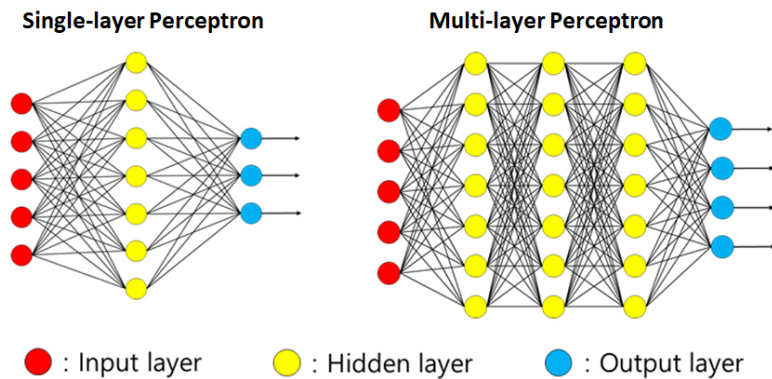


Fig. 1. Single-layer perceptron and multi-layer perceptron (Lee, 2018)

표준화된 입력 변수는 가중치를 곱하고 편향을 더한 후 활성화 함수로 보내지며 대표적인 활성화 함수는 계단함수(Hard limiter), 시그모이드(Sigmoid) 함수, ReLU 함수 등이 있다. 활성화 함수를 통과한 값은 그 다음 은닉층으로 보내지며 다시 가중치를 곱하고 편향을 더한 후 다시 활성화 함수를 통과한 후 출력층으로 보내지게 된다. 이후 출력된 값과 대상값의 차이를 최소화 할 수 있는 방향으로 학습이 반복되는 오차역전파 알고리즘을 통해 학습이 수행된다.

2.3 자료구성 및 학습방법

본 연구에 사용된 ECMWF 과거예측자료(hindcast)는 수평해상도 1.5°×1.5°로 변환한 일별 예측자료이다. ECMWF에서는 Table 1과 같이 월요일과 목요일, 일주일에 두 번씩 향후 46일간의 S2S 예측자료 및 과거예측자료를 반복적으로 생산하는데, 본 연구에서는 2018년 예측을 위해 생산된 과거예측자료 중 목요일마다 초기화된 1999-2011년 기간 동안의 과거예측자료를 앙상블 평균하여 사용하였다. 자료의 연속성과 활용성을 높이기 위해 하나의 초기조건에서 생산된 자료를 최대한 이용하여 6주(42일) 간격의 예측자료(1월 4일, 2월 15일, 3월 29일, 5월 10일, 6월 21일, 8월 2일, 9월 13일, 10월 25일, 12월 6일)에서 각각 42일씩을 추출한 후 연속하여 사용하였다. 이렇게 만들어진 S2S 과거예측자료의 기간은 1999년 1월 4일부터 2011년 1월 3일까지이다. 대상영역 지점은 총 4개로, Fig. 2에 점으로 표시하였다.

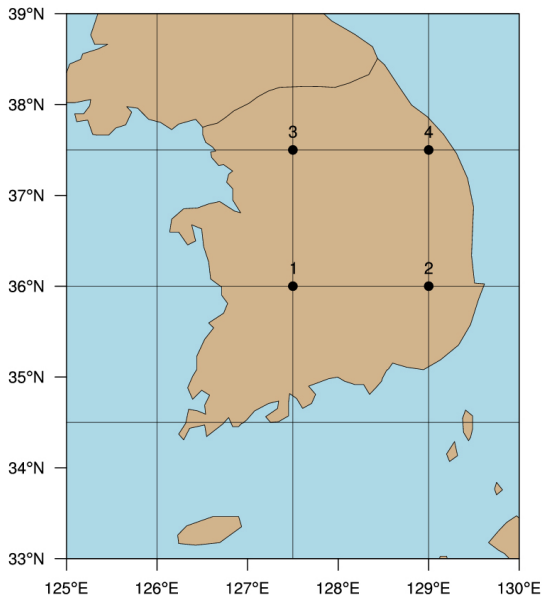


Fig. 2. Grid point over South Korea (1: 127.5°E, 36°N; 2: 129°E, 36°N; 3: 127.5°E, 37.5°N; 4: 129°E, 37.5°N)

입력자료 구축을 위해서 ECMWF의 S2S 예측변수 중 강수량 산정에 활용되는(ECMWF, 2014) 표면 온도(T2M), 850 hPa 기압면에서의 동서방향과 남북방향 바람(U850, V850), 해면기압(MSLP)과 강수(Precipitation) 자료 등 총 5개의 변수를 선택하였으며 최종적으로 풍속과 풍향의 영향을 고려하기 위하여 U850과 V850를 이용하여 풍속 벡터의 크기와 풍향 각도를 계산하여 입력자료를 구축하였다.

관측자료로는 ECMWF에서 제공하는 일별 ERA-interim 재분석 자료를 1.5°×1.5° 격자로 내삽한 후 Fig. 2의 4개 지점에 대한 강수(mm/day)자료를 추출하여 사용하였으며 학습모델 구축에 사용된 자료의 기간은 ECMWF의 재현기간과 동일한 1999년 1월부터 2011년 1월까지의 자료를 이용하였다.

학습 및 검증을 위하여 1999년 1월 4일부터 2011년 1월 3일까지 총 12년 동안의 일(Daily) 자료 중 80%(1999년 1월 4일 ~ 2009년 8월 12일)를 학습에, 20%(2009년 8월 13일 ~ 2011년 1월 3일)를 검증에 활용하였다.

또한 본 연구에서는 강수량이 아닌 강수 유무에 대한 예측성을 평가하기 위하여 모델과 관측 모두 강수량이 0.1 mm/day를 초과한 경우는 1로, 그렇지 않은 경우에는 0으로 값을 변환하여 자료를 재구성하였다.

본 연구에서는 MLP의 활성화 함수, 은닉층 노드의 개수를 산정하기 위하여 시행착오법(Trial and Error)을 적용하였다. 먼저 단일층을 구성하고 활성화 함수로 tanh 함수와 ReLU 함수 두 개의 모델을 구성한 후 은닉층 노드의 개수를 10개씩 증가시켜 학습손실이 가장 작게 나타나는 활성화 함수와 노드의 개수를 취하였다. 그 결과 은닉층의 노드가 200개이며 활성화 함수가 ReLU일 때 가장 우수한 학습 결과를 보여주었다. 동일한 방법으로 두 번째 은닉층 노드 개수는 100개, 활성화 함수는 ReLU로 결정되었으며, 마지막 층에서는 강수의 유무를 판정하기 위하여 이진분류에 최적화된 Sigmoid함수로 설정한 후 강수의 유무를 판단하도록 하였다. 학습을 위한 optimizer는 최근 가장 좋은 성능을 보여준다고 알려진 Adam을 사용하였으며 학습률은 0.01로 설정하였다. 반복 학습 횟수는 최초 10,000회로 설정한 후 반복학습에 따른 학습오차가 10회 이상 감소하지 않는 경우 조기 종료 되도록 설정하여 학습을 진행하였다.

2.4 평가방법

본 연구에서는 양분 예보(Dichotomous Forecast)의 검증에 활용되는 이진 변수 검증(Wilks, 1995)을 통해 학습된 자료의 정확도를 평가한다. Table 2에는 양분예보 기법에 기반한 검증을 수행하기 위하여 예측값과 관측값을 이진 변수로 나타낸 분할표(Contingency Table)을 나타내었다.

Table 2. Contingency table

		Forecasting		
		Yes	No	Total
Observation	Yes	H (Hits)	M (Misses)	H+M
	No	F (False Alarms)	C (Correct rejection)	F+C
	Total	H+F	M+C	.

이진 변수 검증 분할표를 통해 모델의 예측값과 관측값의 비교를 통해 다양한 평가를 수행할 수 있으나 본 연구에서는 탐지 확률(Probability of Detection, *POD*), 편차도(Bias, *B*), 정확도(Accuracy, *A*), 임계성공지수(Critical Success Index, *CSI*)를 통해 학습 결과의 예측성 향상 여부를 평가하였다.

먼저 탐지확률은 현상이 일어났던 표본 수 중에서 정확한 예보가 이루어진 표본의 수로 정의된다. 따라서 탐지 확률 역시 1에 가까울수록 예측 모델의 성능이 좋아 짐을 의미한다. 그러나 이 지수는 예보 실패(F)를 고려하지 않으므로 예보를 남발하는 경우에도 값이 낮아지지 않는다는 특성을 가지고 있으며 계산식은 다음의 Eq. (1)과 같다.

$$\text{탐지 확률}(POD) = \frac{H}{H+M} \quad (1)$$

편차도는 Eq. (2)와 같이 현상의 관측 발생 빈도와 예보 모델에 의한 예측 발생 빈도의 비로 정의된다. 편차도가 1보다 큰 값을 가진다면 실제 현상의 발생 빈도보다 예보 모델에 의한 예측 발생 빈도가 높다는 의미이므로 실제 현상보다 더 많은 수의 예보를 한다는 의미로 해석될 수 있다. 예보가 정확할수록 1에 가까운 값을 가지게 되는데 주의할 점은 편차도는 예보가 얼마나 정확히 관측에 대응하는지를 나타내는 지수가 아니라 관측과 예측의 발생 빈도를 나타내는 지수라는 점이다.

$$\text{편차도}(B) = \frac{H+F}{H+M} \quad (2)$$

정확도는 Eq. (3)에서와 같이 전체 표본 개수에 대해 예측 모델에 의해 정확하게 예보된 표본 개수의 비로 정의된다. 이 지표는 맞힘(H)과 C(미발생 맞힘)을 구분하지 않으므로 건조한 지역에서의 예측정확도를 평가하기 어렵다는 단점을 가지고 있으나 남한지역에 대한 예측 성능을 평가하기에는 적합한 것으로 판단된다.

$$\text{정확도}(A) = \frac{H+C}{H+M+F+C} \quad (3)$$

임계성공지수는 다음 Eq. (4)와 같이 관측과 예측 구별 없이 현상의 발생과 관련된 전체 표본 수(따라서 C는 제외)에 대해 실제로 발생한 현상을 맞춘 비로 정의된다. 따라서 좋은 예보일수록 1에 가까운 값을 나타내게 된다. 임계성공지수는 예보하는 현상에 더욱 초점이 맞추어져 있어 정확도 지수와 함께 보완적으로 활용된다.

$$\text{임계성공지수}(CSI) = \frac{H}{H+M+F} \quad (4)$$

상기 언급된 4가지 지수를 이용한 평가를 수행한 후 학습 이전과 학습 이후의 예측능력 향상 정도를 정량적으로 비교하기 위하여 향상 비율에 대한 계산을 수행하였으며 그 식은 다음 Eq. (5)와 같다.

$$\text{Improvement ratio}(\%) = \frac{\text{Evaluation Index Value of original data}}{\text{Evaluation Index Value of trained data}} \times 100 \quad (5)$$

여기서, Original data는 검증에 활용된 20%(2009년 8월 13일~2011년 1월 3일)기간에 대한 ECMWF의 예측자료를 뜻하며, trained data는 동일 기간에 대해서 학습된 MLP모델을 이용하여 후처리된 값을 의미한다.

3. 결과

본 연구에서 수행된 연구 결과는 Table 3과 같다. 먼저 강수가 발생한 날을 얼마나 잘 예측했는지를 평가할 수 있는 탐지 확률을 살펴보면 ECMWF 모델의 경우 0.8142 (Grid-2)에서 1.0 (Grid-3)의 범위를 보여주고 있다. 이는 강수가 실제 발생한 날에 대해서 ECMWF의 예측값도 강수가 발생할 것으로 예측한 것을 의미하므로 실제 강수가 발생한 날에 강수가 발생할 것이라고 예측하는 예측성이 매우 높은 것을 알 수 있다. 학습된 MLP 모델을 통해 후처리된 결과 역시 실제 강수가 발생한 날에 대해서 강수가 발생할 것으로 예측한 비율은 0.7932 (Grid-4)에서 0.9341 (Grid-2)의 분포를 보여주어 ECMWF모델에 비해 다소 낮은 예측성을 보여주고 있으나 우수한 탐지확률을 가지고 있는 것으로 나타났다.

두 번째로 강수가 발생한 날에 대해서 모델의 예측값이 강수를 얼마나 잘 예측하였는지를 확인할 수 있는 편차도를 살펴보면 ECMWF 모델의 경우 최소 2.2906 (Grid-4)에서 2.7429 (Grid-2)의 분포를 보여주고 있다. 이는 실제 발생한 강수의 일수에 비해 2.3배에서 약 2.7배 가량 강수가 더 많이 발생할

것으로 예측했음을 의미한다. 반면, 학습된 MLP 모델을 통해 후처리 된 결과는 0.9696 (Grid-1)에서 1.0815 (Grid-2)의 분포를 보여줌으로써 실제 강수가 발생한 날에 강수가 발생할 것으로 예측하는 성능이 매우 우수한 것을 확인할 수 있다.

세 번째로 강수의 발생 일과 미발생 일을 얼마나 잘 예측하였는지를 확인할 수 있는 정확도를 살펴보면 후처리 이전의 ECMWF자료는 0.3751 (Grid-1)에서 0.4356 (Grid-4)의 분포를 보여주고 있다. 이는 전체 일 수(H+M+F+C)에 대해서 강수가 발생한 날(H)과 강수가 발생하지 않은 날(C)을 맞춘

비율이 37.51%에서 43.56%에 이른다라는 것을 의미하는 것으로 모델의 예측 대비 실제 관측치가 절반 이상 반대로 나타났다는 것을 의미한다. 그러나 학습된 MLP 모델을 통해 후처리 된 결과는 정확도가 최소 79.59% (Grid-4)에서 최대 92.25% (Grid-2)로 확인됨에 따라 강수의 발생유무를 맞추는 정확도가 대폭 향상 된 것을 확인할 수 있다. 향상 비율을 살펴보면 Grid-4에서 최저 68.9%의 예측성 향상이 나타났으며 Grid-2에서는 153.4%의 예측성 향상이 나타났다.

마지막으로는, 강수라는 현상에 더욱 초점이 맞추어져 있어 정확도 지수와 함께 보완적으로 활용되는 임계성공지수를 살펴보면, ECMWF 원자료의 임계성공지수는 0.3744 (Grid-1)에서 0.4349 (Grid-4)의 분포를 보임으로써 정확도와 비슷한 수준의 예측성능을 보여주었다. 이는 앞에서 살펴본 바와 같이 ECMWF가 강수의 발생 일을 과대하게 예측함으로써 전체 기간에 대해서 강수 발생일 수를 과대하게 예측함으로써 발생한 것으로 판단된다. 학습된 MLP 모델을 통해 후처리 된 결과를 살펴보면 임계성공지수가 0.6286 (Grid-4)에서 0.8142 (Grid-2)의 범위를 보여줌으로써 학습 이전 대비 최대 44.5% (Grid-4)에서 113.7% (Grid-1)의 예측성이 향상된 것을 확인할 수 있다.

Fig. 3에는 Table 3에 나타낸 결과를 보다 쉽게 비교하기 위해 그래프로 나타내었다. 먼저 탐지 확률의 경우 Grid-2를 제외한 나머지 부분에서 ECMWF의 예측결과가 강수의 발생 일을 더 잘 예측하는 것으로 나타나고 있으나 이는 Fig. 3(b)에

Table 3. Summary of results

Index		Grid-1	Grid-2	Grid-3	Grid-4	average
POD	Original	0.9970	0.8142	1.0	0.9974	0.9522
	Trained	0.8754	0.9341	0.8457	0.7932	0.8621
	Ratio (%)	-12.2	14.7	-15.4	-20.5	-9.5
B	Original	2.6596	2.7429	2.5964	2.2906	2.5724
	Trained	0.9696	1.0815	1.0445	1.0550	1.0377
	Ratio (%)	-63.5	-60.6	-59.8	-53.9	-59.7
A	Original	0.3751	0.3637	0.3865	0.4356	0.3902
	Trained	0.9179	0.9225	0.8643	0.7959	0.8752
	Ratio (%)	144.7	153.4	115.9	68.9	124.3
CSI	Original	0.3744	0.3630	0.3851	0.4349	0.3894
	Trained	0.8000	0.8142	0.7054	0.6286	0.7371
	Ratio (%)	113.7	124.3	83.2	44.5	88.5

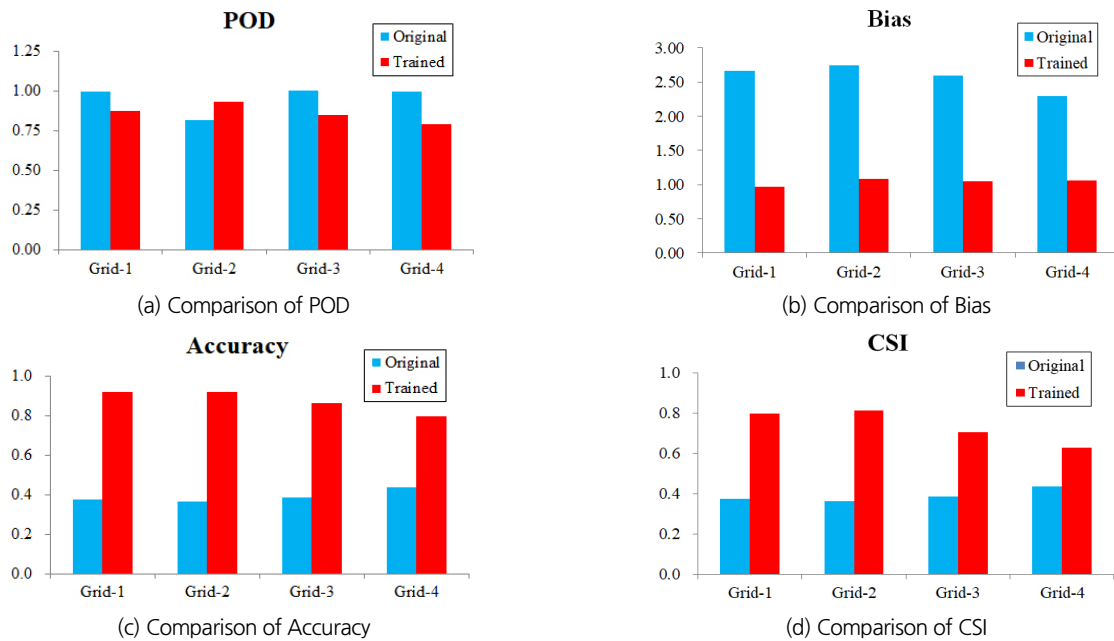


Fig. 3. Comparison of forecasting skills

나타난 바와 같이 ECMWF의 예측결과가 실제 강수 발생 일수에 비해 2배 이상 과대하게 강수가 발생한 것으로 예측하기 때문인 것으로 분석되었다. 또한 Fig. 3(c)에 보여지는 바와 같이 강수의 발생 유무를 예측하는 성능을 평가할 수 있는 예측성의 경우 전 영역에 대해서 학습된 MLP 모델을 통해 후처리된 결과가 ECMWF의 예측결과에 비해 두 배 이상 높은 예측성을 가지는 것을 확인할 수 있다. 마지막으로 Fig. 3(d)에 나타난 강수의 발생 일에 대한 예측성을 평가할 수 있는 임계성공지수도 전 영역에 걸쳐 ECMWF의 예측결과에 비해 학습된 MLP 모델을 통해 후처리된 결과의 예측성이 현저히 높은 것을 확인할 수 있다.

4. 결론 및 향후 연구과제

본 연구에서는 후처리를 통해 ECMWF 모델에서 예측한 S2S 예측자료의 예측성을 향상시키기 위한 방법으로 MLP 기법의 적용성을 평가하였다. 남한 영역에 대한 ECMWF의 과거예측자료 중 1999년 1월 4일부터 2011년 1월 3일까지 총 12년의 기간에 대해서 강수량 산정에 이용되는 지표면 온도 (T2M), 850 hPa 기압면에서의 동서방향과 남북방향 바람 (U850, V850), 해면기압(MSLP) 그리고 예측 강수자료를 이용하여 입력자료를 구축하여 MLP 모델을 학습시켰다. 관측값으로 ERA-Interim 자료가 이용되었으며 전체 자료 기간 중 80%에 해당하는 1999년 1월 4일 ~ 2009년 8월 12일까지의 자료를 모델 학습에 이용하였으며 20%에 해당하는 2009년 8월 13일 ~ 2011년 1월 3일 기간에 대해서 학습된 모델의 적용성을 평가하였다.

원자료인 ECMWF의 예측결과와 학습된 MLP 모델을 통해 후처리된 결과를 비교하기 위하여 양분 예보 검증에 활용되는 탐지확률, 편차, 정확도, 임계성공지수 등 총 4개의 지수들에 대해 예측성 분석을 수행하였다. 분석 결과 탐지확률은 ECMWF의 예측결과가 학습된 MLP 모델을 통해 후처리된 결과보다 우수한 것으로 나타났으나 이는 ECMWF 모델의 예측결과치가 실제 강수의 발생 일수에 비해 2배 이상 과도하게 강수가 발생하는 것으로 예측했기 때문인 것으로 분석되었으며 이는 편차도의 비교결과에서도 확인할 수 있었다.

강수의 발생과 미발생을 얼마나 정확하게 예측할 수 있는지를 평가할 수 있는 정확도 비교 결과 ECMWF의 예측성 대비 학습된 MLP 모델을 통해 후처리된 결과의 결과가 최소 68.9% (Grid-4)에서 최대 153.4% (Grid-2)까지 후처리 대비 평균 124.3% 예측성이 향상된 것을 확인할 수 있었다. 강수가

발생한 날에 대해서만 예측성을 평가하는 임계성공지수의 비교 결과 역시 ECMWF의 예측성 대비 학습된 MLP 모델을 통해 후처리된 결과의 예측성이 평균 88.5%가 향상되었다. 이는 물리 모형의 예측결과와 관측자료를 이용하여 MLP 모델을 통해 후처리를 하는 경우 물리 모형의 예측성을 보다 높게 향상시킬 수 있음을 증명한 결과로 판단된다.

본 연구의 결과는 현재 최대 10일까지의 일기 예보를 생산하고 예보하고 있는 기상청에서 강수의 발생 유무를 판단할 때 참고할 수 있는 자료로 활용할 수 있을 것으로 판단되며 최대 46일까지의 미래에 대해서 강수의 발생 유무가 중요한 농업과 수자원 분야에 활용될 수 있을 것으로 기대된다. 마지막으로 본 연구에서는 강수의 발생 유무에 대한 예측성 향상에 대해서만 연구를 수행하였으므로 향후 S2S 예측자료의 활용성을 보다 높이기 위해 강수의 발생 유무뿐만 아니라 정량적 강수량에 대한 예측성을 향상시키기 위한 심도 깊은 연구가 필요할 것으로 판단된다.

감사의 글

본 연구는 APCC의 지원을 통해 수행되었습니다.

References

- Brunet, G., Shapiro, M., Hoskins, B., Moncrieff, M., Dole, R., Kiladis, G. N., Kirtman, B., Lorenc, A., Mills, B., Morss, R., Polavarapu, S., Rogers, D., Schaake, J., and Shukla, J. (2010). "Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction." *Bulletin of the American Meteorological Society*, Vol. 91, No. 10, pp. 1397-1406.
- Cha, Y. M., and Ahn, J. B. (2005). "Evaluation of artificial neural network correction skill on dynamically downscaled summer rainfall over South Korea." *Journal of the Korean Meteorological Society*, Vol. 41, No. 6, pp. 1125-1135.
- de Andrade, F. M., Coelho, C. A., and Cavalcanti, I. F. (2019). "Global precipitation hindcast quality assessment of the subseasonal to seasonal (S2S) prediction project models." *Climate Dynamics*, Vol. 52, No. 9-10, pp. 5451-5475.
- ECMWF, S. P. (2014). *In IFS documentation CY40R1 Part IV: Physical Processes*. ECMWF: Reading, UK, pp. 111-113.
- Ha, J. H., Lee, Y. H., and Kim, Y. H. (2016). "Forecasting the precipitation of the next day using deep learning." *Journal of Korean Institute of Intelligent Systems*, Vol. 26, No. 2, pp. 93-98.
- Hall, T., Brooks, H. E., and Doswell III, C. A. (1999). "Precipitation forecasting using a neural network." *Weather and forecasting*, Vol. 14, No. 3, pp. 338-345.

- Kang, B. S., and Lee, B. K. (2011). "Application of artificial neural network to improve quantitative precipitation forecasts of meso-scale numerical weather prediction." *Journal of Korea Water Resources Association*, Vol. 44, No. 2, pp. 97-107.
- Kim, H. J., Baek, H. J., Kwon, W. T., and Choi, B. S. (2001). "Long-range forecast of precipitation using an internal arithmetic neural network." *Journal of the Korean Meteorological Society*, Vol. 37, No. 5, pp. 443-452.
- Lee, S. S. (2018). *Development of flood risk analysis technique using seasonal to sub-seasonal data*. APEC Climate Center, Annual Report, p. 9.
- Nair, A., Singh, G., and Mohanty, U. C. (2018). "Prediction of monthly summer monsoon rainfall using global climate models through artificial neural network technique." *Pure and Applied Geophysics*, Vol. 175, No. 1, pp. 403-419.
- Robertson, A. W., Kumar, A., Pena, M., and Vitart, F. (2015). "Improving and promoting subseasonal to seasonal prediction." *Bulletin of the American Meteorological Society*, Vol. 96, No. 3, pp. 49-53.
- Seo, J. H., Lee, Y. H., and Kim, Y. H. (2012). "Feature selection to predict very short-term heavy rainfall based on differential evolution." *Journal of Korean Institute of Intelligent Systems*, Vol. 22, No. 60, pp. 706-714.
- Vitart, F., and Robertson, A. W. (2018). "The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events." *npj Climate and Atmospheric Science*, Vol. 1, No. 1, p. 3.
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., and Hendon, H. (2017). "The subseasonal to seasonal (S2S) prediction project database." *Bulletin of the American Meteorological Society*, Vol. 98, No. 1, pp. 163-173.
- Vitart, F., Robertson, A. W., and Anderson, D. L. T. (2012). "Subseasonal to seasonal prediction project: bridging the gap between weather and climate." *Bulletin of the World Meteorological Organization*, Vol. 61, No. 2, p. 23.
- Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences*. Academic Press, pp. 233-250.
- WMO Publication (2015). *Seamless prediction of the earth system: from minutes to months*.
- World Meteorological Organization (WMO) (2012). "WMO bulletin." *World Meteorological Organization*, Vol. 61, No. 2, p. 48.
- World Meteorological Organization (WMO) (2015). Switzerland, accessed 1 February 2018, <<http://s2sprediction.net>>.
- Xingjian, S. H. I., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2015). "Convolutional LSTM network: A machine learning approach for precipitation nowcasting." *In Advances in neural information processing systems*, pp. 802-810.