



## RESEARCH ARTICLE

# Deep learning-based precipitation bias correction approach for Yin–He global spectral model

Yi-Fan Hu<sup>1</sup>  | Fu-Kang Yin<sup>2</sup>  | Wei-Min Zhang<sup>2</sup>

<sup>1</sup>College of Computer Science and Technology, National University of Defense Technology, Changsha, China

<sup>2</sup>College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China

**Correspondence**

Fu-Kang Yin, College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China.

Email: yinfukang@nudt.edu.cn

**Abstract**

In this paper, a data-driven bias correction approach based on deep learning is proposed, which is appropriate for the Yin–He global spectral model (YHGSM) re-forecasting. The proposed architecture involves four U-Net-based networks estimating the proper bias correction models for YHGSM re-forecasting that consider as correction factors the geopotential, specific humidity, and vertical velocity on three pressure levels from the YHGSM model. The proposed models are then evaluated for their bias correction capability on the 3-h cumulative precipitation over the region of China between 15°–54.5° N, and 63°–122.5° E. The results revealed that U-Net-based models could reduce the root mean squared error (RMSE) and improve the threat scores (TSs), especially for heavy precipitation and rainstorms.

**KEYWORDS**

bias correction, deep learning, precipitation, YHGSM

## 1 | INTRODUCTION

The atmosphere is a dynamic system that exhibits limited predictability, and thus weather forecasts are inherently uncertain. Despite the increasing accuracy of weather forecasts, there is an element of uncertainty in all predictions. The uncertainties in initial conditions and approximations adopted during model formulation impose bias and systematic errors of the numerical weather prediction (NWP) model output. Specifically, weather variables are dependent on local topography and environmental conditions. To remove systematic errors and improve the NWP model output accuracy, several methods have been developed. Bias correction is one of the most important methods, with statistical post-processing (Hemri et al., 2014) being the typical one. For example, the model output statistics (MOS) approach establishes a linear statistical relationship between model predictions

and actual observations to improve forecasting accuracy (Wilks, 2009). The work of J. Wang et al. (2015) employs a Bayesian probability decision bias correction to improve the accuracy of the rainstorm set probability forecasting in the Sichuan Basin of China. Carvalho et al. (2011) developed a Kalman filter to correct the systematic errors for the monthly mean temperature simulated by the PRECIS model.

As the availability of big meteorological data is increasing, exploiting machine learning development and computing performance improvement, data-driven approaches were introduced into NWP. As a result, several machine learning methods have been introduced for meteorological research and have achieved considerable success in weather forecasting (Grover et al., 2015; Hernandez et al., 2016) and data post-processing (Dupuy et al., 2020; Rasp & Lerch, 2018). Weyn et al. (2020) proposed a data-driven global weather forecasting

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Meteorological Applications* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

framework on a cubed sphere using a deep convolutional neural network (CNN). Witt et al. (2020) provided a multi-modal benchmark dataset (called RainBench) and a library PyRain for data-driven extreme precipitation forecasting. Furthermore, P. Larraondo et al. (2019) evaluated several solutions such as random forests, VGG-16, SegNet, and other methods to derive total precipitation and found that U-Net significantly outperforms the competitor techniques. Dupuy et al. (2020) also found that U-Net performs better than other traditional machine learning methods such as random forests and logistic quantile regression. Singh et al. (2021) pointed out that residual learning-based U-Net can unravel physical relationships to target precipitation.

U-Net is a CNN-based model that can flexibly and effectively model non-linear functions (Rasp & Lerch, 2018; Ronneberger et al., 2015). U-Net has been successfully applied in many different fields, with Ronneberger et al. (2015) employing a U-Net model for biomedical segmentation. P. Larraondo et al. (2019) proposed a data-driven approach to parameterize an encoder-decoder type of CNN for precipitation estimation. Their research demonstrates that this type of CNN can interpret the spatial information in the geopotential field and infer total precipitation with a high degree of accuracy. Furthermore, X. Xu et al. (2019) proposed a novel deep convolution framework entitled ordinal boosting autoencoder (OBA) that is appropriate for the bias correction during the numerical precipitation prediction process. Accordingly, P. R. Larraondo et al. (2020) introduced a methodology for optimizing neural network models using a combination of continuous and categorical binary indices in the context of precipitation forecasting. Moreover, Dupuy et al. (2020) suggested a U-Net architecture to perform cloud cover forecast post-processing, while Trebing et al. (2021) proposed a U-Net-based model to nowcast Netherlands' precipitation.

Precipitation is one of the most critical weather conditions affecting people's work and life. Precipitation forecasting depends on many parameters such as temperature, pressure, humidity, wind direction, and wind speed. Due to precipitation's complexity and the uncertainty of NWP, the forecast bias cannot be ignored, especially for heavy precipitation and rainstorms (Shu et al., 2018; Zhang et al., 2019). To compensate for that, a recently developed precipitation forecast bias correction tool was produced by the European Flood Awareness System (EFAS) to improve river discharge forecasts. Furthermore, the research work of Jurlina et al. (2020) of the European Centre for Medium-Range Weather Forecasts (ECMWF) demonstrates that precipitation bias correction can indeed improve flood forecasting. Besides, Zhang

et al. (2019) used a deep learning strategy utilizing eight meteorological factors and improved the root mean square error (RMSE) and the threat scores (TSs) of rainfall forecasts.

The Yin-He global spectral model (YHGSM; Peng et al., 2019, 2020; Wu et al., 2011; Yang et al., 2015; Yin et al., 2018, 2021) re-forecast dataset is released by the National University of Defense Technology (NUDT) for the study of machine learning in meteorological application. The ERA5 re-analysis is used to initialize YHGSM to generate re-forecasts products. YHGSM re-forecasts dataset has a long-term data archive and a relatively stable systematic error. Motivated and inspired by the ongoing research in these areas, we propose a 3-h cumulative precipitation bias correction method based on deep learning for YHGSM. This paper aims to provide a deep learning bias correction method for 3-h cumulative precipitation of YHGSM re-forecasts. The contributions of our work are summarized as follows:

- Four U-Net-based precipitation bias correction models are developed for a YHGSM-based re-forecast operational NWP model. The performance of these models is then evaluated.
- The categorical binary metrics proposed by P. R. Larraondo et al. (2020) are introduced to develop the  $TS_{\text{diff}}$  metric, and then we develop a loss function utilizing the combination of  $TS_{\text{diff}}$  and max absolute error (MAE).
- The experiments indicate that the suggested model improves the prediction accuracy of the 3-h cumulative precipitation, especially for heavy precipitation and rainstorms.

The remainder of this paper is as follows. Section 2 presents the employed dataset, while Section 3 introduces the U-Net-based models, the evaluation scheme, the loss function explored, and the experimental set-up. Section 4 presents the experimental results, and finally, Section 5 concludes this work.

## 2 | DATASET

### 2.1 | Dataset

This paper utilizes the ERA5 (Hersbach et al., 2020) dataset, a global climate re-analysis dataset produced by ECMWF. From the entire dataset, the total precipitation is chosen and converted to 3-h cumulative precipitation representing real precipitation data from January 2015 to May 2020 over China between  $15^{\circ}$ – $54.5^{\circ}$  N and  $63^{\circ}$ – $122.5^{\circ}$  E. Additionally, the geopotential, specific

humidity, vertical velocity, and 3-h cumulative precipitation from YHGSM re-forecasts are employed as correction factors.

The dimension of the dataset is  $80 \times 120$ , where each pixel corresponds to the 3-h cumulative precipitation, recorded eight times per day from 00:00 to 24:00. Since we are only interested in land precipitation in this work, we set the precipitation over the ocean to zero. As a result, the lowest errors are found when combining 1000, 800, and 400 hPa levels geopotential height to derive the total precipitation (P. Larraondo et al., 2019). Hence, we exploit 3-h cumulative precipitation and these three levels of geopotential, humidity, and vertical velocity, constructing a 10-field precipitation feature vector as the input data.

### 3 | METHODOLOGY

#### 3.1 | U-Net-based models

All models presented in this paper (Figure 1) are built upon the U-Net architecture proposed by Ronneberger et al. (2015). U-Net is a U-shaped architecture comprising an encoder and a decoder. The former applies a double convolution and a max-pooling operation for feature extraction, while the decoder comprises the same layers and is used for up-sampling (Dupuy et al., 2020; Trebing

et al., 2021). U-Net achieves a pixel-to-pixel mapping process applied between the input and output image. For a detailed analysis of the U-Net, the reader is referred to the study by Ronneberger et al. (2015).

Attention U-Net (Att-UNet) is a U-Net model using attention modules. Attention is a mechanism that suppresses feature responses within an irrelevant background and directs the network to enhance its attention to task-related important features (Bello et al., 2019; Oktay et al., 2018; Schlemper et al., 2018; Trebing et al., 2021). Several attention mechanisms (Ahmed et al., 2017; Luong et al., 2015; Oktay et al., 2018; S. Woo et al., 2018; K. Xu et al., 2015) have been proposed to achieve appealing performance improvements. In this paper, we build the attention module utilizing the attention gates (AG) proposed by Oktay et al. (2018) (Figure 2a). In attention module, the up-sampled image and the original image information passing through the skip connection are convolved and added and then flow into the activation function, the convolutional layer, and the sigmoid activation function until attention coefficients are obtained. Attention coefficient is a spatial field and has the same dimension as the feature map. The attention coefficient for each grid point is between 0 and 1. The output of the attention module is the product of attention coefficients and eigenvalues at the same position, that is, an element-wise multiplication between attention coefficients and feature maps from skip

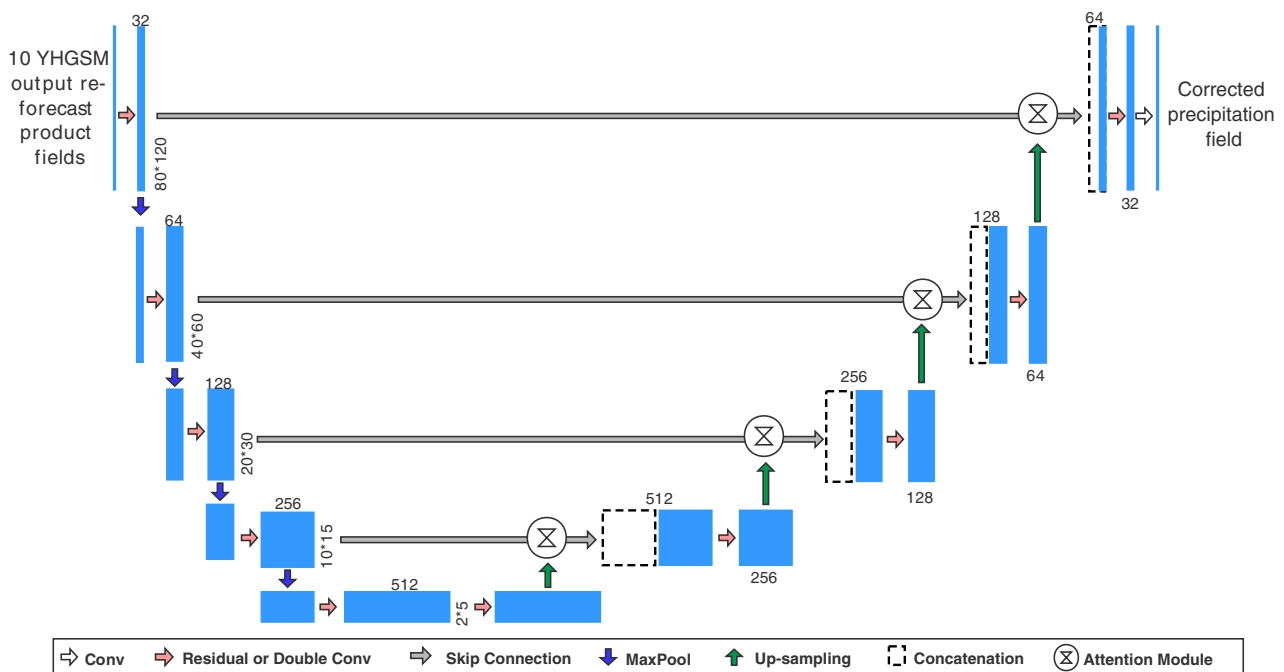


FIGURE 1 The architecture of U-Net based model. When the red arrow indicates double convolution, the original U-Net and Att-UNet are indicated, respectively, according to whether the attention module is used or not. When the red arrow indicates the residual, it indicates Res-UNet, and if using the attention module then it is RA-UNet. YHGSM, Yin-He global spectral model

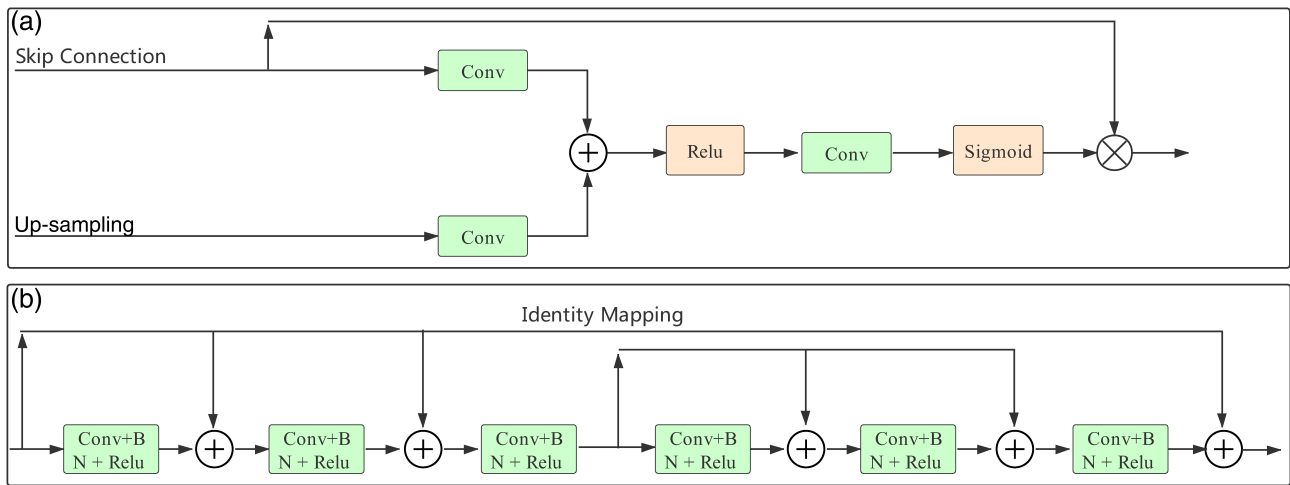


FIGURE 2 Overview of the attention module (a) and the residual block (b)

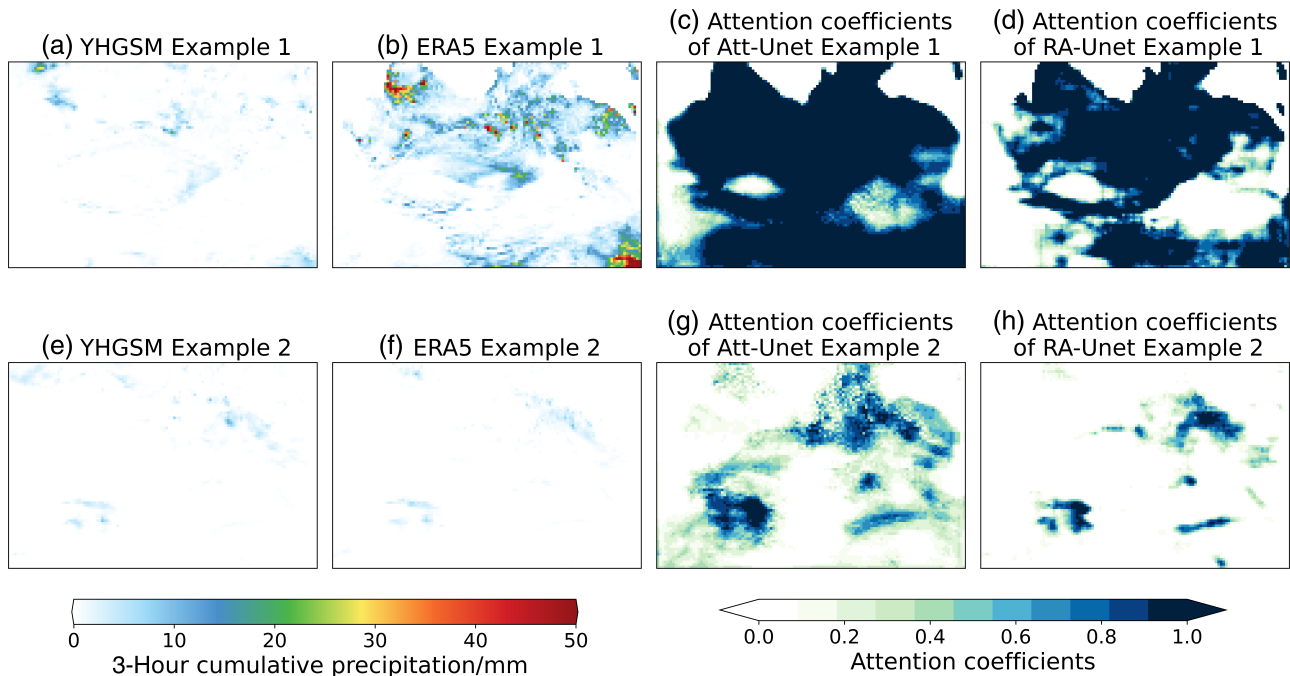


FIGURE 3 Two examples of attention coefficients. From left to right (a-d, e-h): YHGSM predicted precipitation map, ERA5 precipitation map, attention coefficients of Att-UNet, attention coefficients of RA-UNet. YHGSM, Yin-He global spectral model

connection. As shown in Figure 3, the attention mechanism can locate the precipitation areas, and the attention coefficient value of these areas is large, while the attention coefficient of other areas is close to 0. In other words, after multiplying attention coefficients and feature maps, the precipitation area will become more prominent, and the area without precipitation will be suppressed.

Theoretically, the deeper the network, the better the performance. However, deep networks suffer from gradient disappearance in practice, with residuals being a

standard solution to this problem (He et al., 2016a). Therefore, residual learning has also been proposed to learn the residual of the identity mapping. The residual block we use in this work is recurrent residual block (Alom et al., 2019), the structure is illustrated in Figure 2b, where the final feature is the identity mapping added with the convolutional blocks (Alom et al., 2019; He et al., 2016b; Jin et al., 2020). Alternatively, a residual U-Net (Res-UNet) is a U-Net model that uses residual blocks instead of convolution layers.

Residual block can be used with the attention mechanism to enhance performance (He et al., 2016a, 2016b; F. Wang et al., 2017; Jin et al., 2020); Residual attention U-Net (RA-UNet) is the model that uses residual blocks and attention modules at the same time.

or not. And the commonly used evaluation index, TS, is defined as follows:

$$TS = \frac{\text{Hits}}{\text{Hits} + \text{False alarms} + \text{Misses}},$$

### 3.2 | Model evaluation

In precipitation forecasts, a threshold value  $\alpha$  is used to distinguish precipitation magnitudes and whether it rains

TABLE 1 Different combination of a and b for loss function

|   |     |     |     |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| a | 0.5 | 0.5 | 0.5 | 1.0 | 1.0 | 1.0 | 1.0 | 2.0 | 2.0 | 2.0 |
| b | 0.5 | 1.0 | 2.0 | 0.0 | 0.5 | 1.0 | 2.0 | 0.5 | 1.0 | 2.0 |

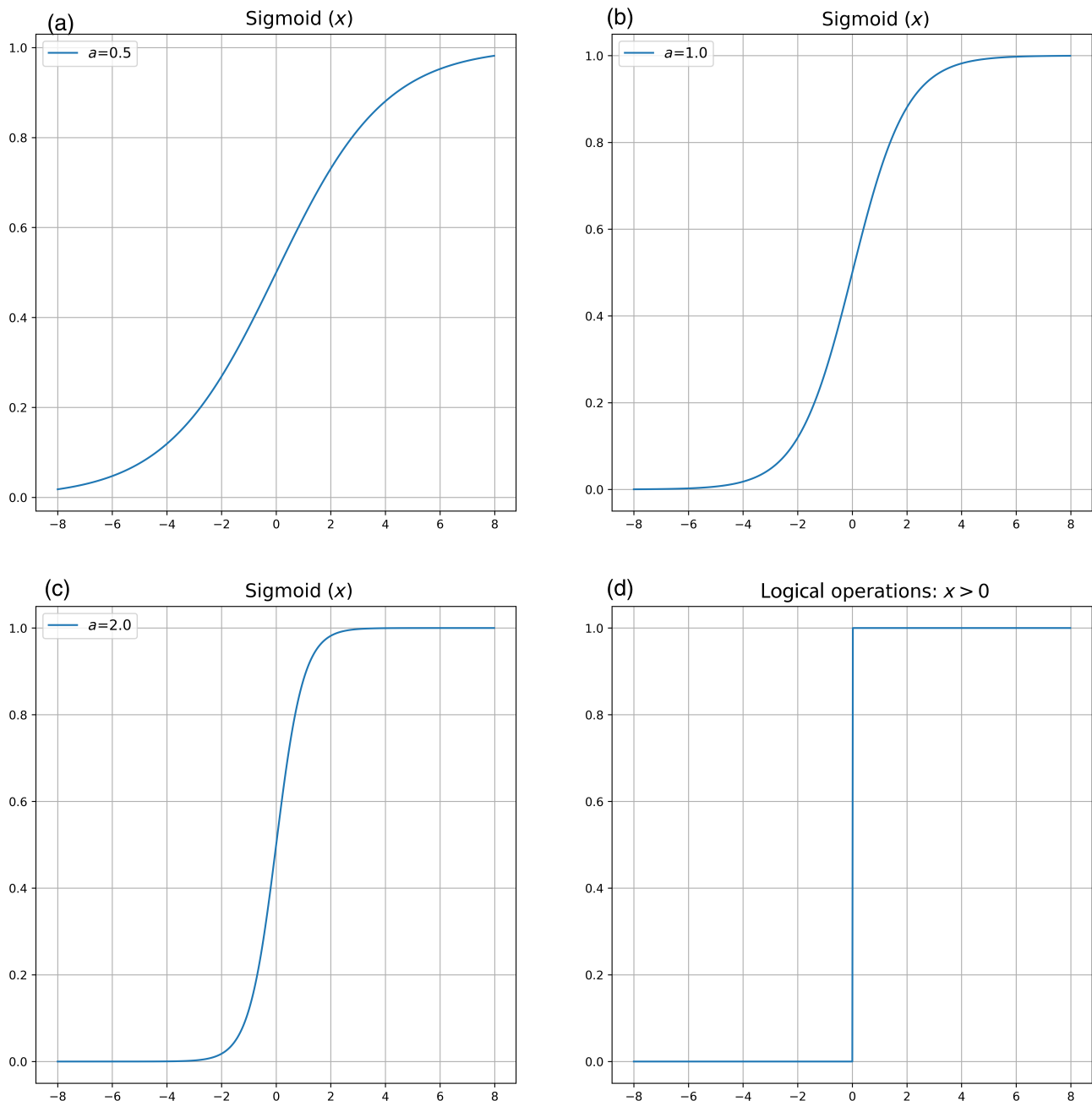


FIGURE 4 The influence of different coefficients, a, on sigmoid function. (a) a = 0.5, (b) a = 1.0, (c) a = 2.0, (d) logical operation

TABLE 2 Precipitation classification (unit mm)

| Precipitation level | No precipitation | Light precipitation | Moderate precipitation | Heavy precipitation | Rainstorm |
|---------------------|------------------|---------------------|------------------------|---------------------|-----------|
| 3 h rainfall        | <0.01            | 0.01–2.9            | 3.0–9.9                | 10.0–19.9           | ≥20.0     |

|           | Model | U-Net            |     | Att-UNet |     | Res-UNet |     | RA-UNet |     |     |     |
|-----------|-------|------------------|-----|----------|-----|----------|-----|---------|-----|-----|-----|
|           |       | Coefficients (h) |     | a        | b   | a        | b   | a       | b   | a   | b   |
| Lead time | 0–24  | 1.0              | 2.0 | 1.0      | 2.0 | 2.0      | 2.0 | 2.0     | 2.0 | 2.0 | 2.0 |
|           | 24–48 | 2.0              | 1.0 | 2.0      | 2.0 | 2.0      | 1.0 | 2.0     | 2.0 | 2.0 | 2.0 |
|           | 48–72 | 2.0              | 2.0 | 2.0      | 2.0 | 2.0      | 2.0 | 2.0     | 2.0 | 2.0 | 2.0 |

TABLE 3 The best combination of loss function coefficients for each model

where

$$\text{Hits} = (\text{observed} > \alpha) \odot (\text{predicted} > \alpha),$$

$$\text{Misses} = (\text{observed} > \alpha) \odot (\text{predicted} < \alpha),$$

$$\text{False Alarms} = (\text{observed} < \alpha) \odot (\text{predicted} > \alpha).$$

for which  $\odot$  means element-wise multiplication (P. R. Larraondo et al., 2020). For logical operations like  $(\text{observed} < \alpha)$  or  $(\text{observed} > \alpha)$ , it is 1 if the statement is true, otherwise it is 0. For each grid point, these indicators can take values of 0 and 1 only, thus also known as categorical binary metrics. These metrics are non-continuous, non-differentiable, and unsuitable for optimizing deep learning models. P. R. Larraondo et al. (2020) proposed an alternative formulation for these categorical binary indices, which are defined as follows:

$$\text{Hits}_{\text{diff}} = (\text{observed} > \alpha) \odot \text{sigmoid}(\text{predicted} - \alpha),$$

$$\text{Misses}_{\text{diff}} = (\text{observed} > \alpha) \odot \text{sigmoid}(-\text{predicted} - \alpha),$$

$$\text{False Alarms}_{\text{diff}} = (\text{observed} < \alpha) \odot \text{sigmoid}(\text{predicted} - \alpha).$$

where  $\text{sigmoid}(x) = \frac{1}{1+e^{-ax}}$ . P. R. Larraondo et al. (2020) use sigmoid function to replace one of the logical operations. Sigmoid is a smooth and differentiable function; the use of sigmoid function represents a smooth transition between the Boolean values at the threshold point (P. R. Larraondo et al., 2020). The value of the new indicators ranges from 0 to 1; they can be used in deep learning models. So, the alternative formulation of TS can be calculated as follows:

$$TS_{\text{diff}} = \frac{\text{Hits}_{\text{diff}}}{\text{Hits}_{\text{diff}} + \text{False Alarms}_{\text{diff}} + \text{Misses}_{\text{diff}}}.$$

The absolute error is also an important indicator to measure the quality of the precipitation forecast. Since YHGSM already has better performance for small magnitude precipitation forecasts but performed poorly for large magnitude precipitation forecasts, the MAE was used in the loss function to improve this deficiency.

The deep learning process needs to reduce the loss function's value. So the loss function used in this paper is a combination of MAE and  $1 - TS_{\text{diff}}$ , that is  $\text{Loss function} = b \cdot (1 - TS_{\text{diff}}) + \text{MAE}$ , where  $\text{MAE} = \max |y^k - p^k|$ , with  $y$  the value of real precipitation, and  $p$  the prediction value.

We use receiver operating characteristics (ROC) and TS (W. Woo & Wong, 2017; Zhang et al., 2019) to decide the best combination of loss function coefficients. As for the evaluation of training, RMSE, false alarm rate (FAR), probability of detection (POD), and bias score (BIAS)

TABLE 4 Number of parameters and RMSE for different bias correction models

| Model    | Parameters | RMSE     |          |          |
|----------|------------|----------|----------|----------|
|          |            | 0–24 h   | 24–48 h  | 48–72 h  |
| YHGSM    | –          | 2.187919 | 2.282786 | 2.308275 |
| U-Net    | 7,862,409  | 1.641556 | 1.189868 | 1.528800 |
| Att-UNet | 9,038,173  | 1.578880 | 1.188916 | 1.535990 |
| Res-UNet | 24,052,809 | 1.523591 | 1.130530 | 1.460506 |
| RA-UNet  | 25,209,245 | 1.506153 | 1.107084 | 1.442939 |

Abbreviations: RMSE, root mean squared error; YHGSM, Yin-He global spectral model.

were used. The RMSE instead of MAE was used because RMSE is more popular for precipitation forecast evaluation and is better for the evaluation of comprehensive performance. For which

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=0}^n (\text{Truth}^k - \text{Output}^k)^2},$$

where  $n$  is the number of samples, Truth is the value of the real precipitation and Output is the value of the prediction. For detailed information on these metrics, please refer to the Supporting Information and W. Woo & Wong (2017).

### 3.3 | Experiments

The dataset contains 15,824 samples from January 2015 to May 2020. We split the dataset randomly into a training set with 11,680 samples, a validation set with 1216 (5 months) samples, and a testing set with 2920 (1 year) samples. YHGSM re-forecasts were used as input data, and the ERA5 were used as labels to train our models. We exploit 3-h cumulative precipitation, 1000, 800, and 400 hPa levels of geopotential, humidity, and vertical velocity, constructing a 10-field feature vector as input data. The output is a  $80 \times 120$  pixels revised precipitation map with values of precipitation intensity.

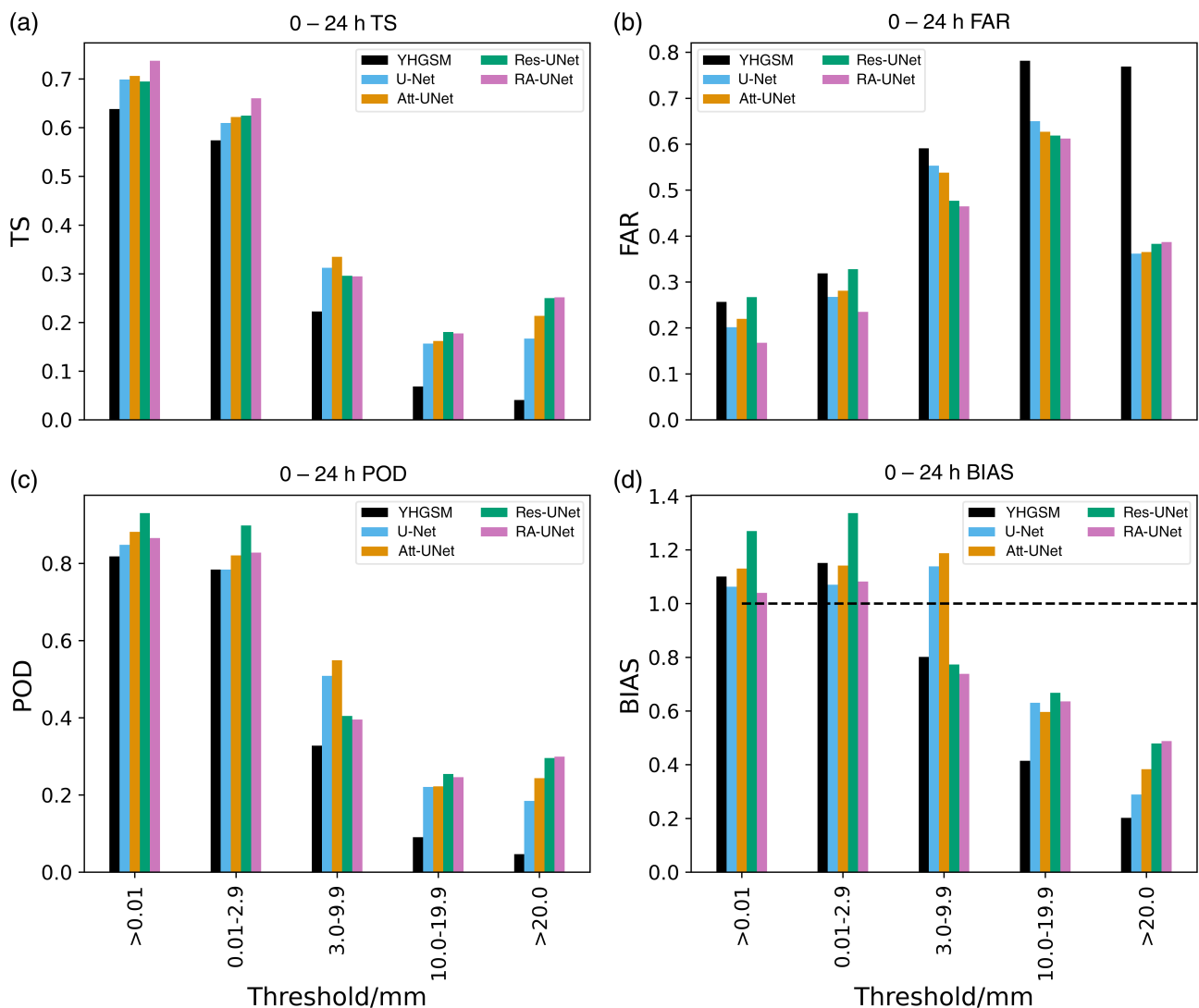


FIGURE 5 Comparison of evaluation scores before and after correction of 0-24 h. (a) TS, (b) FAR, (c) POD, (d) BIAS. BIAS, bias score; FAR, false alarm rate; POD, probability of detection; TS, threat score

For the combined loss function, different coefficients may have different results. The coefficient  $a$  mainly affects the steepness of the sigmoid function. As shown in Figure 4, the larger the coefficient  $a$ , the closer the sigmoid function is to the step function of the logic operation. The coefficient  $b$  changes the proportion of  $1 - TS_{diff}$  and MAE in the loss function. The larger the value of  $b$ , the larger the proportion of  $1 - TS_{diff}$  in the loss function. We tested the performance of 10 different coefficient combinations on the proposed four models. Different combinations of coefficients  $a$  and  $b$  are presented in Table 1. Each model is trained with these 10 combinations of loss functions.

Four models including U-Net, Att-UNet, Res-UNet, and RA-UNet are evaluated for their performance. All models are implemented utilizing Keras and TensorFlow and are trained on a NVidia Quadro RTX 8000 graphics card

(<https://www.nvidia.com/en-us/design-visualization/quadro/rtx-8000/>) with 48Gb of VRAM. The activation function of the network layer is ReLU, and the batch size is 32. All training processes use the same training set and are evaluated on the same test set. Each training session lasts 50 epochs, and the value of  $\alpha$  is a constant 0.01.

## 4 | RESULTS

This section will evaluate the performance of the four U-Net-based models in terms of YHGSM re-forecast bias correction and different lead times. First, the precipitation grades are classified into four classes according to the 3-h cumulative precipitation and distinguish the precipitation presence or absence by 0.01 mm. The specific criteria are shown in Table 2.

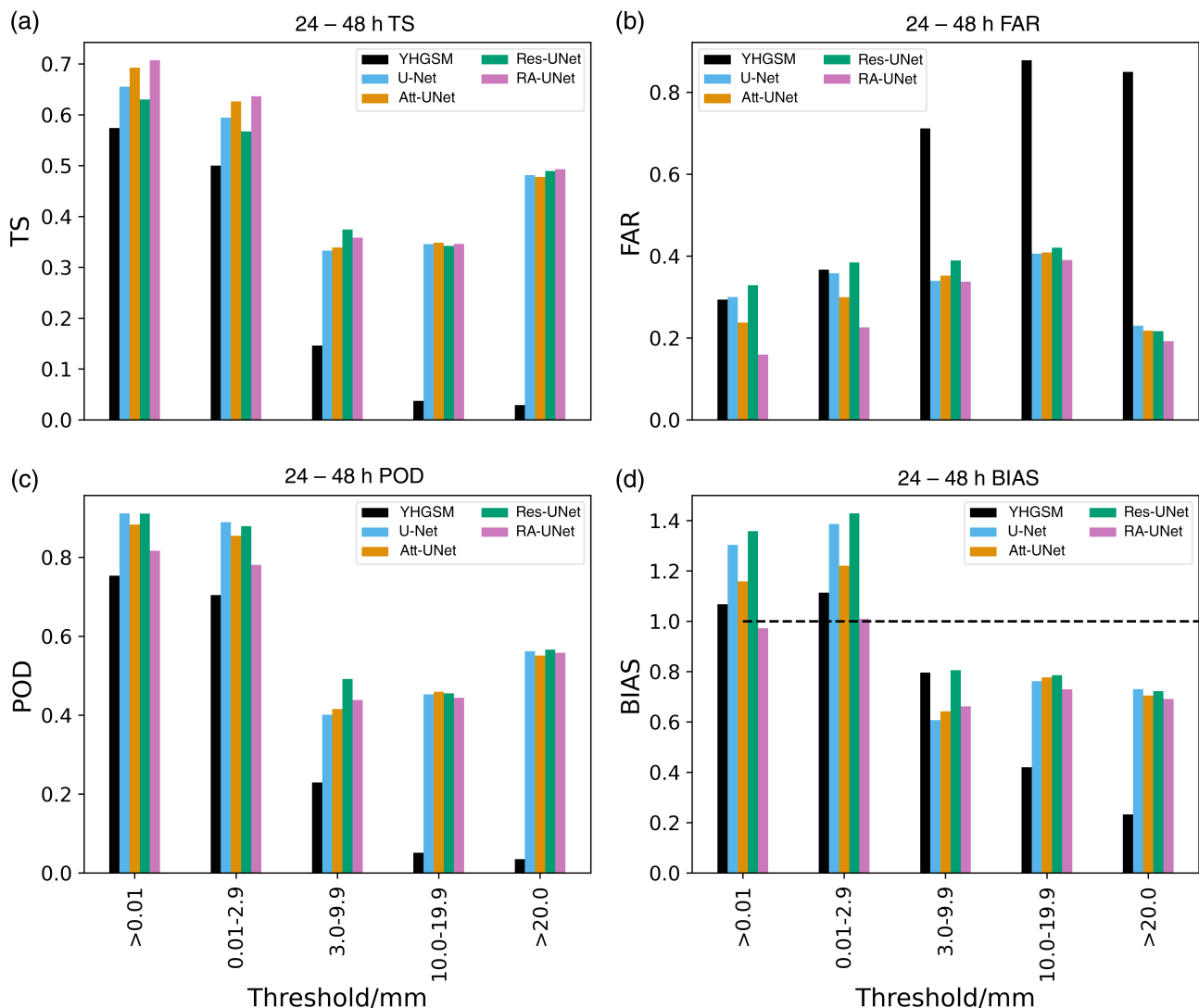


FIGURE 6 Comparison of evaluation scores before and after correction of 24–48 h. (a) TS, (b) FAR, (c) POD, (d) BIAS. BIAS, bias score; FAR, false alarm rate; POD, probability of detection; TS, threat score



To achieve the optimum performance, for each model, the best loss function combination is chosen (Table 3), with the corresponding details listed in the Supporting Information. The optimal combination of loss function coefficients is selected by TS and AUC values, which not only requires the possible maximum AUC value but also requires TS to be more than 6% higher than YHGSM forecasts. Table 4 presents the number of parameters and RMSE for various models under several lead times (0–24, 24–48, and 48–72 h). Our results indicate that the proposed models can improve RMSE. Specifically, among the models evaluated, the RA-UNet has the smallest RMSE affording an RMSE reduction of 31.2%, 51.5%, and 37.5% for lead times of 0–24, 24–48, and 48–72 h, respectively. Res-UNet reduces RMSE by 30.4%, 50.5%, and 36.7% for lead times of 0–24, 24–48, and 48–72 h, respectively. U-Net and Att-UNet also improved

RMSE. Comparing these four models shows that the residual block greater enhances performance than the attention module. The use of the attention module filters the noisy and irrelevant responses through the activation function of neurons (Oktay et al., 2018; Schlemper et al., 2018), that is, background regions without precipitation are downgraded during the backward pass, and the model parameters are updated mainly based on spatial regions with precipitation. But when the feature signals are processed, the ReLU activation function causes the high-dimensional raw signals to collapse into a low-dimensional feature set and potentially impose some features to disappear completely. The residual block increases the feature signal's dimension, and thus many features of the original data are retained. It can be seen from Figure 3 that Res-UNet is better than Att-UNet model in locating the

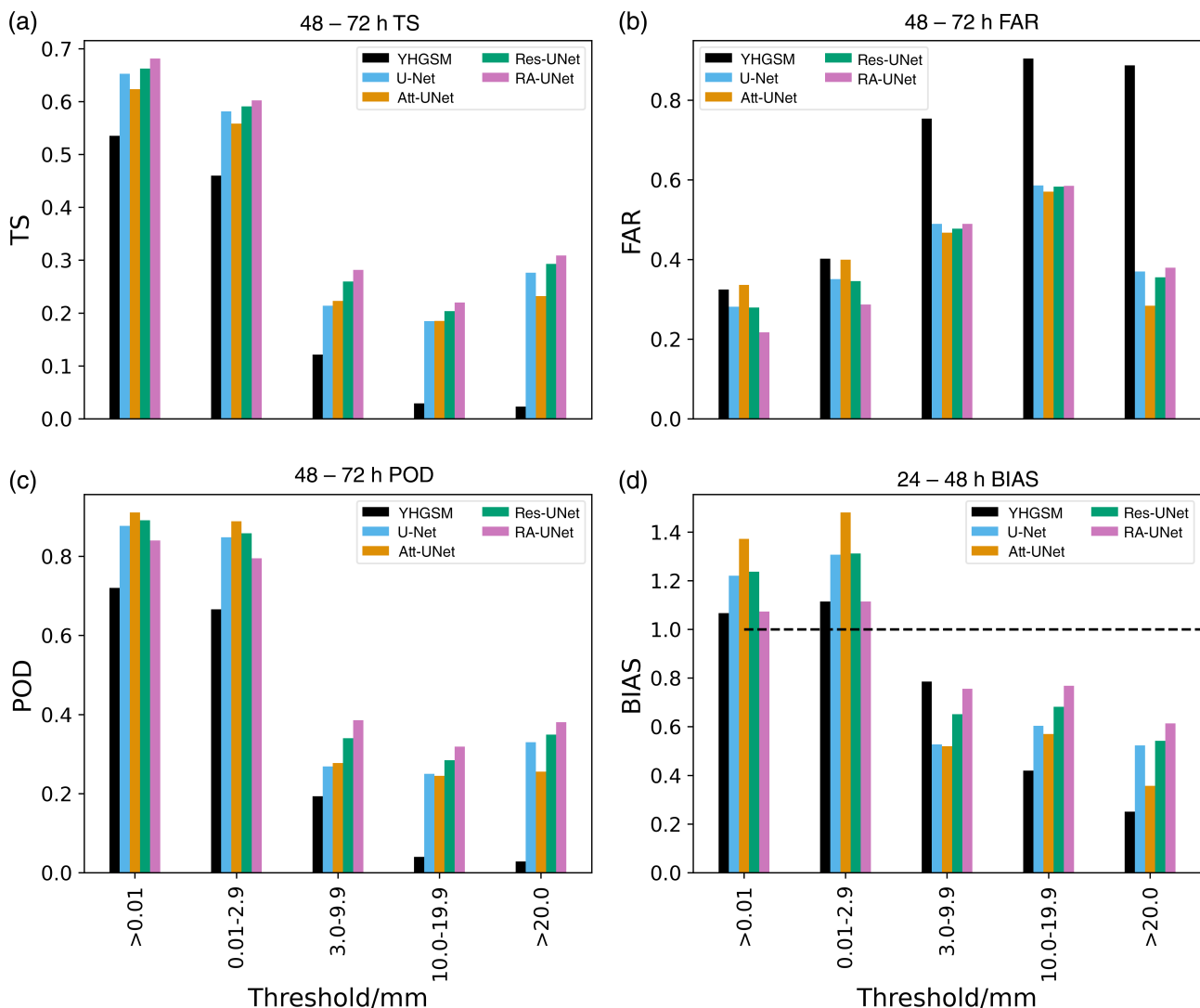
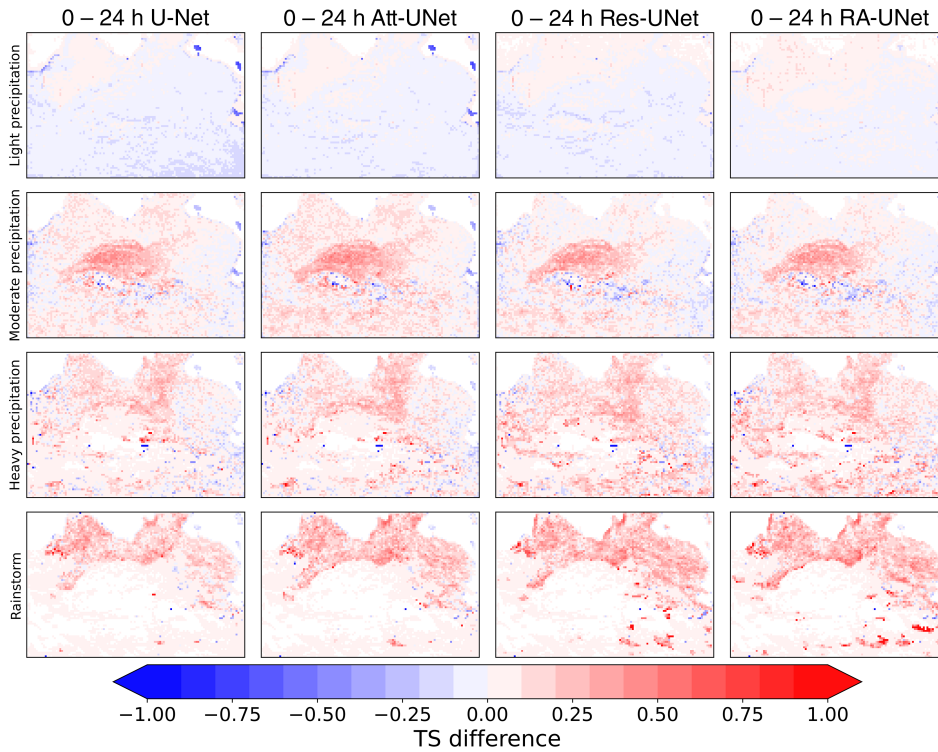


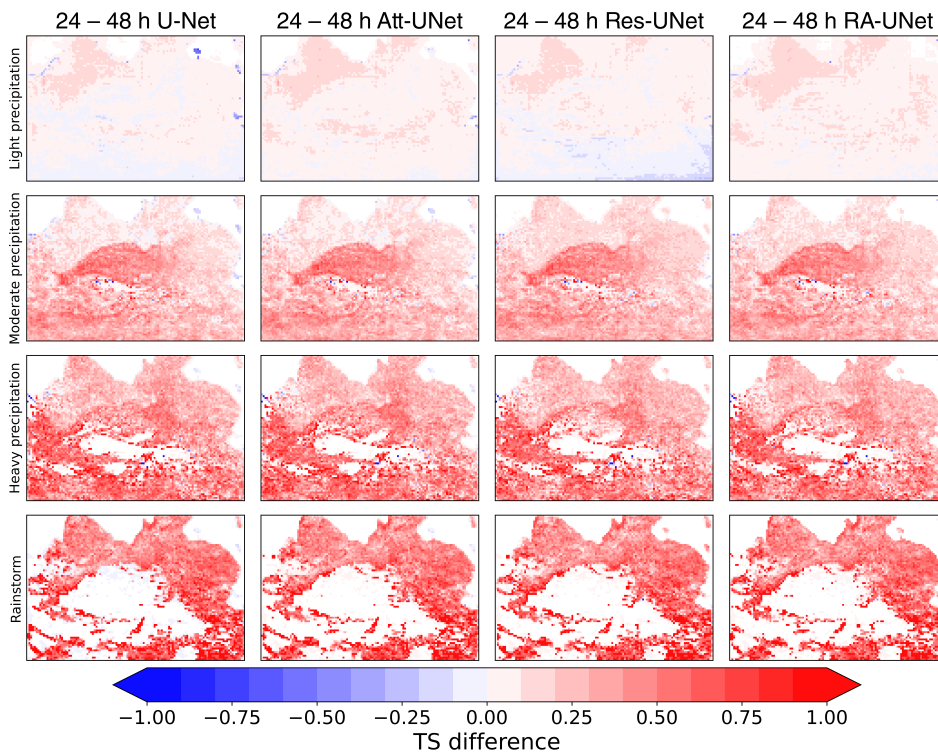
FIGURE 7 Comparison of evaluation scores before and after correction of 48–72 h. (a) TS, (b) FAR, (c) POD, (d) BIAS. BIAS, bias score; FAR, false alarm rate; POD, probability of detection; TS, threat score

precipitation areas, and the residual block can improve the accuracy of feature extraction. Additionally, introducing residuals and attention mechanisms increases the parameter cardinality by more than 100% and 15%, respectively.

Figures 5–7 visualize the performance of the proposed bias correction models for different lead times. As can be seen from Figure 5, U-Net-based models have made significant improvements in TS, FAR, and POD, especially in the case of heavy precipitation and rainstorm.



**FIGURE 8** Difference of TS between four models and YHGSM for lead time of 0–24 h. From left to right: The results of U-Net, Att-UNet, Res-UNet, and RA-UNet, respectively. From top to bottom: The results of light precipitation, moderate precipitation, heavy precipitation, and rainstorm, respectively. TS, threat score; YHGSM, Yin-He global spectral model



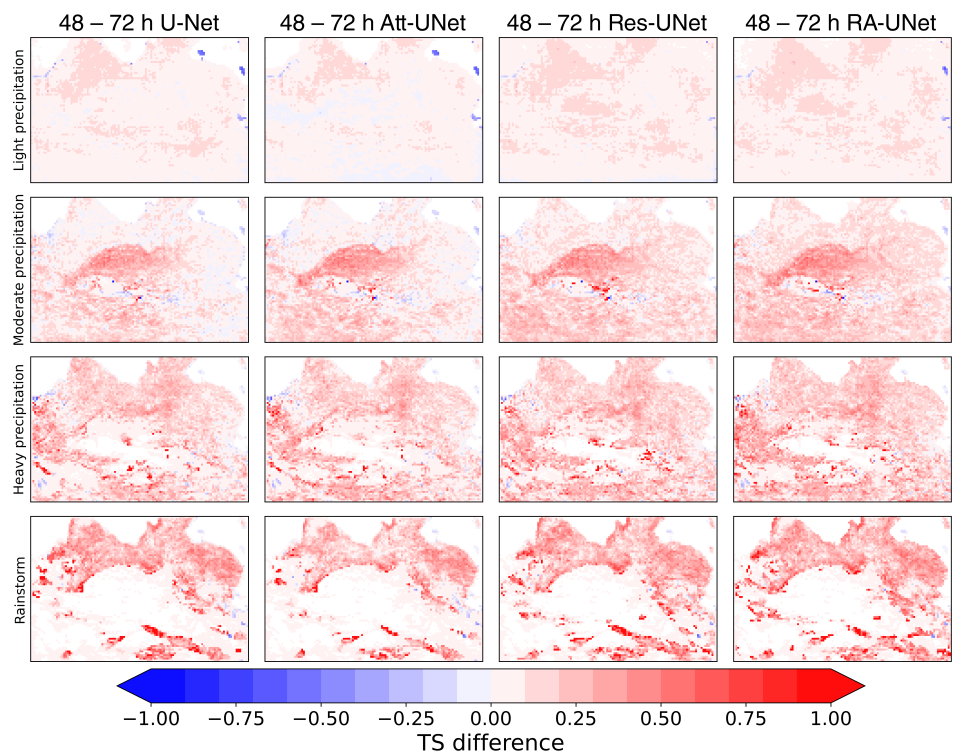
**FIGURE 9** Difference of TS between four models and YHGSM for lead time of 24–48 h. From left to right: The results of U-Net, Att-UNet, Res-UNet, and RA-UNet, respectively. From top to bottom: The results of light precipitation, moderate precipitation, heavy precipitation, and rainstorm, respectively. TS, threat score; YHGSM, Yin-He global spectral model

Compared with YHGSM, Res-UNet and RA-UNet gain an increase of TS by 1.6 times for heavy precipitation. The TS of YHGSM for rainstorms is only 0.04, while RA-UNet increases to 0.25 (6.2 times higher). The results of Figure 5d (BIAS = 1 marked by black dashed line) indicate that the YHGSM deviation increases with the precipitation magnitude. There is a noticeable improvement in BIAS after utilizing RA-UNet except for the moderate precipitation. Similar conclusions can be made from Figure 6 and Figure 7 as Figure 5 for the 24–48 and 48–72 h lead time, respectively. It is important to note that all models manage a TS metric increase of more than five times for heavy precipitation and more than nine times for rainstorms.

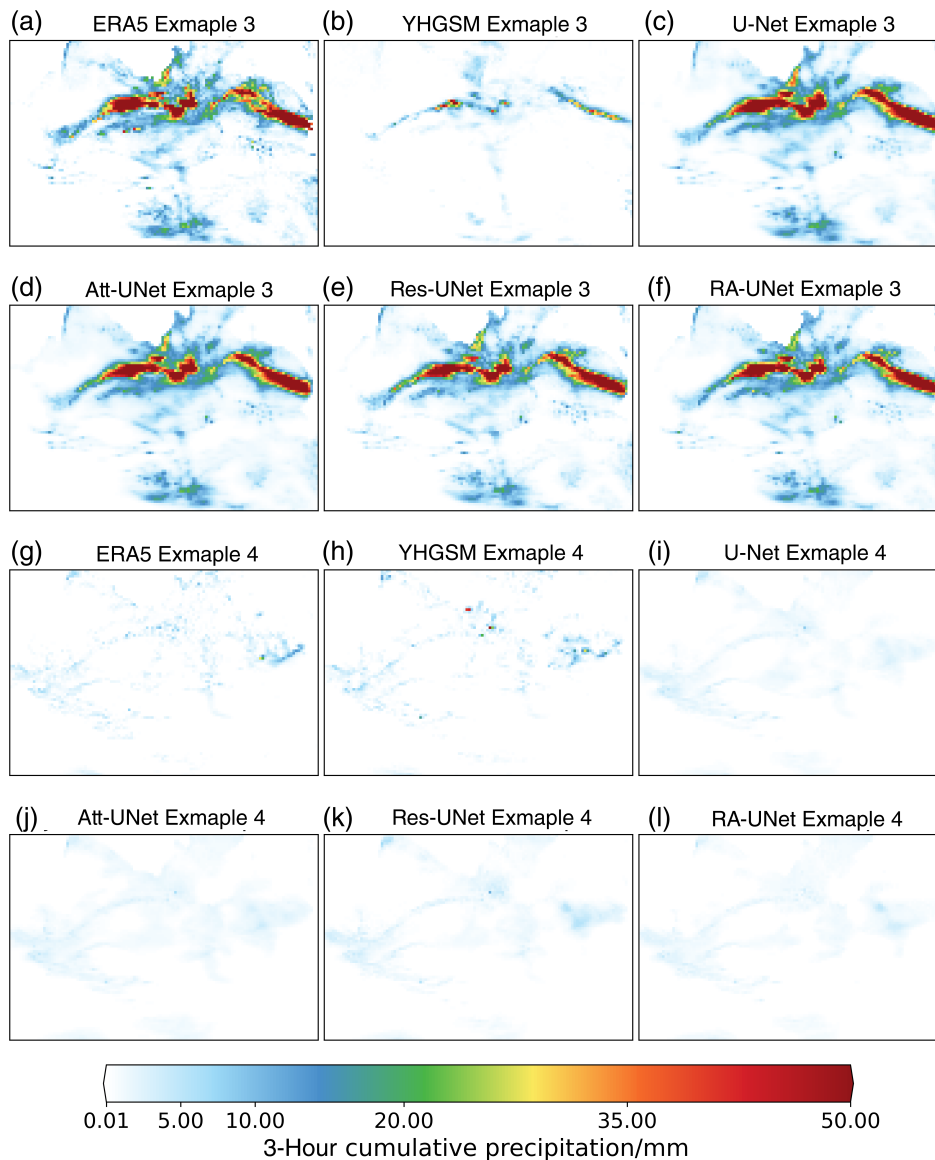
Figures 8–10 show the spatial improvement of TS compared with YHGSM for lead time 0–24, 24–48, and 48–72 h, respectively. The four columns from left to right represent U-Net, Att-UNet, Res-UNet, and RA-UNet, respectively. The rows from top to bottom indicate light precipitation, moderate precipitation, heavy precipitation, and rainstorm. For most areas, our model has significant effects, except for light precipitation with a lead time of 0–24 h. The MAE and area-averaged TS were used in the loss function. And YHGSM has a small MAE in forecasting light precipitation for 0–24 h lead time, so there may be an increase in the average TS, but a decrease in TS in some areas. For moderate to heavy precipitation, all four models show

better correction capability on lead time 24–48 h than the other two lead times. Furthermore, it can be found that RA-UNet has a better generalization ability than the other three models, and the use of an attention module on the residual block can further improve the performance.

Figure 11 provides some examples from the test data, where we sequentially represent the original ERA5 total precipitation field, the YHGSM output, and the output of the four different U-Net-based models. For most cases, and especially for precipitation with a strong correlation between regions like heavy precipitation and rainstorms, the predicted precipitation maps after correction are more closely related to the ERA5 than the YHGSM. For these precipitation cases, the U-Net-based models enhance the distribution and precipitation intensity accuracy with a significant reduction in FAR and an increase in POD. However, for the small magnitude of precipitation and scattered precipitation, the effect of the calculated bias correction is insufficient, as convolution, pooling, and up-sampling impose small-scale detail loss while extracting the main features. Additionally, from our results, we conclude that U-Net-based networks consider the surrounding area and make the prediction smooth. Another reason for blurry scattered precipitation is that by using a deterministic loss function, the model aims to keep the error low with a value that is closest to all possible outcomes of the prediction sequence so that



**FIGURE 10** Difference of TS between four models and YHGSM for lead time of 48–72 h. From left to right: The results of U-Net, Att-UNet, Res-UNet, and RA-UNet, respectively. From top to bottom: The results of light precipitation, moderate precipitation, heavy precipitation, and rainstorm, respectively. TS, threat score; YHGSM, Yin–He global spectral model



**FIGURE 11** Examples of precipitation forecasts by different models

the pixel-level frame prediction is reduced (Babaeizadeh et al., 2017; Denton & Fergus, 2018; Trebing et al., 2021).

## 5 | CONCLUSION

In this paper, four U-Net-based models are applied to correct the 3-h cumulative precipitation of the YHGSM re-forecasts and evaluated for their performance. Employing a strategy that utilizes a loss function with the best weight combination of  $TS_{diff}$  and MAE, we demonstrate that the TS has been improved with the reduction of FAR and the increase of POD, especially for heavy precipitation and rainstorms. What is more, RA-UNet has a better generalization ability than the other three models, and it can reduce the RMSE of the 3-h cumulative precipitation

by 31.2%, 51.5%, and 37.5% for lead times of 0–24, 24–48, and 48–72 h, respectively.

From our trials, we also conclude that the U-Net-based networks may lose some small-scale details while extracting the main image features. On the other hand, residual blocks improve the feature disappearance problem. Accordingly, an attention module imposes the network to pay more attention to areas with precipitation, which increases the accuracy of a certain magnitude but may negatively affect the remaining precipitation magnitude weights.

For future work, we will investigate how to use actual observational data instead of the re-analysis data ERA5, train models for different precipitation magnitudes, and optimal combination of loss function coefficients for better performance.

## CONFLICT OF INTEREST

No conflict of interest exists in the submission of this manuscript, and the manuscript is approved by all authors for publication.

## AUTHOR CONTRIBUTIONS

**Yi-Fan Hu:** Data curation (equal); formal analysis (equal); methodology (equal); validation (equal); visualization (equal); writing – original draft (equal). **Fu-Kang Yin:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); supervision (equal); validation (equal); writing – review and editing (equal). **Wei-Min Zhang:** Formal analysis (equal); funding acquisition (equal); project administration (equal); writing – review and editing (equal).

## DATA AVAILABILITY STATEMENT

ERA5 data sets were obtained from <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>.

## ORCID

Yi-Fan Hu  <https://orcid.org/0000-0003-2840-259X>

Fu-Kang Yin  <https://orcid.org/0000-0002-8353-7000>

## REFERENCES

- Ahmed, K., Keskar, N.S. & Socher, R. (2017) Weighted transformer network for machine translation. Available at: <https://arxiv.org/abs/1711.02132> [Accessed 22nd June 2021].
- Alom, M.Z., Yakopcic, C., Hasan, M., Taha, T.M. & Asari, V.K. (2019) Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *Journal of Medical Imaging*, 6(1), 014006–014006. <https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-6/issue-01/014006/Recurrent-residual-U-Net-for-medical-image-segmentation/10.1117/1.JMI.6.1.014006.full?SSO=1> [Accessed 22nd June 2021].
- Babaeizadeh, M., Finn, C., Erhan, D., Campbell, R. & Levine, S. (2017) Stochastic variational video prediction. In: International conference on learning representations. Available at: <https://openreview.net/forum?id=rk49Mg-CW> [Accessed 22nd June 2021].
- Bello, I., Zoph, B., Le, Q., Vaswani, A. & Shlens, J. (2019). Attention Augmented Convolutional Networks. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3285–3294. <https://doi.org/10.1109/ICCV.2019.00338>.
- Carvalho, J., Assad, E. & Pinto, H. (2011) Kalman filter and correction of the temperatures estimated by PRECIS model. *Atmospheric Research*, 102, 218–226. <https://doi.org/10.1016/j.atmosres.2011.07.007>
- Denton, E. & Fergus, R. (2018) Stochastic video generation with a learned prior. In: International conference on machine learning (ICML'18). Available at: <http://proceedings.mlr.press/v80/denton18a.html> [Accessed 22nd June 2021].
- Dupuy, F., Mestre, O., Serrurier, M., Burdà, V., Zamo, M., Cabrera-Gutiérrez, N. et al. (2020) ARPEGE cloud cover forecast post-processing with convolutional neural network. *Weather and Forecasting*, 36(2), 567–586. <https://doi.org/10.1175/WAF-D-20-0093.1>
- Grover, A., Kapoor, A. & Horvitz, E. (2015) A deep hybrid model for weather forecasting. In: *Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining*, New York, NY: Association for Computing Machinery. pp. 379–386. <https://doi.org/10.1145/2783258.2783275>
- He, K., Zhang, X., Ren, S. & Sun, J. (2016a) Deep residual learning for image recognition. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, Los Alamitos, CA: IEEE. pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- He, K., Zhang, X., Ren, S. & Sun, J. (2016b) Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N. & Welling, M. (Eds.) *Computer vision – ECCV 2016. ECCV 2016. Lecture notes in computer science*, Vol. 9908. Cham: Springer. [https://doi.org/10.1007/978-3-319-46493-0\\_38](https://doi.org/10.1007/978-3-319-46493-0_38)
- Hemri, S., Scheuerer, M., Pappenberger, F., Bogner, K. & Haiden, T. (2014) Trends in the predictive performance of raw ensemble weather forecasts. *Geophysical Research Letters*, 41, 9197–9205. <https://doi.org/10.1002/2014GL062472>
- Hernandez, E., Sanchez-Anguix, V., Julian, V., Palanca, J. & Duque, N. (2016) Rainfall prediction: a deep learning approach. In: Martínez-Álvarez, F., Troncoso, A., Quintián, H. & Corchado, E. (Eds.) *Hybrid artificial intelligent systems. HAIS 2016. Lecture notes in computer science*, Vol. 9648. Cham: Springer. [https://doi.org/10.1007/978-3-319-32034-2\\_13](https://doi.org/10.1007/978-3-319-32034-2_13)
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz, S.J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Jin, Q., Meng, Z., Sun, C., Cui, H. & Su, R. (2020) RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans. *Frontiers in Bioengineering and Biotechnology*, 8, 605132. <https://doi.org/10.3389/fbioe.2020.605132>
- Jurlina, T., Gascón, E., Harrigan, S. & Prudhomme, C. (2020) *Precipitation bias correction improves flood forecasts*. Available at: <https://www.ecmwf.int/en/newsletter/163/news/precipitation-bias-correction-improves-flood-forecasts> [Accessed 22nd June 2021].
- Larraondo, P., Renzullo, L., Inza, I. & Lozano, J. (2019) A data-driven approach to precipitation parameterizations using convolutional encoder-decoder neural networks. Available at: <https://arxiv.org/abs/1903.10274> [Accessed 22nd June 2021].
- Larraondo, P.R., Re Nzullo, L.J., Dijk, A., Inza, I. & Lozano, J.A. (2020) Optimization of deep learning precipitation models using categorical binary metrics. *Journal of Advances in Modeling Earth Systems*, 12(5), e2019MS001909–e2019MS001909. <https://doi.org/10.1029/2019MS001909>
- Luong, M.T., Pham, H. & Manning, C.D. (2015) Effective approaches to attention-based neural machine translation. Available at: <https://arxiv.org/pdf/1508.04025.pdf> [Accessed 22nd June 2021].
- Oktay, O., Schlemper, J., Folgoc, L., Lee, M., Heinrich, M., Misawa, K. et al. (2018) *Attention U-Net: learning where to look for the pancreas*. Available at: <https://arxiv.org/abs/1804.03999v1> [Accessed 22nd June 2021].
- Peng, J., Wu, J., Zhang, W., Zhao, J., Zhang, L. & Yang, J. (2019) A modified nonhydrostatic moist global spectral dynamical core

- using a dry-mass vertical coordinate. *Quarterly Journal of the Royal Meteorological Society*, 145, 2477–2490. <https://doi.org/10.1002/qj.3574>
- Peng, J., Zhao, J., Zhang, W., Zhang, L., Wu, J. & Yang, X. (2020) Towards a dry-mass conserving hydrostatic global spectral dynamical core in a general moist atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 146(732), 3206–3224. <https://doi.org/10.1002/qj.3842>
- Rasp, S. & Lerch, S. (2018) Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, 146(11), 3885–3900. <https://doi.org/10.1175/MWR-D-18-0187.1>
- Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: convolutional networks for biomedical image segmentation. *LNCS*, 9351, 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Schlemper, J., Oktay, O., Chen, L., Matthew, J., Knight, C., Kainz, B. et al. (2018) *Attention-gated networks for improving ultrasound scan plane detection*. Available at: <https://arxiv.org/abs/1804.05338> [Accessed 22nd June 2021].
- Shu, X.U., Wei, Y.H., Xiong, M.M., Wei, L., Observatory, T.M., Center, T.C. et al. (2018) Application of frequency-matching method to ECMWF ensemble statistic fusing prediction products in the Haihe River basin. *Journal of Meteorology and Environment*, 34(4), 11–17.
- Singh, M., Kumar, B., Niyogi, D., Rao, S., Gill, S.S. & Chattopadhyay, R. et al. (2021) Deep learning for improved global precipitation in numerical weather prediction systems. Available at: <https://arxiv.org/abs/2106.12045> [Accessed 22nd June 2021].
- Trebing, K., Stańczyk, T. & Mehrkanon, S. (2021) SmaAt-UNet: precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognition Letters*, 145, 178–186. <https://doi.org/10.1016/j.patrec.2021.01.036>
- Wang, J., Chen, C., Long, K., Feng, L.J.P. & Research, M.M. (2015) Temporal and spatial distribution of short-time heavy rain of Sichuan Basin in summer. *Plateau and Mountain Meteorology Research*, 35, 16–20.
- Wang, F., Jiang, M., Qian, C., Yang, S., Li, C. & Zhang, H. (2017) Residual Attention Network for Image Classification. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6450–6458. Available at: <https://ieeexplore.ieee.org/document/8100166> [Accessed 22nd June 2021].
- Weyn, J.A., Durran, D.R. & Caruana, R. (2020) Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9), e2020MS002109.
- Wilks, D. (2009) Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications*, 16, 361–368. <https://doi.org/10.1002/met.134>
- Witt, C., Tong, C., Zantedeschi, V., De Martini, D., Kalaitzis, F. & Chantry, M. et al. (2020) RainBench: towards global precipitation forecasting from satellite imagery. Available at: <https://arxiv.org/abs/2012.09670> [Accessed xxxx].
- Woo, W. & Wong, W. (2017) Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere*, 8(3), 48. <https://doi.org/10.3390/atmos8030048>
- Woo, S., Park, J., Lee, J.Y. & Kweon, I.S. (2018) CBAM: convolutional block attention module. In: *European conference on computer vision (ECCV)*. Available at: <https://arxiv.org/abs/1807.06521> [Accessed 22nd June 2021].
- Wu, J. P., Zhao, J., Song, J.Q. & Zhang, W.M. (2011) Preliminary design of dynamic framework for global non-hydrostatic spectral model. *Computer Engineering and Design*, 32(10), 3539–3543.
- Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R. et al. (2015) Show, attend and tell: neural image caption generation with visual attention. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, 37, 2048–2057 Available at: <https://arxiv.org/pdf/1502.03044.pdf> [Accessed 22nd June 2021]
- Xu, X., Liu, Y., Chao, H., Luo, Y., Chu, H., Chen, L. et al. (2019) *Towards a precipitation bias corrector against noise and maldistribution*. Available at: <https://arxiv.org/abs/1910.07633> [Accessed 22nd June 2021].
- Yang, J., Song, J., Wu, J., Ren, K. & Leng, H. (2015) A high-order vertical discretization method for a semi-implicit mass-based non-hydrostatic kernel. *Quarterly Journal of the Royal Meteorological Society*, 141, 2880–2885. <https://doi.org/10.1002/qj.2573>
- Yin, F., Wu, G., Wu, J., Zhao, J. & Song, J. (2018) Performance evaluation of the fast spherical harmonic transform algorithm in the Yin–He global spectral model. *Monthly Weather Review*, 146, 3163–3182. <https://doi.org/10.1175/MWR-D-18-0151.1>
- Yin, F., Song, J., Wu, J. & Zhang, W. (2021) An implementation of single-precision fast spherical harmonic transform in Yin–He global spectral model. *Quarterly Journal of the Royal Meteorological Society*, 147, 2323–2334. <https://doi.org/10.1002/qj.4026>
- Zhang, C.-J., Zeng, J., Wang, H.Y., Ma, L.-M. & Chu, H. (2019) Correction model for rainfall forecasts using LSTM with multiple meteorological factors. *Meteorological Applications*, 27(1), e1852–e1852. <https://doi.org/10.1002/met.1852>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Hu, Y.-F., Yin, F.-K., & Zhang, W.-M. (2021). Deep learning-based precipitation bias correction approach for Yin–He global spectral model. *Meteorological Applications*, 28(5), e2032. <https://doi.org/10.1002/met.2032>