

## 1개월 기온 예측자료의 오차 특성 분석 및 보정 기법 연구

김용석, 허지나\*, 김응섭, 심교문, 조세라, 강민구

국립농업과학원 기후변화평가과

(2023년 12월 11일 접수; 2023년 12월 20일 수정; 2023년 12월 21일 수락)

## Error Characteristic Analysis and Correction Technique Study for One-month Temperature Forecast Data

Yongseok Kim, Jina Hur\*, Eung-Sup Kim, Kyo-Moon Shim, Sera Jo, Min-Gu Kang

*Climate Change Assessment Division, National Institute of Agricultural Sciences*

(Received December 11, 2023; Revised December 20, 2023; Accepted December 21, 2023)

### ABSTRACT

In this study, we examined the error characteristic and bias correction method for one-month temperature forecast data produced through joint development between the Rural Development Administration and the Hong Kong University of Science and Technology. For this purpose, hindcast data from 2013 to 2021, weather observation data, and various environmental information were collected and error characteristics under various environmental conditions were analyzed. In the case of maximum and minimum temperatures, the higher the elevation and latitude, the larger the forecast error. On average, the RMSE of the forecast data corrected by the linear regression model and the XGBoost decreased by 0.203, 0.438 (maximum temperature) and 0.069, 0.390 (minimum temperature), respectively, compared to the uncorrected forecast data. Overall, XGBoost showed better error improvement than the linear regression model. Through this study, it was found that errors in prediction data are affected by topographical conditions, and that machine learning methods such as XGBoost can effectively improve errors by considering various environmental factors.

*Key words: One-month forecast, Subseasonal forecast, Machine learning, Bias correction*



\* Corresponding Author : Jina Hur  
(hjn586@korea.kr)

## I. 서 론

1개월 기상예측 자료는 농작물의 생육 환경을 예측하며, 이를 기반으로 시비, 수확 등의 영농 계획을 사전에 수립하는데 활용될 수 있다. 또한 특정 기상 조건에서 발생 가능한 작물의 병해충을 예측하고, 중장기적 대응책을 마련하여 작물의 피해를 최소화하는데 기여할 수 있다. 이와 같이 상세한 1개월 기상예측 정보는 농작물의 중장기적인 생산 및 관리 계획 측면에서 매우 유용하게 활용된다.

농촌진흥청은 홍콩과학기술대학교와의 공동연구를 통해 농업분야에 활용 가능한 1개월 기상 예측 시스템을 개발하고 있다(Ha et al., 2023; Im et al., 2021; Qiu et al., 2020). 이 시스템은 지역기후모형을 활용하여 전지구 기후 예측 자료를 역학적 상세화하여, 한국의 지형적 특성을 고려한 5km 해상도의 1개월 농업기상 예측 자료를 생산하도록 설계되었다. 이렇게 생산된 지역 규모의 농업기상 1개월 예측 정보는 높은 활용성에도 불구하고 계통적 오차들을 포함하고 있어 농업모델의 입력자료로 직접 활용하기에는 제약이 있다(Hur et al., 2023).

농업적 활용성 제고를 위해 다양한 통계적 방법들을 활용하여 중장기 기상 예측자료를 보정하는 연구가 이루어졌다. Jo et al. (2022)는 분위사상법을 이용하여 6개월 예측 정보를 보정하고, 벼 생물계절 예측 모형에 적용함으로써 보정된 기상예측자료가 농업적 활용 분야에서 가지는 부가가치를 지닌다고 제시하였다. Hur et al. (2021)는 평균보정기법을 통해 보정한 6개월 예측 자료와 보정하지 않는 예측 자료를 각각 적산 온도모형에 적용하여, 기상예측 정보에 따른 찰옥수수 수확일 예측의 오차 변화를 보여주었다. 그러나 이러한 연구들은 지점마다의 오차나 전체 지점의 평균적인 오차만을 분석하고 있어, 지형 및 위치를 고려하여 예측자료의 오차 특성을 충분히 제시하고 있지 않았다. 또한 예측자료의 정확도를 높이기 위해 지형 및 위치 특성을 고려한 오차 분석과 이와 관련된 인자를 이용해 요인들 간의 복잡한 비선형 관계를 수식화 하여 예측자료를 보정하는 연구도 이루어져야 할 필요가 있다.

이에 따라 본 연구에서는 1개월 기상 예측 시스템에서 생산된 기온 예측 자료의 오차 특성을 분석하고, 기계학습법을 이용하여 오차를 보정함으로써 오차 개선 정도를 살펴보고자 한다. 이에 따라 1개월 기상 예측 시스템에서 생산된 2013년부터 2021년까지의 과

거 예측(hindcast) 자료를 이용하여 오차 분석을 수행하였으며, 이를 위해 기상관측자료 및 환경자료(해발고도, 위도)를 이용하여 예측자료의 오차 특성을 분석하였으며, 환경인자와 기계학습법인 XGboost를 이용하여 예측 오차의 개선 정도를 살펴보았다.

## II. 재료 및 방법

### 2.1. 자료 수집

기상예측자료의 정확도 분석과 오차를 보정하기 위해 기상예측자료, 기상관측자료, 해발고도, 경사도, 위도 자료를 수집하였다.

#### 2.1.1. 기상예측자료

남한지역에 대한 1개월 기온 예측 자료는 약 55km 격자형 NCEP Climate Forecast System의 전지구 기후 예측 자료(CFSv2, Saha et al., 2014)를 지역기후모형을 활용하여 상세화한 자료이다. 역학적 상세화는 38.00 °N, 127.25 °E를 중심으로 20 km 해상도의 모격자(동북 아시아)와 5 km 해상도의 등격자(한국)을 가지도록 설정한 Weather Research and Forecasting model (WRF) (Skamarock et al., 2008, WRF4.2) 지역기후모형을 활용하였다. 지역기후모형은 민감도 실험(Qiu et al., 2020)을 통해 선정된 물리적 모수화 방법들을 적용하였다. 1개월 예측 자료 생산 방법에 대한 보다 자세한 설명은 Qiu et al. (2020) 및 Hur et al. (2023)에서 확인할 수 있다. 예측자료의 오차 분석을 위해 2013년부터 2021년까지의 과거예측(hindcast) 자료를 활용하였다. 과거예측 자료는 관측자료를 사용하지 않고, 미래예측과 동일한 방법으로 과거기간에 대해 예측한 실험을 의미한다. 오차 분석을 위해 5km 격자형 자료에서 기상관측장비들의 좌표와 가장 가까운 격자점 값을 추출하여 사용하였다.

#### 2.1.2. 기상관측자료

기상청의 기상자료개방포털(data.kma.go.kr)을 통해 2013년부터 2022년까지의 1년 단위로 이동이 없는 종관기상관측장비(ASOS)와 자동기상관측시스템(AWS)의 일별 최고기온자료와 최저기온 자료를 각 장비의 좌표 정보 함께 수집하였다. 연도별로 수집된 관측지점은 2013년 519지점, 2014년 550지점, 2015년 328지점, 2016년 553지점, 2017년 514지점, 2018년 591지점, 2019년 558지점, 2020년 289지점, 2021년 483지

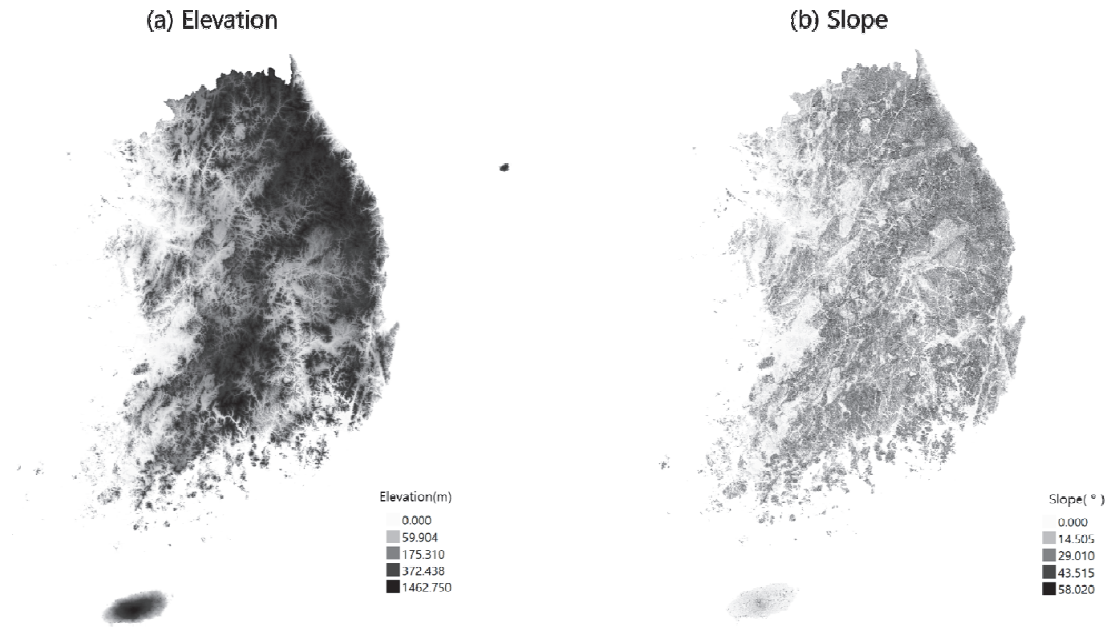


Fig. 1. Terrain Factors Using a Digital Elevation Model.

점을 활용하였다.

### 2.1.3. 지형정보

국토정보지리원에서 제공하는 수치지형도의 등고선과 표고점 정보와 GIS프로그램(ArcGIS10.1)을 이용하여 30m 해상도의 수치표고모델(Digital Elevation Model)을 제작하고 경사도를 생성하였으며, 연도별로 해당되는 기상관측지점의 좌표를 기준으로 해발고도와 경사도 값을 추출하였다.

## 2.2. 방법

기상예측자료의 정확도 향상을 위한 방법을 고안하기 위해 해발고도와 위도, 경사도를 이용하여 기상예측자료의 오차분석을 수행하고 오차를 보정할 변수를 선정한 후 기계학습법을 이용하여 기상예측자료를 보정하는 모형을 구축하였다. 모형을 구축하기 위해 2013년~2021년을 기준으로 기상관측장비에서 수집된 최고기온과 최저기온 자료를 종속변수로 설정하고 기상과거예측자료와 기상관측장비의 위도와 해발고도를 설명변수로 설정하여 기계학습법인 XGBoost를 이용하여 기상예측값을 보정하였다. 설명변수들은 모두 격자형 정보로서 기상관측지점의 좌표값을 이용해 해당 지점의 값을 추출하였고 표준화를 진행한 후 모

형을 구축하였으며, XGBoost의 성능을 비교하기 위해 선형회귀모형도 함께 구축하였다. 모형을 위한 데이터세트는 결측값 및 오류값을 제거한 후 1,362,491 세트를 구축하였으며, 그 중 80%정도를 학습을 위해 1,089,992 세트를 활용했으며, 검증을 위해 20%정도인 272,499세트를 활용하였다. 프로그램은 Google의 Colaboratory (python version: 3.10.12)에서 sklearn 라이브러리를 이용하였다. 이렇게 학습이 이루어진 모형을 이용하여 기상과거예측자료 대신 미래예측자료를 대입한다면 보정된 예측자료를 생산할 수 있을 것으로 예상된다.

### 2.2.1. XGBoost

XGBoost는 Gradient Boosting 알고리즘에 과적합 방지를 위한 기법이 추가된 지도 학습 알고리즘으로 기본 학습기(Base Learner)를 의사결정나무로 하며 잔차를 학습해 나가며 이전 모형을 순차적으로 개선해 나간다.

다음과 같은 과정을 통해 학습이 이루어지며(Chen and Guestrin, 2016; Pan, 2018), 기본식은 식(1)을 통해 예측값이 계산된다.

$$\hat{y} = \phi(x_i) = \sum_{k=1}^k f_k(x_i) \quad (1)$$

**Table 1.** Changes in RMSE of Maximum Temperature According to Slope

(Unit : °C)

Slope	< 5 degrees	5 ~ 10 degrees	> 10 degrees
RMSE	4.706	4.708	4.700

여기서,  $i$ 번째 데이터를 통해  $k$ 번째 의사결정나무의 값이 계산되며, 다음의 목적함수(Objective Function)가 최소가 되도록 데이터를 학습하게 된다.

$$\mathcal{L}(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (2)$$

식(2)에서  $\sum_i l(y_i, \hat{y}_i)$ 는  $i$ 번째 예측값과 정답값의 차이를 합산하는 것을 나타내고 값이 작아지는 방향으로 학습되는데,  $\sum_k \Omega(f_k)$ 의 정규화 요소(Regularization Term)을 나타내는 부분을 통해 특정 파라미터가 기본 학습기가 지나치게 큰 영향력을 갖지 않게 조절해 주는 부분으로 패널티를 부과해 가중치가 커지는 것을 방지한다.

$$\mathcal{L}^j = \sum_{i=1}^n l(y_i, \hat{y}_i^{(j-1)} + f_j(x_i)) + \Omega(f_j) \quad (3)$$

식(3)에서와 같이 연속적으로 훈련(Additive Training)을 반영하여 나타내면,  $i$ 번째 데이터의 정답값  $y_i$ 과  $j-1$ 번째의 의사결정나무 예측값  $\hat{y}_i^{(j-1)}$ 에,  $j$ 번째 의사결정나무를 통해 예측한 값  $f_j(x_i)$ 을 추가하여 정규화요소  $\Omega(f)$ 를 반영하게 된다.

$$\mathcal{L}_{split} = \frac{1}{2} \left[ \frac{(\sum_{i \in L} g_i)^2}{\sum_{i \in L} h_{i+\lambda}} + \frac{(\sum_{i \in R} g_i)^2}{\sum_{i \in R} h_{i+\lambda}} - \frac{(\sum_{i \in L} g_i)^2}{\sum_{i \in L} h_{i+\lambda}} \right] - Y \quad (4)$$

식(4)에서  $g_i$ 는 1차 미분 함수(gradient),  $h_i$ 는 2차

미분 함수(Hessian)을 나타내며,  $L$ 는 현재 마디를, 오른쪽, 왼쪽 자식 마디는 각각  $L_R, L_L$ 로 나타낸다.  $\gamma$ 는 과적합을 방지해 주는 파라미터이며, 의사결정나무의 가치를 확장할 경우 새로운 변수를 기준으로 데이터를 분할 해 나갈 때  $\mathcal{L}_{split}$ 의 감소폭이 큰 쪽으로 분할이 진행된다.

### III. 결 과

#### 3.1. 최고기온

최고기온 예측자료에 대하여 경사도, 해발고도, 위도에 따른 오차 특성을 살펴보았을 때, 경사도의 변화에 따라서는 예측자료의 오차는 특별한 특징이 보이지 않았지만(Table 1), 해발고도(Table 2)와 위도(Table 3)는 증가할수록 오차가 증가하는 경향이 나타났다. 이러한 결과를 바탕으로, 경사도를 제외하고 해발고도와 위도를 이용하여 최고기온의 오차 보정을 수행하였다.

먼저, 기상관측장비의 해발고도별로 50m이하, 50m~100m, 100m~200m, 200m 이상으로 나눠서 최고기온의 오차를 분석하고 보정하였을 때, 50m이하에서의 RMSE는 보정 전 기상예측자료 4.450, 선형회귀모형은 4.351, XGBoost는 4.163로 나타났다(Table 2). 50m 이하의 낮은 고도에서 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.099, 0.287 정도의 RMSE가 감소하는 것을 볼 수 있었다. 50m~100m는 기상예측자료는 RMSE가 4.637로 나타났으며, 선형회귀모형은 4.470, XGBoost는 4.295로 나타났고, 100m~200m는 기상예측자료

**Table 2.** Changes in RMSE of Maximum Temperature According to Elevation

(Unit : °C)

Elevation	Uncorrected Forecast	Corrected Forecast	
		Linear Regression Model	XGBoost
< 50m	4.450	4.351	4.163
50m ~ 100m	4.637	4.470	4.295
100m ~ 200m	4.820	4.508	4.329
> 200m	5.108	4.781	4.382

**Table 3.** Changes in RMSE of Maximum Temperature According to Latitude

(Unit : °C)

Latitude	Uncorrected Forecast	Corrected Forecast	
		Linear Regression Model	XGBoost
< 35 degrees	3.968	3.812	3.635
35~37 degrees	4.771	4.579	4.375
> 37 degrees	4.919	4.681	4.379

**Table 4.** Error Characteristics of Maximum Temperature Forecast Data

(Unit : °C)

Error	Uncorrected Forecast	Corrected Forecast	
		Linear Regression Model	XGBoost
RMSE	4.705	4.502	4.267
ME	-1.245	0.005	0.003

는 RMSE가 4.820로 나타났으며, 선형회귀모형은 4.508, XGBoost는 4.329로 나타났다. 200m 이상은 기상예측자료는 RMSE가 5.108로 나타났으며, 선형회귀모형은 4.781, XGBoost는 4.382로 나타났다. 200m 이상의 높은 고도에서 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.327, 0.726 정도의 RMSE가 감소하는 것을 볼 수 있었다. 해발고도가 높을수록 통계적 보정에 의해 오차 개선이 많이 이루어진 것을 확인할 수 있었으며, 보정 기법을 비교하였을 경우, 모든 해발고도에서 XGBoost가 선형회귀모형 보다 더 많은 최고기온 예측 오차를 감소시켰다.

위도에 따른 최고기온 예측 자료의 오차 특성을 살펴보기 위해 북위 35도이하, 35도~37도, 37도 이상으로 나눠서 오차를 분석하였다(Table 3). 35도 이하에서는 최고기온의 보정 전 예측자료의 RMSE는 3.968로 나타났으며, 선형회귀모형은 3.812, XGBoost는 3.635로 나타났다. 35도 이하 저위도에서 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.156, 0.333 정도의 RMSE가 감소하는 것을 볼 수 있었다. 35도~37도에서는 기상예측자료의 RMSE는 4.771로 나타났으며, 선형회귀모형은 4.579, XGBoost는 4.375로 나타났다. 37도 이상에서는 기상예측자료의 RMSE는 4.919로 나타났으며, 선형회귀모형은 4.681, XGBoost는 4.379로 나타났다. 37도 이상 고위도에서 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.238,

0.540 정도의 RMSE가 감소하는 것을 볼 수 있었다. 위도의 경우도 해발고도와 유사하게 보정기법을 통해 위도가 높을 수록 RMSE 감소폭이 커지는 것을 알 수 있었다. 그리고 모든 위도에서 XGBoost의 경우 일반선형회귀모형에 비해 오차가 더 감소하는 것을 확인할 수 있었다.

전체적인 최고기온 예측자료의 오차와 보정에 따른 오차를 살펴 보았을 때, 보정 전 예측자료(Uncorrected Forecast)의 평균 RMSE는 4.705으로 나타났으며, 선형회귀모형(Linear Regression Model)은 4.502, XGBoost는 4.267으로 나타났다(Table 4). 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.203, 0.438 정도의 RMSE가 감소하는 것을 볼 수 있었다. 그리고 ME (Mean Error)로 분석했을 때, 보정 전은 과소 추정하는 데이터가 많은 경향이 나타났으나 보정 후에는 어느정도 과소 추정이 감소되는 것을 볼 수 있었다.

### 3.2. 최저기온

다음으로 최저기온 예측자료에 대하여 경사도, 해발고도, 위도에 따른 오차 특성을 살펴보았을 때, 최고기온과 유사하게 경사도의 변화에 따라서는 예측자료의 오차는 특별한 특징이 보이지 않았지만(Table 5), 해발고도(Table 6)와 위도(Table 7)는 증가할수록 오차가 증가하는 경향이 나타났다. 이러한 결과를 바탕으로, 경사도를 제외하고 해발고도와 위도를 이용하여 최저기온의 오차 보정을 수행하였다.

**Table 5.** Changes in RMSE of Minimum Temperature According to Slope

(Unit : °C)

Slope	< 5 degrees	5~10 degrees	> 10 degrees
RMSE	3.802	3.710	3.794

**Table 6.** Changes in RMSE of Minimum Temperature According to Elevation

(Unit : °C)

Elevation	Uncorrected Forecast	Corrected Forecast	
		Linear Regression Model	XGBoost
< 50m	3.576	3.520	3.206
50m~100m	3.823	3.688	3.352
100m~200m	3.940	3.851	3.497
> 200m	3.971	3.950	3.651

해발고도에 따른 최저기온 예측 자료의 오차 특성을 살펴보았다(Table 6). 해발고도별 최저기온의 오차를 분석한 결과 50m이하에서의 RMSE는 보정 전 기상예측자료 3.576, 선형회귀모형은 3.520, XGBoost는 3.206로 나타났다. 50m 이하의 낮은 고도에서 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.056, 0.370 정도의 RMSE가 감소하는 것을 볼 수 있었다. 50m~100m는 기상예측자료는 RMSE가 3.823로 나타났으며, 선형회귀모형은 3.688, XGBoost는 3.352으로 나타났다. 100m~200m는 기상예측자료는 RMSE가 3.940로 나타났으며, 선형회귀모형은 3.851, XGBoost는 3.497로 나타났다. 200m 이상은 기상예측자료는 RMSE가 3.971로 나타났으며, 선형회귀모형은 3.950, XGBoost는 3.651로 나타났다. 200m 이상의 높은 고도에서 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.021, 0.320 정도의 RMSE가 감소하는 것을 볼 수 있었다. 최저기온의 경우도 최고기온과 같이 해발고도가 높을수록 예측자료의 오차가 커지는 특성

이 있었으며, 선형회귀모형과 XGBoost로 보정하였을 때 모든 해발고도에서 오차가 감소하는 것을 확인할 수 있었다.

위도에 따른 최저기온 예측 자료의 오차 특성을 살펴보았다(Table 7). 위도별 오차를 분석한 결과, 35도 이하에서는 최저기온의 보정 전 예측자료의 RMSE는 3.331로 나타났으며, 선형회귀모형은 3.330, XGBoost는 3.066로 나타났다. 35도 이하 저위도에서 선형회귀모형의 보정효과가 거의 나타나지 않았으며(RMSE 감소: 0.001), XGBoost로 보정한 예측자료는 보정 전 예측자료보다 각각 0.265 정도의 RMSE가 감소하는 것을 볼 수 있었다. 35도~37도에서는 기상예측자료의 RMSE는 3.713로 나타났으며, 선형회귀모형은 3.635, XGBoost는 3.341로 나타났다. 37도 이상에서는 기상예측자료의 RMSE는 4.054로 나타났으며, 선형회귀모형은 3.969, XGBoost는 3.592로 나타났다. 37도 이상 고위도에서 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.085, 0.462 정도의 RMSE가 감소하는 것을 볼 수 있었다.

**Table 7.** Changes in RMSE of Minimum Temperature According to Latitude

(Unit : °C)

Latitude	Uncorrected Forecast	Corrected Forecast	
		Linear Regression Model	XGBoost
< 35 degrees	3.331	3.330	3.066
35~37 degrees	3.713	3.635	3.341
> 37 degrees	4.054	3.969	3.592

**Table 8.** Error Characteristics of Minimum Temperature Forecast Data

(Unit : °C)

Error	Uncorrected Forecast	Corrected Forecast	
		Linear Regression Model	XGBoost
RMSE	3.779	3.710	3.389
ME	0.653	0.007	0.007

최저기온의 예측자료는 위도가 높을수록 오차가 더 커지는 것을 알 수 있었다. XGBoost로 보정하였을 때 오차가 0.265~0.462 정도 감소하는 반면, 선형회귀모형으로 보정하였을 때 오차가 0.001~0.085 정도로 거의 감소하지 않은 것을 확인할 수 있었다. 전반적으로 XGBoost가 선형회귀모형 보다 예측자료의 RMSE를 더 많이 감소시켰다.

전체적인 최저기온 예측자료의 오차와 보정에 따른 오차를 살펴 보았을 때, 보정 전 예측자료(Uncorrected Forecast)의 평균 RMSE는 3.779로 나타났으며, 선형회귀모형(Linear Regression Model)은 3.710, XGBoost는 3.389으로 나타났다(Table 8). 선형회귀모형과 XGBoost로 보정한 예측자료가 보정 전 예측자료보다 각각 0.069, 0.390 정도의 RMSE가 감소하는 것을 볼 수 있었다. 그리고 ME (Mean Error)로 분석했을 때, 보정 전은 과대 추정하는 데이터가 많은 경향이 나타났으나 보정 후에는 어느정도 과대 추정이 감소되는 것을 볼 수 있었다.

#### IV. 결 론

본 연구에서는 1개월 예측시스템에서 생산된 2013년부터 2021년까지의 과거 예측(hindcast) 자료를 이용하여 최고기온 및 최저기온 예측 자료의 오차 특성을 분석하였다. 기온 예측자료를 기상관측자료를 이용하여 다양한 해발고도 및 위도 조건에서 평가한 결과, 높은 고도 및 위도에서의 예측 오차가 더 큰 경향을 보였다. 이러한 오차 패턴은 1개월 예측자료 보정에 위도와 고도 정보가 중요하다는 것을 의미한다.

기상관측정보와 더불어 환경자료(해발고도, 위도)를 함께 고려하여 선형회귀모형과 XGBoost 기법으로 예측정보의 오차를 보정하고, 각 모형의 오차 개선 정도를 살펴보았다. 평균적으로 최고기온과 최저기온에 대한 보정 전 예측자료의 RMSE는 각각 4.705 및 3.779으로 나타났으며, 최고기온의 오차가 최저기온보다 더 높게 나타났다. 선형회귀모형과 XGBoost를

적용하여 보정한 경우, 최고기온과 최저기온의 RMSE는 각각 0.203 및 0.069(선형회귀모형), 0.438 및 0.390(XGBoost) 정도 감소하였다. XGBoost가 선형회귀모형보다 더 효과적인 보정효과를 나타냈으며, 높은 고도 및 위도에서의 예측 오차 개선 효과가 더 크게 나타나는 경향이 있다. 종합적으로, 이 연구는 1개월 기온 예측자료의 오차가 다양한 지형적 조건의 영향을 받고 있으며, 다양한 기상 및 환경인자들을 예측정보와 함께 XGBoost와 같은 기계학습법에 충분히 학습시킨다면 예측 정보의 계통적인 오차를 개선하는데 효과적일 것으로 분석되었다.

#### 적 요

본 연구에서는 농촌진흥청과 홍콩과학기술대학교의 공동 개발로 생산된 1개월 예측 자료의 오차를 분석하고, 통계적 보정 기법을 활용한 오차 개선 효과를 살펴보고자 하였다. 이를 위해 2013년부터 2021년까지의 과거 예측(hindcast) 자료, 기상관측자료, 다양한 환경정보들을 수집하고 다양한 환경 조건에서의 오차 특성을 분석하였다. 최고기온과 최저기온의 경우, 해발고도와 위도가 높을 수록 예측 오차가 더 크게 나타났다. 평균적으로, 선형회귀모형과 XGBoost로 보정한 예측자료는 보정 전 예측자료보다 각각 0.203, 0.438(최고기온) 및 0.069, 0.390(최저기온) 정도의 RMSE가 감소했으며, 높은 고도와 위도에서의 오차 개선이 더 크게 나타났다. 모든 분석 조건에서 XGBoost가 선형회귀모형보다 우수한 오차 개선 효과를 나타냈다. 본 연구를 통해 예측 자료의 오차가 지형적 조건에 영향을 받는다는 사실을 확인하였고, XGBoost와 같은 기계학습법이 다양한 환경인자들을 고려하여 효과적으로 오차를 개선할 수 있다는 것을 확인하였다.

## 감사의 글

본 연구는 농촌진흥청 국립농업과학원 “농업과학기술연구(과제번호: PJ016772)”의 지원으로 수행되었습니다.

## REFERENCES

- Chen T., C. Guestrin, 2016: Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 785–794.
- Ha, S., Y. T. Kim, E. S. Im, J. Hur, S. Jo, Y. S. Kim, and K. M. Shim, 2023: Impacts of meteorological variables and machine learning algorithms on rice yield prediction in Korea, *International Journal of Biometeorology*, <https://doi.org/10.1007/s00484-023-02544-x>
- Hur J., E.-S. Im, S. Ha, Y.-S. Kim, E.-S. Kim, J. Lee, S. Jo, K.-M. Shim, and M.-G. Kang, 2023: 1-month prediction on rice harvest date in South Korea based on dynamically downscaled temperature, *Korean Journal of Agricultural and Forest Meteorology* **25**(4), 267–275 (in Korean with English abstract)
- Hur J., Y. S. Kim, S. Jo, K. M. Shim, J.-B. Ahn, M.-J. Choi, Y.-H. Kim, M. Kang, and W. J. Choi, 2021: Estimation of Waxy Corn Harvest Date over South Korea Using PNU CGCM-WRF Chain, *Korean Journal of Agricultural and Forest Meteorology* **23**(4), 405–414 (in Korean with English abstract)
- Im, E. S., S. Ha, L. Qiu, J. Hur, S. Jo, and K. M. Shim, 2021: An evaluation of temperature-based agricultural indices over Korea from the high-resolution WRF simulation. *Front Earth Sci* **9**, 656787. <https://doi.org/10.3389/feart.2021.656787>
- Jo S., J. Lee, K. M. Shim, J.-B. Ahn, J. Hur, Y. S. Kim, W. J. Choi, and M. Kang, 2022: The Advanced Bias Correction Method based on Quantile Mapping for Long-Range Ensemble Climate Prediction for Improved Applicability in the Agriculture Field, *Korean Journal of Agricultural and Forest Meteorology* **24**(3), 155–163 (in Korean with English abstract)
- Pan Bingyue 2018: Application of XGBoost algorithm in hourly PM2.5 concentration prediction, *Earth and Environmental Science* **113**, 012127.
- Qiu, L., E. S. Im, J. Hur, and K. M. Shim, 2020: Added value of very high resolution climate simulations over South Korea using WRF modeling system. *Climate Dynamics* **54**(1), 173–189. <https://doi.org/10.1007/s00382-019-04992-x>
- Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H.-y. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. Dool, O. Zhang, W. Wang, M. Chen, and E. Becker, 2014: The NCEP climate forecast system version 2. *Journal of Climate* **27**, 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note, NCAR/TN- 475+STR, 125 pp.