Article ID: 1006-8775(2024)03-0337-14

Advances in Deep-Learning-based Precipitation Nowcasting Techniques

ZHENG Qun (郑 群), LIU Qi (刘 奇), LAO Ping (劳 坪), LU Zhen-ci (卢振赐)
(School of Earth and Space Sciences/CMA-USTC Laboratory of Fengyun Remote Sensing, University of Science and Technology of China, Hefei 230026 China)

Abstract: Precipitation nowcasting, as a crucial component of weather forecasting, focuses on predicting very short-range precipitation, typically within six hours. This approach relies heavily on real-time observations rather than numerical weather models. The core concept involves the spatio-temporal extrapolation of current precipitation fields derived from ground radar echoes and/or satellite images, which was generally actualized by employing computer image or vision techniques. Recently, with stirring breakthroughs in artificial intelligence (AI) techniques, deep learning (DL) methods have been used as the basis for developing novel approaches to precipitation nowcasting. Notable progress has been obtained in recent years, manifesting the strong potential of DL-based nowcasting models for their advantages in both prediction accuracy and computational cost. This paper provides an overview of these precipitation nowcasting approaches, from which two stages along the advancing in this field emerge. Classic models that were established on an elementary neural network dominated in the first stage, while large meteorological models that were based on complex network architectures prevailed in the second. In particular, the nowcasting accuracy of such data-driven models has been greatly increased by imposing suitable physical constraints. The integration of AI models and physical models seems to be a promising way to improve precipitation nowcasting techniques further.

Key words: precipitation nowcasting; deep learning; neural network; classic model; large model
CLC number: P456.1 **Document code:** A
Citation: ZHENG Qun, LIU Qi, LAO Ping, et al. Advances in Deep-Learning-based Precipitation Nowcasting Techniques [J]. Journal of Tropical Meteorology, 2024, 30(3): 337–350, https://doi.org/10.3724/j.1006-8775.2024.028

1 INTRODUCTION

Precipitation nowcasting, dealing especially with rainfall prediction in a very short range (spanning just a few hours), is an important branch of weather forecasting (Burton et al. [1]). It has gained extensive attention in recent years, mainly because of the increasing requirements for accurate prediction of heavy precipitation events that suddenly occur and are likely to cause hydrological disasters such as flooding and waterlogging (Douglas et al. [2]). The two catastrophic rainstorms that occurred in Zhengzhou in July 2021 and in Beijing in July 2023 serve as stark reminders of the importance of this field: they brought record-breaking rainfall, resulting in enormous economic losses and serious casualties. Timely and accurate prediction of these sudden events is crucial for early warning of risks, which plays a pivotal role in government planning for disaster prevention and mitigation (Pulkkinen et al. [3]; Wilson et al. [4]; Sun et al. [5]).

Physics-based numerical weather prediction (NWP) and observation-based extrapolation are two prevalent technical solutions for precipitation nowcasting (Pulkkinen et al. [3]). The radar extrapolation method has demonstrated considerable effectiveness in most cases thanks to the direct information acquired about the rapidly changing precipitation fields (Golding [6]; Cuo et al. [7]), while the NWP method has achieved comparable performances by employing data assimilation, ensemble forecasting, and post-processing corrections (Browning and Collier [8]; Chen et al. [9]; Lin et al. [10]; Golding [6]). Nevertheless, the capability of NWP declines dramatically when forecasting precipitation within one to two hours, especially for quickly developing thunderstorms, which are the main targets of precipitation nowcasting. Moreover, the tremendous computational burdens restrict to some extent the application of the NWP method in precipitation nowcasting (Lin et al. [10]; Kotsuki et al. [11]; Wilson et al. [12]).

Conventionally, radar extrapolation methods are divided into two categories: object-based and pixel-based (Zahraei et al. [13]). The object-based methods consider storm events as independent objects. They are effective in detecting and tracking individual objects of specific thunderstorms (Vila et al. [14]). In contrast, pixel-based methods excel in predicting precipitation in both convective storms and stratocumulus clouds by using full motion fields (Berenguer et al. [15]; Bechini and Chandrasekar [16]). Pixel-based methods have a wider range of applications since they support high spatial resolution. Among them, the most commonly used is the optical flow algorithm, which incorporates computer vision techniques to extrapolate radar images with high flexibility and accuracy. The underlying assumption of

Received 2023-11-20; **Revised** 2024-05-15; **Accepted** 2024-08-15
Funding: National Natural Science Foundation of China (42075075); National Key R&D Program of China (2023YFC3007700); Pre-Research Fund of USTC (YZ2082300006)
Biography: ZHENG Qun, MSc, primarily undertaking research on deep-learning-based precipitation forecast.
Corresponding author: LIU Qi, e-mail: qliu7@ustc.edu.cn

optical flow algorithms is that advections, rather than the complex interactions between dynamics and microphysics, are the most important factor in dominating the development of storms (Bechini and Chandrasekar [16]). Accordingly, only simple-pattern motions can be well perceived by these traditional extrapolation methods. These optical flow methods primarily focus on local changes between consecutive frames to estimate the velocity fields and they usually assume that the motion keeps uniform and continuous (Wang et al. [17]). It is, in principle, difficult for these methods to accurately capture non-linear motions and complex patterns of rapidly changing storms (Mao and Sorteberg [18]; Yu et al. [19]).

In the past decade, artificial intelligence (AI) techniques have flourished thanks to the notable progress in deep learning models, which have also shown significant potential in circumventing various dilemmas in the meteorological field, including the extrapolation problem in precipitation nowcasting (Gagne et al. [20]; Zhang et al. [21]). It has been widely proven that deep learning models have much higher skills in implementing extrapolation of radar images compared with traditional advection-based extrapolations (Inoue and Misumi [22]). Deep learning models are able to extract knowledge directly from a large amount of meteorological data, enabling them to capture comprehensive features of the spatio-temporal variations of rainstorms, including the location and magnitude of precipitation events, as well as their movement (Espeholt et al. [23]). These models can fully utilize frequently refreshed data from ground radar observation, satellite remote sensing, and other facilities (Leinonen et al. [24]). They exhibit evident advantages for nonlinear phenomena such as convective initiation and short-lived heavy precipitation, which are difficult to capture by using traditional methods, either the NWP modeling or the optical-flow extrapolation (Harris et al. [25]). Moreover, in contrast to the enormous computing power demanded by traditional methods, a well-trained deep learning model can generate nowcasts with fine details in just a few seconds by using a single graphics processing unit (GPU) (Chen et al. [9]). Such a notable relief in hardware requirements brought additional advantages, as it largely promoted the practicality and accessibility of deep learning models in real-time forecasting applications. Currently, the application of AI techniques in meteorology is deepening, especially in the field of weather forecasting. In particular, substantial advancements have been made for further improving precipitation nowcasting (Chen and Wang [26]). All these constitute the primary motivation for this review article, which is dedicated to summarizing the important works in this inspiring field and helping witness their rapid development that is still underway.

The purpose of this paper is to present an overview of deep-learning-based precipitation nowcasting algorithms, which were intensively proposed in recent years. The outline of this paper is as follows. Section 2 concisely introduces the representative deep learning networks that are frequently used in developing precipitation nowcasting algorithms. Section 3 provides a comprehensive review of practical algorithms that are specially designed to implement precipitation nowcasting. Discussions are presented in Section 4 with an outlook on subsequent advancements, and concluding remarks are presented in Section 5.

2 DEEP LEARNING FRAMEWORKS

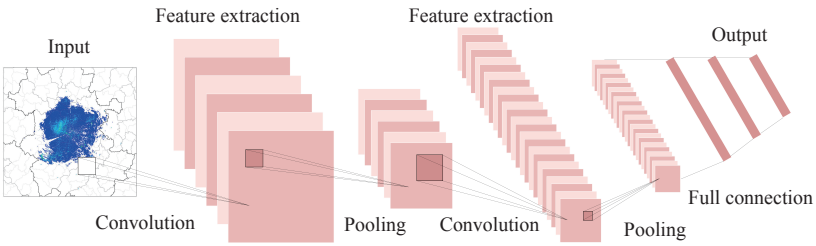
The application of deep learning in precipitation nowcasting primarily benefits from the development of spatio-temporal sequence radar echo extrapolation algorithms. These algorithms have demonstrated efficacy in various domains, including traffic flow prediction (Yao et al. [27]), video action recognition (Donahue et al. [28]), physical scene understanding (Van Steenkiste et al. [29]), and precipitation forecasting.

Currently, there are three main categories of basic neural network frameworks for precipitation nowcasting. The first category utilizes convolutional neural network (CNN) structures to generate sequential images. The second category employs recurrent neural network (RNN) to predict sequence images. The third category is the ConvRNN network which integrates the advantages of both CNN and RNN. ConvRNN-based models excel in capturing spatial correlation and temporal dynamics, enabling precise spatio-temporal sequence predictions across various applications. Furthermore, networks such as Transformer and Generative Adversarial Networks (GAN) have also been applied to precipitation nowcasting in recent years. By harnessing the capabilities of these high-performance models from the AI field, significant advancements have been achieved in precipitation nowcasting, providing more reliable and precise predictions.

2.1 *CNN*

CNN is currently a research hotspot in the fields of speech data analysis and image recognition. First proposed in the 1980s, CNN was inspired by the visual cortex of cats' brains. The shared-weight network structure of CNN makes it more akin to biological neural networks. This network structure effectively reduces the complexity of the network model and the number of parameters, particularly when processing high-dimensional images. This architecture enables direct use of images as inputs to the entire network, thereby effectively circumventing the complex feature extraction and reconstruction processes typical of traditional algorithms.

A CNN consists of a series of basic units stacked between the input and output layers (Fig. 1). Each basic unit typically includes the following layers: the convolutional layer, the pooling layer, and the activation layer. In the convolutional layer, each neuron is connected to a local receptive field that corresponds to the upper layer. Local filters are used to convolve the input data and extract features from the local receptive fields. In the pooling layer, features extracted by the convolutional layer are dimensionally reduced through operations such as



max-pooling or average-pooling. These operations not only obtain low-dimensional data but also enhance the model's ability to resist distortions. The nonlinear operations in the activation layer enhance the CNN's nonlinear fitting capability. Rather than explicitly extracting features between each layer, the model implicitly learns the feature representations of input samples through continuous training. Taking the classification task as an example, features are extracted by various layers, and the final classifier head outputs the probability distribution of the results.

Due to its effectiveness in extracting and predicting features from spatial data, CNN is extremely proficient in processing spatial data, such as remote sensing images and radar images containing precipitation information. Consequently, CNN has seen widespread adoption in precipitation nowcasting in recent years. However, a major limitation of CNN is its general inability to handle sequential data adequately. The foundational premise of the CNN model is based on the assumption of sample independence, which suggests that the relationships among samples cannot be learned. Therefore, CNN-based networks have limited capability in capturing temporal changes in sequential forecasting tasks, such as radar extrapolation, which involves complex time dynamics. Although certain techniques, such as using time frames as channel inputs, can enable CNN to process sequential data, these methods generally fail to capture complex temporal dependencies over time.

2.2 *RNN*

Numerous samples in real-world datasets exhibit correlations with each other. In the case of sequential data, the order of the sample is crucial for practical applications. While CNN-based methods have limitations in handling sequential data, RNN and its derivatives excel in this aspect, especially in capturing long-term dependencies.

An RNN primarily consists of three parts: an input layer, a hidden layer, and an output layer (Fig. 2). Specifically, the network sequentially inputs each unit of the input sequence into the RNN, obtaining the corresponding output unit for that phase and the information passed to the next phase. This process utilizes the correlations within the sequence. Thus, the information from the previous time step is remembered and affects the computation of the current output, establishing connections between the hidden layers. In theory, an RNN can handle sequential data of any length. However, in practical applications, to manage the scale of the input data, it is often assumed that the current state is only influenced by a limited number of preceding states.

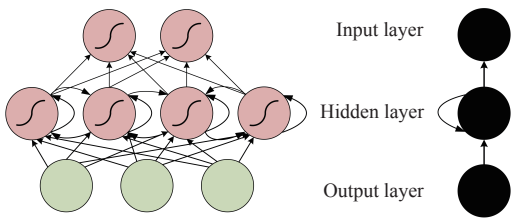


Figure 2. Architecture of a RNN. The curved arrow within the hidden layer represents self-feedback connections, allowing the RNN to exhibit dynamic temporal behavior for a time sequence. The straight arrows depict the feedforward connections between the layers, facilitating the flow of input data through the network to generate an output.

The RNN network has shown good performance in many scenarios, but it suffers from a notable limitation. Inputting a long sequence triggers deep continuous multiplications, resulting in long-term dependencies during the RNN's backpropagation process. On the one hand, this can lead to the vanishing gradient problem. On the other hand, the network's internal complexity can increase to the point of exceeding its memory capacity, resulting in chaotic and ineffective outputs. Over time, significant advancements in RNN technology have led to the development of enhanced models such as the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. These models have demonstrated superior performance in tackling the key challenges encountered by RNNs mentioned in the previous paragraphs.

2.2.1 LSTM

The intricate internal structure of LSTM networks facilitates effective management of long-sequence challenges. The classic unit structure of an LSTM is depicted in Fig. 3. As an enhancement over traditional RNN, LSTM incorporates forget gates (f_t), input gates (i_t), and output gates (o_t) to control the filtering of historical states. This architecture is designed to capture the relevant historical states that impact the current state rather than solely focusing on the most recent state.

The LSTM unit receives three inputs: x_t, a vector representing the input signal at the current time step t. The network also takes the previous output signal from the last step, denoted as vector h_{t-1}, which has the same dimension as x_t. Moreover, it receives the previous memory signal, denoted as vector c_{t-1}, which also has the same dimension as x_t. For each new input word x_t, the LSTM unit generates a new output signal h_t, representing the accumulated

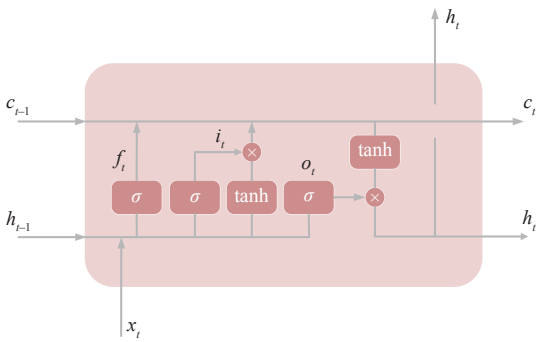


Figure 3. Unit structure of LSTM. x_t denotes the input vector; h_{t-1} denotes the previous hidden state; h_t denotes the new hidden state; c_{t-1} denotes the previous cell state; c_t denotes the new cell state; f_t denotes the forget gate; i_t denotes the input gate; o_t denotes the output gate; σ denotes the sigmoid activation function; tanh denotes the hyperbolic tangent activation function; × denotes the Hadamard product.

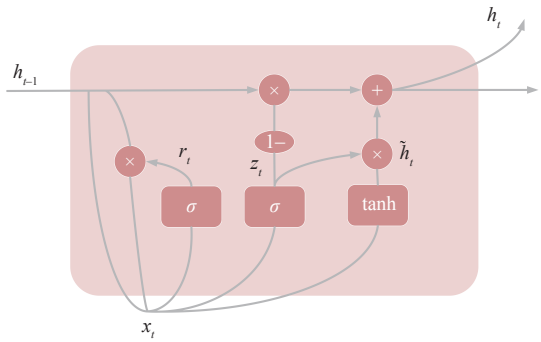


Figure 4. Unit structure of GRU. x_t denotes the input vector; h_{t-1} denotes the previous hidden state; h_t denotes the new hidden state; r_t denotes the reset gate; z_t denotes the update gate; σ denotes the sigmoid activation function; tanh denotes the hyperbolic tangent activation function; × denotes the Hadamard product; + denotes addition; 1− denotes the operation of subtraction by one.

information at that stage. Upon receiving these two signals, the LSTM does not immediately merge them but instead calculates the weights of three gates. By learning the weights of different gates, the LSTM can selectively ignore or emphasize certain inputs based on the current input signals and memory information, thus optimizing the network's ability to learn semantic information from long sequences. LSTM has gained widespread applications in the field of precipitation forecasting in recent years (Kumar et al. [30]; Kang et al. [31]; Li and Yuan [32]).

2.2.2 GRU

As an well-acknowledged variant of LSTM, GRU combines the forget gate and input gate into a single update gate. Despite having fewer parameters and a simpler structure, the GRU achieves performance comparable to that of the LSTM. A typical GRU unit is illustrated in Fig. 4. The GRU primarily comprises a reset gate and an update gate, which manage the flow of long-term information. The reset gate dictates the extent to which past information is utilized in the current phase, whereas the update gate regulates the retention or discarding of previous information in the current and subsequent phases.

The symbols z_t and r_t in the diagram represent the update gate and reset gate, respectively. The update gate controls the degree to which the state information from the previous time step is incorporated into the current state. A higher value of the update gate indicates a greater influence of the previous state information. The reset gate controls the amount of information from the previous state that is transferred to the current candidate set. A smaller reset gate implies that less information from the previous state is retained. GRU has gained widespread applications in the field of precipitation forecasting in recent years (Manokij et al. [33]; Zhang et al. [34]).

2.3 *ConvRNN*

With the increasing application of deep learning networks in precipitation nowcasting, the performance of a single network has gradually become insufficient to meet research demands. Consequently, researchers are exploring the development of integrated networks that leverage the strengths of various models.

One of the most representative networks is ConvLSTM, a type of RNN specifically designed for spatiotemporal prediction. It incorporates convolutional structures in both the input-to-state and state-to-state transitions. ConvLSTM predicts the future state of a specific grid region by considering the input from its local neighboring states and historical states. During this process, it effectively learns and models spatial and temporal representations from the inputs of both CNN and RNN. Given the promising performance of ConvLSTM, numerous variants and enhancements have been developed based on it, collectively referred to as ConvRNN methods in this context.

2.4 *Other classic models for precipitation nowcasting*

In recent years, a variety of deep-learning-based generative models have been explored to improve the accuracy and efficiency of precipitation nowcasting. These models capture and simulate data distributions in their unique ways, offering new perspectives and methods for precipitation nowcasting. The most representative among them are the Generative Adversarial Network (GAN) and Transformer.

GAN is a class of machine learning frameworks designed for unsupervised learning. Each GAN consists of two neural networks: a generator and a discriminator. The generator aims to generate new data instances similar to the training data, while the discriminator works to distinguish between actual and generated instances (Fig. 5). These two models engage in a zero-sum game, training adversarially until the discriminator model is fooled about half the time, indicating that the generator produces plausible examples.

GAN has a wide range of applications. It can learn to capture the statistical distribution of training data, enabling the synthesis of samples from this distribution. Thus, they can be used to generate new examples for image datasets, create realistic photographs, and perform image-to-image

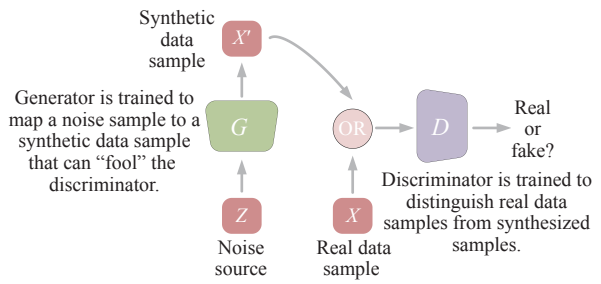


Figure 5. Unit structure of GAN. X' denotes the synthetic data sample; Z denotes the noise source; X denotes the actual data sample; D (discriminator) and G (Generator) are the two models learned during the training process for a GAN.

translation tasks. In the field of precipitation nowcasting, GAN can be used to generate radar images and enhance their details.

Transformer is a deep-learning model introduced by Google in 2017 (Vaswani et al. [35]). This model is predominantly used in the field of natural language processing and computer vision.

The overall structure of the Transformer model consists of two parts (Fig. 6): the Encoder and the Decoder. The Encoder's primary function is to process the input data and create representations that capture contextual relationships within the data. The Decoder's primary role is to generate the output sequence step-by-step. Instead of using recurrence, the model relies entirely on an attention mechanism to identify global dependencies between input and output.

Transformer has demonstrated superior generalization capabilities across a range of natural language processing tasks. This ability also enables it to handle and learn various complex weather patterns in the field of precipitation nowcasting.

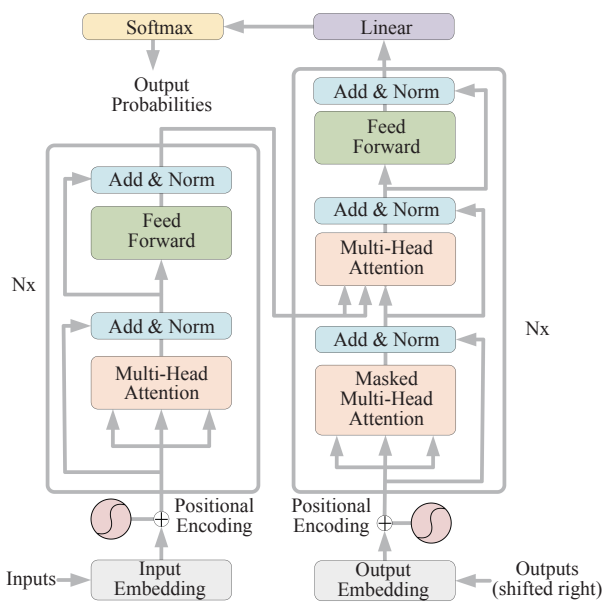


Figure 6. Unit structure of Transformer.

3 DEEP-LEARNING-BASED ALGORITHMS FOR PRECIPITATION NOWCASTING

3.1 *CNN-based algorithms*

The foundational concept of CNN dates back to the 1980s, but it gained significant popularity in 2012 after AlexNet's breakthrough achievement in the ImageNet Challenge. Since then, CNN has been widely applied across various fields, such as image recognition, video analysis, and natural language processing.

Given its ability to automatically extract spatial features from radar and remote sensing images, CNN is being increasingly explored by researchers for short-term precipitation prediction. The application of CNN was initially limited by some realistic obstacles, such as computational resources, data availability, and technology maturity. However, with the growth of computational power and the abundance of data, numerous improved CNN models have been designed to capture fine precipitation features. An important direction of improvement is to enhance data utilization to adapt to the spatiotemporally varying features of the input image. The research from Klein focuses on enhancing conventional CNN by introducing the convolutional kernel that varies with the input samples during the test period (Klein et al. [36]). He proposed a dynamic convolutional neural network (DCNN) and utilized four consecutive 2D radar reflectivity images to predict radar images for the next 10 minutes. Similarly, a dynamic convolutional neural network (DCNN-I) proposed by Shi was also used to solve the prediction problem of radar echo images (Shi et al. [37]). DCNN-I overcomes the limitations of the traditional cross-correlation methods by taking into account the strong correlation between inputs and outputs of the radar echo extrapolation problem, and also the network combines the Dynamic Sub-Network work (DSN) and the Probabilistic Prediction Layer (PPL). The DSN dynamically adjusts the convolution kernel according to different input image sequences, and its output is used as the convolution kernel of the PPL, which improves the flexibility and adaptability of the convolution operation. Likewise, a fully convolutional neural network (FCNN) for rainfall prediction is proposed by Ayzel et al. [38]. Based on radar maps, FCNN adopts four different transformations for data preprocessing and ensures as much spatial resolution as possible during the process of input data. All of these improvements effectively increase data utilization and ultimately achieve higher prediction accuracy.

On the other hand, research on fusing multiple sources of meteorological data (e.g., radar, satellites, ground-based observatories, etc.) is advancing with increasing data availability. For example, the multi-channel three-dimensional cubic successive convolution network (3D-SCN) proposed by Zhang et al. [39] utilizes multiple sources of meteorological data for convective storm initiation and growth prediction. The model takes 3D data as input, including raw 3D radar and meteorological reanalysis data. Each channel consists of

multiple layers of 2D data and is convolved using a cross-channel 3D convolution method, culminating in the use of rainfall detection as a classification problem to determine if the radar echo exceeds 35 dBZ.

It is also worth noting that a method of transformation is currently being used as a popular approach. A study by Agrawal aimed to explore the efficacy of treating precipitation nowcasting as an image-to-image translation problem (Agrawal et al. [40]). Rather than modeling the complex physics involved in the atmospheric evolution of precipitation, a time-consuming and computationally intensive practice, researchers treat this as a data-driven input/output problem. This approach involves feeding in a series of images that depict the recent history of rainfall in a specific area and predicting the rainfall status after one hour as the output. This approach is inspired by the successful application of CNN to image-to-image translation. In such tasks, the CNN learns to map an input image's pixels to some target image's pixels. For example, the target image could explicitly label salient objects in the image, denoise the input, or even just be the original image itself (in which case the CNN is referred to as an autoencoder). It's possible to model precipitation nowcasting in this way as well. Given an image measuring the instantaneous rate of precipitation, let the target training image be the image collected 1 hour after that instant. Due to its strong performance in computational efficiency and result accuracy, this transformation method has evolved into a popular approach in subsequent research.

Compared with traditional extrapolation methods, CNN can utilize extensive historical data to identify complex spatial features, including the shape, intensity, and large-scale spatial patterns of radar images. However, CNN is still not strong enough to extract comprehensive features of the entire radar sequence, such as the rotational features of radar echoes. More importantly, CNN has limited capability in extracting temporal feature changes between adjacent radar images. In this regard, RNN can offer an effective solution.

3.2 *RNN-based algorithms*

RNN has gained widespread adoption in precipitation nowcasting due to its capability to extract temporal features. It can predict future outputs by retaining information from previous inputs, making it particularly suitable for handling sequential data (Fang et al. [41]). However, the original RNN architecture is prone to issues such as gradient vanishing and exploding, especially when training and predicting long sequences. These problems arise from the compounding effect of multiplying gradients across multiple layers during backpropagation. In long sequences, this issue can cause gradients to become extremely small (vanish) or excessively large (explode), hampering the model's ability to learn effectively and thus leading to diminished performance. To address these challenges, enhanced neural network models based on RNN, such as LSTM and GRU, provide more robust solutions (Zang et al. [42]).

LSTM and GRU models, with their specialized gating mechanisms, are better equipped to capture long-term dependencies in sequential data. These enhanced RNN models have been validated in practice to give better results in prediction problems. In a study conducted by Akbari Asanjan, an LSTM model was innovatively applied to generate quantitative precipitation forecasts for the next six hours using remote sensing data. The research thoroughly compared this model's performance with that of conventional RNN and numerical simulations. The results indicate that the LSTM forecasts not only showed superior capabilities over other methods but also exhibited significant potential for global precipitation nowcasting (Akbari et al. [43]).

As the feasibility of improved RNN networks for precipitation nowcasting has been gradually demonstrated, research efforts increasingly focus on enhancing the fundamental network architecture. A spatiotemporal prediction model, termed spatiotemporal long- and short-term memory (ST-LSTM-SA), was introduced, leveraging a self-attentive mechanism to enhance its predictive capabilities. This model is capable of effectively capturing sequence features and fully leveraging short-term spatiotemporal information. Through comprehensive experiments using radar echo sequences in Wuhan, China, this model successfully predicts future radar reflectivity images for the next three hours. The study evaluates the model's forecast performance in terms of image quality and rainfall error, and supplementary experiments on minute-level datasets showcase the superiority of ST-LSTM-SA. These findings and endeavors provide valuable guidance for urban precipitation nowcasting (Liu et al. [44]). Similarly, a GRU model is developed for high-resolution quantitative precipitation nowcasting in Taiwan Island of China, predicting up to three hours ahead. By combining the discriminator and the attention techniques, Ashesh et al. [45] proposed a model that is trained with a dataset containing radar reflectivity and rain rates at 10-minute intervals, predicting hourly accumulated rainfall for the next three hours. This demonstrates that such an RNN-based QPN model holds great potential.

The aforementioned studies integrate advanced techniques such as attention mechanisms and discriminator technologies into RNN models, aiming to enhance the accuracy of precipitation nowcasting. These examples represent only a fraction of the extensive research efforts in this area. Even now, the incorporation of cutting-edge computational techniques into the fundamental architecture of RNN-based models continues to attract significant research interest.

Furthermore, several studies focus on designing solutions tailored to specific needs and scenarios. For example, the HPRNN model is specialized for long-term extrapolation of radar echoes to meet the needs of accurate precipitation prediction. This model combines a hierarchical prediction strategy with a coarse-to-fine

recursion mechanism to minimize the accumulation of prediction errors over time. This approach not only focuses on the basic structure of RNN, but also emphasizes practical problem-solving through specific designs (Jing et al. [46]). It is also noteworthy that models incorporating multiple data sources as input have also achieved highly desirable forecasting results. MetNet, for example, is a model that blends an LSTM encoder with an axial attention decoder. It processes inputs, including radar quantitative precipitation estimates, satellite data, and real-valued features such as latitude, longitude, altitude, and temporal information (hour, day, and month). MetNet provides eight-hour precipitation forecasts and can produce probabilistic precipitation forecast maps (Sønderby et al. [47]).

3.3 *ConvRNN-based algorithms*

While CNN networks have strong spatial feature extraction capabilities and RNN networks excel in temporal feature extraction, neither network type is independently sufficient for processing spatiotemporal data, such as sequential radar images. With the rapid advancement of deep learning, researchers have begun to explore the synergistic effects of combining these two models. This idea has led to the development of spatiotemporal sequence prediction models, marking a significant advancement in precipitation nowcasting.

In 2015, the ConvLSTM model was proposed by Shi (Shi et al. [48]). This hybrid framework can simultaneously establish temporal relationships, extract local spatial features, and demonstrate superior representation capabilities. ConvLSTM represents an evolutionary step in deep learning models for sequential data. By integrating the convolutional layers of CNN with the recurrent layers of LSTM, it offers a more effective approach for tasks requiring both spatial and temporal analysis. This development has consequently led to numerous research outcomes in the field of precipitation nowcasting. The study conducted in 2017 by Shi et al. [49] on the ConvLSTM model revealed a limitation regarding positional invariance within its convolutional recurrent structure. This limitation could potentially lead to a temporally constant data structure, posing challenges as most natural phenomena, including radar echo features, inherently exhibit temporal evolution.

To address this issue, the TrajGRU model was designed to actively learn the changing trajectories in radar images. Compared with ConvLSTM, the TrajGRU model has shown enhanced predictive performance in precipitation nowcasting. It is crucial to meticulously design network architectures to adequately reflect the unique characteristics of radar data. This further underscores the importance of appropriate network architecture in achieving robust performance.

Therefore, subsequent research has primarily focused on two key aspects. First, exploring the specific applications of ConvLSTM and TrajGRU networks in various scenarios. For instance, a data-driven precipitation forecasting model called “DeepRain,” which extends the application of ConvLSTM to process three-dimensional (width, height, and depth) data (Kim et al. [50]). Second, enhancing and modifying the network architectures based on ConvLSTM and TrajGRU. In this regard, the Machine Learning Group at the School of Software, Tsinghua University (THUML), with their contributions from PredRNN in 2017 (Wang et al. [51]), to Memory In Memory (MIM) in 2019 (Wang et al. [52]), and further to MotionRNN in 2021 (Wu et al. [53]), has had a significant impact on the field of precipitation nowcasting.

Initially introduced, the PredRNN model incorporates both temporal and spatial memory units. This allows for horizontal state-to-state memory transmission as well as vertical layer-to-layer memory transfer. Such a configuration effectively models shape deformations and motion trajectories, significantly enhancing the model’s ability to capture and predict radar echoes (Wang et al. [51]). However, the PredRNN model was found to have potential issues with vanishing gradients. To address this, the PredRNN++ model was proposed in 2018, featuring Gradient Highway Units that capture long-term memory dependencies and improve the model’s handling of short-term dynamics and sudden events. The newly designed memory mechanism ensures that each pixel in the final generated frame has a larger receptive field across each time step, endowing the prediction model with enhanced capabilities to handle short-term dynamics and abrupt changes. Additionally, the design and introduction of shortcuts for gradient propagation effectively mitigate the issue of vanishing gradients (Wang et al. [54]).

Subsequently, to learn higher-order non-stationary changes in the accumulation, deformation, or dissipation of radar echoes in precipitation forecasting, the THUML team developed the MIM network. Utilizing differential signals between adjacent recurrent states, MIM employs two cascading, self-updating memory modules to simulate both non-steady state and quasi-steady state characteristics in spatiotemporal dynamics. Subsequently, researchers unified motion modeling as a combination of instantaneous changes and motion trends, introducing MotionGRU units embedded into existing predictive models for enhanced complex motion modeling (Wu et al. [53]). Researchers from the THUML team have continually improved model performance by refining ConvLSTM units, incorporating self-attention mechanisms, and modifying the direction of information flow within the model. The series of progressively updated models has effectively addressed the declining accuracy observed in ConvLSTM and TrajGRU when applied to multi-frame predictions, significantly improving the accuracy in predicting the spatial and temporal movement of radar echoes.

Currently, these research achievements have successfully replaced the existing Severe Weather Automatic Nowcasting (SWAN) system, which is based on traditional optical flow extrapolation algorithms, at the National Meteorological Center of China Meteorological

Administration (CMA). These models are now employed to construct a new generation of short-term and severe weather forecasting service platforms for CMA.

3.4 *Other classic models for precipitation nowcasting*

The most representative GAN-based model in precipitation nowcasting is DGMR (Deep Generative Models of Radar), proposed by DeepMind. In 2021, DeepMind collaborated with the UK Met Office to develop DGMR, a deep learning approach focusing on the nowcasting of rainfall for the next 5 to 90 minutes. Developed based on GAN, the model utilizes two discriminators to facilitate adversarial learning across both spatial and temporal dimensions. Moreover, to enhance accuracy, an extra regularization term was introduced during its training. Through continuous adversarial cycles between the generator and discriminators, both components significantly improved, culminating in the development of the DGMR model. This model can generate detailed and reasonable predictions of future weather information based on past radar data. Compared with other deep learning models, DGMR addressed the issue of forecast ambiguity and surpassed other commonly used precipitation nowcasting methods in predicting rainfall, as well as the structure and intensity of circulation (Ravuri et al. [55]). Furthermore, in the field of precipitation nowcasting, the most common application of the GAN model is to improve the quality of prediction results, such as enabling the model to generate more realistic radar extrapolated images, improving image clarity, and enriching image details (Tian et al. [56]).

Another influential model in this field is the Transformer. Given its strong ability to address gradient vanishing and exploding issues, Transformer has outperformed RNN in many studies. As a result, in recent years, Transformer has gained significant popularity in precipitation nowcasting.

A novel motion-guided global-local aggregation Transformer network for for precipitation nowcasting was proposed by Dong et al. [57]. This model effectively combines spatiotemporal cues at different time scales, thereby strengthening global-local spatiotemporal aggregation urgently required by the extrapolation task. Finally, it achieves superior performance in nowcasting skill scores, precipitation details, and image clarity over existing methods (Dong et al. [57]). As a highly extensible model, it can also be well combined with other models to achieve stronger performance. A novel data-driven predictive model called Attention Augmented TransUNet (AA-TransUNet) was designed by Yang and Mehrkanoon [58]. The TransUNet model that combines the Transformer and UNet models is the core of this research and is further equipped with a Convolutional Block Attention Module (CBAM) and Depthwise-separable Convolutions (DSC). This model not only improves performance but also significantly reduces the number of trainable parameters in the encoder, and it has been successfully applied for precipitation nowcasting (Yang and Mehrkanoon [58]). The increase in model complexity will increase the training cost. How to control the structure of the model as simply as possible while maintaining the model modeling ability is also noticed by people in the application process of Transformer. Performer is a streamlined Transformer framework for precipitation nowcasting that efficiently captures global spatiotemporal dependencies among multiple meteorological elements. The framework implements an encoder-translator-decoder architecture, where the encoder integrates spatial features of multiple elements, the translator models spatiotemporal dynamics, and the decoder combines spatiotemporal information to forecast future precipitation. Without introducing complex structures or strategies, Performer achieves state-of-the-art performance with minimal parameters (Zheng and Liao [59]).

In fact, with the continuous progress of deep learning technology, there are many other advanced models that have rich application scenarios in the field of precipitation nowcasting. Beyond the generative models mentioned, diffusion models have shown great potential to overcome the approximation/optimization errors of current data-driven methods (Pan et al. [60]) and significantly enhance the reliability of forecast uncertainty estimates (Pan et al. [61]). Furthermore, strategies like reinforcement learning and transfer learning are applicable, and some interesting research applications have already been seen.

4 DISCUSSION

4.1 *Limitations and trends of classic models*

Neural network models like CNN and RNN have been widely used as the core for developing practical algorithms of precipitation nowcasting. CNN models have a reputation for their excellent capability in extracting spatial features, yet have obvious defects in processing time serial information. These deficiencies can be mitigated to some extent by RNN, especially through its variants, such as LSTM and GRU, which possess special aptitudes for capturing temporal features. Furthermore, the hybrid models, e.g., ConvRNN and ConvLSTM, were proposed to integrate the distinct capabilities of both CNN and RNN. In this way, the comprehensive features of spatiotemporal variations of precipitation fields can be coherently extracted. The precipitation nowcasting algorithms based on the hybrid models achieve more accurate predictions compared with those relying on a single elementary model that is concentrated on either spatial or temporal features.

However, the current hybrid model-based algorithms also have non-negligible shortcomings. Some are nearly identical to the single-model algorithms due to their succession relation, which includes notable difficulties in predicting convection initiation, underestimating heavy or extreme rainfall, and smoothing prediction with time. Other shortcomings are unique to hybrid models. For instance, because of their relatively complex architecture, the size of required training samples significantly increases. High-quality measurements or data are thus needed to facilitate the establishment of the available

sample set. Meanwhile, more computational resources are demanded due to the enormous increase in training samples. These factors collectively create additional obstacles for the practical applications of hybrid models. In such a context, the challenge is to develop lightweight models that perform comparably to their hybrid counterparts. In this way, basic networks like CNN are still vigorous, since they are very likely to be the main components for actualizing a lightweight model, and in turn developing special algorithms aiming at specific purposes in the scope of precipitation nowcasting. Recent studies have utilized pure convolutional structures to design models, and then use them to process sequential images and extract temporal evolution information. In this way, the complex RNN structure is replaced by the simple CNN structure (Prudden et al. [62]; Hu et al. [63]; Han et al. [64]). Such models have flexible structures and require fewer computational resources for optimization, and have been applied to precipitation nowcasting or other spatiotemporal forecasting tasks (Tan et al. [65]).

4.2 *Rapid growth of meteorological large models*

Note that at present, a new round of revolution is happening for AI. With the advent of the so-called large models that have been applied to so many fields in a very short time, AI technology and the associated diverse applications have seen dramatic progress and are receiving much more widespread attention than ever before. These novel AI models are distinct from the classic ones previously mentioned (CNN, RNN, LSTM, GAN, Transformer, etc.), which are exclusively featured as containing billions or even hundreds of billions of model parameters in the neural networks devised for deep learning. The network architectures of these models are undoubtedly gigantic and sophisticated, thus supporting very strong and flexible capabilities in many applications. They initially had amazing successes in relevant applications of natural language processing and were subsequently introduced into the meteorological field, particularly for improving weather forecasts as expected. Besides short- and medium-range forecasts of ordinary meteorological fields composed of pressure, temperature, moisture, velocity, etc., large models also manifested to be outstanding in meeting specific prediction requirements, such as typhoon paths, extreme weather, and severe thunderstorms (Chen et al. [66]). Real-data-based results reflected that some medium-range forecasts achieve similar or even higher accuracy compared with the most advanced NWP method (Hu et al. [67]).

Large models are now becoming prevalent across the field of weather forecasting in virtue of their excellent performance and exceedingly high efficiency for numerous scenarios in this domain (Andrychowicz et al. [68]). A dozen meteorological large models have been successively proposed by independent teams in the past two years. For example, a forecasting-target network model based on a Transformer structure with an Adaptive Fourier Neural Operator (AFNO) was released in February 2022. It was nominated as FourCastNet and was deemed as a milestone in the development of weather forecast models since its prediction accuracy of the selected meteorological variables, for the first time, became comparable to the advanced operational NWP method, such as the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF), as long as the lead time was limited within 48 hours (Pathak et al. [69]). This global-scale data-driven model has a spatial resolution of about 30 km × 30 km, and could generate a 7-day forecast in less than two seconds. It was reported to be about 45,000 times faster than the NWP method. This is highly beneficial for increasing members in the ensemble forecasts and thus helps to raise confidence about the prediction of extreme events. Compared with the prevailing NWP method, its high speed and low computational cost, as well as the capacity for generating large ensembles, are deemed to have important implications and strong potential.

Another AI-based weather forecasting model, Pangu, was released initially in November 2022 and was dedicated to high-resolution medium-range predictions for global weather variations. This model employs a three-dimensional (3D) Earth-specific Transformer (3DEST) architecture to inject Earth-specific priors into the deep networks. By formulating height into individual dimensions, such a 3D model has the ability to capture the relationship between atmospheric states at different vertical levels, thus gaining notable accuracy, compared with the common 2D model, such as FourCastNet. Also, a hierarchical temporal aggregation skill was used to train a series of models with increasing lead times. In this way, the number of iterations was reduced, and the cumulative forecast errors were alleviated. This model was heavily trained by using 39 years (1979–2017) of global data from the ERA5 reanalysis archive. For lead times ranging from 1 hour up to 7 days, it generated more accurate deterministic forecast results of the selected variables, compared with those from the operational IFS, with speeds increased by at least 10,000 times. The Pangu model has been proven to have a notable advantage in extreme weather forecasting based on a good many experiments of reanalysis data. In particular, when it was initialized by the reanalysis, its capacity for tracking tropical cyclones was also higher than that of the IFS (Bi et al. [70]).

In December 2022, GraphCast was released, dedicated to medium-range global weather forecasts. This deep-learning model was based on Graph Neural Networks (GNN) with a special configuration of encoder-processor-decoder, in which about 40 million parameters were involved. By using the autoregressive scheme similar to that in the NWP system, a very long trajectory of weather states could be generated by feeding its own predictions back in as input. This model was also trained using 39 years of ERA5 reanalysis data. It was reported that the forecasting results from GraphCast outperformed Pangu in most targets. Moreover, it is capable of

generating a global-scale 10-day forecast at 0.25° resolution in less than 60 seconds and has prominent performance for the prediction of high-impact events like tropical cyclone tracking, atmospheric rivers, and extreme temperatures (Lam et al. [71]).

Meteorological large models continued to burst out in the following year. ClimaX, the first AI model aiming at climate projection, was proposed in January 2023, and is a Transformer-based model. In order to alleviate the computational cost and enhance the model's applicability, the Transformer kernel was expanded by using novel encoding and aggregation blocks. Even with the low-resolution version that was based upon limited computational load, ClimaX demonstrates to be much better than the existing data-driven models in the benchmarks of both weather forecasting and climate prediction (Nguyen et al. [72]). Soon after, a new AI weather model nominated as Fengwu was released in April 2023, which is essentially based on multimodal and multitask deep learning networks. This model was reported to outperform GraphCast in the majority of evaluating indices concerned. Particularly, Fengwu got an even higher operation efficiency as it generated global-scale weather forecasts for the next 10 days in just 30 seconds, half the time of GraphCast (Chen et al. [73]). Then, in June 2023, another AI model, Fuxi, was proposed, which employs an efficient *U*-Transformer structure especially for improving forecast duration. This model extends the AI-based forecast up to 15 days for the first time, achieving a breakthrough in the nominal extended-range weather forecast (Chen et al. [74]).

All these meteorological large models (Fig. 7) are based on deep networks with sophisticated architectures, requiring billions of parameters to actualize the expected functions. Classic deep learning neural networks, such as CNN, GNN, GAN, Transformer, etc., are typically employed as the inner components of these so-called large models. Compared with classic models, the development of large models relies even more heavily on the available data of actual weather since they are inherently driven by big data. Collecting massive meteorological data that meet the ultra-high criteria for valid model training is actually difficult. For this reason, so far, reanalysis datasets (e.g., ERA5) have been used as the training samples, which are deemed to represent a good approximation of real meteorological situations. It is worth noting that the reanalysis of meteorological fields is, in practice, generated by the NWP method (e.g., IFS) with the prerequisite of substantial observation data. In this context, the current deep learning models have certain genes from NWP and are thus not absolutely independent. Even so, it has become rather clear that the weather forecasts generated by the AI large models are overall comparable to the NWP in terms of prediction accuracy. Meanwhile, the AI weather forecast definitely overwhelms the NWP in computational cost, which is differed by at least several orders of magnitude of computer time given a common forecast task. Undoubtedly, such a high computational efficiency is vital for the AI model to surpass the NWP and gain further applicability in the future. Nevertheless, the development of meteorological large models seems to be in its very early stages. There is considerable room for subsequent improvements along this technical path and more appropriate applications yet need to be explored.

4.3 *Enhanced AI large model with explicit physics*

As the conventional scheme used for weather forecast, the NWP is based on exact physical theories that are mostly expressed as a set of partial differential equations, and is then actualized by numerical integrations on computers, for which the NWP is generally deemed a physical model. It solves a certain problem in a deduction way via relying on basic principles rather than in an induction way by using abundant accurate data. Therefore, the accumulated historical datasets that have dramatically grown in recent decades contribute little to the foundation of NWP. On the contrary, the AI forecast is a data-driven model that is independent of the acknowledged physical principles. It is noteworthy that the problem of weather forecast differs greatly from other common scenarios of AI applications (e.g., image/video/audio processing and natural language processing) by the fact that it is a real physical problem. Physics is not explicitly contained in the

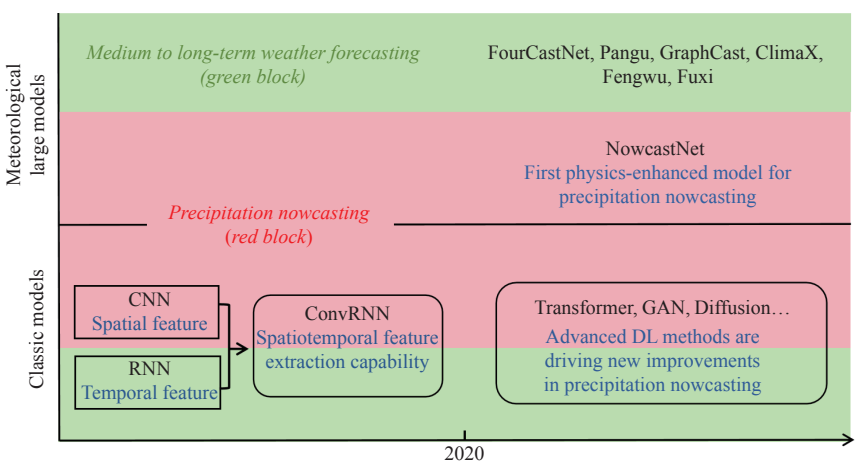


Figure 7. Schematic diagram of model evolution.

AI forecast model, but it is implicitly extracted in the training process from big data of meteorological reality. In other words, in the AI model, physics is essentially involved and transformed as a large number of features that could be learned by computer, rather than expressed as physical theories or mathematical equations that are understandable by the human brain and used naturally as the basis for the NWP, which was designed to simulate the real world in a digital way in virtue of the strong computational power. This constitutes the fundamental difference between AI models, and the NWP method. It is evident that the marked advantage of AI models, i.e., ultra-high computational efficiency, actually arises from the fact that it does not need to take into account the concrete physical processes that are computationally expensive.

However, the lack of explicit physics leads to the inexplicability of the forecasting results from AI models, which is now considered to be the principal challenge of AI-based weather forecasts. Since there are no physical constraints in AI models, fictitious results cannot be completely excluded despite the complexity of deep learning networks. This dilemma of AI models is often depicted as a "black box," as the inner mechanisms for solving a certain problem are invisible and uncertain. Given the core component that is wholly dependent on intensive computations, there is little room for artificial efforts to intentionally adjust the model. In principle, the underlying reasons for success or failure in practical forecasts remain ambiguous. In short, the absence of physics in AI models that is used for circumventing a real physical problem inevitably brings the issue of reliability, which is vital for such a powerful tool. Therefore, the incorporation of physical modules into AI models may be a promising way out, as it could benefit from both the reliability of physical principles and the high computational efficiency of AI. As suggested by de Bézenac et al. [75] and Morrison et al. [78], this, to some extent, represents the transitional zone between one end of the purely physical model and the other end of the purely data-driven model, both of which have been independently designed for simulating the real-word meteorology (De Bézenac et al. [75]; Paganini et al. [76]; Rasp et al. [77]; Morrison et al. [78]).

A new meteorological large model that integrates both physical schemes and AI schemes, nominated as NowcastNet, was released in July 2023 (Zhang et al. [79]). Note that this novel AI model was dedicated especially to the precipitation nowcasting problem (Fig. 7). It was trained using six years of ground radar observations and is capable of generating 3-hour precipitation forecasts every 10 minutes. Based on rigorous verifications, this AI model was proven to significantly outperform traditional radar-extrapolation approaches (Beauchemin and Barron [80]). It has now been deployed into the precipitation nowcasting service platform of the National Meteorological Center, China Meteorological Administration.

The conventional precipitation nowcasting approaches relying on classic neural networks often yield vague results and occasionally fail to predict severe rainstorms (Shi et al. [48]; Kim et al. [50]; Wang et al. [51]). Such defects were mainly attributed to the lack of physical constraints, and this was the primary concern in developing the NowcastNet model. In this model, the precipitation forecast problem is divided into two parts, i.e., a mesoscale forecast based on a deterministic evolution network and a convective-scale forecast driven by a stochastic generative network. The first part, as a deterministic forecast problem, was solved by taking into account two terms, convection and diffusion, which express the physics of precipitation processes in an explicit way. This strategy notably enhances the prediction ability for extreme precipitation since a conservation law as the continuity equation in the basis of NWP was imposed into the deep learning network. Therefore, NowcastNet achieves multiscale nowcasting by conditioning the data-driven generative network on the advection-based evolution network. The precipitation fields, motion fields, and intensity residuals are simultaneously learned by the neural networks based on the new differentiable neural evolution operator. A unified network framework was accordingly established, which allows end-to-end forecast error optimization. All these features contribute to the notable improvement in precipitation nowcasting with lead times up to three hours, as well as high resolution and local details. Especially for those extreme cases that are usually accompanied by both advective and convective processes, NowcastNet has been proven to be quite skillful. Note that only one physical principle was introduced into this framework, i.e., mass conservation. Significant improvements are thus expected once more physical principles could be integrated in a concordant manner to impose more complete constraints on such a deep-learning-based nowcasting system.

5 CONCLUSION

The deep-learning-based precipitation nowcasting techniques were reviewed particularly in this paper, which was motivated by the rapid progress currently occurring in this field. Numerous advancements have been attained by invoking classic deep learning models, while the advent of meteorological large models, as a milestone, brought a distinct solution for this problem and took a large step forward. Furthermore, the addition of constraints from physical principles has been shown to even strengthen the AI nowcasting model, which represents a feasible way for developing a unified meteorological model that integrates physics with AI. Such a new paradigm for circumventing the precipitation nowcasting problem, benefitting from reliable physical theories, growing meteorological observations, advancing AI algorithms, and increasing computational power, has extraordinary advantages compared with previous ones. We are fortunate to witness the beginning of a new era, and significant progress is expected in the foreseeable future.

***Acknowledgements*:** The authors deeply appreciate the valuable suggestions from the two reviewers and the support from the CMA-USTC Laboratory of Fengyun Remote Sensing.

REFERENCES

[1] BURTON R R, BLYTH A M, CUI Z, et al. Satellite-based nowcasting of West African mesoscale storms has skill at up to 4-h lead time [J]. Weather and Forecasting, 2022, 37(4): 445–455, https://doi.org/10.1175/waf-d-21-0051.1

[2] DOUGLAS I, ALAM K, MAGHENDA M, et al. Unjust waters: Climate change, flooding and the urban poor in Africa [J]. Environment and Urbanization, 2008, 20(1): 187–205, https://doi.org/10.1177/0956247808089156

[3] PULKKINEN S, NERINI D, PÉREZ HORTAL A A, et al. Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1) [J]. Geoscientific Model Development, 2019, 12(10): 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019

[4] WILSON J W, CROOK N A, MUELLER C K, et al. Nowcasting thunderstorms: A status report [J]. Bulletin of the American Meteorological Society, 1998, 79(10): 2079–2100, https://doi.org/10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2

[5] SUN J, XUE M, WILSON J W, et al. Use of NWP for nowcasting convective precipitation: Recent progress and challenges [J]. Bulletin of the American Meteorological Society, 2014, 95(3): 409–426, https://doi.org/10.1175/BAMS-D-11-00263.1

[6] GOLDING B. Nimrod: A system for generating automated very short range forecasts [J]. Meteorological Applications, 1998, 5(1): 1–16, https://doi.org/10.1017/S1350482798000577

[7] CUO L, PAGANO T C, WANG Q. A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting [J]. Journal of Hydrometeorology, 2011, 12(5): 713–728, https://doi.org/10.1175/2011JHM1347.1

[8] BROWNING K, COLLIER C. Nowcasting of precipitation systems [J]. Reviews of Geophysics, 1989, 27(3): 345–370, https://doi.org/10.1029/RG027i003p00345

[9] CHEN L, CAO Y, MA L, et al. A deep learning-based methodology for precipitation nowcasting with radar [J]. Earth and Space Science, 2020, 7(2): e2019EA000812, https://doi.org/10.1029/2019EA000812

[10] LIN C, VASIĆ S, KILAMBI A, et al. Precipitation forecast skill of numerical weather prediction models and radar nowcasts [J]. Geophysical Research Letters, 2005, 32(14): L14801, https://doi.org/10.1029/2005GL023451

[11] KOTSUKI S, KUROSAWA K, OTSUKA S, et al. Global precipitation forecasts by merging extrapolation-based nowcast and numerical weather prediction with locally optimized weights [J]. Weather and Forecasting, 2019, 34(3): 701–714, https://doi.org/10.1175/WAF-D-18-0164.1

[12] WILSON J, MEGENHARDT D, PINTO J. NWP and radar extrapolation: Comparisons and explanation of errors [J]. Monthly Weather Review, 2020, 148(12): 4783–4798, https://doi.org/10.1175/MWR-D-20-0221.1

[13] ZAHRAEI A, HSU K-L, SOROOSHIAN S, et al. Short-term quantitative precipitation forecasting using an object-based approach [J]. Journal of Hydrology, 2013, 483: 1–15, https://doi.org/10.1016/j.jhydrol.2012.09.052

[14] VILA D A, MACHADO L A T, LAURENT H, et al. Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) using satellite infrared imagery: Methodology and validation [J]. Weather and Forecasting, 2008, 23(2): 233–245, https://doi.org/10.1175/2007WAF2006121.1

[15] BERENGUER M, SEMPERE-TORRES D, PEGRAM G G. SBMcast–An ensemble nowcasting technique to assess the uncertainty in rainfall forecasts by Lagrangian extrapolation [J]. Journal of Hydrology, 2011, 404(3–4): 226–240, https://doi.org/10.1016/j.jhydrol.2011.04.033

[16] BECHINI R, CHANDRASEKAR V. An enhanced optical flow technique for radar nowcasting of precipitation and winds [J]. Journal of Atmospheric and Oceanic Technology, 2017, 34(12): 2637–2658, https://doi.org/10.1175/JTECH-D-17-0110.1

[17] WANG C, WANG P, WANG D, et al. Nowcasting multicell short-term intense precipitation using graph models and random forests [J]. Monthly Weather Review, 2020, 148(11): 4453–4466, https://doi.org/10.1175/MWR-D-20-0050.1

[18] MAO Y, SORTEBERG A. Improving radar-based precipitation nowcasts with machine learning using an approach based on random forest [J]. Weather and Forecasting, 2020, 35(6): 2461–2478, https://doi.org/10.1175/WAF-D-20-0080.1

[19] YU X, ZHOU X, WANG X. The advances in the nowcasting techniques on thunderstorms and severe convection [J]. Acta Meteorologica Sinica, 2012, 70(3): 311–337, https://doi.org/10.1007/s11783-011-0280-z

[20] GAGNE D J, MCGOVERN A, HAUPT S E, et al. Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles [J]. Weather and Forecasting, 2017, 32(5): 1819–1840, https://doi.org/10.1175/WAF-D-17-0010.1

[21] ZHANG P, ZHANG L, LEUNG H, et al. A deep-learning based precipitation forecasting approach using multiple environmental factors [C]// 2017 IEEE International Congress on Big Data. Boston: IEEE, 2017.

[22] INOUE T, MISUMI R. Learning from precipitation events in the wider domain to improve the performance of a deep learning-based precipitation nowcasting model [J]. Weather and Forecasting, 2022, 37(6): 1013–1026, https://doi.org/10.1175/WAF-D-21-0078.1

[23] ESPEHOLT L, AGRAWAL S, SØNDERBY C, et al. Deep learning for twelve hour precipitation forecasts [J]. Nature Communications, 2022, 13(1): 5145, https://doi.org/10.1038/s41467-022-32483-x

[24] LEINONEN J, HAMANN U, SIDERIS I V, et al. Thunderstorm nowcasting with deep learning: A multi-hazard data fusion model [J]. Geophysical Research Letters, 2023, 50(8): e2022GL101626, https://doi.org/10.1029/2022GL101626

[25] HARRIS L, MCRAE A T, CHANTRY M, et al. A generative deep learning approach to stochastic downscaling of precipitation forecasts [J]. Journal of Advances in Modeling Earth Systems, 2022, 14(10): e2022MS003120, https://doi.org/10.1029/2022MS003120

[26] CHEN G, WANG W C. Short-term precipitation prediction for contiguous United States using deep learning [J]. Geophysical Research Letters, 2022, 49(8): e2022GL097904, https://doi.org/10.1029/2022GL097904

[27] YAO H, TANG X, WEI H, et al. Revisiting spatial-temporal

similarity: A deep learning framework for traffic prediction [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii: Association for the Advancement of Artificial Intelligence, 2019.

[28] DONAHUE J, ANNE HENDRICKS L, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE Computer Society, 2015.

[29] VAN STEENKISTE S, CHANG M, GREFF K, et al. Relational neural expectation maximization: Unsupervised discovery of objects and their interactions [J]. arXiv: 1802.10353, 2018, https://doi.org/10.48550/arXiv.1802.10353

[30] KUMAR D, SINGH A, SAMUI P, et al. Forecasting monthly precipitation using sequential modelling [J]. Hydrological Sciences Journal, 2019, 64(6): 690–700, https://doi.org/10.1080/02626667.2019.1595624

[31] KANG J, WANG H, YUAN F, et al. Prediction of precipitation based on recurrent neural networks in Jingdezhen, Jiangxi Province, China [J]. Atmosphere, 2020, 11(3): 246, https://doi.org/10.3390/atmos11030246

[32] LI J, YUAN X. Daily streamflow forecasts based on cascade long short-term memory (LSTM) model over the Yangtze River Basin [J]. Water, 2023, 15(6): 1019, https://doi.org/10.3390/w15061019

[33] MANOKIJ F, VATEEKUL P, SARINNAPAKORN K. Cascading models of CNN and GRU with autoencoder loss for precipitation forecast in Thailand [J]. ECTI Transactions on Computer and Information Technology, 2021, 15(3): 333–346, https://doi.org/10.37936/ecti-cit.2021153.240957

[34] ZHANG X, DUAN B, HE S, et al. A new precipitation forecast method based on CEEMD-WTD-GRU [J]. Water Supply, 2022, 22(4): 4120–4132, https://doi.org/10.2166/ws.2022.037

[35] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Advances in Neural Information Processing Systems 30. Honolulu: Neural Information Processing Systems Foundation, Inc (NeurIPS), 2017.

[36] KLEIN B, WOLF L, AFEK Y. A dynamic convolutional layer for short range weather prediction [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: Institute of Electrical and Electronics Engineers, 2015.

[37] SHI E, LI Q, GU D, et al. Weather radar echo extrapolation method based on convolutional neural networks [J]. Journal of Computer Applications, 2018, 38(3): 661, https://doi.org/10.11772/j.issn.1001-9081.2017082098

[38] AYZEL G, HEISTERMANN M, SOROKIN A, et al. All convolutional neural networks for radar-based precipitation nowcasting [J]. Procedia Computer Science, 2019, 150: 186–192, https://doi.org/10.5194/gmd-13-2631-2020

[39] ZHANG W, HAN L, SUN J, et al. Application of multi-channel 3D-cube successive convolution network for convective storm nowcasting [C]// 2019 IEEE International Conference on Big Data (Big Data). Los Angeles: Institute of Electrical and Electronics Engineers, 2019.

[40] AGRAWAL S, BARRINGTON L, BROMBERG C, et al. Machine learning for precipitation nowcasting from radar images [J]. arXiv: 1912.12132, 2019, https://doi.org/10.48550/arXiv.1912.12132

[41] FANG W, CHEN Y, XUE Q. Survey on research of RNN-based spatio-temporal sequence prediction algorithms [J]. Journal on Big Data, 2021, 3(3): 97–110, https://doi.org/10.32604/jbd.2021.016993

[42] ZANG Z, BAO X, LI Y, et al. A modified RNN-based deep learning method for prediction of atmospheric visibility [J]. Remote Sensing, 2023, 15(3): 553, https://doi.org/10.3390/rs15030553

[43] AKBARI A A, YANG T, HSU K, et al. Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks [J]. Journal of Geophysical Research: Atmospheres, 2018, 123(22): 12,543–12,563, https://doi.org/10.1029/2018jd028375

[44] LIU J, XU L, CHEN N. A spatiotemporal deep learning model ST-LSTM-SA for hourly rainfall forecasting using radar echo images [J]. Journal of Hydrology, 2022, 609: 127748, https://doi.org/10.1016/j.jhydrol.2022.127748

[45] ASHESH A, CHANG C T, CHEN B F, et al. Accurate and clear quantitative precipitation nowcasting based on a deep learning model with consecutive attention and rain-map discrimination [J]. Artificial Intelligence for the Earth Systems, 2022, 1(3): e210005, https://doi.org/10.1175/AIES-D-21-0005.1

[46] JING J, LI Q, PENG X, et al. HPRNN: A hierarchical sequence prediction model for long-term weather radar echo extrapolation [C]// ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: Institute of Electrical and Electronics Engineers, 2020.

[47] SØNDERBY C K, ESPEHOLT L, HEEK J, et al. Metnet: A neural weather model for precipitation forecasting [J]. arXiv: 2003.12140, 2020, https://doi.org/10.48550/arXiv.2003.12140

[48] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [C]// Advances in Neural Information Processing Systems. Montreal: Neural Information Processing Systems, 2015.

[49] SHI X, GAO Z, LAUSEN L, et al. Deep learning for precipitation nowcasting: A benchmark and a new model [C]// Advances in Neural Information Processing Systems. California: Neural Information Processing Systems, 2017.

[50] KIM S, HONG S, JOH M, et al. Deeprain: ConvLSTM network for precipitation prediction using multichannel radar data [J]. arXiv: 1711.02316, 2017, https://doi.org/10.48550/arXiv.1711.02316

[51] WANG Y, LONG M, WANG J, et al. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal LSTMs [C]// Advances in Neural Information Processing Systems. California: Neural Information Processing Systems, 2017.

[52] WANG Y, ZHANG J, ZHU H, et al. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. California: Institute of Electrical and Electronics Engineers, 2019.

[53] WU H, YAO Z, WANG J, et al. MotionRNN: A flexible model for video prediction with spacetime-varying motions [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Montreal: Institute of Electrical and Electronics Engineers, 2021.

[54] WANG Y, GAO Z, LONG M, et al. PredRNN++: Towards

a resolution of the deep-in-time dilemma in spatiotemporal predictive learning [C]// International Conference on Machine Learning. Stockholm: Proceedings of Machine Learning Research, 2018: 5123–5132.

[55] RAVURI S, LENC K, WILLSON M, et al. Skilful precipitation nowcasting using deep generative models of radar [J]. Nature, 2021, 597(7878): 672–677, https://doi.org/10.1038/s41586-021-03854-z

[56] TIAN L, LI X, YE Y, et al. A generative adversarial gated recurrent unit model for precipitation nowcasting [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(4): 601–605, https://doi.org/10.1109/LGRS.2019.2926776

[57] DONG X, ZHAO Z, WANG Y, et al. Motion-guided global–local aggregation transformer network for precipitation nowcasting [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1–16, https://doi.org/10.1109/TGRS.2022.3217639

[58] YANG Y, MEHRKANOON S. Aa-transunet: Attention augmented transunet for nowcasting tasks [C]// 2022 International Joint Conference on Neural Networks. Padova: Institute of Electrical and Electronics Engineers, 2022.

[59] ZHENG Y, LIAO C. Transformer-based nowcasting model of severe convective weather [C]// Fifth International Conference on Geoscience and Remote Sensing Mapping (ICGRSM 2023). Lianyungang: Jiangsu Ocean University, 2024.

[60] PAN B, WANG L, ZHANG F, et al. Probabilistic diffusion model for stochastic parameterization-A case example of numerical precipitation estimation [Z]. ESS Open Archive, 2023, https://doi.org/10.22541/essoar.170158335.56592781/v1

[61] PAN B, HAI J, CHEN X, et al. Reliable precipitation nowcasting using probabilistic diffusion model [Z]. ESS Open Archive, 2023, https://doi.org/10.22541/essoar.169945499.97460779/v1

[62] PRUDDEN R, ADAMS S, KANGIN D, et al. A review of radar-based nowcasting of precipitation and applicable machine learning techniques [Z]. arXiv: 2005.04988, 2020, https://doi.org/10.48550/arXiv.2005.04988

[63] HU Y, CHEN L, WANG Z, et al. Towards a more realistic and detailed deep-learning-based radar echo extrapolation method [J]. Remote Sensing, 2021, 14(1): 24, https://doi.org/10.3390/rs14010024

[64] HAN L, LIANG H, CHEN H, et al. Convective precipitation nowcasting using U-Net Model [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1–8, https://doi.org/10.1109/TGRS.2021.3100847

[65] TAN C, GAO Z, LI S, et al. Simvp: Towards simple yet powerful spatiotemporal predictive learning [J]. arXiv: 2211.12509, 2022, https://doi.org/10.48550/arXiv.2211.12509

[66] CHEN L, DU F, HU Y, et al. SwinRDM: Integrate SwinRNN with diffusion model towards high-resolution and high-quality weather forecasting [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Washington: Association for the Advancement of Artificial Intelligence, 2023.

[67] HU Y, CHEN L, WANG Z, et al. SwinVRNN: A data-driven ensemble forecasting model via learned distribution perturbation [J]. Journal of Advances in Modeling Earth Systems, 2023, 15(2): e2022MS003211, https://doi.org/10.1029/2022MS003211

[68] ANDRYCHOWICZ M, ESPEHOLT L, LI D, et al. Deep learning for day forecasts from sparse observations [J]. arXiv: 2306.06079, 2023, https://doi.org/10.48550/arXiv.2306.06079

[69] PATHAK J, SUBRAMANIAN S, HARRINGTON P, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators [J]. arXiv: 2202.11214, 2022, https://doi.org/10.48550/arXiv.2202.11214

[70] BI K, XIE L, ZHANG H, et al. Accurate medium-range global weather forecasting with 3D neural networks [J]. Nature, 2023, 619(7970): 533–538, https://doi.org/10.1038/s41586-023-06185-3

[71] LAM R, SANCHEZ-GONZALEZ A, WILLSON M, et al. Learning skillful medium-range global weather forecasting [J]. Science, 2023, 382(6677): 1416–1421, https://doi.org/10.1126/science.adi2336

[72] NGUYEN T, BRANDSTETTER J, KAPOOR A, et al. ClimaX: A foundation model for weather and climate [J]. arXiv: 2301.10343, 2023, https://doi.org/10.48550/arXiv.2301.10343

[73] CHEN K, HAN T, GONG J, et al. FengWu: Pushing the skillful global medium-range weather forecast beyond 10 days lead [J]. arXiv: 2304.02948, 2023, https://doi.org/10.48550/arXiv.2304.02948

[74] CHEN L, ZHONG X, ZHANG F, et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast [J]. npj Climate and Atmospheric Science, 2023, 6(1): 190, https://doi.org/10.1038/s41612-023-00512-1

[75] DE BÉZENAC E, PAJOT A, GALLINARI P. Deep learning for physical processes: Incorporating prior scientific knowledge [J]. Journal of Statistical Mechanics: Theory and Experiment, 2019, 2019(12): 124009, https://doi.org/10.1088/1742-5468/ab3195

[76] PAGANINI M, DE OLIVEIRA L, NACHMAN B. CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks [J]. Physical Review D, 2018, 97(1): 014021, https://doi.org/10.1103/PhysRevD.97.014021

[77] RASP S, PRITCHARD M S, GENTINE P. Deep learning to represent subgrid processes in climate models [J]. Proceedings of the National Academy of Sciences, 2018, 115(39): 9684–9689, https://doi.org/10.1073/pnas.1810286115

[78] MORRISON H, VAN LIER-WALQUI M, FRIDLIND A M, et al. Confronting the challenge of modeling cloud and precipitation microphysics [J]. Journal of Advances in Modeling Earth Systems, 2020, 12(8): e2019MS001689, https://doi.org/10.1029/2019MS001689

[79] ZHANG Y, LONG M, CHEN K, et al. Skilful nowcasting of extreme precipitation with NowcastNet [J]. Nature, 2023, 619(7970): 526–532, https://doi.org/10.1038/s41586-023-06184-4

[80] BEAUCHEMIN S S, BARRON J L. The computation of optical flow [J]. ACM Computing Surveys (CSUR), 1995, 27(3): 433–466, https://doi.org/10.1145/212094.212141

Citation: ZHENG Qun, LIU Qi, LAO Ping, et al. Advances in Deep-Learning-based Precipitation Nowcasting Techniques [J]. Journal of Tropical Meteorology, 2024, 30(3): 337–350, https://doi.org/10.3724/j.1006-8775.2024.028