International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 3
ISSN: 2088-5334

Precipitation Probability Prediction through NWP Bias Correction for South Korea Using Random Forest

Yun Am Seo [a,*], Jieun Cha [b]

[a] Department of Data Science, Jeju National University, 102 Jejudaehak-ro, Jeju-si, 63243, Republic of Korea
[b] AI Meteorological Research Division, National Institute of Meteorological Sciences, Seogwipo-si, 63568, Republic of Korea
Corresponding author: *seoya@jejunu.ac.kr

Abstract— **This study presents the results of an effort to improve the forecast of precipitation (> 0.1 mm/hr or > 0.1 mm/3hr) in the Local Data Assimilation and Prediction System (LDAPS) and the Global Data Assimilation and Prediction System (GDAPS) by applying the Random Forest (RF) model in South Korea. LDAPS and GDAPS are Numerical Weather Prediction (NWP) models operated by the Korea Meteorological Administration (KMA) for weather forecasting. GDAPS operates the Unified Model (UM) and the Korean Integrated Model (KIM). This study used weather forecast data from LDAPS, GDAPS/KIM, and GDAPS/UM. Precipitation forecasts from LDAPS and GDAPS were corrected by RF training with rain gauge observations from about 685 stations. Approximately 35 selected NWP model output variables were used as inputs to the RF training. To reflect recent trends in biases between observations and NWP, the precipitation probability prediction model was designed for real-time learning using a sliding window technique. In addition, the precipitation data had a data imbalance problem with more precipitation cases than non-precipitation cases, so an under-sampling method was applied to solve this problem. Comparing the performance of the proposed method with NWP in predicting precipitation, the CSI was improved by 14.7-23.1% (LDAPS), 33.9% (GDAPS/KIM), and 6.7%-38% (GDAPS/UM) over NWP, and the accuracy was also better. In future research, automating the sampling rate selection to reflect recent weather trends when under-sampling is likely to improve forecast performance.**

Keywords— **Precipitation forecast; imbalanced data; sliding window; bias correction.**

Manuscript received 10 Aug. 2022; revised 29 Dec. 2022; accepted 20 Apr. 2023. Date of publication 30 Jun. 2023.

I. INTRODUCTION

Today, precipitation forecasting is a critical factor affecting not only individual lives but also significantly impacting society as a whole, including water management, agriculture, logistics, and transportation. With the onset of climate change, extreme weather events such as torrential rains, heat waves, heavy snowfalls, floods, and droughts are occurring more frequently around the world, and as the scale of natural disasters increases, so does the demand for accurate precipitation forecasts.

Very short-term forecasts, which predict weather conditions within six hours of the present, are relatively accurate because they are based on ground observations, radar observations, and high-resolution models. However, short-term and medium-term forecasts, which cover days to weeks, depend mostly on numerical weather prediction (NWP) models, and the performance of the NWP models determines the accuracy of the forecast. Therefore, meteorological agencies worldwide develop, continuously improve, and operate the NWP system to suit their environment. However, due to limitations such as errors in the numerical model itself, initial value problems, atmospheric boundary conditions, atmospheric nonlinearity, and physical process parameterization, the forecast data of numerical models always have errors [1]–[3]. Since the model does not reflect changes in the regional and seasonal occurrence mechanisms of precipitation, precipitation prediction can increase the uncertainty of precipitation prediction [4], [5]. Therefore, statistical post-processing methods have been applied in many countries to reduce the uncertainty of precipitation forecasts from numerical models and improve their forecast performance [6]. The most representative statistical models are Model Output Statistics (MOS) and Bayesian Model Averaging (BMA), which reduce the systematic error of the model from the relationship between actual observations and numerical model predictions. They are known to effectively calibrate forecasts from today to 10 days in the future [7]-[9].

However, because they use linear relationships, they require significant tuning by forecasters depending on the forecast data distribution [3].

Recently, machine learning methods have been actively applied to NWP post-processing because of their ability to handle complex nonlinear processes. In a related study, Herman and Schumacher [10] and Loken et al. [11] developed a post-processing model of a numerical weather prediction model using a Random Forest (RF) model and applied a deep neural network for uncertainty diagnosis and bias correction of NWP forecasts. Ko et al. [12] developed a post-processing model for precipitation correction in NWP using XGboost (Extreme Gradient Boosting). In addition, there are other methods such as Quantile Regression, Gradient Boosted Regression Tree, Neural Network, Long Short-Term Memory, ELM (Extreme Learning Machine), and Genetic Programming for post-processing of NWP data [13] –[19].

There are many factors to consider when correcting NWP forecast results, but this study focused on two factors that significantly impact the correction results. The first factor is the difficulty of dealing with heterogeneity resulting from changes in the characteristics of NWP forecast data due to improvements in NWP physical processes and data assimilation processes without reproducing historical data using a new model [20]. In addition, when training a model by accumulating training data over a long period, it is difficult to account for changes in the bias between observations and NWP forecasts due to rapid changes in weather conditions. To address these issues, Allen et al. [21] suggested that learning a small set of recent data at regular intervals would be effective. Therefore, in this study, a model was designed using a Sliding Window Technique (SWT) to learn recent data at regular intervals.

Second, precipitation, which is the focus of this study, is an imbalanced dataset with a dominance of no-precipitation cases. This imbalance significantly affects the prediction performance of the learning model. Random Over-sampling (ROS) and Random Under-sampling (RUS) are representative methods dealing with data imbalance, but there is no universally best method. Therefore, it is necessary to tailor the method to each training domain [22], [23]. The ROS method may add unnecessary noise to the training data and cause overfitting due to overlap, while the RUS method may lose important data [24]. In this study, we applied the RUS method to avoid overfitting due to overlapping data and changes in the physical properties of the NWP prediction data.

This paper aims to predict the probability of precipitation through the bias correction of LDAPS/UM (Local Data Assimilation and Prediction System/Unified Model), GDAPS/UM (Global Data Assimilation and Prediction System/Unified Model), and GDAPS/KIM (Global Data Assimilation and Prediction System/Korean Integrated Model), which are NWP models operated by the Korea Meteorological Administration (KMA). RF was used to train the probability prediction model, and an SWT and RUS method was used to select the training data. The specific data and research methods are described in Section 2, the results and discussion in Section 3, and the conclusion in Section 4.

II. Materials and Method

A. Materials

In this study, the numerical model data used to develop a precipitation probability prediction model are LDAPS/UM, GDAPS/KIM, and GDAPS/UM. The LDAPS/UM has a spatial resolution of 1.5 km, and the vertical layers are provided at 25 hPa intervals for 1000-850 hPa and 50 hPa intervals for 800-200 hPa. It also consists of 602 grids from east to west and 781 grids from north to south. LDAPS/UM is run four times a day (0000, 0600, 1200, 1800 UTC) and provides a 48-hour forecast with 1-hour intervals. The input variables used for training are LDAPS surface and pressure levels (Table 1), latitude, longitude, forecast time, and local time.

TABLE I
List of input variables in the LDAPS

Surface level variables	Pressure level variables
Total Downward SW Flux (Surface)	850 hPa Vertical Velocity
Outgoing LW Flux (TOA)	700 hPa Vertical Velocity
Large-scale precipitation	500 hPa Vertical Velocity
Large-scale precipitation rate	850 hPa Temperature
Richardson Number	700 hPa Temperature
U-wind (10m)	500 hPa Temperature
V-wind (10m)	850 hPa Relative Humidity
Sensible Heat Flux (Surface)	700 hPa Relative Humidity
Latent Heat Flux (Surface)	500 hPa Relative Humidity
Temperature (1.5m)	850 hPa U- wind
Minimum Temperature (1.5m)	700 hPa U- wind
Maximum Temperature (1.5m)	850 hPa V- wind
Relative Humidity (1.5m)	700 hPa V- wind
Dewpoint Temperature (1.5m)	500 hPa Geopotential Height
Total Cloud	
Cloud Top Height	
Mean Sea Level Pressure	※ SW: Short-Wave
(Surface)	
Pressure (Surface)	LW: Long-Wave
K-index	

The ground truth used was the 1-hour cumulative precipitation observed by the KMA's Automated Synoptic Observation System (ASOS) and Automatic Weather Station (AWS). Observational data were collected from 705 stations (Figure 1).

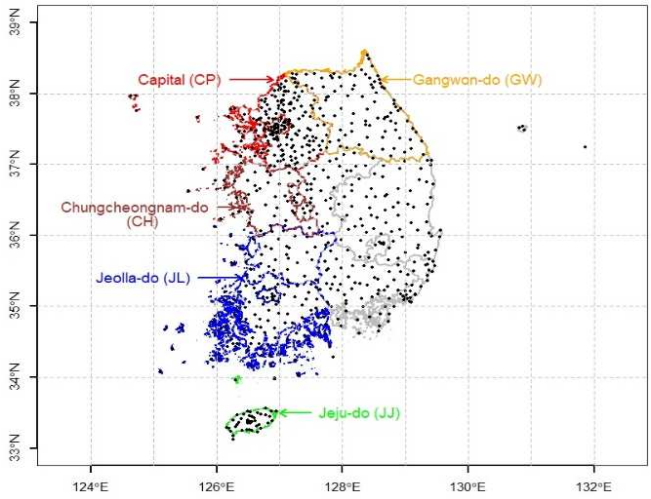


Fig. 1 Training sites in South Korea, Jeju Island (JJ), Gangwon-do (GW), Chungcheongnam-do (CH), Jeolla-do (JL), and the Capital Area (CP)

Data from grid points with the closest Euclidean distance were used. Then, LDAPS learning and verification were performed for the summer (June-August) of 2019-2021, and the precipitation threshold for the precipitation and non-precipitation was set to 0.1 mm/hr based on the observed precipitation.

The input variables of the precipitation probability prediction model using GDAPS forecast data are shown in Table 2. Latitude, longitude, forecast time, and local time were also used. The ground truth of the output variable is the observed 3-hour cumulative precipitation (threshold 0.1 mm/3hr). GDAPS/KIM and GDAPS/UM are run four times daily (0000, 0600, 1200, 1800 UTC) and provide an 87-hour forecast at 3-hour intervals. The KIM model has a spatial resolution of 12 km, and the UM model has a spatial resolution of 10 km. The GDAPS-based models were trained and tested using data from KIM's 2021 summer season (June-August) and UM's 2020-2021 summer season (June-August).

TABLE II
LIST OF INPUT VARIABLES IN THE GDAPS

Surface level variables	Pressure level variables
Large-scale Prcp.	850 hPa Temp.
U-wind (10m)	700 hPa Temp.
V-wind (10m)	500 hPa Temp.
Sensible Heat Flux (Surface)	850 hPa Relative Humidity
Temp. (1.5m)	700 hPa Relative Humidity
Relative Humidity (1.5m)	500 hPa Relative Humidity
Dewpoint Temp. (1.5m)	850 hPa U-wind
Pressure (Surface)	700 hPa U-wind
Geometric Height	850 hPa V-wind
Low Cloud Cover	700 hPa V-wind
Medium Cloud Cover	850 hPa WB. Potential Temp
High Cloud Cover	700 hPa WB. Potential Temp
Total Cloud Cover	500 hPa WB. Potential Temp
Geopotential Height	700hPa Geopotential Height
Downward LW Rad. Flux	
Downward LW Rad. Flux (x)	
Large Scale Water Prcp. Validation	
Convective Water Prcp.	※ Temp.: Temperature
K-index	Prcp.: Precipitation
Total Prcp. Rate	LW: Long-Wave
Convective Prcp.	

B. Methods

The learning and prediction system of the precipitation probability prediction model based on NWP (Numerical Weather Prediction) consists of preprocessing, model learning, and inference stages, as shown in Figure 2. In the preprocessing stage, the SWT is used to select the learning target data to reflect the bias trend according to the change of the numerical model and the recent weather change, and under-sampling is applied to solve the data imbalance problem of precipitation and non-precipitation. In the model learning stage, a precipitation probability prediction model is learned using RF, and finally, the precipitation probability is predicted using the trained model.

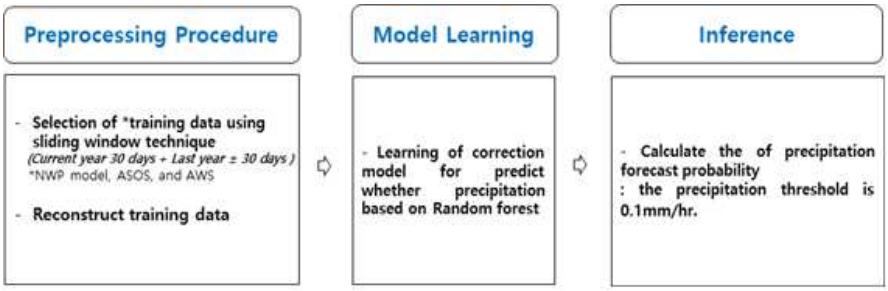
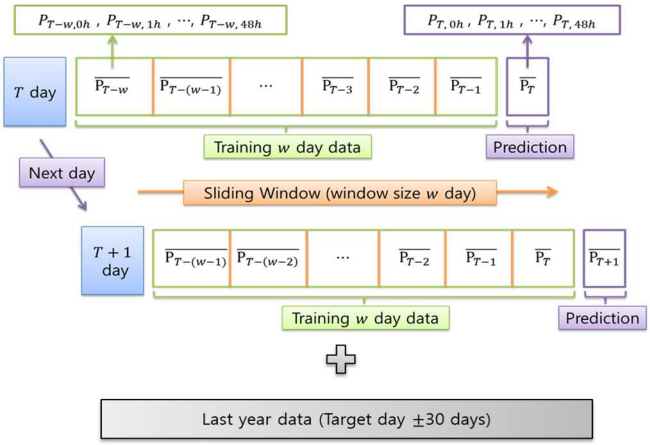


Fig. 2 Overall framework of precipitation probability prediction model for learning system based on NWP model

SWT in preprocessing for real-time learning is effective for modeling time series data. When applied to data whose characteristics change over time, such as seasonal changes in weather, it can improve the performance of learning models [25]-[29]. In this study, the training data is selected by selecting historical data equal to the Window Size (WS) from the forecast target date, and on the next day, the new data is included in the training, and the data outside the WS is excluded from the training. In a sliding window, the WS significantly impacts the amount of training data and the training time. Since meteorological phenomena such as precipitation have very different characteristics in different seasons, it is necessary to select an appropriate WS, because if the WS is very large, it is possible to learn data from seasons that are less relevant to the present. The WS used in this study is 30 days based on the results of NIMS (National Institute of Meteorological Sciences) [30]. In addition, it is difficult to learn enough precipitation patterns that may occur in the future using only the current year's data, so ±30 days were used based on the previous year's forecast target date (Fig. 3).



w: window size, T: target day, P: training data

Fig. 3 Schematic diagram of the sliding window technique

The training data selected by SWT used RUS to mitigate the imbalance between precipitation and non-precipitation

data. The target data of this study consists of about 87-93% (based on observed precipitation for June-August 2019-2021) non-precipitation cases. Therefore, most precipitation data is extracted, and only some non-precipitation data is extracted. In this case, the proportion of non-precipitation cases extracted in RUS greatly affects the tendency of the learning model to overestimate or underestimate the prediction. In this study, we experimented with the following two methods depending on the extraction rate of non-precipitation cases (Figure 4). The first method extracts 85% or 95% of the precipitation data in the training data, and the second method extracts 30% or 40% of the non-precipitation cases (hereafter referred to as Method 1). This method does not consider the relative distribution of precipitation and non-precipitation in the training data, which may have the disadvantage that the training model is more likely to underestimate precipitation when it is relatively distributed. To compensate for the disadvantages of Method 1, the second method is to extract the number of precipitation cases multiplied by a certain ratio so that the number of precipitation cases is proportional to the number of rainfall cases when extracting precipitation data (hereafter referred to as Method 2). The training data extracted by Method 2 can be expected to have better prediction performance than Method 1 by having a constant ratio between precipitation and non-precipitation.

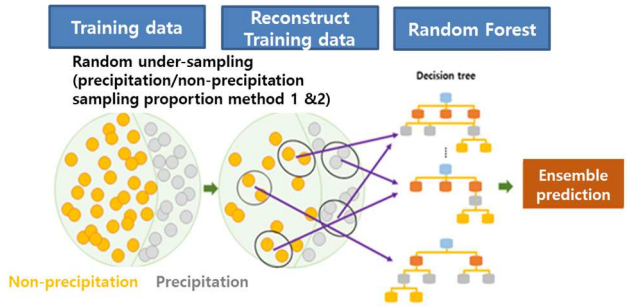


Fig. 4 Schematic diagram for training a precipitation probability prediction model

The RF model used in the training phase of the precipitation probability prediction model is an ensemble model that trains multiple decision tree models and combines the results to make a prediction [31]. RF is generally less computationally expensive to train than other machine learning methods because it favors overfitting the training model and requires fewer parameters to be estimated. Moreover, it is often used for learning predictive or classification models in meteorology and other fields because it guarantees a certain good performance [32] - [34]. In this study, since we have the constraint of being able to quickly learn new data selected by SWT every day, a model with a large number of parameters to be estimated or a large computational cost is not appropriate, so we used RF. The number of trees in RF is 500, and other initial parameters [30].

To evaluate the performance of the trained model, we used the verification metrics shown in Table 3, which are commonly used to evaluate the performance of precipitation forecasting. In the verification metric, ACC represents the average agreement between forecast and observation, and CSI is ACC subtracted by the value of C, which is not related to precipitation. POD is the ratio of the accuracy of the forecast to the observations, with ACC, CSI, and POD ranging from 0 to 1, with a perfect forecast having a value of 1. FAR is the percentage of wrong precipitation forecasts, ranging from 0 to 1, with a perfect forecast having a value of 0. Finally, bias ranges from 0 to infinity and has a value of 1 when the range of predicted precipitation is the same as the range of observed precipitation. Bias has nothing to do with accuracy but the relative skewness of the observed and predicted values. Therefore, a bias less than 1 indicates a tendency to underestimate, and a bias greater than 1 indicates a tendency to overestimate.

TABLE III
THE VERIFICATION METRICS FOR THE PREDICTION OF PRECIPITATION IN TWO CATEGORICAL DATA

Verification metrics	Formula	Range
Accuracy	$\mathrm{ACC} = \frac{H+C}{H+M+F+C}$	$(0 \le \mathrm{ACC} \le 1)$
Critical success index	$\mathrm{CSI} = \frac{H}{H+M+F}$	$(0 \le \mathrm{CSI} \le 1)$
False alarm ratio	$\mathrm{FAR} = 1 - \frac{H}{H+F}$	$(0 \le \mathrm{FAR} \le 1)$
Probability of detection	$\mathrm{POD} = \frac{H}{H+M}$	$(0 \le \mathrm{POD} \le 1)$
Bias	$\mathrm{Bias} = \frac{H+F}{H+M}$	$(0 \le \mathrm{Bias})$

H: Hits, M: Misses, F: False alarms, C: Correct negatives

III. Results and Discussion

This section describes the performance and characteristics of the precipitation probability prediction models trained for each NWP by applying the method proposed in this study to the following LDAPS/UM, GDAPS/KIM, and GDAPS/UM forecast data.

A. LDAPS/UM-based Precipitation Probability Prediction Model

The ground truth for training the LDAPS/UM-based precipitation probability prediction model was labeled as precipitation if the observed precipitation was greater than or equal to 0.1 mm/1hr, and as no precipitation otherwise. The output of the trained model, the probability of precipitation, was labeled as precipitation if it was greater than or equal to 0.5.

TABLE IV
RESULTS FOR THE SAMPLING PROPORTION (NPR: NON-PRECIPITATION, PR: PRECIPITATION, METHOD 1) EXPERIMENT FOR SOUTH KOREA DURING THE SUMMER MONTHS (JUNE-AUGUST) FROM 2019-2021

Year	Sampling proportion (npr / pr)	CSI	POD	FAR	Bias	ACC
2019	LDAPS / UM	0.314	0.423	0.450	0.769	0.925
	30 / 85 (%)	0.371	0.544	0.461	1.010	0.925
	30 / 95 (%)	0.374	0.566	0.476	1.081	0.923
	40 / 85 (%)	0.358	0.482	0.419	0.831	0.929
	40 / 95 (%)	**0.364**	**0.505**	**0.434**	**0.892**	**0.928**
2020	LDAPS / UM	0.339	0.453	0.427	0.791	0.880
	30 / 85 (%)	0.396	0.644	0.494	1.272	0.866
	30 / 95 (%)	0.395	0.668	0.508	1.357	0.861
	40 / 85 (%)	0.389	0.573	0.452	1.046	0.878
	40 / 95 (%)	**0.393**	**0.600**	**0.467**	**1.126**	**0.874**
2021	LDAPS / UM	0.286	0.398	0.494	0.786	0.914
	30 / 85 (%)	0.360	0.536	0.478	1.027	0.917
	30 / 95 (%)	0.361	0.557	0.493	1.100	0.914
	40 / 85 (%)	0.349	0.479	0.436	0.850	0.922
	40 / 95 (%)	**0.354**	**0.501**	**0.452**	**0.913**	**0.920**

Table 4 shows the experimental results for sampling Method 1 (non-precipitation sampling proportion 30 and 40%, precipitation sampling proportion 85 and 95%) on LDAPS/UM data produced four times a day during the summer months of 2019-2021. For all sampling proportions, the LDAPS/UM model outperforms the LDAPS model.

Specifically, CSI and POD were dominant in all years when the non-precipitation proportion was 30%, while FAR and ACC were dominant when the non-precipitation proportion was 40%. However, bias tended to be overestimated when the non-precipitation proportion was set to 30%. For precipitation, 95% was always better than 85% for all verification metrics. Overall, the best performance for each metric was achieved when the precipitation proportion was set to 95%, and the precipitation rate was set to 40%. Compared to the performance of the LDAPS model, CSI was improved by 15.9-23.8%, POD by 19.4-32.5%, and Bias by 16.0-42.4%. This means that even if the training data is under-sampled and reconstructed, the NWP model data can be well corrected.

Precipitation is characterized by different seasons but also by different regions. While the model proposed in this study performed better than LDAPS/UM for the whole of South Korea, when CSI and POD were analyzed for each of the five rainy administrative regions in South Korea, some regions performed worse than the original data (Table 5). This was analyzed because, in years and regions with relatively more rainless days than precipitation days, sampling Method 1 did not alleviate the data imbalance problem because the rainless cases were extracted relatively more than the precipitation cases. Therefore, sampling Method 2 was used to compensate for this problem.

TABLE V
THE RATE OF IMPROVEMENT OF THE VERIFICATION METRICS COMPARED TO LDAPS/UM (PERIODS AS SHOWN IN TABLE IV). UNIT: %.

Year	Jeju-do		Gang won-do		Jeolla-do		Chung cheong-do		Capital area	
	CSI	POD	CSI	POD	CSI	POD	CSI	POD	CSI	POD
2019	15.0	22.8	8.6	9.3	9.7	13.4	**-0.1**	**-2.7**	**-12.6**	**-17.1**
2020	28.0	49.9	9.8	16.9	15.5	34.1	1.8	8.8	10.6	21.5
2021	20.1	27.4	14.7	5.7	38.6	49.6	18.4	17.8	**-0.7**	**-11.9**

For sampling Method 2, we kept the precipitation proportion fixed at 95% and sampled the number of non-precipitation cases multiplied by 2.5, 3, 3.5, and 4 times the number of precipitation cases, respectively. The results are shown in Table 6.

TABLE VI
RESULTS FOR THE SAMPLING RATIO (NON-PRECIPITATION CASE: PRECIPITATION CASE × RATIO, METHOD 2) EXPERIMENT FOR SOUTH KOREA (PERIODS AS SHOWN IN TABLE IV).

Year	Non-precipitation sampling ratio	CSI	POD	FAR	Bias	ACC
2019	LDAPS / UM	0.314	0.423	0.450	0.769	0.925
	2.5	0.373	0.653	0.534	1.400	0.911
	3	0.376	0.619	0.510	1.264	0.915
	3.5	**0.376**	**0.587**	**0.489**	**1.148**	**0.921**
	4	0.374	0.559	0.470	1.054	0.924
2020	LDAPS / UM	0.339	0.453	0.427	0.791	0.880
	2.5	0.392	0.681	0.520	1.417	0.857
	3	0.392	0.639	0.496	1.270	0.866
	3.5	**0.389**	**0.602**	**0.476**	**1.149**	**0.872**
	4	0.384	0.569	0.458	1.050	0.876
2021	LDAPS / UM	0.286	0.398	0.494	0.786	0.914
	2.5	0.363	0.553	0.485	1.074	0.916
	3	0.359	0.517	0.459	0.956	0.920
	3.5	**0.352**	**0.486**	**0.439**	**0.867**	**0.922**
	4	0.345	0.462	0.424	0.801	0.923

For South Korea, similar to Method 1, the verification metrics are improved over the LDAPS/UM model in all cases. When each verification metric is evaluated comprehensively, the non-precipitation case performs best, with 3.5 times the performance of the precipitation case.

When comparing the performance of sampling method 1 and Method 2 by region, it was found that the performance of Method 2 improved in regions where Method 1 did not perform as well as LDAPS/UM (Figure 5). Therefore, it was analyzed that it is effective to sample non-precipitation cases in proportion to precipitation cases, which are a minority group.

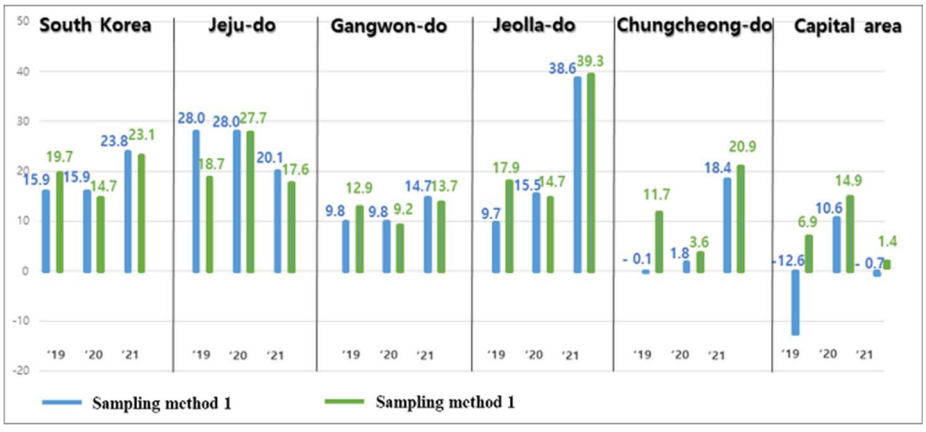


Fig. 5 The rate of CSI improvement for sampling methods 1 and 2 over LDASP/UM, with periods as shown in Table 4. Unit: %

Figure 6 shows the CSI and Bias by forecast time for LDAPS/UM and the proposed model. CSI shows that the proposed model outperforms the LDAPS model at all forecast times. On the other hand, the bias tends to overestimate compared to LDAPS/UM, but the bias of the constrained model is closer to 1 than LDAPS/UM, indicating that it compensates well for the bias of NWP precipitation.

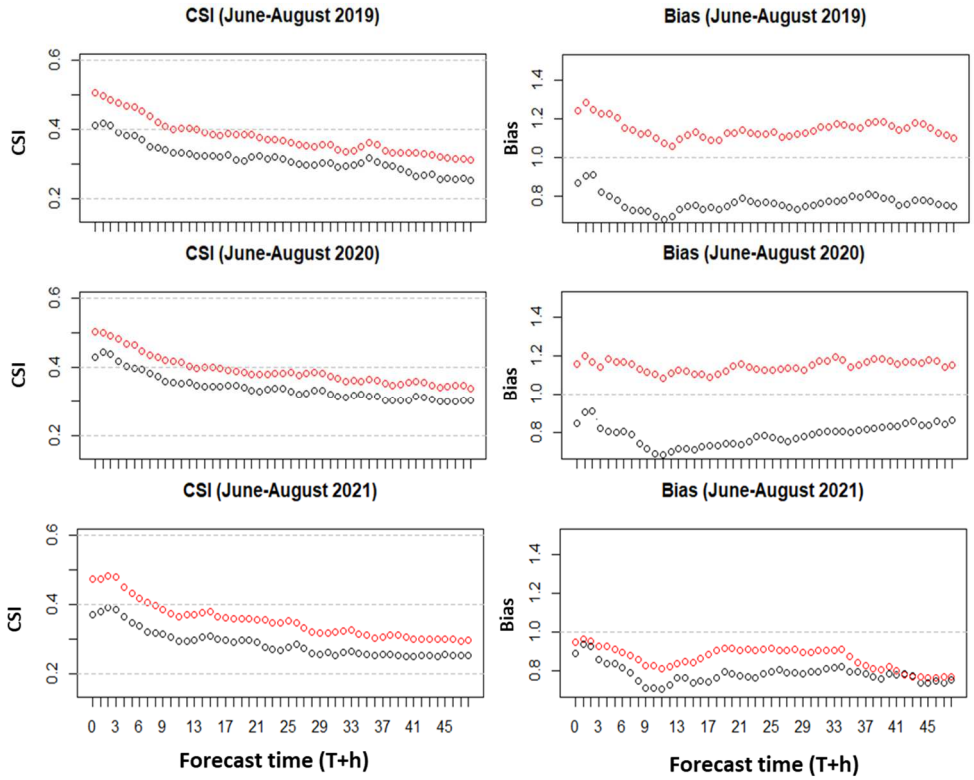


Fig. 6 CSI and Bias comparison of the proposed model and LDAPS/UM for South Korea at each prediction time. The red line shows the proposed model, and the black line shows LDAPS/UM

B. GDAPS-based Precipitation Probability Prediction Model

It shows an improvement in CSI over the GDAPS/KIM model in all cases. However, for POD, the performance of POD decreases as the proportion of non-precipitation cases increases, with a decrease of -13.9% for 1.5x, -22.1% for 2x, and -28.4% for 2.5x. For bias, doubling the number of non-precipitation cases was the most dominant method, with performance improvements of 86.8% for 1.5x, 95.5% for 2x, and 86.7% for 2.5x. When each verification metric was evaluated together, it was found that the non-precipitation case performed best when the precipitation case was doubled. The difference between the best performances of GDAPS/KIM and LDAPS/UM indicates that the distribution of precipitation and non-precipitation is different for other NWP models. It is necessary to experiment with sampling proportions of finer units in the future.

TABLE VII
RESULTS FOR THE SAMPLING RATIO (NON-PRECIPITATION CASE: PRECIPITATION CASE × RATIO, METHOD 2) EXPERIMENT FOR SOUTH KOREA DURING THE SUMMER MONTHS (JUNE-AUGUST) FROM 2021

Year	Non-precipitation sampling ratio	CSI	POD	FAR	Bias	ACC
2021	GDAPS / KIM	0.290	0.697	0.668	2.100	0.772
	1.5	0.388	0.600	0.476	1.145	0.874
	2	**0.386**	**0.543**	**0.429**	**0.951**	**0.884**
	2.5	0.377	0.499	0.392	0.820	0.890

Table 8 shows the results of the precipitation probability prediction model using GDAPS/UM data. For 1.5x and 2x the 2020 precipitation cases, the CSI improvement over GDAPS/UM is 9.2% and 6.7%, respectively, FAR is 31.3% and 21.2%, ACC is 7.9% and 9.4%, and bias corrects the tendency to overestimate close to 1. However, FAR underperformed GDAPS/UM at both 1.5x and 2x. For 2021, using 1.5x and 2x precipitation cases, the CSI improvement over the model was 37.6% and 38.0%, respectively, FAR was 39.0% and 57.1%, and ACC was 15.9% and 17.8%. In both cases, the CSI improvements were almost identical, but given the other verification metrics, it concluded that doubling was optimal. As with the LDAPS/UM results, the reason for the higher performance improvement in 2021 compared to 2020 is largely due to the percentage of precipitation each year. This suggests that we should further devise a method to resolve the data imbalance considering the current or recent precipitation distribution.

TABLE VIII
RESULTS FOR THE SAMPLING RATIO (NON-PRECIPITATION CASE: PRECIPITATION CASE × RATIO, METHOD 2) EXPERIMENT FOR SOUTH KOREA DURING THE SUMMER MONTHS (JUNE-AUGUST) FROM 2020-2021

Year	Non-precipitation sampling ratio	CSI	POD	FAR	Bias	ACC
2020	GDAPS / UM	0.390	0.758	0.554	1.699	0.756
	1.5	0.426	0.664	0.457	1.222	0.816
	2	**0.416**	**0.597**	**0.422**	**1.032**	**0.827**
2021	GDAPS / UM	0.279	0.725	0.688	2.322	0.749
	1.5	0.384	0.617	0.495	1.222	0.868
	2	**0.385**	**0.550**	**0.438**	**0.979**	**0.882**

IV. CONCLUSION

This study developed a model to predict the probability of precipitation by post-processing the forecast data of the NWP model for more accurate precipitation prediction. The

proposed model applies SWT to reflect the recent bias between observations and NWP, and the RUS method alleviates the data imbalance between precipitation and non-precipitation. When sampling via RUS, the sampling proportion of precipitation cases was 95%, and the sampling proportion of non-precipitation cases was proportional to precipitation cases. The extracted data are trained in real time immediately after RF produces the daily NWP data, and the probability of precipitation on the target day is predicted. In this study, when the proposed method is applied to LDAPS/UM, GDAPS/KIM, and GDAPS/UM operated by KMA, the bias with observation is reduced and the prediction performance is improved compared to NWP for precipitation and no-precipitation prediction. The prediction performance of the proposed model improves the CSI by 14.7-23.1% compared to LDAPS/UM for South Korea, 33.9% for GDAPS/KIM, and 6.7%-38% for GDAPS/UM.

With RUS sampling, the sampling rate significantly impacts model performance. Weather phenomena such as precipitation have different characteristics depending on the season and region, and their patterns vary from year to year. The optimal sampling rate may be different for each region and season. Estimating the appropriate sampling rate will help your model perform better. This could be done by automating the selection of an appropriate sampling rate that takes into account differences in the probability distribution of precipitation across the training set and the location and scale parameters of the probability distribution of recent precipitation and by using AutoML to automate the selection of the sampling rate.

In this study, a precipitation threshold of 0.1 mm was used. If the precipitation threshold range is extended to predict heavy precipitation, the data imbalance will be more severe because heavy precipitation has a much smaller number of cases than weak precipitation. Therefore, a mixed sampling method using RUS for the majority population and data augmentation techniques for the minority population will be effective. Data augmentation has been used to address data imbalance using generative models, noise transformers, etc. Future research could improve forecast performance by improving sampling methods by expanding the range of precipitation thresholds and developing automated methods for selecting sampling rates.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2022-00166370). Cha's work is supported by Developing AI Technology for Weather Forecasting (KMA2021-00121).

REFERENCES

[1] A. Jabbari, and D.H. Bae, "Application of artificial neural networks for accuracy enhancements of real-time flood forecasting in the Imjin basin". *Water*, vol.10, no. 11, pp. 1626, 2018.

[2] T. Zhang, W. Lin, Y. Lin, M. Zhang, H. Yu, K. Cao, and W. Xue, "Prediction of tropical cyclone genesis from mesoscale convective systems using machine learning", *Weather and Forecasting*, vol. 34, no. 4, pp. 1035-1049, 2019.

[3] S.E. Haupt, W. Chapman, S.V. Adams, and C. Kirkwood, "Combining artificial intelligence with physics-based methods for probabilistic renewable energy forecasting", *Energies*, Vol. 13, pp. 1979, 2020.

[4] J. Y. Lee, M. Kwon, K. S. Yun, S. K. Min, I. H. Park, Y. G. Ham, E.K. Jin, J.H. Kim, K.H. Seo, W. Kim, S.Y. Yim, and J. H. Yoon, "The long-term variability of changma in the East Asian summer monsoon system: A review and revisit", *Asia-Pacific Journal of Atmospheric Sciences*, vol. 53, pp. 257-272, 2017.

[5] H.J. Song, and B.J. Sohn, "Polarizing rain types linked to June drought in the Korean peninsula over last 20 years" *International Journal of Climatology*, vol . 40, pp. 2173-2182, 2020.

[6] G.M. Carter, J.P. Dallavalle, and H.R. Glahn, "Statistical forecasts based on the National Meteorological Center's numerical weather prediction system", Weather and Forecast, vol. 4, pp. 401–412, 1989.

[7] A.E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, "Using bayesian model averaging to calibrate forecast ensembles", *Monthly Weather Review*, col. 133, pp. 1155-1174, 2005.

[8] J.M Sloughter, T. Gneiting, and A. Raftery, "Probabilistic wind speed forecasting using ensembles and Bayesian model averaging", *Journal of the American Statistical Association*, vol. 105, pp.25–35, 2010.

[9] S. Hemri, M. Scheuerer, F. Pappenberger, K. Bogner, and T. Haiden, "Trends in the predictive performance of raw ensemble weather forecasts", *Geophys. Res. Lett.*, vol. 41, pp. 9197–9205, 2014.

[10] G.R. Herman, and R.S. Schumacher, "Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests", *Monthly Weather Review*, vol. 146, pp. 1571-1600, 2018.

[11] E.D. Loken, A.J. Clark, A. McGovern, M. Flora, and K. Knopfmeier, "Post-processing next-day ensemble probabilistic precipitation forecasts using random forests", *Weather and Forecasting*, vol. 34, pp. 2017–2044, 2019.

[12] C.M. Ko, Y.Y. Jeong, Y.M. Lee, and B.S. Kim, "The development of a Quantitative Precipitation Forecast Correction Technique Based on Machine Learning for Hydrological Application", *Atmosphere*, vol. 11, no. 1, pp. 111, 2020.

[13] M. Taillardat, O. Mestre, M. Zamo, and P. Naveau, "Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics". *Monthly Weather Review,* vol. 144, pp. 2375–2393, 2016.

[14] K. Bakker, Whan, K. Knap, and M. Schmeits, "Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation" *Solar Energy*, vol. 191, pp. 138-150, 2019.

[15] D. Cho, C. Yoo, J. Im, and D.H. Cha, "Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas", *Earth and Space Science,* vol. 7, no. 4, pp. e2019EA000740, 2020, doi: 10.1029/2019EA000740.

[16] Z. Tian, S. Li, and Y. Wang, "A prediction approach using ensemble empirical mode decomposition-permutation entropy and regularized extreme learning machine for short-term wind speed" Wind Energy, vol. 23, no. 2, 177-206, 2020.

[17] C. Kirkwood, T. Economou, H. Odbert, and N. Pugeault, "A framework for probabilistic weather forecast post-processing across models and lead times using machine learning" *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, pp. 20200099, 2021.

[18] P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T.Hoefler, "Deep learning for post-processing ensemble weather forecasts", *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, pp. 20200092, 2021.

[19] D. Cho, C. Yoo, B. Son, J. Im, D. Yoon, and D. H. Cha, "A novel ensemble learning for post-processing of NWP Model's next-day maximum air temperature forecast in summer using deep learning and statistical approaches". *Weather and Climate Extremes*, vol. 35, no. 100410, 2022.

[20] S. Vannitsem, J.B. Bremnes, J. Demaeyer, G.R. Evans, J. Flowerdew, and S. Hemri, "Statistical post-processing for weather forecasts: Review, challenges, and avenues in a big data world", *Bulletin of the American Meteorological Society*, vol. 102, no. 3, pp. E681-E699, 2020.

[21] N.B. Allen, B.W. Nelson, D. Brent, and R.P. Auerbach, "Short-term prediction of suicidal thoughts and behaviors in adolescents: Can recent developments in technology and computational science provide a breakthrough?", *Journal of affective disorders*, vol. 250, no. 163-169, 2019.

[22] J.L. Leevy, T.M. Khoshgoftaar, R.A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data" *Journal of Big Data*, vol. 5, no.1, pp.1-30, 2018.

[23] T. Sasada, Z.Liu, T. Baba, K. Hatano, Y. Kimura, "A resampling method for imbalanced datasets considering noise and overlap", *Procedia Computer Science*, vol. 176, pp. 420-429, 2020.

[24] G.U. Park, and I. Jung, "Comparison of resampling methods for dealing with imbalanced data in binary classification problem", *The Korean Journal of Applied Statistics*, vol. 32, no. 3) 349-374, 2019.
[25] M. Alam, and M. Amjad, "Weather forecasting using parallel and distributed analytics approaches on big data clouds", *Journal of Statistics and Management Systems*, vol. 22, no. 4, pp. 791-799, 2019.
[26] A. Bhatt, W. Ongsakul, and J.G. Singh, "Sliding window approach with first-order differencing for very short-term solar irradiance forecasting using deep learning models", *Sustainable Energy Technologies and Assessments*, vol. 50, pp.101864, 2022.
[27] D. R. Garrido, and M. S. Lorenzo, "Application of the Sliding Window Method to the Short Range Prediction System for the Correction of Precipitation Forecast Errors", *Environmental Sciences Proceedings*, vol. 19, no.1, pp. 53, 2022.
[28] M.S. Saravanan, "Prediction of Temperature for Next Three Days Using Decision Tree Algorithm by Comparing Sliding Window Algorithm for Better Accuracy", *ECS Transactions*, vol. 107, no. 1, pp. 14097, 2022.
[29] C. Chen, Q. Zhang, M.H. Kashani, C. Jun, S.M. Bateni, S.S. Band, S.S. Dash, and K.W. Chau, "Forecast of rainfall distribution based on fixed sliding window long short-term memory", *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 248-261, 2022.
[30] H.S. Lee, H.S. Park, S.Y. Kim, J.H. Park, C.K. Park, Y.A. Seo, I.K. Kim, S.Y. Roh, J.S. Park, H.J. Song, M.K. Hong, and Y.S. Ryu, "Development of the AI technique for the prediction of rainfall optimized over the Korean peninsula", National Institute of Meteorological Sciences, Korea, Tech. Rep. 11-1360620-000209-10, 2020.
[31] L. Breiman, "Random forests", *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.
[32] A.J. Hill, G.R. Herman, and R.S. Schumacher, "Forecasting severe weather with random forests", *Monthly Weather Review*, vol. 148, no. 5, pp. 2135-2161, 2020.
[33] Y. He, C. Chen, B. Li, and Z. Zhang, "Prediction of near-surface air temperature in glacier regions using ERA5 data and the random forest regression method", *Remote Sensing Applications: Society and Environment*, vol. 28, pp. 100824, 2022.
[34] E.D. Loken, A.J. Clark, and A. McGovern, "Comparing and interpreting differently designed random forests for next-day severe weather hazard prediction", *Weather and Forecasting*, vol. 37, no. 6, pp. 871-899, 2022.