

# Improved Precipitation Nowcasting Through a Deep Learning Model Based on Three-Dimensional Cloud Structures

Yuankang Ye<sup>id</sup>, Feng Gao, Wei Cheng<sup>id</sup>, Chang Liu<sup>id</sup>, Shaoqing Zhang, and Shudong Wang<sup>id</sup>

**Abstract**—Precipitation nowcasting pertains to the localized forecasting of rainfall over a brief time horizon, characterized by precise estimates of both coverage and intensity. This capability holds particular significance in various societal applications, including agriculture, aviation safety, and transportation. However, since traditional methods mainly by extrapolating radar echo in 2-D space, cannot accurately and sufficiently represent the spatiotemporal state of clouds in the vertical direction, the accuracy of precipitation nowcasting using weather radar has reached a bottleneck. A new deep learning precipitation nowcasting model called 3dCloudNet is designed and evaluated in this study. The 3dCloudNet incorporates historical 3-D radar echo sequences obtained from weather radar data to improve the accuracy and reliability of precipitation nowcasting. By capturing both horizontal and vertical motion patterns of clouds at various altitude levels, this model demonstrates an enhanced capability in detecting and distinguishing regions prone to severe convective weather events. The experimental results show that the model better captures the cloud's motion patterns and trends, and therefore has a noteworthy ability to detect and distinguish areas that may lead to severe convective weather. This study provides a step toward further improving the accuracy of precipitation nowcasting.

Manuscript received 29 November 2023; revised 11 April 2024; accepted 14 May 2024. Date of publication 24 May 2024; date of current version 4 June 2024. This work was supported in part by the Science and Technology Innovation Project of Laoshan Laboratory under Grant LSKJ202202200 and Grant LSKJ202202202; in part by the National Natural Science Foundation of China under Grant 41830964; in part by the Shandong “Taishan” Scholar Program, Key Research and Development Plan of Heilongjiang Province under Grant 2022ZX01A15; and in part by the Qingdao Science and Technology Planning Park Cultivation Program Project under Grant 23-1-5-yqy-11-qy. (Corresponding author: Chang Liu.)

Yuankang Ye is with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: heuyuankang@hrbeu.edu.cn).

Feng Gao is with the College of Intelligent System Science and Engineering, Harbin Engineering University, Harbin 150001, China, and also with the Qingdao Innovation and Development Center, Harbin Engineering University, Qingdao 266400, China (e-mail: gaofeng19@hrbeu.edu.cn).

Wei Cheng is with the Beijing Institute of Applied Meteorology, Beijing 100029, China (e-mail: chengw@mail.iap.ac.cn).

Chang Liu is with the College of Intelligent System Science and Engineering, Harbin Engineering University, Harbin 150001, China, and also with Qingdao Hatran Ocean Intelligence Technology Company Ltd., Qingdao 266400, China (e-mail: liuchang407@hrbeu.edu.cn).

Shaoqing Zhang is with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China, also with the Key Laboratory of Physical Oceanography, MOE, Institute for Advanced Ocean Study, Frontiers Science Center for Deep Ocean Multispheres and Earth System (DOMES), and the College of Ocean and Atmosphere, Ocean University of China, Qingdao 266100, China, and also with the Beijing Institute of Applied Meteorology, Beijing 100029, China (e-mail: szhang@ouc.edu.cn).

Shudong Wang is with the Public Meteorological Service Center, China Meteorological Administration, Beijing 100081, China (e-mail: wangs@cma.gov.cn).

Digital Object Identifier 10.1109/TGRS.2024.3404062

**Index Terms**—Deep learning, precipitation nowcasting, spatiotemporal sequence forecasting, weather radar.

## I. INTRODUCTION

PRECIPITATION nowcasting involves forecasting the intensity of rainfall at a high resolution within a specific local area in the short term (e.g., 0–6 h). Accurate nowcasting is crucial for production and lifestyle decisions [1], including crop production, flood warnings, and air traffic control, which rely on weather processes. Previously, meteorological researchers forecast future atmospheric motion and weather phenomena for a given period based on numerical weather prediction (NWP). They would create a set of partial differential equations (PDEs) describing weather evolution processes, such as fluid mechanics and thermodynamics equations, starting from the initial state [2], [3]. However, for complex and severe weather dynamic systems like cloud motion [4], [5] and convective weather [6], [7], researchers cannot fully understand their inherent change rules and can only use approximate functions to solve these weather processes [8], [9]. As a result, the most advanced precipitation nowcasting systems commonly employ radar echo extrapolation methods, which offer higher spatiotemporal resolution and are more computationally cost-effective [10], [11].

The conventional method for precipitation nowcasting, which is based on radar echo extrapolation, utilizes the optical flow technique [12], [13]. Initially, the optical flow method assumes that the luminosity of the sequential images remains unchanged over time and subsequently assumes that the dynamic process has temporal continuity, meaning that the motion is slight and mild. By computing the optical flow field between successive frames, capturing the motion pattern of every pixel in the radar echo image, and the future motion state can be determined. Atmospheric motion entails generation, dissipation, and change processes [14], [15] that render it challenging to meet the prerequisites assumed by the optical flow method. Consequently, applying the optical flow method for promptly developing convective rainfall predictions is usually restricted.

As deep learning continues to develop [16], researchers have started to explore weather forecasting methods based on artificial intelligence to solve the high-dimensional, complex, and nonlinear precipitation nowcasting problem using deep neural networks [17], [18], [19]. They have abstracted the

precipitation nowcasting problem based on radar echo extrapolation into a spatiotemporal data prediction problem and have designed appropriate deep prediction networks to predict the most probable future precipitation spatiotemporal state [20]. Currently, precipitation nowcasting models based on deep learning can be classified into two main categories: convolutional recurrent neural networks (ConvRNNs) [21], [22], [23], [24], [25], [26] and nonrecurrent neural network [27], [28], [29], [30]. For instance, ConvLSTM [20] employs convolutional operations to replace the fully connected operations in the structure of fully connected long–short-term memory neural network (FC-LSTM), thus capturing temporal and spatial variations and outperforming the previously optimal optical flow extrapolation method in the business field. Meanwhile, PredRNN [22] posits that spatiotemporal prediction learning should express temporal and spatial change patterns in a unified memory unit and proposes a new LSTM unit, the ST-LSTM, which can vertically traverse stacked LSTM layers and horizontally connect all LSTM states. This approach achieves state-of-the-art performance in multiple spatiotemporal datasets, including the weather radar dataset. Various models have been developed based on PredRNN to tackle problems such as vanishing gradients, nonstationarity, and complex spatial–temporal motion. These models include PredRNN++ [31], memory in memory (MIM) [32], and MotionRNN [25]. However, spatiotemporal prediction models based on ConvRNNs often suffer from the problem of vanishing gradients. Moreover, the ConvRNNs structure hinders the direct application of historical spatiotemporal features to future prediction tasks, as these features can only be passed on to the hidden unit through memory units. This issue negatively impacts high-resolution image generation tasks. Precipitation nowcasting models based on convolutional neural networks typically use U-Net [33] architecture as a benchmark. RainNet [27], which combines U-Net and SegNet [34] structures, has outperformed other precipitation prediction models. SmaAtUNet [28], which is based on the U-Net architecture, achieves comparable predictive performance by incorporating attention modules [35] and depth-wise separable convolutions [36] while using fewer parameters than other models. In addition, some researchers use deep generative models for deep learning-based precipitation nowcasting [37], while others combine physical principles with deep learning to construct forecasting models [38].

Precipitation nowcasting models have made significant advancements in training and prediction by leveraging 2-D weather radar images, which include composite reflectivity (CR) and constant altitude plan position indicator (CAPPI). However, in reality, the spatial motion of cloud clusters is 3-D, encompassing more complex motion transformations such as expansion, dissipation, and deformation. In addition, 2-D radar echoes at a single altitude layer still lack a comprehensive understanding of the internal structure, evolution, and characteristics of entire cloud clusters. There is a lack of direct observation of the subtle physical changes and interactions occurring within the cloud clusters, posing challenges in terms of accuracy and comprehensiveness. Relying solely on 2-D

radar echo reflectivity images is insufficient for accurately capturing the spatiotemporal characteristics of clouds in the vertical direction. Fig. 1 illustrates the radar reflectivity at various altitude levels during a rainfall event. A comprehensive consideration of the reflectivity states at various altitude levels will provide a more accurate understanding of the precipitation process. Research on extrapolating radar echoes using multidimensional data has not been as thoroughly studied as extrapolating 2-D radar images. In related studies, FURENet [39] incorporated polarimetric radar variables  $K_{DP}$  and  $Z_{DR}$  into the model, which provides additional microphysical and dynamic structural information about storms, and achieved more reliable predictive results than single-variable inputs. 3DCNN [40] and 3D-ConvLSTM [41] proposed two forecasting models that rely on 3-D convolution and 3-D ConvRNN for radar reflectivity data, which offers a more informative future weather forecasting system. However, these methods have certain limitations, including a small spatiotemporal range or high computational resource requirements.

Previously, ConvRNNs and nonrecurrent neural networks commonly utilized convolution kernels with fixed sizes to capture local spatial information of pixels. However, this approach was unable to establish the correlation between local and global features, which led to inadequate modeling of the overall spatiotemporal motion state. The transformer [42] structure has gained popularity in the field of natural language processing due to its remarkable ability to model global dependencies [43]. Later, the visual transformer [44] introduced the transformer structure to the field of computer vision, where it has exhibited remarkable performance across diverse tasks. Nevertheless, incorporating the self-attention module of the transformer directly into computer vision tasks during the initial design of Vit resulted in increased computational complexity of the model [45]. In addition, the absence of inductive bias and translational invariance [46], [47] of CNN in Vit necessitates an extensive amount of data for model training, which is challenging to obtain in the field of remote sensing.

This study proposes two improvements to optimize the deep learning-based model for precipitation nowcasting. First, we fuse radar echo data with varying altitude levels into the feature encoding module to enhance the model's performance in precipitation nowcasting. It is crucial to emphasize that, owing to the prevalence of the majority of clouds within the troposphere [48], which essentially encompasses the vast majority of atmospheric moisture, this experiment primarily investigates the influence of 3-D cloud structures within the troposphere (altitude range: 0.5–9 km) on radar echo extrapolation based on deep learning methodologies. Second, we introduce a large kernel spatiotemporal attention (LK-STA) module specifically designed for capturing spatiotemporal movement, enabling the model to effectively model long-range dependencies and adaptivity. Based on these enhancements, we introduce a novel precipitation nowcasting model named 3dCloudNet. Through extensive experiments on the same dataset, our proposed model demonstrates superior performance compared to other state-of-the-art models, as evaluated

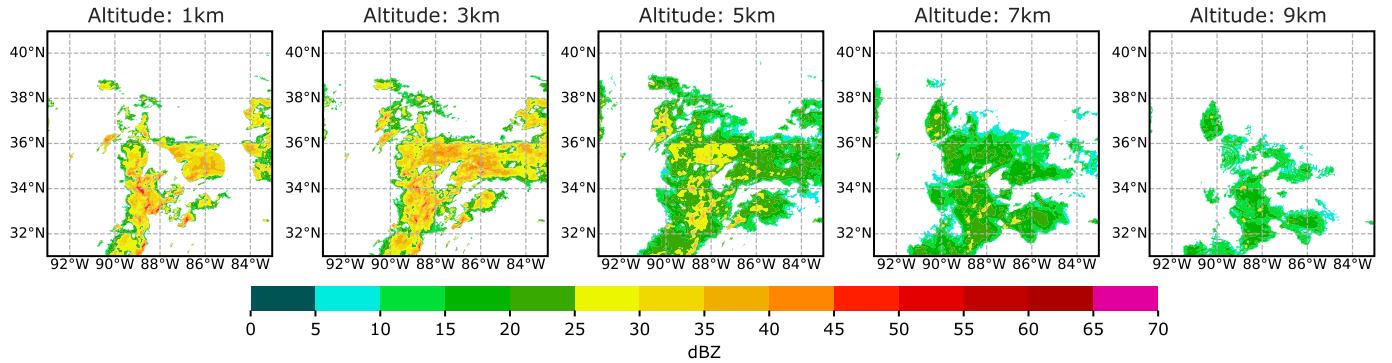


Fig. 1. Radar reflectivity data at various altitude levels collected during a single precipitation event.

by several metrics. This article presents the following main contributions.

- 1) Utilizing radar echo data from different altitude levels and maximizing the utilization of the 3-D cloud structure, our proposed model enhances precipitation nowcasting accuracy.
- 2) Our proposed model introduces a novel module that combines CNN and self-attention, efficiently capturing spatiotemporal information. This integration enables the model to learn intricate patterns and dependencies within the data, enhancing its ability to make accurate predictions.
- 3) We compare our model against existing approaches using the same dataset. The results demonstrate its superior performance across multiple evaluation metrics, showcasing improvements in both the precision and timeliness of weather warnings.

The remaining sections in this article are structured as follows. Section II presents the methodology and format used to construct the dataset for this study. Section III provides a detailed description of the proposed 3dCloudNet model. In Section IV, we present and analyze the experimental results of the comparative experiments. Section V is dedicated to discussing and analyzing the limitations. Finally, Section VI summarizes the main research content presented in this study.

## II. DATASET

The conventional forecasting method using radar echo images involves extrapolating echo and mapping calculations from echo intensity to precipitation intensity. The primary focus of this study is on the echo extrapolation stage, aiming to investigate whether a new modeling method can enhance performance in this task. This study utilized WSR-88D radar data from approximately 300 severe weather events spanning three years from 2019 to 2021 as the research dataset. The data source was obtained from the ds841.6 dataset product of the National Center for Atmospheric Research (NCAR) research data archive [49]. The dataset covers a range from 124.99°W to 66.01°E in longitude and from 49.99°N to 24.01°S in latitude, encompassing most of the USA mainland and its diverse terrain. This scope encompasses multiple geographical regions, seasons, and climate conditions, providing a rich dataset to support our research objectives. We emphasize that the dataset we used is publicly available and can be accessed

at the URL: <https://doi.org/10.5065/2B46-1A97>. This data encompasses 3-D radar echo data from varying altitude levels, covering regions with high echo intensity that may trigger severe convective weather events. By processing and analyzing this data, a new 3-D radar echo dataset was constructed for evaluating the performance of the proposed precipitation nowcasting model.

Specifically, this 3-D dataset has a temporal resolution of 5 min. The horizontal extent in latitude and longitude varies due to the different impact areas of severe convective weather, the gridded data ranges from 480 to 2424, with a spatial resolution of approximately  $0.02^\circ \times 0.02^\circ$  after gridization. The vertical altitude levels span from 0.5 to 9 km and are divided into 16 vertical layers (with a resolution of 0.5 km between 0.5 and 7 km, and a resolution of 1 km between 7 and 9 km). Due to storage and GPU limitations, and to fully utilize the existing dataset, this experiment extracts 3-D radar reflectivity data with a spatial resolution of  $480 \times 480$  pixels from the northeast, southeast, northwest, and southwest corners of each event. The spatial resolution is then interpolated to  $240 \times 240$  pixels. We utilized a sliding window method to construct the dataset, ensuring a minimum time interval of 15 min between each radar echo sequence to avoid excessive data duplication. Each radar echo sequence comprises 3-D radar echo data for 24 time steps, with a time interval of 5 min between each step. Thus, each sequence represents 2 h of 3-D radar echo data. Finally, we divided the dataset into training, validation, and testing sets for model training and testing. It's worth emphasizing that the training, validation, and testing sets do not overlap in the constructed time periods. Table I presents detailed information on the datasets used in this article.  $T$  represents the length of input and output sequence time steps,  $N_i$  represents the number of altitude levels in the input data, and  $N_o$  represents the number of altitude levels in the output data.  $C$ ,  $H$ , and  $W$  denote each input radar echo image's channel number, height, and width, respectively. The period represents the time range over which the dataset is constructed, and the number represents the total number of sequences in the constructed dataset.

## III. METHOD

### A. Motivation

In this study, we define continuous radar echo images as continuous matrices  $X_{1:T} = [X_1, X_2, \dots, X_T]$ ,  $X_t \in$

TABLE I  
DIMENSIONS AND DISTRIBUTION OF THE DATASET

Name	$(T, N_i, C, H, W)$	$(T, N_o, C, H, W)$	Period	Number
Training	(12, 16, 1, 240, 240)	(12, 1, 1, 240, 240)	2019.01-2020.12	35706
Validation	(12, 16, 1, 240, 240)	(12, 1, 1, 240, 240)	2021.01-2021.06	6623
testing	(12, 16, 1, 240, 240)	(12, 1, 1, 240, 240)	2021.07-2021.12	9800

$\mathbb{R}^{C \times H \times W}$ , the variables  $C$ ,  $H$ , and  $W$  refer to the number of channels, height, and width of each radar echo image, respectively. The following equation describes the proposed 3-D cloud-assisted method for precipitation nowcasting:

$$\bar{\mathcal{X}}_{t+1:t+f} = \arg \max_{\mathcal{X}_{t+1:t+f}} \mathcal{P}(\mathcal{X}_{t+1:t+f} | \mathcal{X}_{t-h:t}, \mathcal{A}_{t-h:t}^k). \quad (1)$$

In this equation, precipitation nowcasting based on radar echo extrapolation is regarded as a spatiotemporal prediction problem, where  $t$  denotes the current time. The model aims to generate the future most probable spatiotemporal states  $\bar{\mathcal{X}}$  for the succeeding  $f$  time steps. The previous  $h$  time steps' data is used as the input for the model before the current time  $t$ , where  $\mathcal{X}$  denotes the image data of a specific altitude level, it is important to emphasize that most precipitation nowcasting methods based on radar reflectivity extrapolation use data from the radar located at an altitude of 3 km. Therefore, in this article,  $\mathcal{X}$  represents radar reflectivity data at the 3 km altitude level, and  $\mathcal{A}^k$  signifies the radar echo image data (excluding the 3 km altitude level) of the  $k$ th altitude level.

### B. Overview

The 3dCloudNet model, proposed in this study and depicted in Fig. 2, comprises three primary components. The encoder block [see Fig. 2(a)] extracts historical altitude-specific radar echo information, transforming it into high-level semantic features. It integrates high-level semantic features from radar echo images at various altitudes, extracted from the auxiliary prediction block. The spatiotemporal attention block [see Fig. 2(b)], a pivotal network component, captures continuous spatiotemporal state changes, effectively establishing correlations among spatiotemporal information. Through the utilization of both global context and local attention mechanisms, this module enhances prediction performance and identifies meaningful patterns in the data. Finally, the decoder block [see Fig. 2(c)] interprets the spatiotemporal state within a high-dimensional space, generating predicted sequence images to achieve cloud image predictions in the future.

1) *Encoder Block*: Initially, the module utilizes continuous convolution and downsampling operations to extract high-dimensional features from the input sequence of 3 km radar echo images. The features initially extracted are directly concatenated with the decoder block, ensuring the seamless transfer of information. Second, the auxiliary prediction block gathers feature information from radar echo images at  $k$  different altitude levels through continuous convolution and downsampling operations. Subsequently, convolution is applied to further extract features from the various altitude levels, capturing valuable information crucial for predicting future spatiotemporal states. Finally, feature aggregation is achieved through a  $1 \times 1$  convolution of the high-dimensional

features from 3 km radar echo images and those obtained from radar echo images at other altitude levels. Detailed specifications of this module are presented in Table II.

Equation (2) provides a clearer representation of the data flow direction within the encoder block. Specifically,  $X_r^0$  denotes the input 3 km radar echo image data while  $X_r^k$  refers to the radar echo data corresponding to the  $k$ th ( $k \leq 15$ ) altitude level input within the auxiliary prediction block. The term Conv2d represents the most frequently utilized convolutional layer in this model, with a kernel size of 3 and a stride of 1. An illustrative instance can be found in the layer of conv2d-1 as detailed in Table II. Similarly, Conv2d corresponds to the layer named conv2d-2 in Table II, while Conv2d corresponds to the layer referred to as conv2d-3 in Table II. In addition,  $2 \times$  represents data flowing through the subsequent two modules

$$\begin{aligned} X_c^0 &= 2 \times \text{SiLU}(\text{BatchNorm}(\text{Conv2d}(X_r^0))) \\ X_c^k &= 2 \times \text{SiLU}(\text{BatchNorm}(\text{Conv2d}(X_r^k))) \\ X_o &= 2 \times \text{Conv2d}(X_c^0) \\ X_o^k &= 2 \times \text{Conv2d}(X_c^k) \\ X_p &= \text{Conv2d}(\text{Concat}(X_o^1, X_o^2, \dots, X_o^k)) \\ X_e &= \text{Conv2d}(\text{Concat}(X_o, X_p)). \end{aligned} \quad (2)$$

2) *Spatiotemporal Attention Block*: The attention mechanism is a selection process that aids in identifying relevant features while ignoring interfering factors such as noise during the training process. Inspired by transformer [42] and VAN [50], the LK-STA module generates dynamic distance spatiotemporal attention maps by using depth-wise convolution and depth-wise dilation convolution [51] to construct the attention map for the spatial dimension and channel convolution to construct the attention map for the temporal dimension, to enable the model to recognize feature importance independently. Table III presents detailed information about the LK-STA module, and the dim in this table depends on hyperparameters.

Equation (3) clarifies the data flow within the LK-STA module, where  $X$  denotes the input feature matrix and  $\odot$  represents the element-wise product. The modules dw-conv and dw-d-conv can both be located with their respective parameters in Table III

$$\begin{aligned} X_{\text{map}} &= X \odot \text{Conv2d}(\text{dw-d-conv}(\text{dw-conv}(X))) \\ \text{LK-STA} &= \text{Softmax}(X_{\text{map}}) \odot X. \end{aligned} \quad (3)$$

Equation (4) demonstrates the working mechanism of the data flow in other parts of the block. In this equation,  $X$  denotes the input feature of the attention and feed-forward

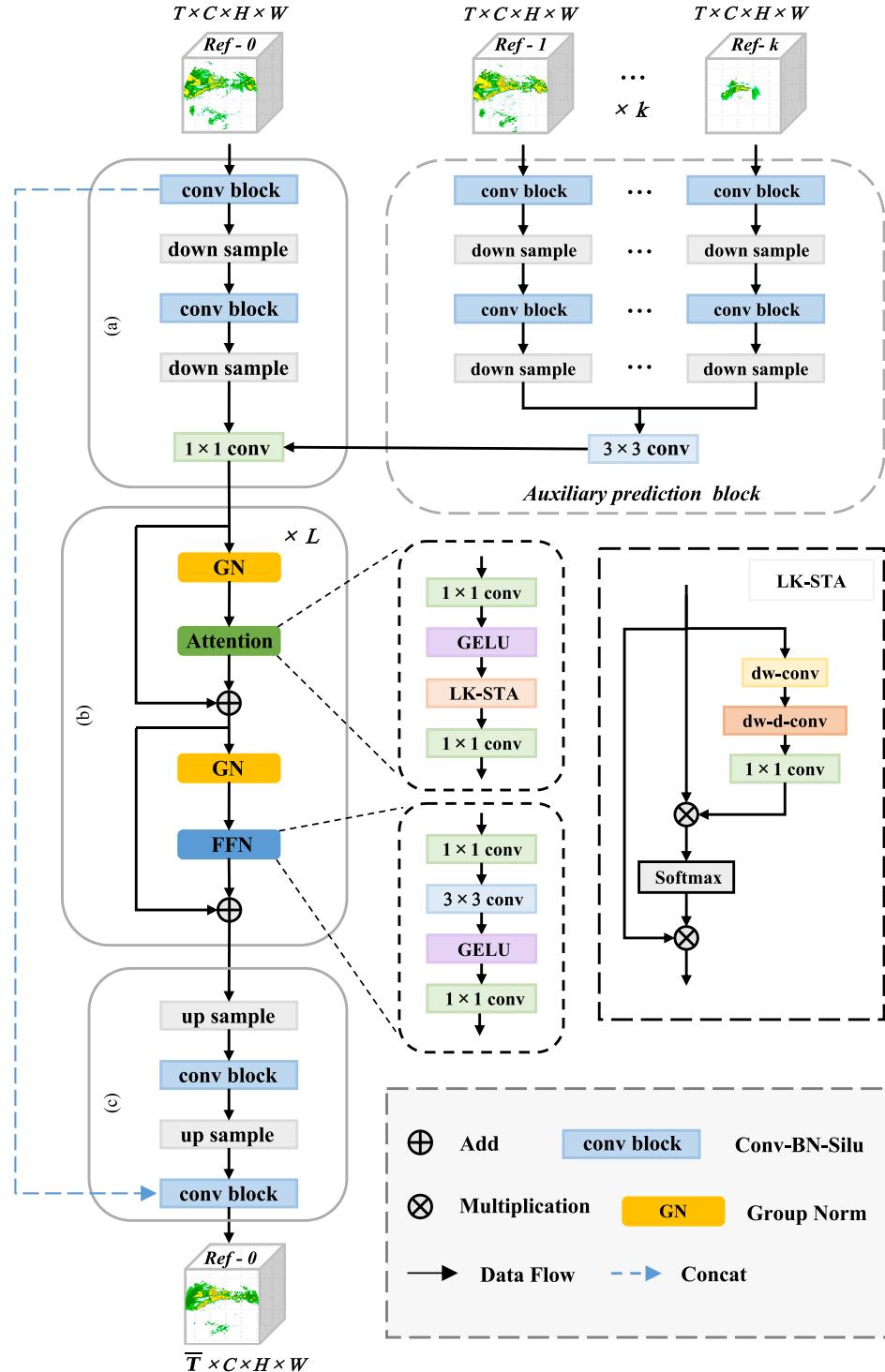


Fig. 2. Overall framework of 3dCloudNet. Radar echo sequence data ref-0 (altitude: 3 km) is fed into (a) encoder block, while auxiliary data ref-1 to ref- $k$  (altitude: 0.5–9 km, excluding 3 km) is directed to the auxiliary prediction block. Both modules employ convolutional layers and downsampling operations for spatiotemporal feature extraction, followed by spatiotemporal feature fusion in (b) spatiotemporal attention block. Subsequently, the extracted features traverse through (c) decoder block. In its final layer, shallow features from (a) encoder block enhance spatiotemporal decoding for forecasting future states of the ref-0 radar echo (altitude: 3 km).

network (FFN) structures.  $X_e$  refers to the output feature of the encoder block in (2). The variable  $L$  indicates the number of layers present in the spatiotemporal attention block

$$\text{Attention} = \text{Conv2d}_{\text{kernel}=1}\left(\text{LK-STA}\left(\text{GELU}\left(\text{Conv2d}_{\text{kernel}=1}(X)\right)\right)\right)$$

$$X_a = X_e + \text{Attention}(\text{GroupNorm}(X_e))$$

$$\text{FFN} = \text{Conv2d}_{\text{kernel}=1}\left(\text{GELU}\left(\text{Conv2d}_{\text{kernel}=1}\left(\text{Conv2d}_{\text{kernel}=1}(X)\right)\right)\right)$$

$$X_f = X_a + \text{FFN}(\text{GroupNorm}(X_a))$$

$$X_{\text{sta}} = L \times X_f. \quad (4)$$

3) Decoder Block: This module uses continuous convolution and upsampling operations to decode the high-dimensional spatiotemporal features learned from

TABLE II  
DETAILS OF THE ENCODER BLOCK

Module	Layer	Kernel	Stride	Padding
conv block	Conv2d-1	(3, 3)	(1, 1)	(1, 1)
down sampling	Conv2d-2	(3, 3)	(2, 2)	(1, 1)
1 × 1 conv	Conv2d-3	(1, 1)	(1, 1)	(0, 0)

TABLE III  
DETAILS OF THE LK-STA

module	layer	kernel	stride	padding	groups	dilation
dw-conv	Conv2d-4	(5, 5)	(1, 1)	(2, 2)	dim	1
dw-d-conv	Conv2d-5	(7, 7)	(1, 1)	(9, 9)	dim	3
1 × 1 conv	Conv2d-6	(1, 1)	(1, 1)	(0, 0)	1	1

TABLE IV

DETAILS OF THE DECODER BLOCK

Module	Layer	Kernel	Stride	Padding
upsampling	ConvTranspose2d	(3, 3)	(2, 2)	(1, 1)

the spatiotemporal attention block. In the last layer, a concatenation operation is directly performed on the low-dimensional spatiotemporal information extracted in the first layer of the encoder block to obtain the final output. Detailed information about this module can be found in Table IV.

The data flow mechanism of this module is demonstrated in (5), while  $X_{\text{sta}}$  represents the output feature of the spatiotemporal attention block in (4),  $X_c^0$  refers to the features of the input 3 km radar echo image after undergoing two convolution operations in (2) and output represents the forecast motion state of the future 3 km radar echo

$$\begin{aligned} X_{u1} &= \text{ConvTranspose2d}(X_{\text{sta}}) \\ X_d &= 2 \times \text{SiLU}(\text{BatchNorm}(\text{Conv2d}(X_{u1}))) \\ X_{u2} &= \text{ConvTranspose2d}(X_d) \\ \text{output} &= 2 \times \text{SiLU}(\text{BatchNorm}(\text{Conv2d}(X_{u2} + X_c^0))). \end{aligned} \quad (5)$$

#### IV. EXPERIMENT

This section aims to evaluate the 3dCloudNet both quantitatively and qualitatively. First, we provide detailed information about the model training and the metrics used in the evaluation process. Second, we compare the performance of 3dCloudNet with other models that have demonstrated excellent results in the same field. Finally, we conduct comparative experiments to explore different ways of implementing 3-D cloud-assisted techniques.

##### A. Implementation Details

This study adopts competitive models in the field of precipitation nowcasting, specifically the ConvLSTM [20] and SmaAT-UNet [28] models. Furthermore, the SimVP [30] model, demonstrating exceptional performance in spatiotemporal prediction, is also designated as the baseline model

for 2-D echo extrapolation in this study. To ensure a fair comparison of the experimental results, all models used the MSE loss function during the training phase. Furthermore, during the data preprocessing stage, all data were normalized using the max-min normalization method to ensure they fell within the range of (0, 1). For the 3dCloudNet training, the spatiotemporal attention block was stacked four layers high. The initial learning rate was set to 0.0005, and the adamW [52] optimizer was used for optimization. The batch size was set to 4, and the training was terminated after 20 epochs. All the methods were implemented using PyTorch and were run on a single NVIDIA A100 GPU.

##### B. Evaluation Index

This study employs six quantitative metrics to evaluate the proposed model and other comparative models. Mean squared error (MES) and mean absolute error (MAE) assess the overall predictive performance and precision. In addition, the critical success index (CSI), probability of detection (POD), Heidke skill score (HSS), and false alarm ratio (FAR) serve as conventional indicators for evaluating meteorological forecast accuracy across various models. CSI quantifies the consistency between prediction and observation, measuring the ratio of successfully predicted events to the total possible events. CSI values range from 0 to 1, with 1 indicating perfect consistency and 0 signifying a lack of successful predictions. The calculation formula is as follows:

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \quad (6)$$

The concept of POD pertains to quantifying the ratio of successfully detected events to those that occurred in reality. The POD value varies between 0 and 1, with 1 denoting flawless detection and 0 signifying a failure in detection. The calculation formula is articulated as follows:

$$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (7)$$

The FAR is a metric that measures the proportion of falsely alerted events to the total predicted events. The FAR value, ranging from 0 to 1, represents the likelihood of false alarms; 0 indicates no false alarms, while 1 signifies all predictions are incorrect. The formula for calculation is

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (8)$$

The HSS is a quantitative metric assessing a model's predictive accuracy compared to random prediction. HSS values are bounded between  $-\infty$  and 1, with 1 denoting flawless predictive skill, 0 indicating performance akin to random prediction, and negative values signifying performance inferior to random prediction. The computation formula for HSS is outlined below

$$\begin{aligned} A &= (\text{TP} + \text{FN}) \times (\text{FN} + \text{TN}) \\ B &= (\text{TP} + \text{FP}) \times (\text{FP} + \text{TN}) \\ \text{HSS} &= \frac{2 \times (\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{A + B}. \end{aligned} \quad (9)$$

TABLE V  
CONFUSION MATRIX

		Forecast > Threshold	
		Positive	Negative
Observed > Threshold	Positive	$TN$	$FP$
	Negative	$FN$	$TP$

Table V shows the calculation methods for TN, FP, FN, and TP. The threshold represents the value set after the radar echo intensity (unit: dBZ) has been discretized into a grid. A grid point is considered a positive case if its value exceeds the threshold, and a negative case if its value is below the threshold.

### C. Experimental Results and Analysis

This article initiates a comprehensive quantitative assessment of the proposed 3dCloudNet structural design. Throughout this evaluation, the nomenclature 3dCloudNet w/o is employed to signify the variant of 3dCloudNet devoid of the auxiliary prediction block. Analogous to comparable control experiments, this variant exclusively conducts extrapolation utilizing the 2-D radar echoes situated at an altitude of 3 km. Conversely, the 3dCloudNet variant embodies the exploitation of the 3-D cloud structure beneath the troposphere (ranging from 0.5 to 9 km, totaling 15 layers) to facilitate the extrapolation of radar echoes at the 3 km altitude. This methodological approach is undertaken with the explicit objective of appraising the efficacy of the model structure advanced in this study. The resultant empirical findings are succinctly delineated in Table VI, wherein  $CSI^{20}$ ,  $CSI^{30}$ , and  $CSI^{40}$  denote, respectively, the computed values of the CSI index within the confusion matrix at thresholds of 20, 30, and 40. Notably, the application of **Bold** and underlined formatting serves to demarcate the optimal and suboptimal outcomes discerned within the metrics.

Based on the experimental results in Table VI, the proposed 3dCloudNet w/o model in this study outperforms other comparative models in both MSE and MAE. This observation indicates that the model exhibits higher accuracy and stability in nowcasting forecasting, which is crucial for enhancing the reliability of predictive models in practical applications. This not only validates the rationality of the proposed model structure in this article but also underscores the outstanding performance of the LK-STA module redesigned relative to the SimVP model. Furthermore, the integration of the Auxiliary prediction block into the 3dCloudNet w/o model, resulting in the 3dCloudNet, significantly surpasses other models without 3-D cloud assistance in key metrics of the precipitation nowcasting, such as CSI, POD, FAR, and HSS. Particularly noteworthy is the substantial improvement in  $CSI^{30}$  and  $CSI^{40}$ , with increases of 6.302% and 10.39%, respectively. This implies that, with the assistance of the 3-D cloud structure, the model excels in detecting and controlling high-intensity rainfall events. Further observation of the results indicates that 3dCloudNet demonstrates more robust performance in terms of both POD and FAR, exhibiting a higher probability

of detecting rainfall events and successfully reducing false alarm rates. This contributes to enhancing the operational effectiveness of predictions.

To enable a comprehensive comparison between the proposed model and the control model in terms of performance, Fig. 3 provides a detailed analysis of MSE and MAE metrics across the succeeding 12-time intervals. Smaller MSE and MAE values signify a closer correspondence between the predicted and actual images, indicating a higher level of predictive accuracy. Upon careful examination of Fig. 3, it is apparent that the 3dCloudNet model, integrated with 3-D cloud assistance, demonstrates superior performance in both MSE and MAE metrics compared to alternative models. Particularly noteworthy is the conspicuous trend showing an augmented overall performance with longer prediction horizons. This empirical observation further substantiates the crucial role of 3-D cloud-assisted prediction in elevating predictive accuracy.

In order to conduct a thorough assessment of the low dBZ regions (potential rainfall occurrence areas) and high dBZ regions (regions prone to triggering severe convective weather), this study employs three meteorological thresholds for meticulous comparison, specifically set at 20, 30, and 40 dBZ. Grounded in a unified observational perspective, this article delves into a comprehensive temporal analysis of the CSI, POD, FAR, and HSS metrics over the next 12-time intervals. The methodological framework not only facilitates the precise characterization of rainfall features in low dBZ regions but also enables an effective analysis of intense convective phenomena in high dBZ areas. Through the incorporation of these three distinct dBZ thresholds of varying intensities, a holistic understanding of the evolving processes associated with diverse meteorological phenomena is achieved, thereby providing a more detailed foundation for future weather predictions. Simultaneously, the temporal comparison of evaluation metrics such as CSI, POD, FAR, and HSS allows for the capture of dynamic variations in these meteorological features over future time intervals, offering robust support for the refinement of more accurate weather prediction models.

The CSI index, through a comprehensive consideration of both prediction accuracy and omission rate, has emerged as a holistic indicator for evaluating forecasting effectiveness. In the context of this study, Fig. 4 elucidates the superior performance exhibited by the proposed 3dCloudNet model in the precipitation nowcasting and severe convective weather alerts, across varying reflectivity thresholds. The empirical evidence presented herein substantiates the claim that the incorporation of 3-D cloud-assisted precipitation nowcasting, as introduced in this article, significantly augments forecasting models and enhances warning systems.

POD serves as a metric representing the likelihood of precipitation in meteorological forecasting, higher POD values correspond to an increased probability of accurate predictions. However, relying solely on POD may inadequately gauge the precision of a model, as elevated POD values may coincide with false alarms associated with nonprecipitation echoes. Consequently, it is imperative to incorporate FAR into the evaluation, denoting the probability of erroneously identifying

TABLE VI  
QUANTITATIVE EVALUATION OF THE MODEL ON FOUR INDICATORS

Metrics \ Model	ConvLSTM	SmaAtUnet	SimVP	3dCloudNet w/o	3dCloudNet
MSE( $10^{-3}$ )	20.20	20.74	19.35	<u>18.79</u>	<b>17.16</b>
MAE( $10^{-3}$ )	124.5	117.3	126.4	<u>110.4</u>	<b>103.3</b>
CSI <sup>20</sup>	0.5333	0.5342	0.5483	<u>0.5673</u>	<b>0.5917</b>
POD <sup>20</sup>	0.8026	0.7985	<u>0.8038</u>	<u>0.7995</u>	<b>0.8138</b>
FAR <sup>20</sup>	0.4095	0.4067	0.3883	<u>0.3613</u>	<b>0.3354</b>
HSS <sup>20</sup>	0.6558	0.6548	0.6707	<u>0.6888</u>	<b>0.7114</b>
CSI <sup>30</sup>	0.3786	0.3805	0.3944	<u>0.4221</u>	<b>0.4487</b>
POD <sup>30</sup>	0.6874	0.7012	<u>0.7119</u>	<u>0.7074</u>	<b>0.7201</b>
FAR <sup>30</sup>	0.5799	0.5831	0.5661	<u>0.5246</u>	<b>0.4902</b>
HSS <sup>30</sup>	0.5030	0.5018	0.5212	<u>0.5516</u>	<b>0.5804</b>
CSI <sup>40</sup>	0.1833	0.1736	0.1876	<u>0.2098</u>	<b>0.2316</b>
POD <sup>40</sup>	0.5703	0.6379	0.6381	<u>0.6445</u>	<b>0.6449</b>
FAR <sup>40</sup>	0.8419	0.8487	0.8316	<u>0.8118</u>	<b>0.7865</b>
HSS <sup>40</sup>	0.2664	0.2495	0.2736	<u>0.3029</u>	<b>0.3316</b>

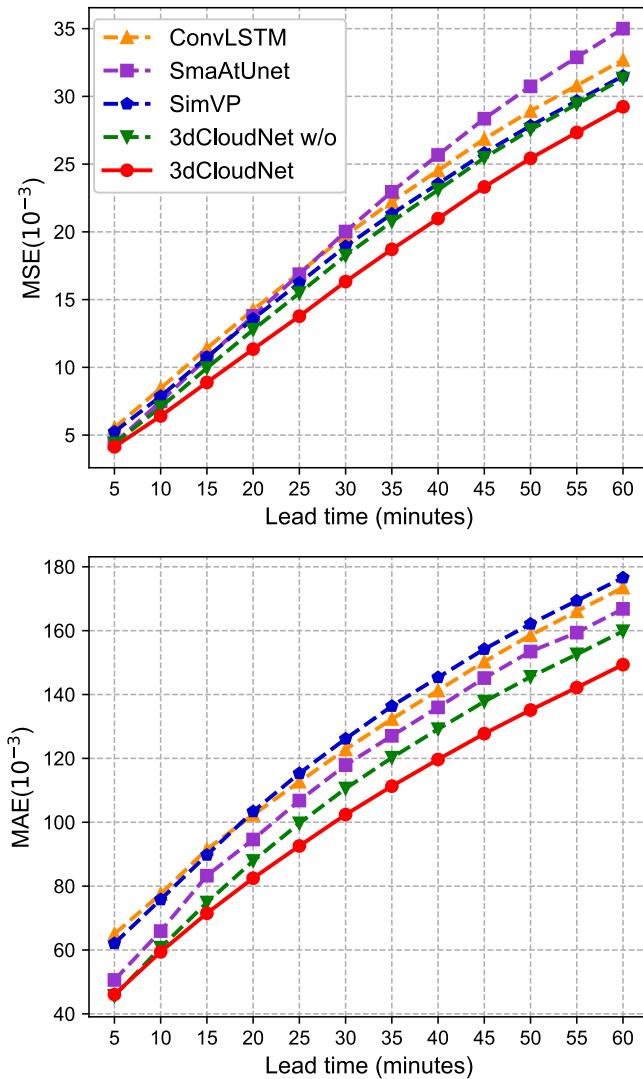


Fig. 3. Frame-wise results of MSE and MAE indicators in the test set.

nonprecipitation echoes as precipitation echoes. Exceptional predictive models should exhibit both high POD and low FAR and through a synergistic analysis of these two metrics,

a more comprehensive evaluation of model performance can be attained.

To comprehensively compare the POD and FAR of all models, we have depicted curves in Fig. 5 illustrating the temporal variations of POD and FAR at different thresholds. While certain models exhibit superior hit rates, such as SimVP, it is noteworthy that this is accompanied by higher false alarm rates, indicating a tendency for these models to broadly misclassify nonprecipitation echoes as precipitation echoes. The proposed model, 3dCloudNet w/o, which does not incorporate the 3-D cloud-assistive structural design, effectively reduces false alarm rates. However, shortcomings persist in terms of hit rates. Upon introducing the 3-D cloud-assistive structure, the model maintains low false alarm rates while concurrently increasing hit rates. Experimental results substantiate the rationale and reliability of incorporating the 3-D cloud-assistive structure for improved rainfall forecasting.

In the field of precipitation nowcasting, the HSS serves as a comparative metric for assessing the model's predictive accuracy against actual observations. By effectively utilizing true positives and true negatives, as well as applying appropriate penalties for false positives and false negatives, the HSS measures the model's discriminative capability for both rainfall and nonrainfall events, including accuracy in predicting extreme rainfall events. The temporal evolution of the HSS metric under various thresholds is delineated in Fig. 6. It is evident that the proposed 3dCloudNet w/o and the 3dCloudNet with 3-D cloud assistance exhibit superior overall predictive performance. With increasing time, these models demonstrate a slower decline in overall indicators. This is crucial for the precision and reliability of warning systems, aiding in the timely implementation of necessary preventive measures and enhancing the capacity to respond to extreme weather events.

In order to visually demonstrate the forecasting capabilities of the proposed model in this article, in Fig. 7, we map the grayscale image variations representing representative weather processes into the RGB space to visualize radar frames. To provide a more direct display, we plot the |Truth-Predict| as

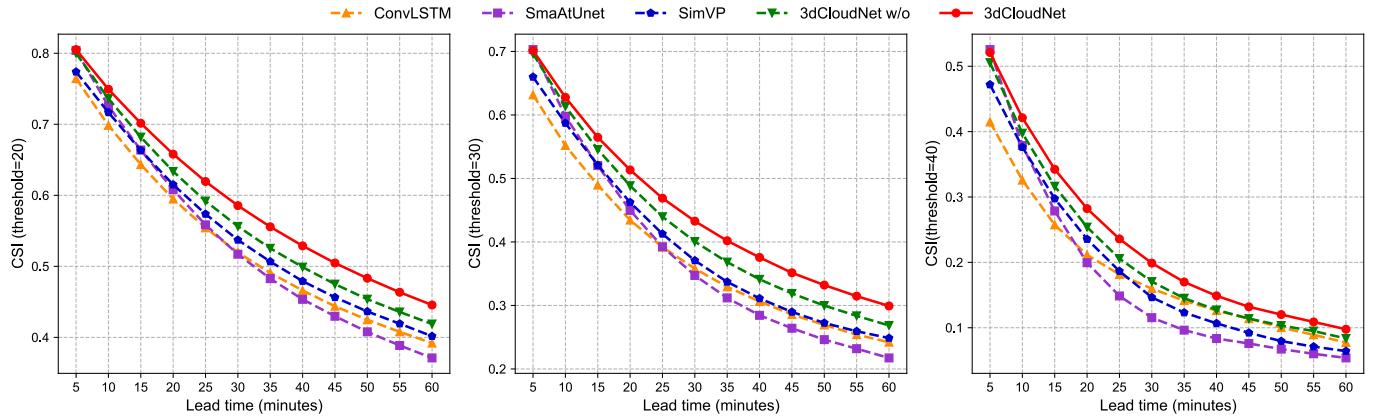
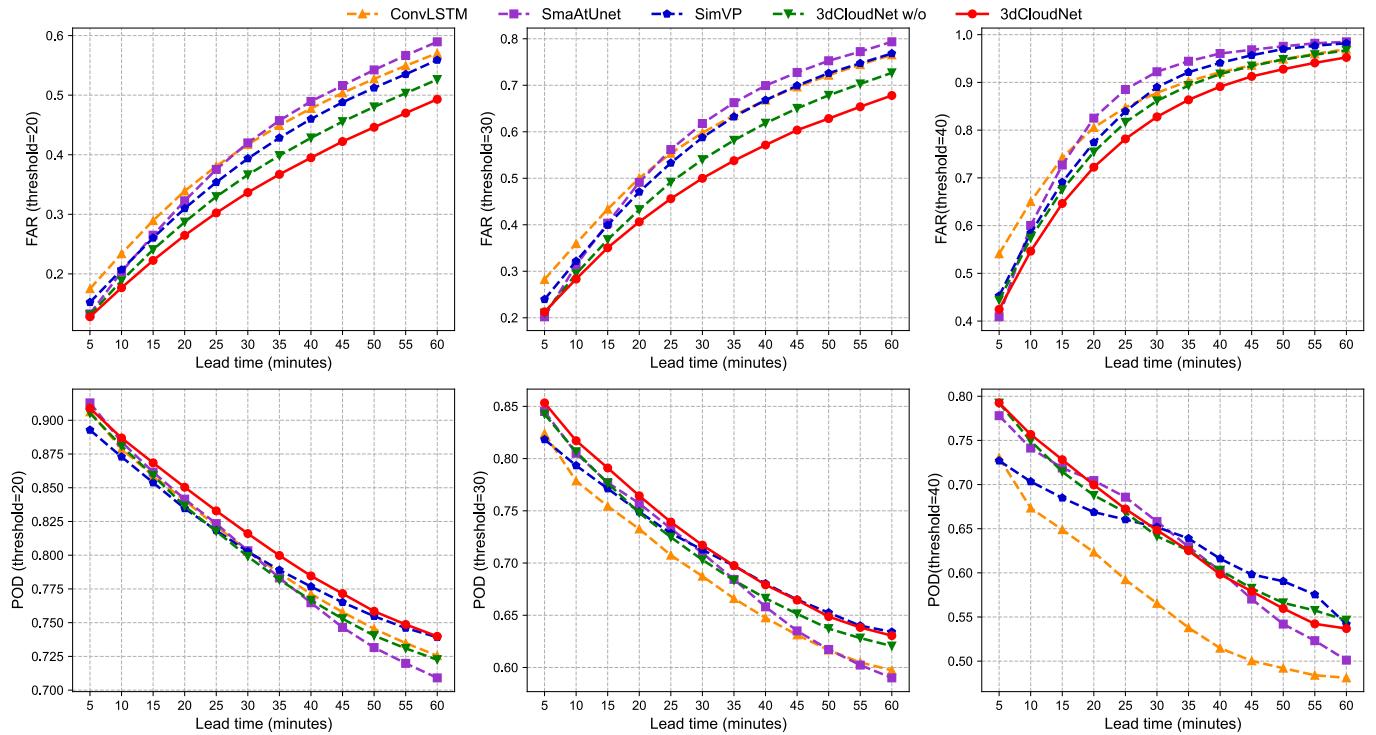
Fig. 4. Frame-wise results of  $\text{CSI}^{20}$ ,  $\text{CSI}^{30}$ , and  $\text{CSI}^{40}$  indicators in the test set.

Fig. 5. Frame-wise results of FAR and POD indicators in the test set.

an error image. Regions with echo values exceeding 30 dBZ typically indicate the potential occurrence of severe convective weather. As evident in Fig. 7, in areas with high echo values, the 3dCloudNet utilizing the proposed 3-D cloud-assisted approach exhibits superior overall motion state prediction capabilities. Particularly noteworthy is the closer alignment of the overall prediction trend of our model with the actual values over time, consistent with the error plot. This provides more robust support for forecasters in assessing aspects such as short-term severe convective rainfall.

The preceding text undertook a comprehensive quantitative and qualitative assessment of the efficacy of 3-D cloud assistance on the test dataset. This study further probes the influence of the 3-D cloud structure on radar echo extrapolation specifically at the 3 km altitude layer. The experiment categorizes the assisted predictive altitude layers into three

distinct ranges: low (L, spanning 0.5–2.5 km, totaling five layers), middle (M, spanning 3.5–5.5 km, totaling five layers), and high (H, spanning 6–9 km, totaling five layers). Three separate experimental sets are executed to examine the impact of various altitude layers' 3-D cloud structures on the accuracy of extrapolating radar echoes beneath the radar echo map at the 3 km altitude level.

Subsequently, three additional sets of experiments are introduced: low-middle (L-M, spanning 0.5–5.5 km, totaling ten layers), middle-high (M-H, spanning 3.5–9 km, totaling ten layers), and low-middle-high (L-M-H, spanning 0.5–9 km, totaling fifteen layers). The objective is to analyze the influence of different altitude layers on the precision of radar echo extrapolation predictions utilizing deep learning techniques, with the goal of identifying the optimal combination of altitude layers for the 3-D cloud-assisted predictive model. The

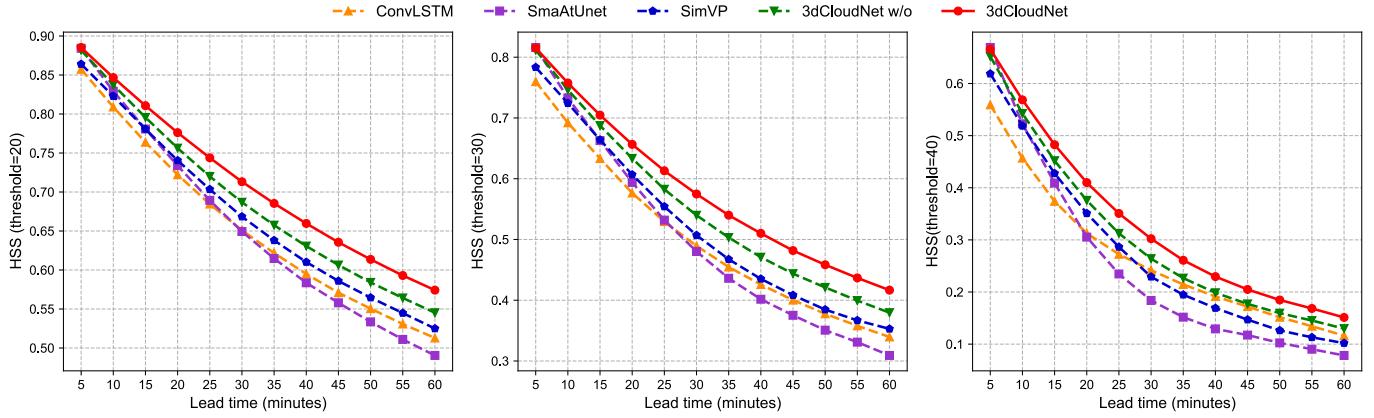
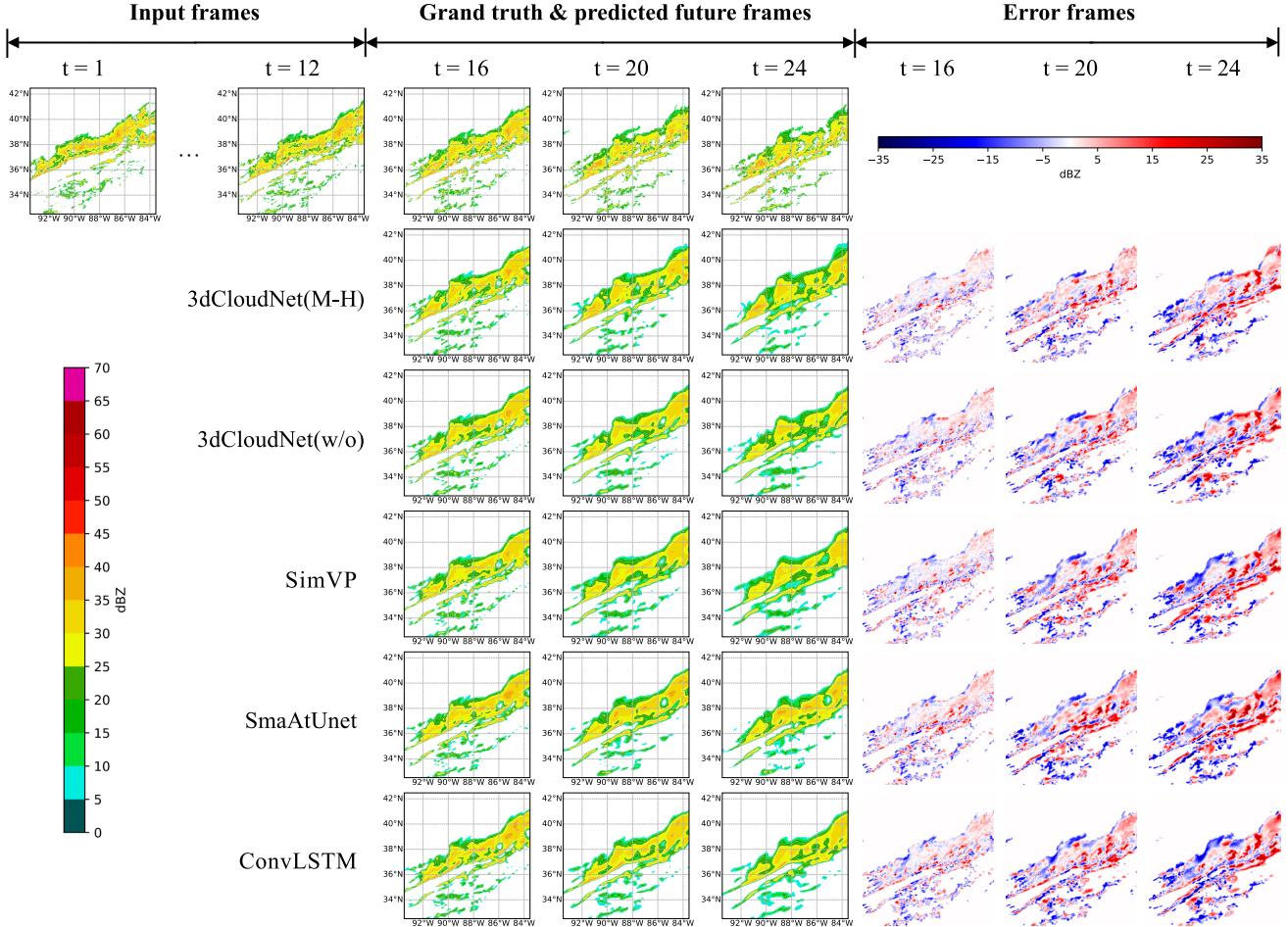
Fig. 6. Frame-wise results of  $HSS^{20}$ ,  $HSS^{30}$ , and  $HSS^{40}$  indicators in the test set.

Fig. 7. Performance of different comparison models in an example of 1-h radar echo prediction.

experimental scope and hyperparameter configurations align with those employed in prior experiments. Detailed findings from the experiments are documented in Table VII.

According to the experimental results data presented in Table VII, we still use solely the 2-D radar echoes at the 3 km altitude for extrapolation as the baseline, without incorporating the structural assistance of 3-D cloud data. Initially, we consider the auxiliary prediction using the 3-D cloud structure at lower atmospheric heights (altitude range: 0.5–2.5 km, including the planetary boundary layer). Analysis of the experimental outcomes suggests that the 3-D cloud

structure at these lower altitudes did not notably facilitate the extrapolation of radar echoes at the 3 km altitude. The analysis suggests that this may be attributed to the near-surface layers of the troposphere within the planetary boundary layer, where the atmospheric state is susceptible to the influence of heat and momentum transfer from the Earth's surface. Moreover, the geographic location during different intense convective weather events in this experiment varies, leading to weaker performance of 3dCloudNet (L) compared to 3dCloudNet (w/o) across multiple indicators. The middle layer (altitude range: 3.5–5.5 km) and the high layer (altitude range: 6–9 km)

TABLE VII  
INFLUENCE OF VARIOUS ALTITUDE LEVELS RADAR ECHOES ON PRECIPITATION NOWCASTING PERFORMANCE

Metrics \ Model	3dCloudNet w/o	3dCloudNet(L)	3dCloudNet(M)	3dCloudNet(H)	3dCloudNet(L-M)	3dCloudNet(M-H)	3dCloudNet(L-M-H)
MSE( $10^{-3}$ )	18.79	18.86	17.47	17.24	<u>17.13</u>	<b>17.03</b>	17.16
MAE( $10^{-3}$ )	110.4	121.0	<u>105.3</u>	106.6	108.4	107.15	<b>103.3</b>
CSI <sup>20</sup>	0.5673	0.5630	0.5894	0.5887	0.5884	<b>0.5979</b>	<u>0.5917</u>
POD <sup>20</sup>	0.7995	0.7982	0.8058	<u>0.8106</u>	0.8079	0.8032	<b>0.8138</b>
FAR <sup>20</sup>	0.3613	0.3661	<u>0.3347</u>	0.3367	0.3365	<b>0.3201</b>	0.3354
HSS <sup>20</sup>	0.6888	0.6846	0.7093	0.7091	0.7085	<b>0.7180</b>	<u>0.7114</u>
CSI <sup>30</sup>	0.4221	0.4191	0.4413	<u>0.4492</u>	0.4421	<b>0.4549</b>	0.4487
POD <sup>30</sup>	0.7074	0.6978	<u>0.7162</u>	0.6924	0.7036	0.7038	<b>0.7201</b>
FAR <sup>30</sup>	0.5246	0.5217	0.4999	<b>0.4700</b>	0.4875	<u>0.4702</u>	0.4902
HSS <sup>30</sup>	0.5516	0.5483	0.5721	<u>0.5815</u>	0.5729	<b>0.5873</b>	0.5804
CSI <sup>40</sup>	0.2098	0.2203	0.2318	<u>0.2330</u>	<b>0.2332</b>	0.2232	0.2316
POD <sup>40</sup>	0.6445	0.6111	0.6341	<u>0.6135</u>	0.6013	<b>0.6564</b>	<u>0.6449</u>
FAR <sup>40</sup>	0.8118	0.7942	0.7841	<b>0.7655</b>	<u>0.7764</u>	0.7918	<u>0.7865</u>
HSS <sup>40</sup>	0.3029	0.3169	<u>0.3316</u>	0.3311	<b>0.3324</b>	0.3194	<u>0.3316</u>

meteorologically belong to the free atmosphere, experiencing comparatively less influence from the Earth's surface. Analysis of the experimental results indicates that utilizing the M and H layers as 3-D cloud-assisted prediction data significantly enhances the overall radar echo extrapolation metrics. This underscores the effectiveness of employing deep learning methodologies to more efficiently extract information related to the movement states of clouds.

In a similar vein, the training of the model with 3-D cloud structures across multiple altitude layers has unveiled the superior performance of 3dCloudNet (L-M), 3dCloudNet (M-H), and 3dCloudNet (L-M-H) across various metrics when compared to the exclusive utilization of 3dCloudNet (L), 3dCloudNet (M), and 3dCloudNet (H) derived from five altitude layers. A meticulous analysis revealed that 3dCloudNet (M-H) when excluding the near-surface layer, achieved optimal outcomes across nearly all metrics. Conversely, the incorporation of the near-surface layer in 3dCloudNet (L-M) and 3dCloudNet (L-M-H) resulted in declines in specific metrics compared to 3dCloudNet (M) and 3dCloudNet (M-H). These findings strongly imply that the inclusion of the near-surface layer may introduce perturbations to the overall prediction outcomes during the training process of deep learning models. Overall, the experiments conducted in this study underscore that the exclusive use of M-H altitude layers (altitude range: 3.5–9 km) in modeling 3-D cloud structures significantly enhances the prediction of radar echo motion states at the 3 km altitude, manifesting as the most favorable enhancement effects.

We redesigned the experiment in the same dataset as before to further assess the impact of 3-D cloud structures on radar echo extrapolation for longer-term forecasting. This experiment employed a shorter input time span of six time steps, coupled with an extended forecasting period of 18 time steps, effectively predicting the next 1.5 h using data from the preceding 0.5 h. The result representation of the experiment on the test set is presented in Fig. 8.

As evident from Fig. 8, the combination of a shorter input period and a longer forecast duration poses challenges to all models. Nevertheless, across multiple evaluation metrics, the 3dCloudNet model, which incorporates 3-D cloud structure information, demonstrates superior forecasting capabilities.

## V. DISCUSSION

This study introduces and evaluates a new deep-learning model for precipitation nowcasting using 3-D cloud data. The proposed 3dCloudNet model incorporates 3-D cloud structures and attention-based spatiotemporal feature fusion modules, and these enhancements capture the motion patterns and evolution trends of clouds more effectively. The experimental results show that the 3dCloudNet model surpasses other competitive models in terms of multiple evaluation indicators, thus confirming the feasibility and rationality of the proposed model structure. In addition, the 3dCloudNet model improves performance in predicting future high-intensity echoes, indicating its sensitivity and ability to distinguish regions that could lead to severe convective weather.

It is crucial to acknowledge the limitations and practical considerations of applying the proposed model. One consideration is to carefully input all radar echo altitude level data into the model. It is recommended to experimentally select appropriate altitude levels as auxiliary data for precipitation nowcasting. Factors such as geographical location and radar data quality should be considered. By adopting this approach, the model ensures the utilization of relevant and trustworthy data sources, enhancing the accuracy and effectiveness of the nowcasting model. Meanwhile, in the encoder module of this study, we observed redundant content in the input's low-level semantic information. For example, clouds at different altitude levels exhibit similar motion tendencies within the same weather process. Hence, we adopted the mid-level fusion approach for information integration. Experimental results validated the effectiveness of this fusion method. However, it is necessary to evaluate whether mid-level fusion is the optimal approach for fusion. Future research should explore alternative fusion methods for different altitude levels of information.

While this study has demonstrated the superiority of the 3dCloudNet model for precipitation nowcasting using 3-D cloud data, it is essential to acknowledge the limitations of using only one source of information. Weather forecasting is a complex process that involves multiple physical factors and parameters, and incorporating additional physical information beyond radar data may further improve the accuracy of predictions. In future research, we plan to extend the

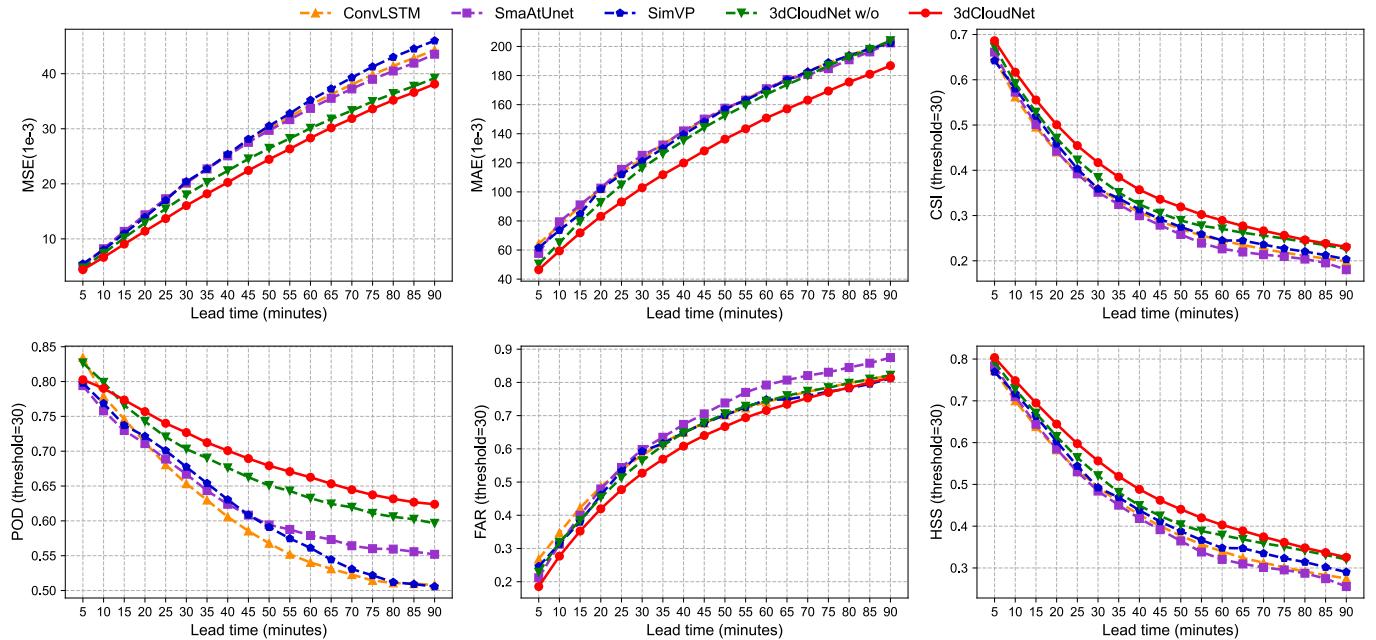


Fig. 8. Frame-wise performance of different comparison models in longer forecast results.

3dCloudNet model to forecast other meteorological events and explore the possibility of integrating additional data sources to enhance predictive capabilities. In addition to utilizing 3-D cloud structures for predictive assistance, we will focus on integrating various data sources, including but not limited to satellite observations, ground-based observations, and meteorological radar, to obtain more comprehensive and accurate meteorological information. By leveraging the strengths of different data sources, we can overcome the limitations of individual sources and enhance the model's understanding and predictive capabilities of meteorological events.

## VI. CONCLUSION

Precisely forecasting rainfall intensity for a short duration can provide crucial information for warning of hazardous weather conditions and making emergency management decisions. Earlier studies on deep learning-based techniques mainly relied on historical 2-D radar echo images to estimate the possibility of rainfall over a short period, and little focus was given to investigating the impact of 3-D clouds on radar echo prediction. Consequently, this study presents a state-of-the-art neural network architecture, 3dCloudNet, with two distinctive features. First, it employs an innovative spatiotemporal image attention module to fuse and extract spatiotemporal information. Second, it integrates 3-D cloud structures into the model to facilitate prediction. Specifically, the proposed model conducts several convolutional and down-sampling operations on radar echo images at various altitudes to extract spatiotemporal features. Subsequently, an attention-based spatiotemporal feature fusion module is utilized to combine the extracted features. Finally, future spatiotemporal motion states are predicted using a feature decoding network.

This study provides an objective quantitative and qualitative evaluation of multiple models using the same weather radar dataset. The experimental results demonstrate that the

proposed model structure in this study outperforms other competitive models on multiple evaluation indicators when using 2-D data input, showing the feasibility and rationality of the proposed model structure. In addition, this study incorporates a 3-D cloud-assisted structure into the model. Comparative experiments indicate that this structure is better equipped to comprehend clouds' motion law and evolution trends, improving predictions' accuracy and reliability. The fully developed 3dCloudNet model proposed in this study better predicts future high-intensity echoes, indicating that the model has better sensitivity and discriminative power for regions that may lead to severe convective weather. The quantitative and qualitative evaluation of experimental results demonstrates that the 3dCloudNet model proposed in this study has significant superiority in precipitation nowcasting. Finally, we explored the impact of the 3-D structure of clouds at different altitudes on radar echo extrapolation through experiments. We analyzed the advantages of 3-D cloud assistance in precipitation nowcasting and examined potential uncertainties that may arise. In summary, we extend the current understanding of precipitation nowcasting methodologies, offering advancements in data utilization, model architecture, and overall predictive performance.

## REFERENCES

- [1] A. McGovern et al., "Using artificial intelligence to improve real-time decision-making for high-impact weather," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 10, pp. 2073–2090, Oct. 2017.
- [2] J. Sun et al., "Use of NWP for nowcasting convective precipitation: Recent progress and challenges," *Bull. Amer. Meteorol. Soc.*, vol. 95, no. 3, pp. 409–426, Mar. 2014.
- [3] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, Sep. 2015.
- [4] J. Fang, K. Huang, M. Du, Z. Zhang, R. Cao, and F. Yi, "Investigation on cloud vertical structures based on Ka-band cloud radar observations at Wuhan in Central China," *Atmos. Res.*, vol. 281, Jan. 2023, Art. no. 106492.

- [5] H. Letu et al., "Surface solar radiation compositions observed from Himawari-8/9 and Fengyun-4 series," *Bull. Amer. Meteorol. Soc.*, vol. 104, no. 10, pp. E1772–E1789, Oct. 2023.
- [6] H. L. Kuo, "Further studies of the parameterization of the influence of cumulus convection on large-scale flow," *J. Atmos. Sci.*, vol. 31, no. 5, pp. 1232–1240, Jul. 1974.
- [7] M. K. Tippett, J. T. Allen, V. A. Gensini, and H. E. Brooks, "Climate and hazardous convective weather," *Current Climate Change Rep.*, vol. 1, no. 2, pp. 60–73, Jun. 2015.
- [8] P. Hess and N. Boers, "Deep learning for improving numerical weather prediction of heavy rainfall," *J. Adv. Model. Earth Syst.*, vol. 14, no. 3, Mar. 2022, Art. no. e2021MS002765.
- [9] G. Hu et al., "Progress, challenges, and future steps in data assimilation for convection-permitting numerical weather prediction: Report on the virtual meeting held on 10 and 12 November 2021," *Atmos. Sci. Lett.*, vol. 24, no. 1, p. e1130, Jan. 2023.
- [10] N. E. H. Bowler, C. E. Pierce, and A. Seed, "Development of a precipitation nowcasting algorithm based upon optical flow techniques," *J. Hydrol.*, vol. 288, nos. 1–2, pp. 74–91, Mar. 2004.
- [11] H. Sakaino, "Spatio-temporal image pattern prediction method based on a physical model with time-varying optical flow," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 3023–3036, May 2013.
- [12] W. Woo and W. Wong, "Application of optical flow techniques to rainfall nowcasting," in *Proc. 27th Conf. Severe Local Storms*, 2014.
- [13] W.-C. Woo and W.-K. Wong, "Operational application of optical flow techniques to radar-based rainfall nowcasting," *Atmosphere*, vol. 8, no. 3, p. 48, Feb. 2017.
- [14] A. J. Simmons and B. J. Hoskins, "The life cycles of some nonlinear baroclinic waves," *J. Atmos. Sci.*, vol. 35, no. 3, pp. 414–432, Mar. 1978.
- [15] T. Zhang et al., "Estimating hourly surface shortwave radiation over Northeast of the Tibetan Plateau by assimilating Himawari-8 cloud optical thickness," *Geosci. Lett.*, vol. 11, no. 1, p. 1, Jan. 2024.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [17] Z. Gao et al., "Earthformer: Exploring space-time transformers for Earth system forecasting," 2022, *arXiv:2207.05833*.
- [18] M. R. Ehsani, A. Zarei, H. V. Gupta, K. Barnard, E. Lyons, and A. Behrangi, "Nowcasting-nets: Representation learning to mitigate latency gap of satellite precipitation products using convolutional and recurrent neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4706021.
- [19] Z. Ma, H. Zhang, and J. Liu, "MM-RNN: A multimodal RNN for precipitation nowcasting," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4101914.
- [20] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 802–810.
- [21] X. Shi et al., "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5617–5627.
- [22] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal lstms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 879–888.
- [23] Y. Wang, L. Jiang, M.-H. Yang, L.-J. Li, M. Long, and L. Fei-Fei, "Eidetic 3D LSTM: A model for video prediction and beyond," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 193–206.
- [24] Z. Lin, M. Li, Z. Zheng, Y. Cheng, and C. Yuan, "Self-attention ConvLSTM for spatiotemporal prediction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, Apr. 2020, pp. 11531–11538.
- [25] H. Wu, Z. Yao, J. Wang, and M. Long, "MotionRNN: A flexible model for video prediction with spacetime-varying motions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15435–15444.
- [26] Z. Ma, H. Zhang, and J. Liu, "PrecipLSTM: A meteorological spatiotemporal LSTM for precipitation nowcasting," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4109108.
- [27] G. Ayzel, T. Scheffer, and M. Heistermann, "RainNet v1.0: A convolutional neural network for radar-based precipitation nowcasting," *Geosci. Model Develop.*, vol. 13, no. 6, pp. 2631–2644, Jun. 2020.
- [28] K. Trebing, T. Stańczyk, and S. Mehrkanon, "SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture," *Pattern Recognit. Lett.*, vol. 145, pp. 178–186, May 2021.
- [29] C. Wang, P. Wang, P. Wang, B. Xue, and D. Wang, "Using conditional generative adversarial 3-D convolutional neural network for precise radar extrapolation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5735–5749, 2021.
- [30] Z. Gao, C. Tan, L. Wu, and S. Z. Li, "SimVP: Simpler yet better video prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 3170–3180.
- [31] Y. Wang, Z. Gao, M. Long, J. Wang, and P. S. Yu, "PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning," in *Proc. PMLR Int. Conf. Mach. Learn.*, Jul. 2018, pp. 5123–5132.
- [32] Y. Wang, J. Zhang, H. Zhu, M. Long, J. Wang, and P. S. Yu, "Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9154–9162.
- [33] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, Munich, Germany. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [34] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [35] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.
- [36] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
- [37] S. Ravuri et al., "Skilful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, Sep. 2021.
- [38] Y. Zhang et al., "Skilful nowcasting of extreme precipitation with NowcastNet," *Nature*, vol. 619, no. 7970, pp. 526–532, Jul. 2023.
- [39] X. Pan, Y. Lu, K. Zhao, H. Huang, M. Wang, and H. Chen, "Improving nowcasting of convective development by incorporating polarimetric radar variables into a deep-learning model," *Geophys. Res. Lett.*, vol. 48, no. 21, Nov. 2021, Art. no. e2021GL095302.
- [40] D.-K. Kim et al., "Improving precipitation nowcasting using a three-dimensional convolutional neural network model from multi parameter phased array weather radar observations," *Atmos. Res.*, vol. 262, Nov. 2021, Art. no. 105774.
- [41] N. Sun, Z. Zhou, Q. Li, and J. Jing, "Three-dimensional gridded radar echo extrapolation for convective storm nowcasting based on 3D-ConvLSTM model," *Remote Sens.*, vol. 14, no. 17, p. 4256, Aug. 2022.
- [42] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.
- [43] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [44] A. Dosovitskiy et al., "An image is worth  $16 \times 16$  words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [45] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
- [46] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2018, pp. 7132–7141.
- [47] J. Dai et al., "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.
- [48] R. B. Stull, *An Introduction to Boundary Layer Meteorology*, vol. 13. Berlin, Germany: Springer, 1988.
- [49] School of Meteorology, University of Oklahoma, "GridRad-severe—Three-dimensional gridded NEXRAD WSR-88D radar data for severe events," Res. Data Arch. Nat. Center Atmos. Res., Comput. Inf. Syst. Lab., Boulder, CO, USA, 2021. Accessed: Jun. 15, 2023, doi: [10.5065/2B46-1A97](https://doi.org/10.5065/2B46-1A97).
- [50] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, "Visual attention network," 2022, *arXiv:2202.09741*.
- [51] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [52] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.



**Yuankang Ye** received the bachelor's degree in automation from Harbin Engineering University, Harbin, China, in 2020, where he is currently pursuing the Ph.D. degree in intelligent weather forecasting with the School of Intelligent Science and Engineering.

Subsequently, he conducted research at the School of Intelligent Science and Engineering Technology, Harbin Engineering University, from 2020 to 2022, during the master's program. His research focused on processing weather radar data and conducting short-term and near-term rainfall predictions.



**Chang Liu** received the Ph.D. degree from the College of Automation, Harbin Engineering University, Harbin, China, in 2013.

He served at the College of Automation, Harbin Engineering University. From 2015 to 2016, he was a Visiting Scholar with the University of Wisconsin-Madison, Madison, WI, USA, and NOAA/GFDL, Princeton University, Princeton, NJ, USA. He is currently an Associate Professor with the College of Intelligent System Science and Engineering, Harbin Engineering University. His research interests include marine big data information and ocean data assimilation.



**Feng Gao** received the Ph.D. degree from the College of Automation, Harbin Engineering University, Harbin, China, in 2013.

She served at the College of Automation, Harbin Engineering University, where she is currently an Associate Professor with the College of Intelligent System Science and Engineering. Her research interests include marine big data information and ocean data assimilation.



**Shaoqing Zhang** received the Ph.D. degree from the Department of Meteorology, Florida State University, Tallahassee, FL, USA, in 2000.

From 2002 to 2008, he was a Research Scientist with UCAR/GFDL, Princeton University, Princeton, NJ, USA. From 2008 to 2016, he was a Federal Researcher and a Professor at the NOAA/GFDL. Since 2016, he has been a Distinguished Professor with the Key Laboratory of Physical Oceanography, Ministry of Education, Ocean University of China (POL/OUC), Qingdao, China, an Outstanding Scientist with the Qingdao Pilot National Laboratory for Marine Science and Technology (QNLML), Qingdao, and the Leading PI of High-Resolution Earth System Modeling and Prediction Initiative at QNLML. His major research interests include coupled Earth system modeling, seamless weather-climate predictability, coupled model data assimilation, and parameter estimation.



**Wei Cheng** received the Ph.D. degree from the Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, in 2012.

He is a Senior Engineer with the Beijing Institute of Applied Meteorology, Beijing. His major research interests include satellite data assimilation, regional numerical weather forecasting, and artificial intelligence in meteorological research.



**Shudong Wang** received the Ph.D. degree from the Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, in 2011.

He served at the Public Meteorological Service Center, China Meteorological Administration, Beijing. His major research interests include weather observation data quality control and radar and satellite data application.