



RESEARCH ARTICLE

10.1029/2025JH000675

Key Points:

- VGG-like deep learning model improves higher +6-hr heavy TC remote precipitation forecasting accuracy by 18% than the control experiment
- Incorporating physical knowledge into the model enhances attention allocation of the SE-Block to topographic features by 21.74%
- Henan heavy rainfall (2021) shows a 0.44 TS increase in +6-hr heavy TC remote precipitation forecasting than the control experiment

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Zhang,
zhangaoq3@mail.sysu.edu.cn

Citation:

Xiao, S., Zhang, A., Gong, J., Chen, Y., Chen, S., & Li, W. (2025). An explainable deep learning method with squeeze and excitation block for 6-hour forecast of tropical cyclone remote precipitation. *Journal of Geophysical Research: Machine Learning and Computation*, 2, e2025JH000675. <https://doi.org/10.1029/2025JH000675>

Received 5 MAR 2025

Accepted 19 AUG 2025

Author Contributions:

Conceptualization: Shiqi Xiao, Aoqi Zhang

Funding acquisition: Aoqi Zhang, Yilun Chen, Shumin Chen, Weibiao Li

Investigation: Shiqi Xiao, Jing Gong

Methodology: Shiqi Xiao

Project administration: Aoqi Zhang

Software: Shiqi Xiao

Writing – original draft: Shiqi Xiao

Writing – review & editing: Aoqi Zhang, Yilun Chen, Shumin Chen, Weibiao Li

© 2025 The Author(s). *Journal of Geophysical Research: Machine Learning and Computation* published by Wiley Periodicals LLC on behalf of American Geophysical Union.

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.

An Explainable Deep Learning Method With Squeeze and Excitation Block for 6-Hour Forecast of Tropical Cyclone Remote Precipitation

Shiqi Xiao^{1,2} , Aoqi Zhang¹ , Jing Gong³, Yilun Chen¹ , Shumin Chen⁴, and Weibiao Li^{1,4}
¹School of Atmospheric Sciences and Key Laboratory of Tropical Atmosphere-Ocean System, Ministry of Education, Sun Yat-Sen University, Zhuhai, China, ²Guangzhou Institute of Tropical and Marine Meteorology, China Meteorological Administration, Guangzhou, China, ³School of Software Engineering, Ministry of Education, Sun Yat-Sen University, Zhuhai, China, ⁴Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China

Abstract This study presents a deep learning framework based on a VGG-like architecture for forecasting +6-hr tropical cyclone remote precipitation (TCRP), a major precipitation phenomenon in East Asia. A three-stage optimization strategy was implemented: (a) K-fold cross-validation was used to select 10 optimal ensemble members, which were further optimized; (b) atmospheric physics knowledge was integrated into activation function design and objective functions, resulting in a 19.23% increase in the Squeeze and Excitation Block's (SE-Block) mean attention allocation to topographic features during the study period (2021–2023); (c) multi-modal inputs combined with custom objective functions enhanced heavy TCRP prediction accuracy by 0.18 compared to a control experiment. The July 2021 Henan extreme rainfall event demonstrated that replacing traditional activation functions with physics-informed designs and using custom objective functions increased SE-Block attention allocation to topographic features by 21.74% and improved heavy TCRP forecasting accuracy by 0.44 for optimal models compared to the control experiment. Results showed superior performance in heavy TCRP prediction and enhanced attention allocation to topographic features during extreme rainfall events, exceeding average values in 2021–2023. This work provides an explainable deep learning approach for TCRP forecasting, advancing both methodology and physical interpretability, suggesting promising solutions for extreme precipitation forecast through hybrid approaches combining atmospheric physics with data-driven techniques.

Plain Language Summary This study developed a deep learning model using a VGG-like architecture for forecasting +6-hr tropical cyclone remote precipitation (TCRP), a major East Asian precipitation phenomenon. Using multi-modal inputs, integrating atmospheric physics into activation functions and objectives achieved a 0.18 increase in heavy TCRP prediction compared to a control model and enhanced model attention distribution to topographic feature by 19.23%. In the 2021 Henan extreme rainfall event, replacing traditional activation functions with physics-informed designs boosted topographic feature attention by 21.74% and improved heavy TCRP accuracy by 0.44. The study demonstrates superior performance in predicting heavy TCRP events and highlights an explainable deep learning approach that combines atmospheric physics with data-driven techniques, offering promising solutions for extreme precipitation forecast.

1. Introduction

Precipitation is one of the most important processes in atmospheric water cycle that are tightly associated with moisture transport, whose extremes result in natural hazards, especially the floods, landslides, and soil erosion (Gimeno-Sotelo & Gimeno, 2023; Konapala et al., 2020; Lu & Lall, 2017; Sellars et al., 2017). These events led to significant losses: over 50% (~1.09 Million) deaths and ~\$3.4 Trillion in economic loss globally (Douris et al., 2022; L. Zhang et al., 2023), especially in developing regions. Extreme precipitation has been increasing over the past six decades in dry and wet regions, resulting in catastrophic natural hazards (Barlow et al., 2019; Donat et al., 2016). Previous studies show that precipitation extremes increase with stronger moisture transport, which increases by 6%–7% K^{−1} with surface temperature, and 8% K^{−1} for integral over the whole atmosphere under the global warming effect, according to Clausius-Clapeyron relation (Houze, 2014a; Koutsoyiannis, 2012). Unlike moderate increase in long-term precipitation, there is sharp increase in extreme precipitation, likely due to stronger convective activities (Houze, 2014b; Luo et al., 2023). The essential role of precipitation in the present and the future, requires the intelligent prediction methods to improve forecast performance.

Precipitation is tightly associated with diverse weather systems and events, which is the result of microphysical processes (Houze, 2014a; A. Zhang et al., 2022). To investigate the association between intense moisture transport and extreme precipitation, identification methods of atmospheric rivers (ARs), mesoscale convective systems, and tropical cyclones (TCs) are developed (Cheng et al., 2022; M. Pan & Lu, 2020; L. Zhang et al., 2023). Apart from precipitation associated with weather events above, precipitation associated with TC-induced remote moisture transport attracts increasing attention in recent decades, as Beijing heavy rainfall in 2012, Yangtze Delta heavy rainfall in 2018 and Henan extreme rainfall in 2021 have caused floods and economic loss (Chen et al., 2010; Rao et al., 2025). This type of precipitation falls outside TC circulations (roughly 700 km depending on the size of TCs), with TCs the source of moisture transport (Chen et al., 2010; Sun et al., 2017; Xu et al., 2022). Following the definitions above, the first objective identification algorithm of TC induced remote moisture transport was developed to quantify the moisture transport characteristics of TC induced remote moisture transport clusters (TRCs) (Xiao, Zhang, Chen, & Li, 2024), which is further used to find TRC tracks that are classified into five types featuring orientations and find the significant Gaussian relation between precipitation and moisture transport height, and sigmoid relation between precipitation and moisture transport intensity for each type (Xiao, Zhang, Chen, & Li, 2024).

Numerical weather prediction methods, mainly consisting of atmospheric governing equations, are widely used in weather agencies in many countries (Bauer et al., 2015). The numerical prediction framework integrates data assimilation, the specification of initial and boundary conditions, and the computational solution of governing equations—primarily the Navier-Stokes equations (Ershkov et al., 2021; Larios & Victor, 2024; Zhou et al., 2024). The numerical prediction method has made significant progress, including advancements in ensemble predictions, parameterization schemes and finer resolutions (Gao et al., 2021; Wu et al., 2022). Weather prediction for Research and Forecasting (WRF) is one of the most widely used research software package for numerical weather prediction, which are used in various studies to investigate effects on TC-induced remote precipitation. A recent study set different levels of water vapor saturated and found that 300–400 hPa and 900–1,000 hPa moisture can result in peak values of precipitation over Yangtze River Delta induced by Typhoon “Mangkut” (Liu et al., 2023). Studies using the WRF to investigate Henan extreme rainfall in July 2021 show that Typhoon “In-fa” plays a significant role on generating jet-streak and transporting moisture to arc-shaped convergence zone, and Typhoon “Cempaka” also plays a role in maintaining Huang-Huai cyclone and south-easterly flow (Rao et al., 2025; Xu et al., 2022; Yin et al., 2022). The Integrated Forecast System (IFS) in the European Center for Medium-range Weather Forecast (ECMWF, or EC), similar to the WRF, is another numerical weather prediction system that powers 6-hr forecast appeared in this study, whose historical forecast records starting from 2015 are archived in the Interactive Grand Global Ensemble (TIGGE) (Buizza et al., 2018; Swinbank et al., 2016).

However, governing-equation-based numerical modeling require increasing computational cost with more parameterization and finer resolution, which limits the number of ensembles (G. Chen & Wang, 2022; Espenholt et al., 2022). Fortunately, deep learning models have attracted considerable attention over the past few years, achieving superior performance in generic precipitation nowcast with more parameters and increased complexity for higher accuracy (Cao et al., 2023; Kim et al., 2024). One of the notable breakthroughs is PhyDNet, a model integrating a physics-constrained Long Short-Term Memory (LSTM) network. Originally developed for predicting moving object trajectories in the MNIST data set (Guen & Thome, 2020), PhyDNet has been capable of scoping with challenging task of precipitation forecast (Schultz et al., 2021), demonstrating significant progress in physics-guided deep learning for weather forecasting. In 2023, the recent Pangu-Weather model outperforms the numerical weather prediction model the in higher accuracy and shorter prediction time using 33.7 billion learnable parameters (Bi et al., 2023). Another metamodel Fuxi has achieved longer lead time than ECMWF in Z500 (9.25–10.5 days) and T2M (10.25–14.5 days) by only using 3 billion parameters (Lei Chen et al., 2023). A small deep learning model with radar data as input has achieved higher accuracy in 3-hr precipitation nowcasting than Deep Generative Model for Radar (DGMR) using numerical modeling equations as modules of the model. Moreover, there are further achievements in forecast for special weather events (Ravuri et al., 2021; Y. Zhang et al., 2023). TC track forecasting and AR forecast is achieved by GraphCast by converting numerical modeling grid cells into huge digraphs (Peng et al., 2023). The innovatively designed event-specific models achieve promising forecast of precipitation induced by the ARs, mesoscale convective systems and the TCs by integrating their objective identification methods and minimum outer ellipse algorithm (L. Zhang et al., 2024). However, there is lack in forecasting the TCRP combining objective identification methods and deep learning models.

Despite the significant achievements in skillful precipitation using deep learning models, there remains a challenge of revealing the forecast mechanisms of machine learning, also called the black-box problem (McGovern et al., 2019). To make the AI models more white-box, explainable methods are developed, and these are demanding and highly concerned. Random forest is one of the most interpretable machine learning models, and these are used for feature importances analysis for prediction of droughts using the inputs and outputs of soil moisture (Huang et al., 2023). However, deep learning models becomes more uninterpretable with higher complexity including more layers, functions and modules (Hussain, 2019). Fortunately, solutions to more white-box deep learning models are proposed like surrogate models, information extraction, feature importance, model representation, showing pivotal intermediate variables and parameters (also called learned weights) (Huang et al., 2024). One of the most feasible methods to increase the transparency of deep learning models is to design explainable attention modules or layers, which has successfully interpreted wave fusion in ideal computational experiments (Briden & Norouzi, 2021; Shakin et al., 2023; R. Yang et al., 2024).

Combining the lack of AI-based TCRP forecast and high demand of explainability of deep learning models, we conducted the TCRP forecast method using attention-based deep learning models. Moreover, an explainable method based on attention module is applied. Section 2 presents the data source, model selection, training strategies and the explainability method. Section 3 shows the results of the model performance, interannual mean of precipitation amounts and explainability features, and case analysis on Henan extreme rainfall in July 2021 using explainable deep learning models. Section 4 concludes all our findings and discusses the implications of this study.

2. Data and Methods

Deep learning has achieved remarkable progress in precipitation forecasting. However, critical limitations persist in existing research. First, the majority of models (e.g., ConvLSTM, U-Net) rely solely on single-source precipitation data, failing to adequately incorporate synergistic interactions among multiple physical fields. Second, forecasting studies targeting specific weather systems remain notably scarce such as the TCRP. To address these gaps, this study innovatively develops a multi-source data fusion framework based on an enhanced VGG encoder-decoder architecture. A total of 150 ensemble members were generated through 5-fold cross-validation, with the top 10 optimal members selected via TS_{25} scoring in the TCRP for post-processing refinement.

In this study, we systematically designed four pivotal experimental groups, which are specifically aimed to improve the performance on TCRP forecast, including multi-source data inputs (atmospheric variables, moisture transport and relevant masks, GPM precipitation fields, and topography), SE-Block attention mechanisms, atmosphere-informed activation functions and spatiotemporal weighted mean squared error (WMSE). Forecast results of TCRP by models above are evaluated in both pixel-based and neighborhood-based metrics. Moreover, an interpretability analysis method based on channel attention weights was developed to quantitatively reveal the attention distribution of SE-Block on multi-physical fields in the TCRP forecasting.

The methodology breaks through the single-source data limitations of traditional precipitation extrapolation models while establishing a novel technical pathway for refined extreme weather forecasting through deep integration of atmospheric knowledge and data-driven approaches. The framework demonstrates significant potential for enhancing TC remote precipitation early warning capabilities in operational forecasting systems, offering both methodological advancements and practical applications.

2.1. Data

Developed by European Center for Medium Range weather forecast (ECMWF), ERA5 reanalysis data is in higher resolution in $0.25^\circ \times 0.25^\circ$ spatially and 1 hr temporally (Hersbach et al., 2020). We used variables in July–August–September (JAS) during 2001–2023 at multiple pressure levels with temporal resolution of 6 hrs. Three-dimensional temperature, geopotential height, relative humidity and wind field are used for input variables of precipitation forecast. Three-dimensional Wind-field and specific humidity are used for moisture transport calculation and objective identification of Tropical Cyclone induced remote moisture transport. Precipitation forecast results from TIGGE retrieval provided by the ECMWF (TIGGE_EC) is used as the numerical weather prediction benchmark due to its wide acceptance and applications to compare its performance with all deep learning methods in this study (Buizza et al., 2018; Swinbank et al., 2016).

The best-track TC data set during 2001–2023 is retrieved from the Shanghai Typhoon Institute of the China Meteorological Administration (CMA). We used 6-hourly TC positions to identify strong moisture transport of TCs, which is called TC clusters (TCCs), as well as TC remote moisture transport, which are TRCs that split from TCCs (Xiao, Zhang, Chen, & Li, 2024; Ying et al., 2014).

The Integrated Multisatellite Retrievals for Global Precipitation Measurement (GPM-IMERG) is funded by NASA and the Japanese Aerospace Exploration Agency (JAXA), which offers globally integrated precipitation observational data (Pradhan et al., 2022). This product enhances the scope of the Tropical Rainfall Measuring Mission, extending polar observations. We analyzed the calibrated precipitation rate from the final precipitation L3 V07 product in JAS during 2001–2023, at $0.1^\circ \times 0.1^\circ$ spatial and 0.5-hr temporal resolutions.

The digital elevation model (DEM) ETOPO2v2c_f4 is developed by the National Ocean and Atmospheric Administration (NOAA) National Center for Environmental Information (NCEI), which is a 64-bit float number network saved in netCDF format, with $2' \times 2'$ ($1/30^\circ \times 1/30^\circ$) resolution (National Geophysical Data Center, 2001).

2.2. Identification Method of TC Remote Precipitation

We used the objective identification method to identify TRCs and their tracks. A concise overview of this identification method is provided. The identification of intense moisture transport is revised from PanLu algorithm, and then the clusters are segmented from intense moisture transport using maximum gradient method (Y. Chen et al., 2020; M. Pan & Lu, 2019). Clusters are then matched with TC tracks and assigned as TC clusters (TCCs), and those clusters that split from TCCs based on temporal digraphs and yet to be identified as TCCs and over 700 km away from the TCC at the same timestep are identified as TRCs and otherwise seed TRCs (Xiao, Zhang, Chen, & Li, 2024). These clusters are depicted as masks with pixels inside TRCs as true for further use in TRC-related features. The TCRP refers to cumulative precipitation in the next 6-hr timestep over the region of the TRC. There are 253 TRCs in 2001–2020 and 51 TRCs in 2021–2023.

2.3. Deep Learning Architecture

The state-of-the-art deep learning model PhyDNet has been successfully used in precipitation nowcasting and has shown superior performance to numerical weather prediction models, thanks to the dynamical feature extraction using physical prognostic equations in the PhyDNet (Schultz et al., 2021). In this study, the model PhyDNet was directly trained 30 epochs with learning rate of 1×10^{-3} using the GPM-IMERG in the past 10 hr (10 frames) as input to predict the TCRP in the next 10 hr (10 frames) to be consistent with the architecture in the previous study (Guen & Thome, 2020). We calculated the sum of the first six frames of the output as 6-hr precipitation forecast to compare the performance with that of our deep learning models mentioned afterward in the evaluation stage.

We revised and applied the ChenNet to forecast the TCRP as well as other precipitations over East Asia in Figure 1 (G. Chen & Wang, 2022), which is based on VGG-like architecture named after a team in ImageNet Large-Scale Vision Recognition Challenge (ILSVRC-2014) (Simonyan & Zisserman, 2014). This model consists of two main components: an encoder with downsampling operations and a decoder with upsampling layers. Each stage of network includes four to five VGG-like blocks depending on data types, which serve as base units for both encoding and decoding processes. TA Squeeze and Excitation Block (SE-Block) was introduced to enhance feature representation when all four types of input variables are used (see Table 1).

The first type of input data that the model processes is 6-hourly (0000/0600/1200/1800 UTC) ERA5 fields in multiple pressure levels including temperature (T), geopotential heights (z), relative humidity (RH), zonal, and meridional wind components (u , v), with the array size of $5 \times 30 \times 193 \times 193$. The first dimension of arrays represents the number of channels. The vertical pressures vary from 30 to 1000 hPa (30 vertical grids), and the horizontal range of training region (TR) is 102°E – 150°E , 8°N – 56°N (see Figure S1 in Supporting Information S1). The second type consists of moisture transport height characterized by the median height of moisture transport (H_{50}), moisture transport intensity characterized by the Integrated water Vapor Transport (IVT) and the TRC mask (as in the “TRC” in Figure 1), all of which processed from the ERA5 data set. The H_{50} is derived from the vertical structure of the IVT. The TRC masks are derived from the IVT based on the objective identification method in the previous study (Xiao, Zhang, Chen, & Li, 2024). The IVT is derived from specific humidity (q) zonal and meridional wind components (u , v). There are two arrays for the second type input, each size of $2 \times 193 \times 193$.

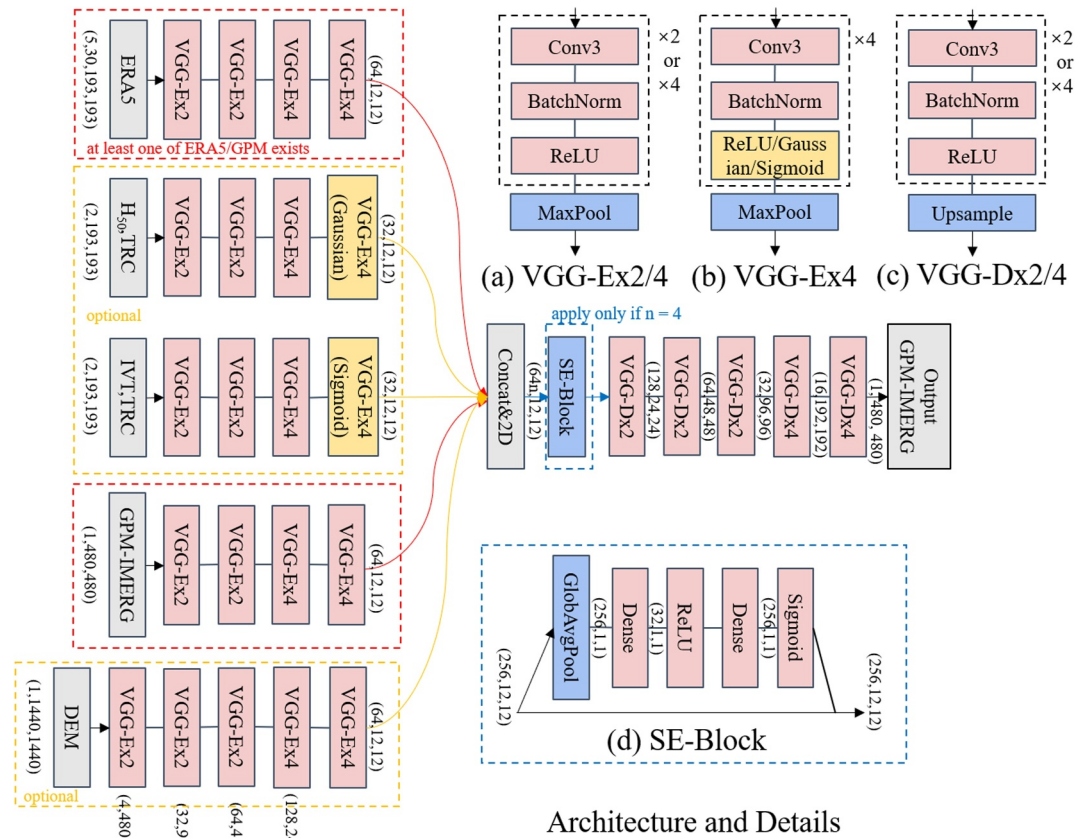


Figure 1. The VGG-like deep learning network used in this study. Gray boxes represent inputs, outputs, concatenation, and 2D processing operations. Red boxes represent VGG-like encoder and decoder, convolutional layers, batch normalization layers, and non-atmospheric-informed activation functions; the blue boxes represent pooling layer and upsampling layers; yellow boxes represent atmosphere-informed activation functions. At least one type of variable is input variable in the red dashed box; yellow dashed box represents optional data types; SE-Block in the blue dashed box applies only if there are four types of input variables ($n = 4$). (a) Details of the VGG-Ex2 block where the layers in black dashed box repeat twice. (b) VGG-Ex4 block that is specifically used in the optional TRC-related features. (c) VGG-Dx2/4 showing the same architecture as (a) but existing in the decoder block. (d) Architecture details of the SE-Block.

The first array consists of H_{50} and the TRC mask that occupies the first dimension of the array, and the second array consists of the IVT and the TRC mask. The third type of data is GPM-IMERG cumulative precipitation in the past 6 hrs derived from calibrated half-hourly rainfall rate, with the array size $1 \times 480 \times 480$. The last type is one type of the DEM data ETOPO, which is time-independent with array size $1 \times 1,440 \times 1,440$. The range of

Table 1
The List of Data and Variables Used in This Study

Data types	Variables	Derived from	Array size
ERA5	Temperature (T); Geopotential heights (z); Relative humidity (RH); Zonal and meridional wind components (u, v)	N/A	$5 \times 30 \times 193 \times 193$
TRCf	First part: H_{50} and TRC mask Second part: IVT and TRC mask	H_{50} and TRC mask derived from the IVT; IVT derived from Specific humidity (q), Zonal and meridional wind components (u, v)	$2 \times 193 \times 193$, $2 \times 193 \times 193$
GPM	6-hourly precipitation	Half-hourly calibrated rainfall rate	$1 \times 480 \times 480$
DEM	Elevation (in m)	N/A	$1 \times 1440 \times 1440$

training data are all the TR among all types of data (each abbreviated as ERA5, TRCf, GPM, and DEM), roughly ordered from high to low vertical levels. At least one data type of ERA5 and GPM is mandatory shown in the red dashed boxes, and the remaining data types are optional shown in yellow dashed boxes. The output variable is the GPM-IMERG cumulative precipitation in the next 6 hrs with the array size $1 \times 480 \times 480$. The lead time of these VGG-like models are 6 hrs.

The training set covers JAS in 2001–2020 with 7,360 samples in total, and the testing set covers JAS in 2021–2023, with the sample size of 1,104. Part of the training set split as validation set for 5-fold cross validation. The number of years share the same remainder mod 5 for each fold. For instance, the validation set of the fifth fold consists of the year 2005, 2010, 2015, and 2020. This training strategy can contain multiple phases of atmospheric oscillation similar to El-Nino. We trained 30 epochs for each fold and save each fold and each epoch, in total 150 parameter sets (also called “members”) for precipitation forecast and further fine tune. This methodology resembles ensemble weather forecast to select “good” members for combination, and also resembles ensemble machine learning (Lin et al., 2024; Zängl, 2023).

The VGG-like blocks are the base blocks throughout all stages of deep learning network, which consists of the first-half encoder with downsampling and the second-half decoder with upsampling. The convolutional layer is 3D if the data type is ERA5 and otherwise 2D. In the first stage of our model, each encoder block (see Figure 1a) consists of convolution, batch normalization and Rectified Linear Unit (ReLU) activation function assigning negatives as zeros, repeating two or four times depending on the depth of encoder phase or decoder phase, then followed by a maxpooling layer. The kernel size of all other maxpooling layer is two, except for the first (five) encoder block VGG-Ex2 in the GPM, the first (three) and the second (five) encoder block VGG-Ex2 in the DEM for the alignment of the array sizes in Figure 1. The batch normalization layers ensure a stable statistical distribution of activation values in the whole training process to allow the deep learning models to be less stringent during initialization, thereby reaching high computing performance and less computational cost necessary for the training phase. For the data type TRCf, the last activation function right before concatenation and 2D operation (as in Figure 1b) is standard Gaussian function (H_{50} and the TRC mask) and sigmoid function (IVT and the TRC mask), consistent with non-linear fitting functions in the past research, which is now called atmospheric-informed activation functions (Xiao, Zhang, Chen, & Li, 2024).

In the second stage, each decoder block shares the same repetitions as that of encoder blocks, but with upsampling layer at the end of the block (see Figure 1c). The scale factor of the all other upsampling layer is two, except for 2.5 for the final upsampling layer. The optional attention module is called Squeeze and Excitation Block (SE-Block), which squeezes the inputs by global average pooling for each channel and excites them through two dense layers to dimensions of $32 \times 1 \times 1$ and then $256 \times 1 \times 1$, each followed by a ReLU layer to and finally sigmoid activation function for channel-wise multiplication to rescale the input array (see Figure 1d) to filter and rearrange the signals of these channels (such as the RGB and infrared channels in the pictures) (Hu et al., 2018). This module operates only when the input variables consist of four types.

2.4. Objective Functions and Evaluation Metrics

We used mean squared error (MSE) to evaluate the similarity between forecasted values and GPM-IMERG values at the same timestep in the testing set. Additionally, the Threat Score (TS, also called critical success index CSI) is used to evaluate the hit rate of precipitation over a set threshold (Han et al., 2022; X. Pan et al., 2021). These two metrics are used to evaluate the forecast of all precipitation and the TCRP of each cross-validation member and that of fine-tuned forecast results combining those members. Neighborhood threat score (NTS) is a metric that considers any of events that forecasts true or false to compare with the centered “ground truth” true or false within a 3×3 window for each pixel, which is only used after the postprocessed results. The formula is shown as the following:

$$\text{MSE} = \frac{1}{n} \sum_{i,j} (\hat{P}_{ij} - P_{ij})^2, \quad (1)$$

$$\text{WMSE} = \frac{\sum_{i,j} f_{ij} (\hat{P}_{ij} - P_{ij})^2}{\sum_{i,j} f_{ij}}, \quad (2)$$

$$\text{TS}_p = \frac{\text{NA}}{\text{NA} + \text{NB} + \text{NC}}, \quad (3)$$

$$\text{NTS}_p = \frac{\text{NA}'}{\text{NA}' + \text{NB}' + \text{NC}'}, \quad (4)$$

Where $\hat{P}_{ij}$ is the precipitation forecast values in the longitude-latitude grid (i,j) , P_{ij} is the GPM “ground truth” in 0.1° except for evaluation on the TIGGE_EC, where the resolution is bilinearly interpolated into 0.25° , n is the number of longitude-latitude grids. WMSE is weighted MSE function, only used as a modified objective function with the coefficient f_{ij} that favors high climatological TRC frequency in JAS during 2001–2020 (see Figure S2 in Supporting Information S1) and instantaneous heavy precipitation in “ground truth.” TS_p refers to Threat Score over precipitation threshold $p \in \{1, 5, 10, 25\}$ in mm, with better forecast with larger TS_p , which is only used as evaluation metrics. NA represents the number of true positives, NB represents the number of false positives, NC represents the number of false negatives. NA' represents the number of true positives but with true center “ground truth” and any of the true forecast value within the 3×3 window, and so are NB' and NC' . We evaluated forecast results in the JAS during 2021–2023 with the range of the evaluate region (ER) over 104° – 136°E , 16° – 48°N .

2.5. Post Processing Method

The frequency of none and light precipitation is much higher than that of heavy precipitation, causing trained deep learning models to underestimate the values of heavy precipitation, especially the more extreme TCRP concerned by public (Xiao, Zhang, Chen, & Li, 2024). The tuning strategy is widely used to improve the overall performance in machine learning predictions (G. Chen & Wang, 2022). Therefore, we selected at most 10 best models from cross-validation with the smallest MSE and $\text{TS}_{25} > 0$ and tuned the precipitation to enhance the overall performance especially for the heavy precipitation. The precipitation is tuned following the equations to minimize the MSE of tuned precipitation:

$$\overline{P_w} = \sum_k w_k \hat{P}_k, \quad (5)$$

$$\text{TP} = \left(1 + \frac{a \overline{P_w}}{P_{\max}}\right)^b \overline{P_w}, \quad (6)$$

$$\text{MSE}_{\text{TP}} = \sum_{i,j} (\text{TP}_{ij} - P_{ij})^2, \quad (7)$$

$$\text{WMSE}_{\text{TP}} = \frac{\sum_{i,j} f_{ij} (\text{TP}_{ij} - P_{ij})^2}{\sum_{i,j} f_{ij}}, \quad (8)$$

$$\mathbf{w}, a, b = \underset{\mathbf{w}, a, b}{\text{argmin}} \text{MSE}_{\text{TP}}, \quad (9)$$

Where $\overline{P_w}$ is the weighted average values of forecasted precipitation, w_k are learned weight parameters for the tuning process as components of the vector $\mathbf{w}$. TP is tune cumulative precipitation in 6 hrs for each training sample, $P_{\max}$ is the maximum precipitation over the training sample, a and b are learned parameters for the multiple and the exponent. We minimize the objective function that is consistent with that during the training of VGG-like models to find the corresponding optimal parameters $\mathbf{w}$, a , and b , subject to constraints that the sum of the $\mathbf{w}$ components equals to one, $a \geq 0$ and $0 \leq b \leq \exp(1)$ to ensure that the maximum tuned precipitation increases and get closer to the maximum GPM “ground truth.”

2.6. Explainability Method of SE-Block

SE-Block consists of global average pooling on each channel and dense (linear) layers followed by ReLU and finally sigmoid activation function to skew and readjust the coefficients of each channel to obtain the rescale vector $\mathbf{R}$ with size of $256 \times 1 \times 1$ (Hu et al., 2018), which is used for channel-wise multiplication to rescale the input array $\mathbf{x}$ in $256 \times 12 \times 12$ (see Figure 1d), followed by VGG decoder blocks to predict 6-hourly cumulative precipitation. Meanwhile, explainability has been a challenging issue on black-box and complex deep learning models instead of relatively simple and white-box linear regression and random forest (Hussain, 2019; Salih & Wang, 2024). Inspired by recent successful applications on SE-Block, we used explainability method on SE-Block in VGG-like models to further investigate the attention distribution on four types of input data, assuming the same number of channel and bytes represent the same amount of information (Briden & Norouzi, 2021; Shakin et al., 2023).

For all the components of $\mathbf{R}$, we derived the attention distribution to ERA5, TRCf, GPM, and DEM data type by summing up the partial derivative of $\mathbf{R}$'s to $\mathbf{x}$'s of each channel $\frac{\partial R_i}{\partial x_j}$, where x_j ($1 \leq j \leq 256$) is obtained from global average pooling (Figure 1d). Every 64 channels represent the information of ERA5, TRCf, GPM, and DEM, respectively. We used Equation 11 to calculate coefficients, which is the expanded version in Equation 10 using the principle of backward propagation, which was initially used in objective function optimization (details are in the Supporting Information S1).

$$\text{coef}_{kj} = \sum_i \frac{\partial R_i}{\partial x_j} = \sum_i \frac{\partial R_i}{\partial y_1} \frac{\partial y_1}{\partial f_{\text{ReLU}}} \frac{\partial f_{\text{ReLU}}}{\partial y_0} \frac{\partial y_0}{\partial x_j}, \quad (10)$$

$$\text{coef}_{kj} = \sum_i \frac{\partial R_i}{\partial x_j} = \sum_i R_i (1 - R_i) \left(\sum_h w_{1ih} w_{0hj} I \left(\sum_j w_{0hj} x_j > 0 \right) \right), \quad (11)$$

Where coef_{kj} is the sum of x_j coefficients for the selected model k by summing up $\frac{\partial R_i}{\partial x_j}$, the partial derivative of the rescaled results R_i to x_j , and the remaining multiples are partial derivatives for each layer in the SE-Block following the chain rule. R_i and y_1 represents the output and input of sigmoid activation function, respectively. w_{1ih} and w_{0hj} represent the second and the first layer parameter matrix, respectively. $I(y_0 > 0)$ represent identity function, one for the true statement and zero for the false statement. The tuned coefficients over each of 256 channels are calculated similarly to tuned precipitation in Equation 5.

Feature importances (FI), mostly applied in random forest and also applied in attention-based deep learning networks, offers a straightforward contribution percentages of each input variable or feature widely used in studies (Huang et al., 2023; L. Zhang et al., 2024). As an analog and combining the sum of squares in wind speed calculation and distance calculation in Cartesian coordinates, we summed up the square of coefficients over each channel to obtain the FI values in SE-Block for each training, validating or testing sample.

$$\text{FI}_l = \frac{\sum_{j=64l+1}^{64(l+1)} \overline{\text{coef}_j^2}}{\sum_{j=1}^{256} \overline{\text{coef}_j^2}}, \quad (12)$$

Where $\overline{\text{coef}_j^2}$ is square of the coefficient of tuned results at channel j using similar principle to that of Equation 5 derived from coef_{kj} , l is the number corresponding to data types from 0 to 3, increasing with ERA5/TRCf/GPM/DEM. The calculation of FI of tuned precipitation results follows the calculation of the weighted average vector $\mathbf{w}$ in the post processing phase.

2.7. Experimental Settings

We set deep learning groups of VGG-like models (as in Table 2), which aims to improve the performance on the TCRP, to investigate the effect of the following effects on models' performance on TCRP forecast, including the diversity/number of data types, existence of SE-Block, atmosphere-informed activation functions right before

Table 2

The Experiments Consisting of Model Numbers, Model Names, Input Data, Whether SE-Block Operates, Last Activation in the TRCf Encoding Block and Objective Functions

Experiments		Model number	Model name	Input data	SE-block	Last activation function in the TRCf encoding block	Objective function
Number of input data types		1	VGG_E	ERA5	No	ReLU	MSE
		2	VGG_P	GPM	No	ReLU	MSE
		3	EP	ERA5, GPM	No	ReLU	MSE
		4	EPT	ERA5, GPM, DEM	No	ReLU	MSE
Replacing atmosphere-informed activation functions	Adding SE-Block	5	ERPT	ERA5, TRCf, GPM, DEM	No	ReLU	MSE
		6	ERPT_SE	ERA5, TRCf, GPM, DEM	Yes	ReLU	MSE
	Modifying the objective function	7	ERPT_SEG	ERA5, TRCf, GPM, DEM	Yes	Gaussian, Sigmoid	MSE
		8	ERPT_SEGW	ERA5, TRCf, GPM, DEM	Yes	Gaussian, Sigmoid	WMSE

data concatenation in Figure 1 and the custom objective function combining TRC frequency and “ground truth” precipitation levels.

The first group is the first control group of the number of data types, whose input data is only ERA5 (No. 1: VGG_E) following the previous study (G. Chen & Wang, 2022). The second group is the second control group, with the input data only GPM Precipitation (No. 2: VGG_P) to be consistent with commonly used deep learning-based precipitation forecast models only with precipitation itself (Cao et al., 2023; Ravuri et al., 2021; Y. Zhang et al., 2023). The rest are experimental groups consisting of VGG-like models with input variables ERA5 and GPM data (No. 3: EP), the VGG-like model with input variables ERA5, GPM, and Topographical DEM data (No. 4: EPT), and the VGG-like model with input variables ERA5, TRCf, GPM, and DEM data (No. 5: ERPT). For the effects of SE-Block, No. 5 is the control group and the VGG-like model with SE-Block is the experimental group (No. 6: ERPT_SE). To investigate the effect of atmosphere-informed activation functions, No. 6 is the control group with activation function ReLU for VGG-Ex4 in Figure 1c, and the experimental group is the VGG-like model with SE-Block and the atmospheric-informed activation functions of Gaussian and sigmoid in VGG-Ex4 in Figure 1c (No. 7: ERPT_SEG) following the fitting relationship between moisture transport characteristics and precipitation (Xiao, Zhang, Chen, Chen, et al., 2024). The effect of the objective function is investigated by comparing the model No. 7 control group and the ChenNet with SE-Block, activation function of Gaussian and sigmoid and the modified objective function WMSE (No. 8: ERPT_SEGW).

3. Results

3.1. Evaluation of Precipitation Forecasting Performance in Deep Learning Models

The model VGG_P is used to determine the optimal learning rate among 2.5×10^{-4} , 5×10^{-4} and 1×10^{-3} (see Figure S3 in Supporting Information S1), as lines of VGG_P are the smoothest among all deep learning groups due to homogeneous precipitation inputs and outputs. The results show that VGG_P with learning rate of 5×10^{-4} converges to optimal values at 10–20 epochs, right in the middle of 30 epochs, which is why we chose 5×10^{-4} as the learning rate for remaining deep learning groups (see Figure S3 in Supporting Information S1).

We selected the most optimal PhyDNet model, which is at 20th epoch (see Figure S4 in Supporting Information S1), and then selected up to 10 best “members” generated from cross validation consisting of 5 folds and 30 epochs for each VGG experiment (see Table S1 in Supporting Information S1), to fine-tune the optimal objective functions of VGG models, whose overall performance is shown in Table 3. In general, all VGG-like models in this study show 0.05–0.2 higher TS and NTS score (higher performance) over all thresholds on TCRP forecast than that of TIGGE_EC, but 23.3%–43.6% higher MSE (lower performance) than that of TIGGE_EC (54.71). Despite of shorter time range of past precipitation inputs, coarser resolution and fewer frames of VGG_like models (one frame) than that of PhyDNet (six frames), all VGG_like models in general show similar model performance to that of PhyDNet, validating the capability of our deep learning models in forecasting the TCRP in the next 6 hrs. The deep learning model ERPT_SE (with SE-Block) achieves the best performance on TCRP in the five metrics of TS₅, TS₁₀, NTS₁, NTS₅, and NTS₁₀ with the values above 0.46, improved by 0.05–0.07 compared to VGG_E, and 0.04–0.08 higher

Table 3
Performance of Models on TCRP in MSE Metric and TS Metrics Over 1, 5, 10 and 25 mm Threshold

Model	MSE	TS ₁	TS ₅	TS ₁₀	TS ₂₅	NTS ₁	NTS ₅	NTS ₁₀	NTS ₂₅
TIGGE_EC	54.71	0.4426	0.3933	0.2830	0.1413	0.3085	0.3121	0.2742	0.1377
PhyDNet	74.38	0.4706	0.4683	0.4718	0.2997	0.4007	0.4117	0.4293	0.2980
VGG_E	78.54	0.5082	0.4671	0.4239	0.0719	0.4591	0.4235	0.4110	0.0857
VGG_P	74.26	0.5470	0.4791	0.4771	0.1453	0.4947	0.4295	0.4491	0.1620
EP	71.90	0.5409	0.5105	0.4680	0.1714	0.4875	0.4612	0.4421	0.2039
EPT	73.64	0.5514	0.5266	0.4708	0.1260	0.4974	0.4751	0.4532	0.1397
ERPT	68.81	0.5248	0.5090	0.4893	0.2458	0.4741	0.4578	0.4509	0.2701
ERPT_SE	68.74	0.5502	0.5283	0.4936	0.2061	0.4982	0.4782	0.4669	0.2238
ERPT_SEG	69.71	0.5183	0.4835	0.4922	0.2012	0.4661	0.4294	0.4506	0.2163
ERPT_SEGW	67.50	0.4967	0.4765	0.4813	0.2360	0.4476	0.4236	0.4362	0.2425

Note. Models that perform the best are bolded in each metric. Only deep learning-based models are bolded if they perform the best in a metric.

than that of PhyDNet, implying the best performance mainly over middle TCRP thresholds due to the presence of the SE-Block. The model ERPT, with most diverse types and addition of TRC-related features including TRC masks, performs the best in two metrics of TS₂₅ and NTS₂₅ with the values above 0.24, improved by 0.17–0.19 compared to VGG_E, implying the superior performance in extreme TCRP to all other models in this study except to PhyDNet, with TS₂₅ and NTS₂₅ almost 0.3. The model EPT (three types of data input) achieves the highest TS₁. The model ERPT_SEGW achieves the lowest MSE (67.50) among all this study's deep learning models, 9.25% smaller than that of the state-of-the-art PhyDNet, but still 23.3% larger than the MSE of the TIGGE_EC. It is worth noting that this model also achieves the third highest TS₂₅ and NTS₂₅ among all this study's deep learning models, and improved by 0.16 compared to VGG_E, implying that modifying the objective function to WMSE almost restores the negative effects of adding SE-Block and replacing activation functions on performance on extreme TCRP. The model VGG_P, whose inputs and outputs are precipitation only, improves most significantly in TS₂₅ and NTS₂₅ by almost 100% compared to VGG_E. Results above show the significant effect on TCRP performance, especially extreme precipitation by diverse data inputs (also with the addition of TRCf variables with TRC mask), adding SE-Block and modifying objective function to WMSE.

3.2. Interannual Analysis in the Testing Set

After the overall performance over the ER is analyzed, Figure 2 further depicts the spatial distribution of annual mean TCRP of GPM-IMERG “ground truth”, TIGGE_EC and VGG-like models in the testing set. Large-scale features are mostly conveyed by PhyDNet and all VGG-like models in this study, consistent with previous findings that deep learning models show promising precipitation values in general statistics (Ayzel et al., 2020; Wang et al., 2023; S. Yang & Yuan, 2023). The annual mean TCRP forecasted from PhyDNet and VGG-like models show high similarity of values and spatial distribution to that of GPM-IMERG. There is overestimation of light precipitation events generated from PhyDNet by 183% and eight VGG-like models by 123%–197% for annual mean TCRP compared to that of GPM-IMERG over the ER (Figure 2a), which is likely the reason of higher MSE than that of TIGGE_EC (Table 3). This overestimation is unlike slight and negligible underestimation of TIGGE_EC by 1.5% (Figure 2b). VGG_E and the EPT achieve the smallest annual mean TCRP in central China with annual TCRP forecast values 1.59 and 1.45 mm yr^{−1} over the ER, respectively (Figures 2d and 2g). The area of 30–50 mm yr^{−1} in the central part is slightly larger for four models VGG_P, EP, ERPT, and ERPT_SEG (Figures 2e, 2f, 2h, and 2j). It is worth noting that ERPT_SE and ERPT_SEGW nowcast over 50 mm yr^{−1} of annual mean TCRP in central China like that of GPM-IMERG, implying the effect of Henan torrential rainfall in July 2021 (Figures 2i and 2k). Meanwhile, the annual mean TCRP of the ERPT_SE and ERPT_SEGW are 1.54 and 1.94 mm yr^{−1} over the ER, respectively.

To further reveal the prediction mechanisms underlying the models with SE-Block (ERPT_SE, ERPT_SEG and ERPT_SEGW), the interannual mean coefficients and intermediate output X's (as in Equation 8) in JAS during

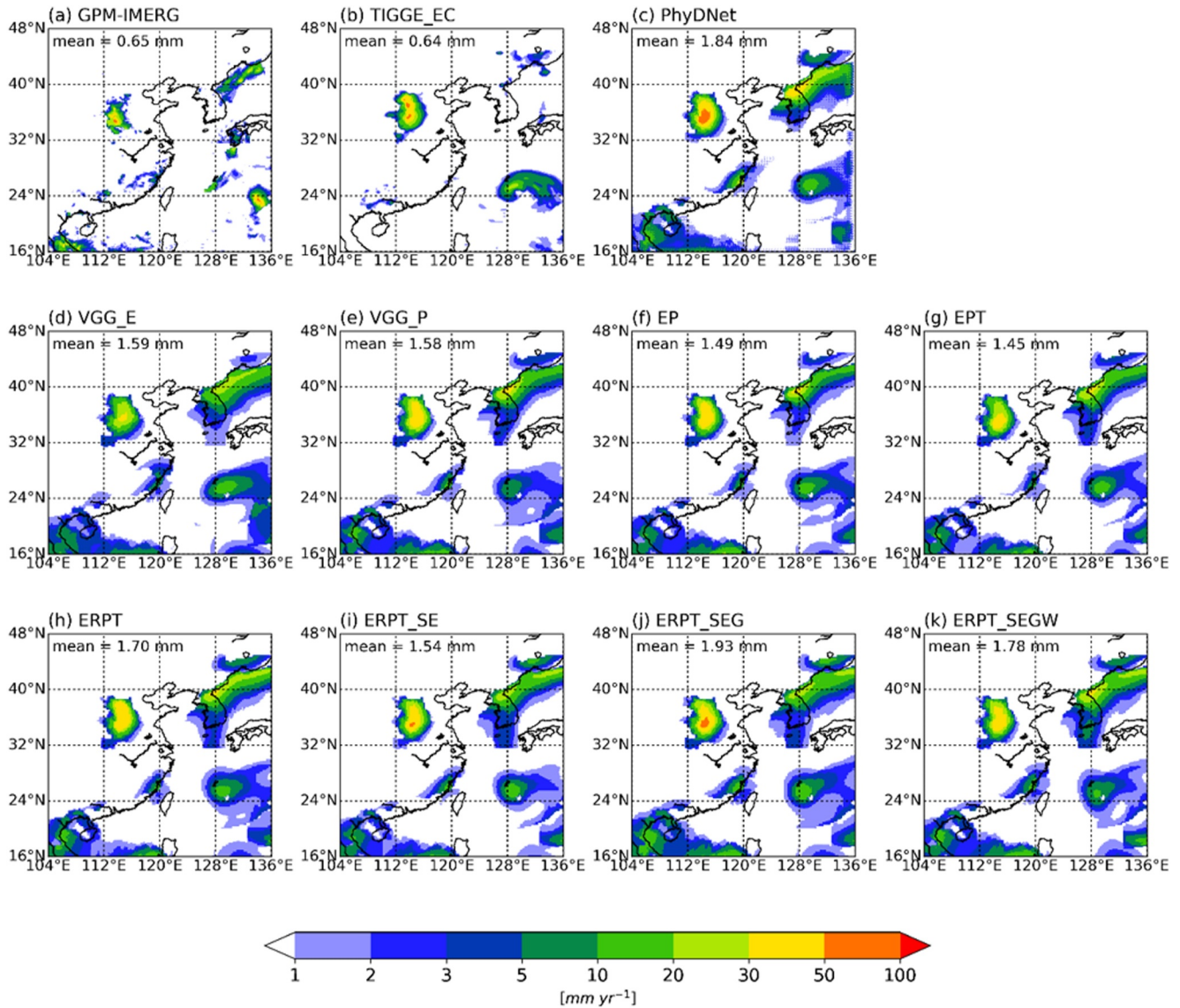


Figure 2. Geographical distribution of annual mean TCRP of (a) GPM-IMERG, (b) TIGGE_EC, (c) PhyDNet, and (d–k) eight VGG-like models in the testing set. Upper-left text in each panel shows the annual mean TCRP over the ER.

2021–2023 (time range of the testing set) are analyzed, as the dot product of the coefficients and X's in 256-dimensional (channels) space affects the values of R's, which further determines the results of precipitation nowcast as shown in Figure 2 and their performances shown in Table 3. Roughly 80% of coefficients are observed as negative in ERTP_SE (Figure 3a). The largest variations of channel-wise coefficients are observed in the channels of TRCf (TRC features) and GPM (precipitation in the past 6 hrs), ranging from -0.2 to 0.4 , which further validates that the FIs of the TRCf and the GPM are the largest (39.31% and 56.89%) among four types of data. Moreover, X's GPM-related channels are in average the largest (0.5–0.8) among four types of data, followed by the TRCf (0.4–0.8), ERA5 (0–1) and DEM (0–0.3) (Figure 3b), implying the information of TRC features and past precipitation are amplified relatively to ERA5 and DEM for precipitation forecasting. Replacing ReLU activation functions with atmospheric-informed Gaussian and sigmoid homogenizes the coefficients and X's of all channels (see Figures 3c and 3d). It is worth noting that there are lower FIs in TRCf and GPM and higher FI in DEM, implying that the replacement of activation function in TRC features can reduce implicit topographical features of TRCf and GPM, making SE-Block distribute more attention to topographical information (DEM in 14.40%) and 3D atmospheric information (ERA5 in 5.15%). Modifying objective function to WMSE makes SE-

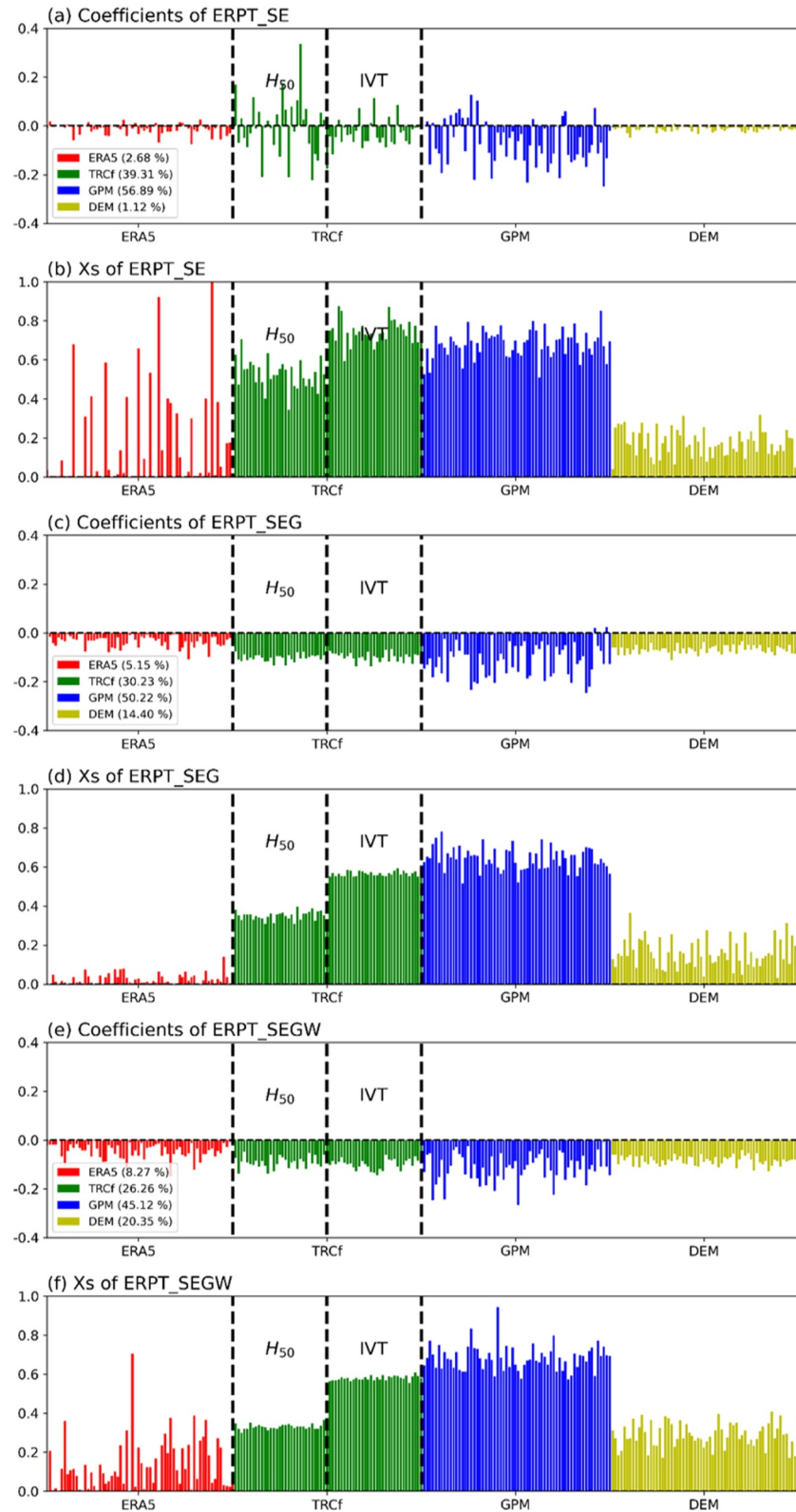


Figure 3. Interannual mean coefficients (a, c, e) to multiply intermediate variables X_s (b, d) of ERTP_SE (a, b), ERTP_SEG (c, d) and ERTP_SEGW (e, f). ERA5, TRCF, GPM, and DEM corresponds to four input data types in Figure 1. Black dashed lines show the channels of moisture transport height and intensity, respectively.

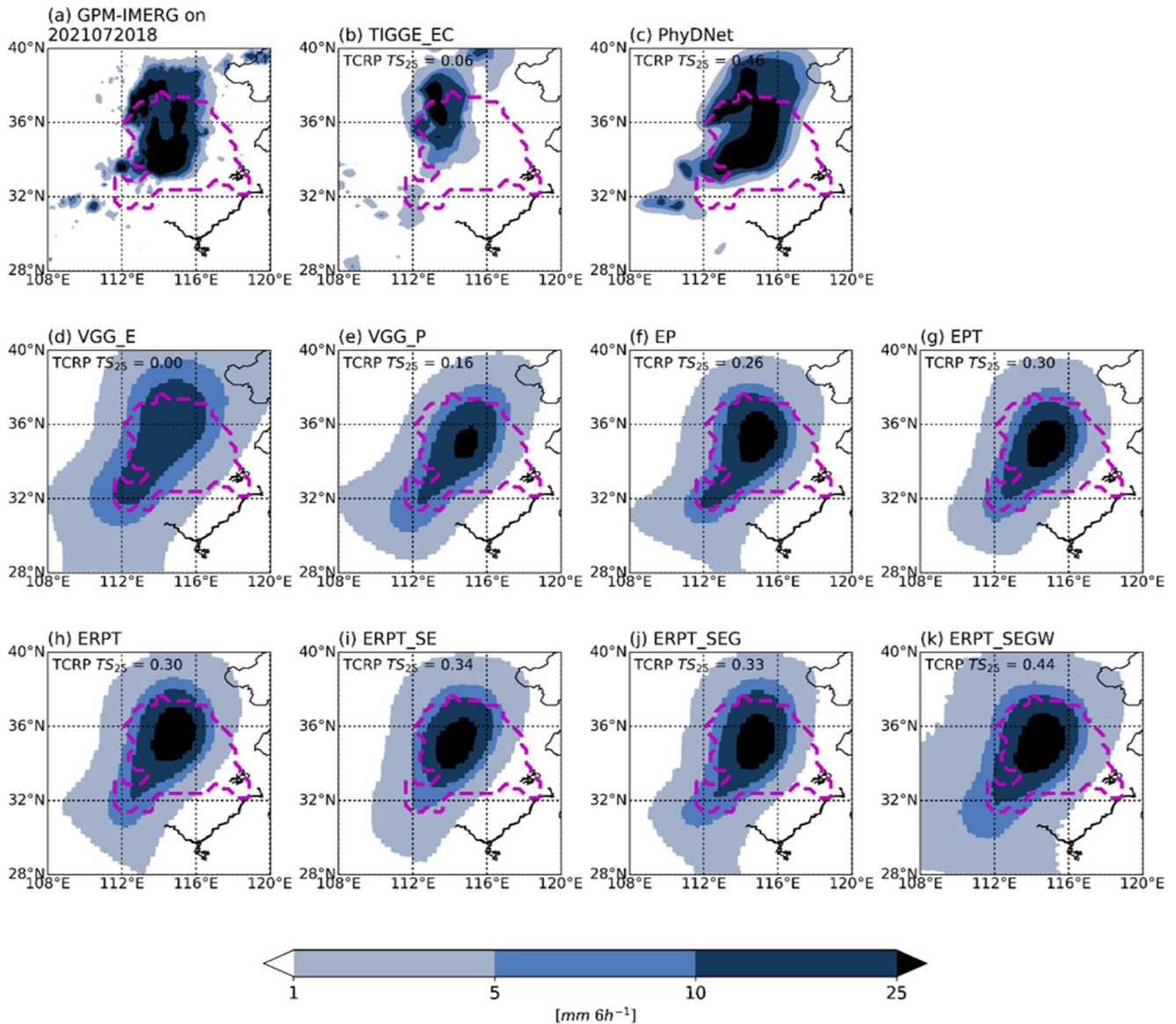


Figure 4. 6-Hourly cumulative precipitation of (a) GPM-IMERG, (b) TIGGE_EC, (c) PhyDNet, and (d–k) VGG-like models starting from 20 July 2021 at 1800 UTC. The dashed line represents the boundary of TRC–“In-fa.”

Block pay more attention to ERA5 (8.27%) and DEM (20.35%) by removing implicit ERA5 and DEM information in TRCf and GPM data (Figures 3e and 3f). The same analysis will be conducted in the case of Typhoon “In-fa.”

3.3. Performance Analysis in the TCRP Induced by Typhoon “In-fa” (2021) in Henan Province

We further analyzed the case of torrential rainfall in Henan Province on 20 July 2021 from 1800 UTC, which was The TCRP induced by Typhoon “In-fa” in Figure 4. Compared to GPM-IMERG “ground truth” (Figure 4a), the area of precipitation forecast values over 25 mm is significantly underestimated in the TIGGE_EC (Figure 4b). Results show that The TS_{25} of TCRP surrounding Henan Province generated from PhyDNet is 0.46 (Figure 4c). Meanwhile, precipitation generated from PhyDNet shows detailed features in the southwest and the heavy TCRP (over 25 mm) in the central region (Figure 4c). All deep learning models including PhyDNet and all the VGG-like models in this study can simulate precipitation values above 25 mm within TRC–“In-fa” in central China except for VGG_E with large precipitation area and low spatial peak precipitation values less than 25 mm (Figure 4d).

The area of heavy precipitation (25 mm) is low in VGG_P (Figure 4e), with TS_{25} of the TCRP 0.16 surrounding Henan Province, implying deep learning models with precipitation the only variable of the inputs and outputs tend to underestimate the precipitation amount and the area of heavy precipitation. The area of heavy precipitation generated by ERPT is the largest among the first five models with TS_{25} of the TCRP 0.30, which are used to investigate the effects of the number of input data (Figure 4h). Each of effects such as presence of SE-Block, the replacement of activation functions and the modified WMSE objective function makes the area of precipitation larger and more similar to that of GPM-IMERG “ground truth” (Figures 4i–4k). Among all VGG-like models in this study, the largest TS_{25} of TCRP surrounding Henan Province is generated from ERPT_SEGW with the value of 0.44 (Figure 4k). Scattered precipitation regions (as shown in the “ground truth”) in the northeast and southwest are interpreted as tails of the central precipitation region in all VGG-like models in this study, implying that there is large potential to improve the detailed precipitation features of VGG-like models in the future.

Explainability analysis using SE-Block is also conducted in the case of Henan heavy rainfall as in Figure 5. The overall results of this specific event are close to the interannual mean values in Figure 3. However, the anomalies from the interannual mean are unnegligible. Roughly 80% of the coefficients in ERPT_SE are in positive anomaly by 0–0.015, with less than 10% of interannual mean (Figure 5). In this same model, SE-Block distribute 36.22% of the attention to TRC-related features, 3.09% less than interannual mean. ERA5 and DEM are distributed by 1% less compared to the interannual mean. Roughly two thirds of X's is larger than interannual mean by -0.2 to 0.2 , ranging from -40% to 25% in percentage (Figure 5b). The coefficients in ERPT_SEG are mostly in positive anomaly with 1%–3% less (Figure 5c) than negative interannual mean values, which results in less than 0.1% deviation from interannual mean FIs in JAS during 2021–2023 (Figure 3c). This is likely due to less variant X's (Figures 3d and 5d), the only variable affecting the coefficients through ReLU function in Equation 9 that results from the replacements of activation function according to nonlinear relation between moisture transport height/intensity and precipitation (Xiao, Zhang, Chen, & Li, 2024). There are less than 17% of anomalies of X's in ERPT_SEG ranging from -0.1 to 0.2 , with roughly two thirds of X's in positive anomalies. Moreover, there are less than 2% of interannual mean coefficients in anomalies of ERPT_SEGW and less than 1% deviation from the interannual mean FIs of ERA5 and TRCf, but over 2% deviation from the mean FIs of GPM and DEM (Figure 5e) and also larger temporal variations (see Figure S5 in Supporting Information S1), implying that the sensitivity of attention distribution to GPM and DEM to Henan extreme rainfall partly restores when the objective function is modified to WMSE, which is likely why the TS_{25} of the TCRP restore in ERPT_SEGW (Table 3). Roughly 80% of X's over 256 channels are larger than interannual mean, with anomalies less than 20% of interannual mean (Figure 5f). The overall results show that replacing the activation of TRC-related features with atmospheric-informed Gaussian and sigmoid functions and modifying the objective function to WMSE can result in more event-based anomalies from interannual mean in coefficient, FIs and X's.

It is essential to further illustrate the input meteorological variables during Typhoon “In-fa” induced remote rainfall. The circulation pattern on 20 July 2021 at 1800 UTC and the moisture transport characteristics is shown in Figure 6. The geopotential height at 500 hPa is relatively high in the northeastern part due to the presence of Typhoon “In-fa” over western North Pacific (see Figure 6a). The region of TRC–“In-fa” highly overlaps to the heavy rainfall over Henan Province, which further validates the accuracy of TRC objective identification algorithm in the previous study (Xiao, Zhang, Chen, Chen, et al., 2024). The moisture transport height H_{50} is generally high over the subtropical high and the southwestern plateau, whereas low over the eastern coastal region, especially the moisture transport path from In-fa to the TRC–“In-fa.” (see Figure 6b) The H_{50} over TRC–“In-fa” is 1.9–3.0 km, showing the northwestern high and the southeastern low, which highly coincides with the local topography. The strongest transport intensity is in the region of TRC–“In-fa” apart from the region of In-fa as in the southeastern ocean. The overall circulation patterns and moisture transport characteristics have validated the importance of TRC-related features (H_{50} , IVT, and TRC masks) as input variables in the deep learning models given extremely limited cases of TCRP events even in 20-year timescale.

4. Conclusions

This study enhances the forecast of the TCRP by improving the VGG-like model, incorporating additional input data types, adding SE-Block, adjusting activation functions based on meteorological knowledge, and modifying the objective function. By using explainable methods like feature importance analysis, the model's transparency increases to analyze the mechanisms of higher TCRP predictability using distribution allocation to multi-modal input data, inspired by random forest and deep learning studies (Salih & Wang, 2024; L. Zhang et al., 2024). The

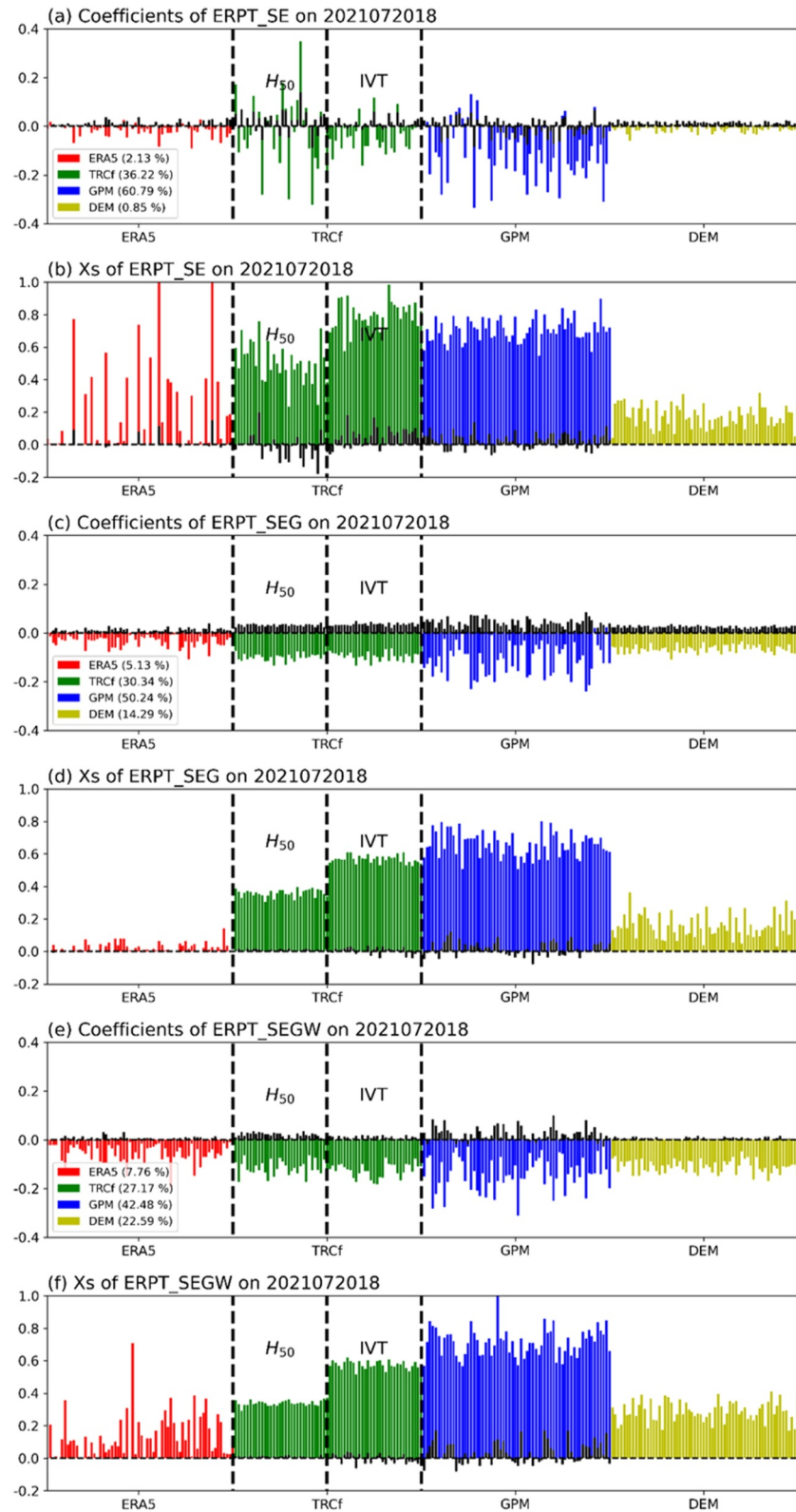


Figure 5. Same as Figure 3 but on 20 July 2021 at 1800 UTC. Black bars show the deviation from to interannual mean in JAS during 2021–2023. The black bars in (a, c and e) are magnified by the factor of 10.

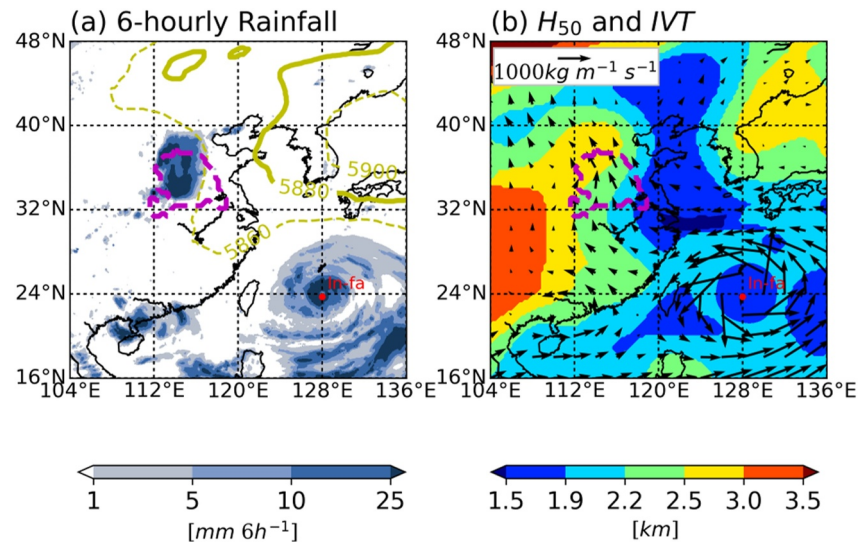


Figure 6. Horizontal distributions of (a) 6-hourly rainfall overlapped with 500 hPa geopotential height and (b) H_{50} (medium moisture transport height) overlapped with moisture transport intensity (black arrows) at 1800 UTC on 20 July 2021. The magenta dashed contour represents the boundary of TRC-“In-fa” and the red dot represents the track of Typhoon “In-fa.”

results show significant improvements in precipitation prediction accuracy and stability, particularly highlighted by TS metrics such as TS_1 and TS_5 , demonstrating strong performance at low and medium precipitation levels. The following are the main conclusions in this study:

- In general, the TCRP performance of all this study's deep learning models are better than that of TIGGE_EC. The model ERPT is the best at extreme TCRP forecast among VGG-like models in this study, implying the positive effects of adding multi-source data, especially TRCf variables with TRC masks. Adding SE-Block can enhance middle-range TCRP forecast, but the performance of the ERPT_SE decays. ERPT_SEGW achieves the smallest MSE, 9.25% smaller than PhyDNet, whereas also achieving near-best performance in extreme TCRP among VGG-like models in this study. Despite shorter time coverage, coarser resolution and fewer frames of the ERPT_SE than PhyDNet, the MSE and the TS of TCRP over middle (10 and 5 mm) and lower (1 mm) thresholds still show superior performance to that of PhyDNet.
- Interannual mean TCRP is overestimated for all VGG-like models in this study, which is the opposite of TIGGE_EC. Modifying objective function to WMSE (ERPT_SEGW) can improve the TCRP performance. Explainability analysis shows that replacing atmospheric-informed functions can make FIs of ERA5 and DEM larger (for ERPT_SEG) without significantly paying less attention to TRCf, which is the core mechanism to improve TCRP forecast of VGG-like models.
- The TS of VGG-like models during Henan heavy rainfall in July 2021 is significantly higher than that of TIGGE_EC, especially the model ERPT_SEGW that is 0.44 higher than that of VGG_E and TIGGE_EC, implying positive overall effects of TCRP improvement strategies. This TS score of ERPT_SEGW is only 0.02 smaller than that of PhyDNet. Precipitation distributions generated from PhyDNet show detailed features. This Explainability analysis in July 2021 show that there are larger temporal variations of FIs, co-efficients and intermediate outputs in GPM and DEM of the model ERPT_SEGW. Those analyses above and TRC-related feature analysis show that GPM, DEM and TRCf are important for forecasting the TCRP.

Although these achievements are notable, limitations of VGG-like models still remain. The detailed precipitation features in GPM-IMERG “ground truth” are grasped as overly smoothed tails of large-scale features in our forecasting results, which prevents the further decrease of MSE. TRC predictions for over two timesteps (lead time of at least 12 hr) need to be developed for precipitation forecast for over two timesteps, so that the effects of the attention modules are validated more adequately.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

We appreciate the ECMWF for providing ERA5 reanalysis data available at Hersbach et al. (2023), the CMA for providing TC best-track data available at Ying et al. (2014), the NASA and JAXA for providing the GPM-IMERG precipitation observational data available at Huffman et al. (2023), and the NCAR NCEI for providing the ETOPO topographical data (National Geophysical Data Center, 2001). We also appreciate authors of PhyDNet article for providing open-source scripts available at Guen and Thome (2020).

Acknowledgments

This study is funded by the Guangdong Major Project of Basic and Applied Basic Research (Grant 2020B0301030004), the National Natural Science Foundation of China (Grants 42475085), the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (Grant SML2024SP013), and the Open Project of China Meteorological Administration Basin Heavy Rainfall key Laboratory (2023BHR-Z04).

References

- Ayzel, G., Scheffer, T., & Heistermann, M. (2020). RainNet v1.0: A convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, 13(6), 2631–2644. <https://doi.org/10.5194/gmd-13-2631-2020>
- Barlow, M., Gutowski, W. J., Gyakum, J. R., Katz, R. W., Lim, Y.-K., Schumacher, R. S., et al. (2019). North American extreme precipitation events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, 53(11), 6835–6875. <https://doi.org/10.1007/s00382-019-04958-z>
- Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567), 47–55. <https://doi.org/10.1038/nature14956>
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970), 533–538. <https://doi.org/10.1038/s41586-023-06185-3>
- Briden, M., & Norouzi, N. (2021). WaveFusion Squeeze-and-Excitation: Towards an accurate and explainable deep learning framework in neuroscience. *Annu Int Conf IEEE Eng Med Biol Soc*, 2021, 1092–1095. <https://doi.org/10.1109/embc46164.2021.9630605>
- Buizza, R., Balmaseda, M. A., Brown, A., English, S., Forbes, R., Geer, A., et al. (2018). *The development and evaluation process followed at ECMWF to upgrade the Integrated Forecasting System (IFS)*. European Centre for Medium Range Weather Forecasts.
- Cao, Y., Zhang, D., Zheng, X., Shan, H., & Zhang, J. (2023). Mutual information based reweighting for precipitation nowcasting. In *Paper presented at the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. <https://doi.org/10.1109/ICASSP49357.2023.10095169>
- Chen, G., & Wang, W. C. (2022). Short-term precipitation prediction for contiguous United States using deep learning. *Geophysical Research Letters*, 49(8), e2022GL097904. <https://doi.org/10.1029/2022GL097904>
- Chen, L., Li, Y., & Cheng, Z. (2010). An overview of research and forecasting on rainfall associated with landfalling tropical cyclones. *Advances in Atmospheric Sciences*, 27(5), 967–976. <https://doi.org/10.1029/2022GL097904>
- Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). FuXi: A Cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, 6(1), 190. <https://doi.org/10.1038/s41612-023-00512-1>
- Chen, Y., Chen, G., Cui, C., Zhang, A., Wan, R., Zhou, S., et al. (2020). Retrieval of the vertical evolution of the cloud effective radius from the Chinese FY-4 (Feng Yun 4) next-generation geostationary satellites. *Atmospheric Chemistry and Physics*, 20(2), 1131–1145. <https://doi.org/10.5194/acp-20-1131-2020>
- Cheng, T. F., Dong, Q., Dai, L., & Lu, M. (2022). A dual regime of mesoscale convective systems in the East Asian monsoon annual cycle. *Journal of Geophysical Research: Atmospheres*, 127(13), e2022JD036523. <https://doi.org/10.1029/2022jd036523>
- Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world’s dry and wet regions. *Nature Climate Change*, 6(5), 508–513. <https://doi.org/10.1038/nclimate2941>
- Douris, J., Alexeeva, V., Shaw, B., & Ikeda, R. (2022). *WMO Atlas of mortality and economic losses from weather, climate, and water extremes (1970–2021)*. World Meteorological Organization Geneva.
- Ershkov, S. V., Prosviryakov, E. Y., Burmasheva, N. V., & Christianto, V. (2021). Towards understanding the algorithms for solving the Navier–Stokes equations. *Fluid Dynamics Research*, 53(4), 044501. <https://doi.org/10.1088/1873-7005/ac10f0>
- Espoholt, L., Agrawal, S., Sonderby, C., Kumar, M., Heek, J., Bromberg, C., et al. (2022). Deep learning for twelve hour precipitation forecasts. *Nature Communications*, 13(1), 5145. <https://doi.org/10.1038/s41467-022-32483-x>
- Gao, K., Harris, L., Zhou, L., Bender, M., & Morin, M. (2021). On the sensitivity of hurricane intensity and structure to horizontal tracer advection schemes in FV3. *Journal of the Atmospheric Sciences*, 78(9), 3007–3021. <https://doi.org/10.1175/jas-d-20-0331.1>
- Gimeno-Sotelo, L., & Gimeno, L. (2023). Where does the link between atmospheric moisture transport and extreme precipitation matter? *Weather and Climate Extremes*, 39, 100536. <https://doi.org/10.1016/j.wace.2022.100536>
- Guen, V. L., & Thome, N. (2020). Disentangling physical dynamics from unknown factors for unsupervised video prediction [Software]. In *Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Github. Retrieved from <https://github.com/vincent-leguen/PhyDNet>. Accessed on 21-06-2025.
- Han, L., Liang, H., Chen, H., Zhang, W., & Ge, Y. (2022). Convective precipitation nowcasting using U-Net model. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–8. <https://doi.org/10.1109/tgrs.2021.3100847>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). ERA5 hourly data on pressure levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.bd0915c6>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, 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>
- Houze, R. A. (2014a). Cloud microphysics. In *Cloud dynamics* (pp. 47–76).
- Houze, R. A. (2014b). Mesoscale convective systems. In *Cloud dynamics* (pp. 237–286).
- Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Paper presented at the proceedings of the IEEE conference on computer vision and pattern recognition*. <https://doi.org/10.1109/iccvt60459.2023.10441407>
- Huang, F., Jiang, S., Li, L., Zhang, Y., Zhang, Y., Zhang, R., et al. (2024). Applications of explainable artificial intelligence. *Earth system science*. <https://doi.org/10.48550/arXiv.2406.11882>
- Huang, F., Zhang, Y., Zhang, Y., Nourani, V., Li, Q., Li, L., & Shanguan, W. (2023). Towards interpreting machine learning models for predicting soil moisture droughts. *Environmental Research Letters*, 18(7), 074002. <https://doi.org/10.1088/1748-9326/acdbe0>
- Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2023). GPM IMERG late precipitation L3 half hourly 0.1 degree x 0.1 degree V07, Greenbelt, MD [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC)*. <https://doi.org/10.5067/GPM/IMERG/3B-HH/07>. Accessed on 08-06-2024.
- Hussain, J. (2019). Deep learning Black box problem. <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-393479>

- Kim, W., Jeong, C.-H., & Kim, S. (2024). Improvements in deep learning-based precipitation nowcasting using major atmospheric factors with radar rain rate. *Computers & Geosciences*, 184, 105529. <https://doi.org/10.1016/j.cageo.2024.105529>
- Konapala, G., Mishra, A. K., Wada, Y., & Mann, M. E. (2020). Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications*, 11(1), 3044. <https://doi.org/10.1038/s41467-020-16757-w>
- Koutsoyiannis, D. (2012). Clausius–Clapeyron equation and saturation vapour pressure: Simple theory reconciled with practice. *European Journal of Physics*, 33(2), 295–305. <https://doi.org/10.1088/0143-0807/33/2/295>
- Larios, A., & Victor, C. (2024). Continuous data assimilation for the 3D and higher-dimensional Navier–Stokes equations with higher-order fractional diffusion. *Journal of Mathematical Analysis and Applications*, 540(1), 128644. <https://doi.org/10.1016/j.jmaa.2024.128644>
- Lin, K.-C., Chen, W.-T., Chang, P.-L., Ye, Z.-Y., & Tsai, C.-C. (2024). Enhancing the rainfall forecasting accuracy of ensemble numerical prediction systems via convolutional neural networks. *Artificial Intelligence for the Earth Systems*, 3(4). <https://doi.org/10.1175/aies-d-23-0105.1>
- Liu, T., Chen, Y., Chen, S., Li, W., & Zhang, A. (2023). Mechanisms of the transport height of water vapor by tropical cyclones on heavy rainfall. *Weather and Climate Extremes*, 41, 100587. <https://doi.org/10.1016/j.wace.2023.100587>
- Lu, M., & Lall, U. (2017). Tropical moisture exports, extreme precipitation and floods in Northeastern US. *Earth Science Research*, 6(2), 91. <https://doi.org/10.5539/esr.v6n2p91>
- Luo, Y., Gao, Y., & Chen, Y. (2023). New insights on the convective and microphysical characteristics of heavy rainfall in monsoon coastal areas (South China). *Mausam*, 74(2), 361–374. <https://doi.org/10.54302/mausam.v74i2.5981>
- McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, 100(11), 2175–2199. <https://doi.org/10.1175/bams-d-18-0195.1>
- National Geophysical Data Center/NESDIS/NOAA/U.S. Department of Commerce. (2001). ETOPO2, global 2 Arc-minute ocean depth and land elevation from the US National Geophysical Data Center (NGDC) [Dataset]. *Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory*. <https://doi.org/10.5065/D6668B75>
- Pan, M., & Lu, M. (2019). A novel atmospheric river identification algorithm. *Water Resources Research*, 55(7), 6069–6087. <https://doi.org/10.1029/2018wr024407>
- Pan, M., & Lu, M. (2020). East Asia atmospheric river catalog: Annual cycle, transition mechanism, and precipitation. *Geophysical Research Letters*, 47(15), e2020GL089477. <https://doi.org/10.1029/2020gl089477>
- Pan, X., Lu, Y., Zhao, K., Huang, H., Wang, M., & Chen, H. (2021). Improving nowcasting of convective development by incorporating polarimetric Radar variables into a deep-learning model. *Geophysical Research Letters*, 48(21), e2021GL095302. <https://doi.org/10.1029/2021gl095302>
- Peng, X., Li, Q., Chen, L., Ning, X., Chu, H., & Liu, J. (2023). A structured graph neural network for improving the numerical weather prediction of rainfall. *Journal of Geophysical Research: Atmospheres*, 128(22), e2023JD039011. <https://doi.org/10.1029/2023jd039011>
- Pradhan, R. K., Markonis, Y., Vargas Godoy, M. R., Villalba-Pradas, A., Andreadis, K. M., Nikolopoulos, E. I., et al. (2022). Review of GPM IMERG performance: A global perspective. *Remote Sensing of Environment*, 268, 112754. <https://doi.org/10.1016/j.rse.2021.112754>
- Rao, C., Chen, G., & Ran, L. (2025). An extreme predecessor rain event in central China amplified by upper-level jet streak. *Monthly Weather Review*, 153(1), 3–22. <https://doi.org/10.1175/mwr-d-24-0024.1>
- Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878), 672–677. <https://doi.org/10.1038/s41586-021-03854-z>
- Salih, A. M., & Wang, Y. (2024). Are linear regression models white box and interpretable? <https://doi.org/10.48550/arXiv.2407.12177>
- Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., et al. (2021). Can deep learning beat numerical weather prediction? *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 379(2194), 20200097. <https://doi.org/10.1098/rsta.2020.0097>
- Sellers, S. L., Kawzenuk, B., Nguyen, P., Ralph, F. M., & Sorooshian, S. (2017). Genesis, pathways, and terminations of intense global water vapor transport in association with large-scale climate patterns. *Geophysical Research Letters*, 44(24). <https://doi.org/10.1002/2017gl075495>
- Shakin, M. M. U. S., Akter, F., Hasan, S. M. M., Hossain, M. R., Srizon, A. Y., Islam, R., & Akter, S. S. (2023). Squeeze and excitation attention meets modified EfficientNetB0 architecture: Multi-class brain tumor classification using explainable Artificial Intelligence. In *Paper presented at the 2023 26th International Conference on Computer and Information Technology (ICCIT)*. <https://doi.org/10.1109/iccit60459.2023.10441407>
- Simonyan, K., & Zisserman, A. (2014). Very Deep convolutional networks for large-scale image recognition. <https://doi.org/10.48550/arXiv.1409.1556>
- Sun, Y., Zhong, Z., Li, T., Yi, L., Hu, Y., Wan, H., et al. (2017). Impact of Ocean warming on tropical cyclone size and its destructiveness. *Scientific Reports*, 7(1), 8154. <https://doi.org/10.1038/s41598-017-08533-6>
- Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., et al. (2016). The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, 97(1), 49–67. <https://doi.org/10.1175/bams-d-13-00191.1>
- Wang, Y., Wu, H., Zhang, J., Gao, Z., Wang, J., Yu, P. S., & Long, M. (2023). PredRNN: A recurrent neural network for spatiotemporal predictive learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2), 2208–2225. <https://doi.org/10.1109/tpami.2022.3165153>
- Wu, Z., Zhang, Y., Zhang, L., Zheng, H., & Huang, X. (2022). A comparison of convective and stratiform precipitation microphysics of the record-breaking typhoon In-Fa (2021). *Remote Sensing*, 14(2), 344. <https://doi.org/10.3390/rs14020344>
- Xiao, S., Zhang, A., Chen, Y., & Li, W. (2024b). Objective identification of tropical cyclone-induced remote moisture transport using digraphs. *Quarterly Journal of the Royal Meteorological Society*, 150(758), 559–575. <https://doi.org/10.1002/qj.4612>
- Xiao, S., Zhang, A., Chen, Y., Chen, S., & Li, W. (2024a). How does tropical cyclone-induced remote moisture transport affect precipitation over east Asia. *Geophysical Research Letters*, 51(21), e2024GL110285. <https://doi.org/10.1029/2024gl110285>
- Xu, H., Duan, Y., & Xu, X. (2022). Indirect effects of binary typhoons on an extreme rainfall event in Henan Province, China from 19 to 21 July 2021: 1. Ensemble-based analysis. *Journal of Geophysical Research: Atmospheres*, 127(10), e2021JD036265. <https://doi.org/10.1029/2021jd036265>
- Yang, R., Hu, J., Li, Z., Mu, J., Yu, T., Xia, J., et al. (2024). Interpretable machine learning for weather and climate prediction: A review. *Atmospheric Environment*, 338, 120797. <https://doi.org/10.1016/j.atmosenv.2024.120797>
- Yang, S., & Yuan, H. (2023). A customized multi-scale deep learning framework for storm nowcasting. *Geophysical Research Letters*, 50(13), e2023GL103979. <https://doi.org/10.1029/2023gl103979>
- Yin, L., Ping, F., Mao, J., & Jin, S. (2022). Analysis on precipitation efficiency of the “21.7” Henan extremely heavy rainfall event. *Advances in Atmospheric Sciences*, 40(3), 374–392. <https://doi.org/10.1007/s00376-022-2054-x>

- Ying, M., Zhang, W., Yu, H., Lu, X., Feng, J., Fan, Y., et al. (2014). An overview of the China meteorological administration tropical cyclone database [Dataset]. *Journal of Atmospheric and Oceanic Technology*, 31(2), 287–301. <https://doi.org/10.1175/jtech-d-12-00119.1>
- Zängl, G. (2023). Adaptive tuning of uncertain parameters in a numerical weather prediction model based upon data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 149(756), 2861–2880. <https://doi.org/10.1002/qj.4535>
- Zhang, A., Chen, Y., Pan, X., Hu, Y., Chen, S., & Li, W. (2022). Precipitation microphysics of tropical cyclones over Northeast China in 2020. *Remote Sensing*, 14(9), 2188. <https://doi.org/10.3390/rs14092188>
- Zhang, L., Cheng, T. F., Lu, M., Xiong, R., & Gan, J. (2023a). Tropical cyclone stalling shifts northward and brings increasing flood risks to East Asian Coast. *Geophysical Research Letters*, 50(10), e2022GL102509. <https://doi.org/10.1029/2022gl102509>
- Zhang, L., Zhao, Y., Cen, Y., & Lu, M. (2024). Deep learning-based precipitation simulation for tropical cyclones, mesoscale convective systems, and atmospheric Rivers in east Asia. *Journal of Geophysical Research: Atmospheres*, 129(20), e2024JD041914. <https://doi.org/10.1029/2024jd041914>
- Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., & Wang, J. (2023b). Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619(7970), 526–532. <https://doi.org/10.1038/s41586-023-06184-4>
- Zhou, W., Li, J., Yan, Z., Shen, Z., Wu, B., Wang, B., et al. (2024). Progress and future prospects of decadal prediction and data assimilation: A review. *Atmospheric and Oceanic Science Letters*, 17(1), 100441. <https://doi.org/10.1016/j.aosl.2023.100441>