remote sensing

MDPI

Article

Spatiotemporal Dual-Polarization Radar Quantitative Precipitation Estimation Using Convolutional Neural Networks

Chenhao Xie [1,2,3], Yonghua Zhang [4], Yi Zhang [1,2,3,*], Haonan Chen [5], Li Li [6], Yuxin Chen [1,2,3] and Jian Sun [7,8]

1 Guangzhou Meteorological Integrated Security Center, Guangzhou 511430, China
2 Collaborative Innovation Center for Smart Meteorology of Guangzhou, Guangzhou 511430, China
3 GBA Academy of Meteorological Research, Guangzhou 510640, China
4 Guangdong Meteorological Service Center, Guangzhou 510640, China
5 Artificial Intelligence and Remote Sensing Laboratory, Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA
6 Guangzhou Ecological and Agricultural Meteorological Center, Guangzhou 511430, China
7 HuaYun METSTAR Radar (Beijing) Company Ltd., Beijing 100094, China
8 Key Laboratory of Intelligent Meteorological Observation Technology, China Meteorological Administration, Beijing 100081, China
* Correspondence: zhangy@gd121.cn

Highlights

What are the main findings?

- The time-resolved 3D baseline had a lower overall MAE (1.321 versus 1.524 mm h^{-1}) and RMSE (3.088 versus 3.307 mm h^{-1}) than the temporally averaged 2D baseline.
- 3DHybrid-QPENet combined ConvLSTM-based temporal and multiscale representations and had the lowest overall MAE (1.270 mm h^{-1}) and RMSE (2.971 mm h^{-1}), with a Bias of −0.537 mm h^{-1} and a CC of 0.866.

What are the implications of the main findings?

- Consecutive radar scans retain scan-to-scan information that is lost through temporal averaging in hourly radar QPE.
- For rainfall rates above 50 mm h^{-1}, 3DHybrid-QPENet reduced estimation errors relative to the other evaluated methods, although its Bias of −20.031 mm h^{-1} shows that underestimation persists.

Abstract

Hourly quantitative precipitation estimation (QPE) from weather radar remains difficult during intense rainfall because fixed empirical relations and temporal averaging may not represent rapid scan-to-scan changes and localized echo structure. We evaluated five convolutional neural network configurations using S-band dual-polarization radar and rain-gauge observations in Guangzhou, China. A locally calibrated composite dual-polarization estimator (Composite-DP) provided a conventional reference. Each sample contains Z_H, Z_{DR}, and K_{DP} from ten consecutive 0.5° scans over the preceding hour in a 3 × 10 × 21 × 21 tensor. Three-dimensional convolution spans time and two horizontal dimensions, without vertical-volume input. The dataset from 2018 to 2021 comprised 3,008,738 matched samples with rainfall ≥ 0.1 mm h^{-1}, including 2,611,457 for training, 149,221 from six 2020 events for validation, and 141,428 from six 2021 events for testing. The models included a temporally averaged 2D baseline (2DPoly-QPENet), a time-resolved 3D baseline (3DPoly-QPENet), separate temporal (3DTime-QPENet) and multiscale (3DEcho-QPENet) extensions, and their fusion (3DHybrid-QPENet). The temporal extension represented the encoded ten-step sequence with a convolutional long short-term memory (ConvLSTM) layer, whereas the multiscale extension used an encoder–decoder pathway. On the independent test set,

Check for updates

Academic Editors: Filomena Romano and Zhiqun Hu

Received: 3 October 2026
Revised: 6 November 2026
Accepted: 10 November 2026
Published: 17 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Composite-DP yielded an RMSE of 4.139 mm h^{-1}. Compared with 2DPoly-QPENet, 3DPoly-QPENet had a lower mean absolute error (MAE; 1.321 versus 1.524 mm h^{-1}) and root-mean-square error (RMSE; 3.088 versus 3.307 mm h^{-1}), while its correlation coefficient (CC) was higher (0.860 versus 0.832). Among the individual extensions, 3DEcho-QPENet had lower overall errors than 3DTime-QPENet. Their fusion, 3DHybrid-QPENet, achieved the lowest overall MAE (1.270 mm h^{-1}) and RMSE (2.971 mm h^{-1}), with a Bias of −0.537 mm h^{-1} and the highest CC (0.866). Across the baseline and extension comparisons, the time-resolved 3D representation and the joint temporal and multiscale configuration were associated with lower hourly QPE errors. 3DHybrid-QPENet also had the lowest relative error (RE), MAE, and RMSE in all five rainfall-intensity classes. Above 50 mm h^{-1}, however, its Bias remained −20.031 mm h^{-1}, indicating persistent underestimation. Generalization beyond the study region remains to be evaluated.

Keywords: quantitative precipitation estimation; dual-polarization weather radar; precipitation remote sensing; spatiotemporal convolutional neural network; time-resolved radar observations; multiscale echo representation; deep learning; short-duration heavy rainfall

1. Introduction

Reliable quantitative precipitation estimation (QPE) supports hydrological applications and the monitoring of short-duration heavy rainfall in densely populated urban areas. Guangzhou is a coastal megacity in South China with a subtropical monsoon climate, and much of its annual rainfall occurs during the flood season from May to August [1]. Short-duration, high-intensity rainfall can generate sharp spatial gradients and localized rainfall maxima. Rapid urbanization and increasing impervious surface coverage can further intensify pluvial flood impacts [2]. These conditions require rainfall estimates with broad spatial coverage, high temporal resolution, and low estimation error.

Rain gauges provide direct point measurements but cannot fully resolve the spatial organization of convective rainfall when station spacing is larger than the scale of individual rain cells. Weather radar provides frequent observations over broad areas and is therefore widely used for QPE. Conventional radar QPE often relies on an empirical relation between radar reflectivity factor (Z) and rainfall rate (R), such as the Marshall–Palmer Z–R relation [3]. Fixed coefficients, however, may not adequately represent variations in drop size distribution or precipitation type, while radar-based rainfall estimates can also be affected by calibration errors, attenuation, beam blockage, and sampling differences between radar and gauges [4]. Dual-polarization weather radar provides additional measurements, including horizontal reflectivity factor (Z_H), differential reflectivity (Z_{DR}), and specific differential phase (K_{DP}), which describe complementary properties of hydrometeors. In particular, K_{DP} is less sensitive than Z_H to reflectivity calibration errors and attenuation, while multiparameter relations combining Z_H, Z_{DR}, and K_{DP} can complement reflectivity-only Z–R estimates [4].

Data-driven methods have consequently been introduced to extract precipitation-related information from radar observations and to learn nonlinear mappings between polarimetric radar variables and surface rainfall. In related radar-imagery applications, supervised learning has been used to detect rain in X-band marine radar observations [5], while unsupervised learning has been used to distinguish rain-affected and low-backscatter regions [6]. For weather-radar QPE, machine-learning estimators based on operational dual-polarization radar data have produced lower errors than conventional empirical relations in several evaluations [7]. Deep neural networks and convolutional neural networks (CNNs)

have subsequently been used to combine polarimetric variables, local radar neighborhoods, and rain-gauge observations [8,9]. Sequence-based QPE models have further extended the representation of temporal radar information [10], while convolutional long short-term memory (ConvLSTM) networks have been applied to precipitation prediction using multichannel radar data [11]. Other deep-learning studies have examined radar QPE using spatial and multi-parameter representations [12,13], and graph neural networks have been introduced to represent nonlocal relationships within radar QPE data [14]. These studies show that learned mappings can improve radar QPE, but performance still depends strongly on how temporal, spatial, and polarimetric information is organized before and within the network.

Many CNN-based QPE models use two-dimensional (2D) convolutions and represent measurements from different scans or radar variables as stacked channels, or compress the temporal dimension before feature extraction. Such representations are computationally convenient, but temporal averaging can remove scan-to-scan changes associated with the growth, movement, and decay of precipitation echoes. Three-dimensional (3D) convolution provides a general mechanism for learning local spatiotemporal features [15]. Three-dimensional CNNs combined with recurrent structures have also been used to model radar echo evolution in precipitation nowcasting [16]. However, nowcasting predicts future radar or precipitation fields, whereas QPE estimates rainfall for an observed period; results from one task therefore cannot be treated as direct evidence for the other. For the time-resolved QPE models developed here, ten single-elevation 0.5° scans form the temporal dimension of the input. Three-dimensional convolution operates across time and the two horizontal dimensions of the radar neighborhood ($T \times H \times W$), while Z_H, Z_{DR}, and K_{DP} serve as the input channels.

A second issue is the representation of precipitation structures at different spatial scales. Repeated convolution and pooling can enlarge the receptive field but may weaken localized maxima and rainband boundaries. Encoder–decoder networks with skip connections provide a way to combine hierarchical context with finer spatial information [17,18]. At the same time, radar–gauge datasets are usually dominated by light-rain samples, with relatively few intense-rainfall observations. This long-tailed distribution makes overall performance alone insufficient for evaluating a QPE model. Comparisons should therefore consider both aggregate statistics and errors across rainfall-intensity classes, particularly the bias and error magnitude for heavy rainfall.

This study addresses two related representation challenges in hourly QPE using observations from the Guangzhou S-band dual-polarization radar and a dense rain-gauge network: retaining scan-to-scan evolution in dual-polarization radar observations and integrating this temporal information with multiscale horizontal echo structure. Each radar–gauge sample contains Z_H, Z_{DR}, and K_{DP} from ten consecutive 0.5° scans over a 21 × 21 gauge-centered neighborhood, forming a 3 × 10 × 21 × 21 input tensor. Observations from 2018 to 2021 are used to construct the dataset, with six complete precipitation events from 2021 held out for independent testing. A temporally averaged 2D baseline and a time-resolved 3D baseline are first compared under the same data partition, training objective, and evaluation procedure. Among the additional time-resolved models, 3DTime-QPENet combines a global representation of the encoded tensor with a ConvLSTM representation of all ten encoded time steps, whereas 3DEcho-QPENet uses an encoder–decoder pathway with skip connections to represent multiscale echo structure. 3DHybrid-QPENet combines these global, ConvLSTM, and multiscale representations for hourly rainfall estimation.

The main contributions of this study are as follows:

1. A time-resolved dual-polarization radar–gauge representation is constructed from ten consecutive low-elevation scans, preserving scan-to-scan changes that are removed by temporal averaging. Under a common experimental protocol, its performance is compared with that of a temporally averaged 2D representation. This comparison evaluates the combined change in input organization and convolutional dimensionality.
2. 3DHybrid-QPENet combines the global and ConvLSTM representations used for temporal modeling with a skip-connected encoder–decoder representation of multiscale echo structure. Comparisons with 3DTime-QPENet and 3DEcho-QPENet under the same experimental protocol evaluate the temporal, multiscale, and joint model configurations.
3. The models are evaluated on an independent test set comprising six held-out 2021 events using overall and rainfall-intensity-stratified metrics, with Composite-DP serving as the conventional dual-polarization reference. The stratified results characterize the remaining bias at high rainfall intensities, while paired event-level comparisons examine whether the RMSE differences between the hybrid model and the two individual extensions are consistent across the test events. Two later intense-rainfall cases provide complementary spatial comparisons.

The remainder of this paper is organized as follows. Section 2 describes the study area, radar and rain-gauge data, sample construction, QPE methods, training strategy, and evaluation metrics. Section 3 presents the overall, rainfall-intensity-dependent, scatter-density, and case-based results. Section 4 discusses the interpretation and limitations of these results, and Section 5 summarizes the main conclusions.

2. Materials and Methods

2.1. Study Area and Observational Data

The study domain was defined as the area within a 100 km radius of the Guangzhou S-band dual-polarization weather radar (CINRAD/SAD; HuaYun METSTAR Radar (Beijing) Company, Ltd., Beijing, China) in South China. Only observations from the 0.5° plan position indicator (PPI) scan were used. The 0.5° scan was available at approximately 6 min intervals, with a native range-gate spacing of 250 m.

Radar observations acquired from 1 January 2018 to 31 December 2021 were paired with hourly precipitation records from national and regional automatic rain-gauge stations over the same period. A total of 1353 rain-gauge stations were available within the study domain (Figure 1). The gauge observations consisted of 1 h precipitation accumulations in millimeters. For model development and presentation of the QPE fields, these observations were expressed as the corresponding hourly precipitation rate in mm h^{-1}. The quality control and radar–gauge matching procedures are described in Sections 2.2 and 2.3, respectively.

For the case-based analysis, two later rainfall events were selected because complete radar and rain-gauge observations were available for both and the events exhibited different spatial rainfall patterns. The corresponding hourly QPE fields ended at 11:30 Beijing Time (BJT; UTC+8) on 18 August 2023 and 10:30 BJT on 21 April 2024. Both events occurred outside the 2018–2021 model-development period and were not used for neural-network training, validation, or checkpoint selection. They were used to compare the spatial rainfall estimates from Composite-DP and the five neural-network QPE models.

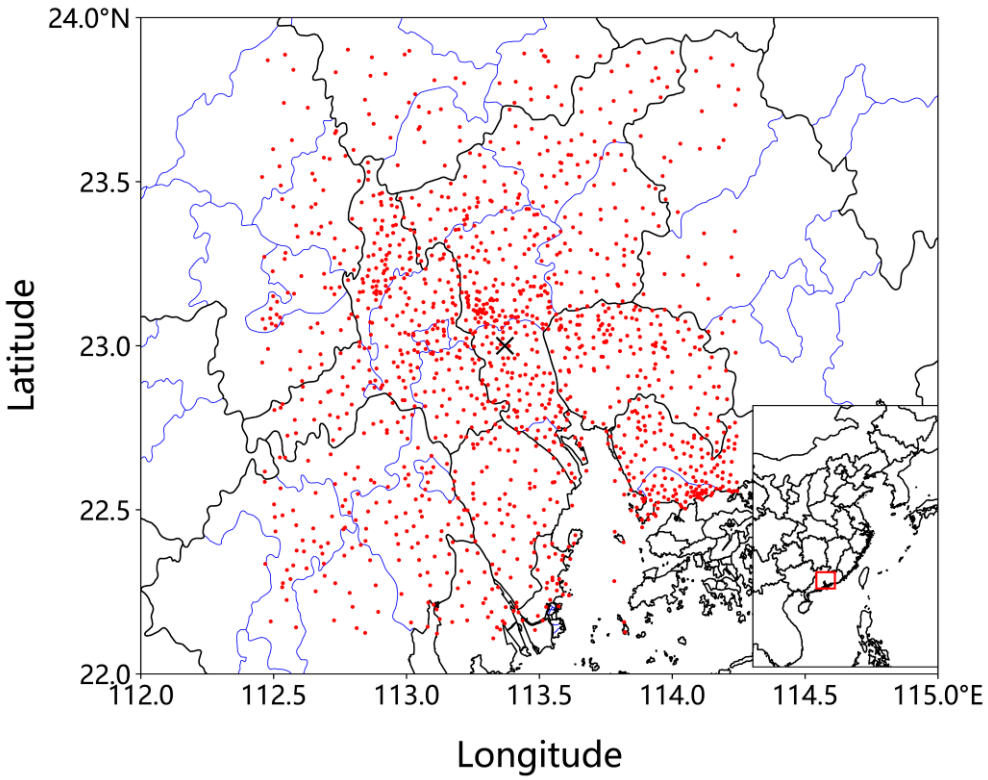


Figure 1. Study area and spatial distribution of automatic rain-gauge stations within the 100 km coverage radius of the Guangzhou S-band dual-polarization weather radar.The blue lines delineate district- and county-level administrative boundaries, the black cross marks the radar location, and the red dots denote the 1353 national and regional rain-gauge stations available for radar–gauge matching. The inset shows the location of the study area in China.

2.2. Radar and Rain-Gauge Quality Control

Only the 0.5° PPI scans described in Section 2.1 were retained for model input because the lowest available elevation reduces the vertical separation between the radar sampling layer and the surface rain gauges [19]. Persistent partial beam blockage affected the radar-relative azimuth sectors of 269–272°, 278–281°, 287–293°, and 295–298°. Accordingly, 59 of the 1353 available rain-gauge stations (4.36%) within these sectors were excluded before radar–gauge matching to prevent blocked low-level radar observations from entering the model-development and evaluation samples. This screening reduced station coverage in the affected directions and may therefore limit the spatial and topographic representativeness of the dataset in those sectors.

Before sample construction, the radar and hourly rain-gauge observations underwent routine operational quality control. Radar processing included the removal of isolated and nonmeteorological echoes, spatial filtering of Z_H and Z_{DR}, processing of Φ_{DP}, and estimation of K_{DP} [20,21]. Hourly rain-gauge records that did not pass the operational quality-control procedure were excluded. The three radar variables retained as model inputs were horizontal reflectivity Z_H (dBZ), differential reflectivity Z_{DR} (dB), and specific differential phase K_{DP} (° km^{-1}). Differential phase Φ_{DP} (°) was used only as an intermediate variable for deriving K_{DP}.

Following quality control, a binary validity mask was generated to identify invalid or missing radar entries. The normalization parameters for each radar variable were calculated using valid entries only. After normalization, the masked positions were assigned a value of zero to preserve fixed input dimensions. Consequently, invalid and missing entries did

not affect the normalization statistics, and zeros at the masked positions served only as numerical fill values.

2.3. Time-Resolved Radar–Gauge Sample Construction

Each sample paired an hourly rain-gauge observation with a sequence of preceding low-elevation radar measurements. The end time of the 1 h gauge accumulation was used as the reference time. Ten consecutive 0.5° PPI scans, acquired at nominal intervals of 6 min, were selected from the preceding 1 h window, as illustrated in Figure 2. Only samples containing all ten scans were retained, giving each sample a fixed temporal dimension.

Before sample construction, the 0.5° PPI fields were mapped to a regular Cartesian grid with a horizontal spacing of 250 m. For each rain-gauge station, a 21 × 21 grid-cell neighborhood centered on the corresponding grid point was extracted from every selected scan, giving a nominal horizontal extent of 5.25 km × 5.25 km. The same spatial window was extracted for horizontal reflectivity Z_{H}, differential reflectivity Z_{DR}, and specific differential phase K_{DP} [22]. The neighborhood provides local spatial context around each station and reduces sensitivity to the point–area representativeness difference between rain-gauge and radar observations. The gauge-centered window remained fixed throughout the ten-scan sequence, so the input represents local scan-to-scan evolution around the station rather than the full displacement of a moving precipitation system. Rapidly advecting echoes may therefore enter or leave the sampled neighborhood during the accumulation period.

The resulting input was organized as a four-dimensional tensor with dimensions $C \times T \times H \times W = 3 \times 10 \times 21 \times 21$, where $C = 3$ denotes the three radar variables, $T = 10$ denotes the consecutive radar scans, and $H = W = 21$ denotes the spatial neighborhood. The corresponding label was the hourly precipitation rate at the rain-gauge station, expressed in mm h^{-1} from the 1 h gauge accumulation. Thus, the three convolutional dimensions correspond to time and the two horizontal spatial dimensions. After quality control, temporal-completeness screening, and radar–gauge matching, 3,008,738 valid samples were retained.

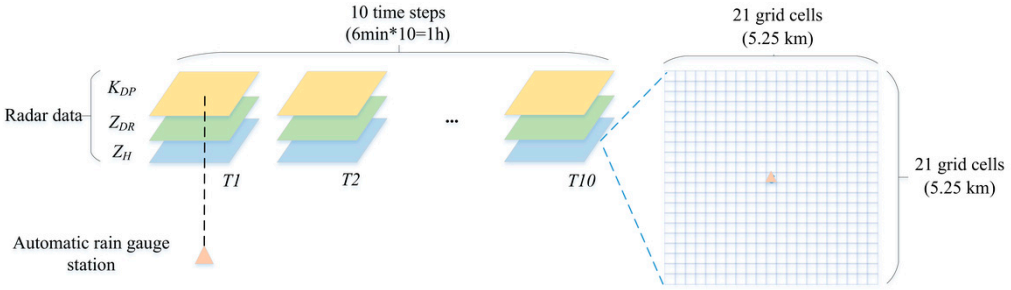


Figure 2. Construction of a time-resolved radar–gauge sample. Before neighborhood extraction, the 0.5° PPI fields were mapped to a regular Cartesian grid with a horizontal spacing of 250 m. Horizontal reflectivity Z_{H}, differential reflectivity Z_{DR}, and specific differential phase K_{DP} were then extracted from ten consecutive 0.5° PPI scans within the 1 h window ending at the rain-gauge timestamp. For each variable and scan, a 21 × 21 grid-cell neighborhood centered on the gauge location and covering a nominal horizontal extent of 5.25 km × 5.25 km was retained, producing an input tensor with dimensions 3 × 10 × 21 × 21. The target is the corresponding hourly precipitation rate (mm h^{-1}) obtained from the 1 h rain-gauge accumulation.

2.4. Independent Test Set and Development-Set Partitioning

To characterize the long-tailed distribution of the precipitation samples and evaluate model performance across rainfall intensities, the hourly precipitation target R_h was divided into five intensity classes: $0.1 \leq R_h \leq 5$, $5 < R_h \leq 15$, $15 < R_h \leq 20$, $20 < R_h \leq 50$, and

$R_h > 50$ mm h^{-1}. Of the 3,008,738 retained samples, 87.24%, 9.56%, 1.31%, 1.76%, and 0.12% fell within these classes, respectively; the highest-intensity class contained 3700 samples.

To reduce the possibility that temporally adjacent scans or radar–gauge samples from the same precipitation system would be shared across subsets, the dataset was partitioned by complete precipitation events. Six events occurring between March and August 2020 were assigned to validation, and six events occurring between February and December 2021 were held out for independent testing; all remaining eligible samples were used for training. All radar–gauge samples from a given event were assigned to the same subset, and a 24 h exclusion buffer was applied on both sides of each validation and test event. Samples within the buffer periods were excluded from model development and evaluation. The resulting subsets contained 2,611,457 training samples, 149,221 validation samples, and 141,428 test samples; a total of 106,632 samples within the buffer periods were excluded.

The training subset was used to optimize model parameters. The validation subset was used to monitor model performance, adjust the learning rate, apply early stopping, and select the final model checkpoint. The independent test subset was evaluated only after the model configuration and checkpoint had been fixed. Because all subsets were drawn from the same radar–gauge network, the independent test assesses previously unseen precipitation events within the study region, whereas transfer to new gauge locations was not evaluated. The variable-specific normalization parameters were calculated from valid entries in the training subset and then applied unchanged to the validation and independent test subsets.

2.5. QPE Models

Five neural-network models were evaluated using the same radar–gauge samples, precipitation targets, and model-development protocol. The model family comprised a temporally averaged two-dimensional baseline (2DPoly-QPENet), a time-resolved three-dimensional baseline (3DPoly-QPENet), and a temporal model that combines a globally pooled representation with a ConvLSTM representation derived from the encoded feature sequence (3DTime-QPENet). The other two were a multiscale encoder–decoder model (3DEcho-QPENet) and a hybrid model that integrates the global and ConvLSTM representations with multiscale encoder–decoder features (3DHybrid-QPENet). Each model produced a scalar estimate of the hourly precipitation rate at the corresponding rain-gauge location. Figures 3–7 show the network architectures.

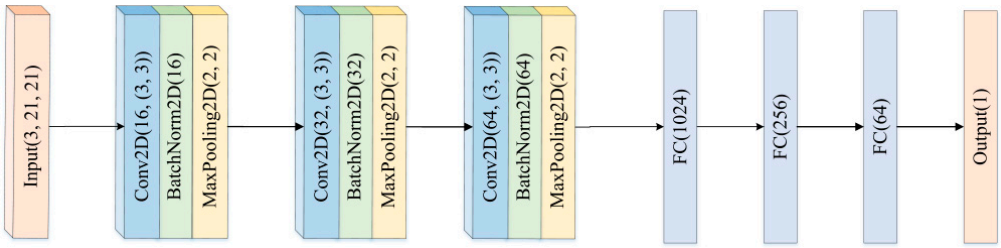


Figure 3. Architecture of 2DPoly-QPENet. The original ten-scan radar tensor is averaged along the temporal dimension to form a 3 × 21 × 21 input. Three Conv2D–BatchNorm–MaxPool2D blocks use 16, 32, and 64 output channels, respectively, followed by fully connected layers with 1024, 256, and 64 units. The one-unit output layer produces the hourly precipitation-rate estimate.

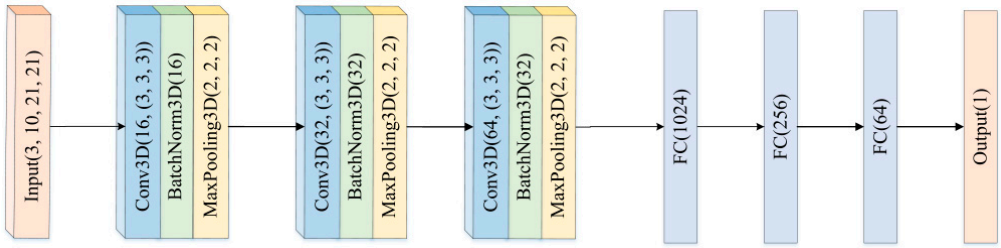


Figure 4. Architecture of 3DPoly-QPENet. The model receives the complete 3 × 10 × 21 × 21 radar tensor. Three Conv3D–BatchNorm3D–MaxPool3D blocks use 16, 32, and 64 output channels, respectively, with 3 × 3 × 3 convolutional kernels and 2 × 2 × 2 pooling windows. The extracted features are passed through fully connected layers with 1024, 256, and 64 units before the one-unit hourly precipitation-rate output.

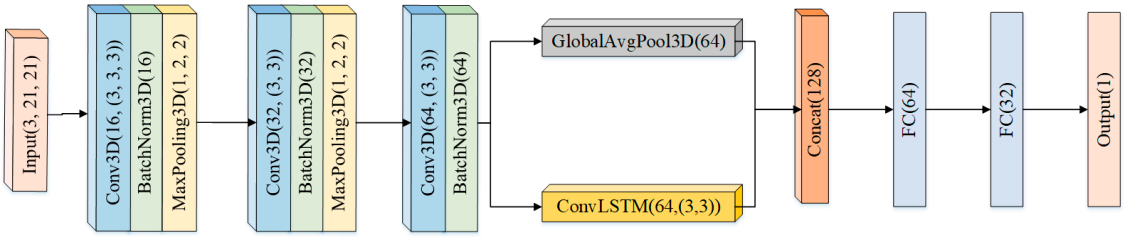


Figure 5. Architecture of 3DTime-QPENet. The 3 × 10 × 21 × 21 input passes through three Conv3D–BatchNorm3D stages with 16, 32, and 64 output channels. Max pooling with a 1 × 2 × 2 window follows the first two stages. The resulting ten-step feature sequence is summarized in parallel by global average pooling and a ConvLSTM layer with 64 hidden channels. Their 64-dimensional representations are concatenated and passed through fully connected layers with 64 and 32 units before the one-unit hourly precipitation-rate output.

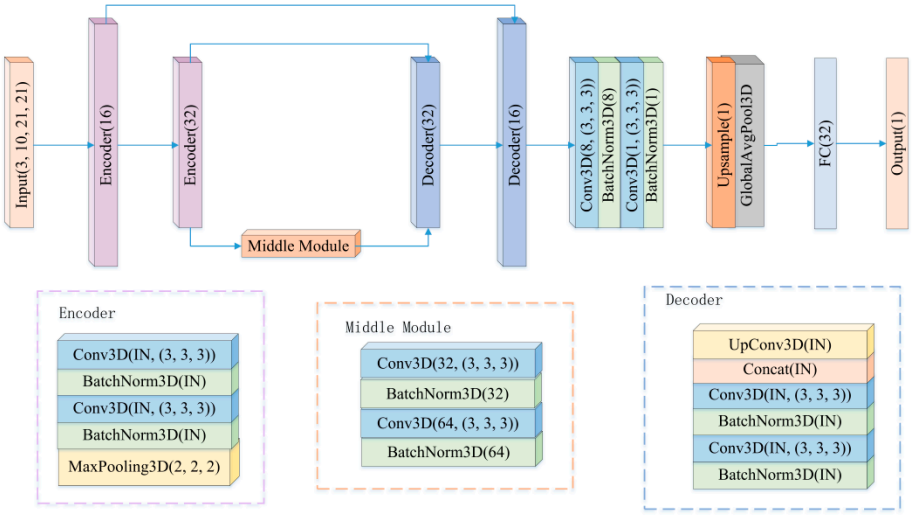


Figure 6. Architecture of 3DEcho-QPENet. The 3 × 10 × 21 × 21 input passes through two encoder stages with 16 and 32 output channels and a middle module with 32- and 64-channel convolutions. Two decoder stages with 32 and 16 output channels restore feature resolution and combine encoder and decoder features through skip connections. Subsequent convolutions reduce the channel number to 8 and then 1, followed by upsampling, global average pooling, a 32-unit fully connected layer, and a one-unit hourly precipitation-rate output.

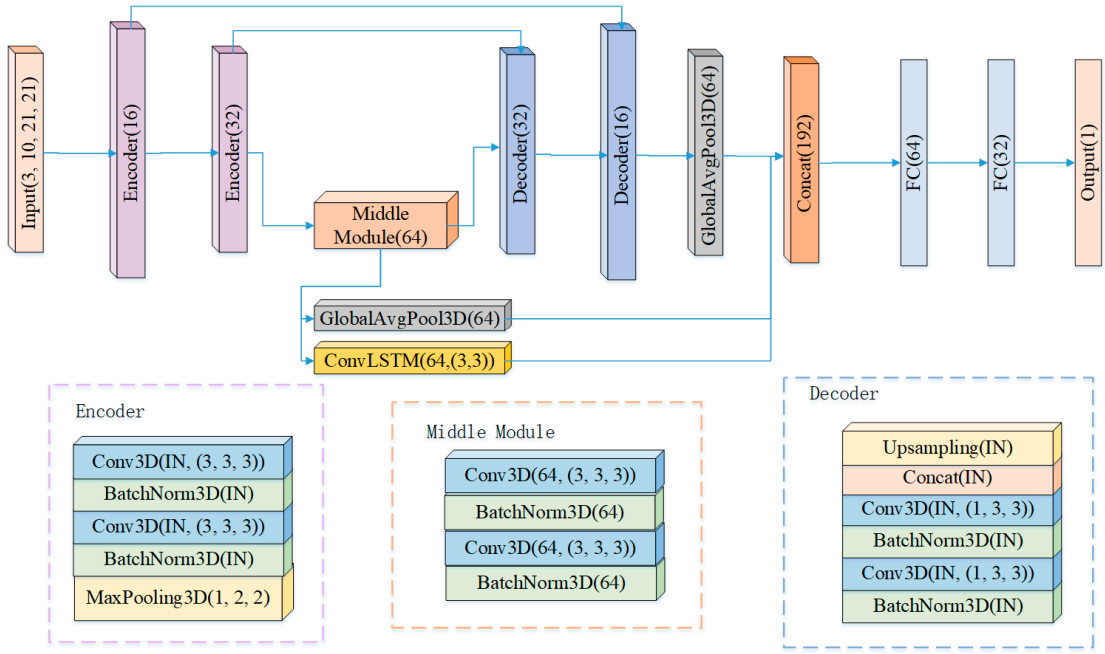


Figure 7. Architecture of 3DHybrid-QPENet. The 3 × 10 × 21 × 21 input passes through two encoder stages with 16 and 32 output channels and a 64-channel middle module. The encoder stages use 1 × 2 × 2 max pooling, and skip connections link the encoder outputs to decoder stages with 32 and 16 output channels. The middle-module tensor is also passed to global-average-pooling and ConvLSTM branches. The three 64-dimensional representations are concatenated into a 192-dimensional vector and passed through fully connected layers with 64 and 32 units before the one-unit hourly precipitation-rate output.

2.5.1. 2DPoly-QPENet and 3DPoly-QPENet Baselines

The 2DPoly-QPENet was used as the two-dimensional convolutional baseline. For this model, the original 3 × 10 × 21 × 21 radar tensor was averaged along the temporal dimension, producing an input with dimensions 3 × 21 × 21. This transformation retained the mean radar information over the ten scans while removing their scan-to-scan order.

The network contained three successive two-dimensional convolutional blocks. The blocks used 16, 32, and 64 output channels, respectively, with a 3 × 3 convolutional kernel in each block. Each convolutional layer was followed by batch normalization and 2 × 2 max pooling. The feature map produced by the third block was flattened and passed through fully connected layers containing 1024, 256, and 64 units. The final output layer contained one unit and produced the estimated hourly precipitation rate.

The 3DPoly-QPENet was constructed as the three-dimensional baseline using the complete 3 × 10 × 21 × 21 input tensor. Its overall architecture followed the 2D baseline, while the convolutional and pooling operations were extended across the temporal and two horizontal spatial dimensions. The network contained three Conv3D–BatchNorm3D–MaxPool3D blocks with 16, 32, and 64 output channels. Each convolution used a 3 × 3 × 3 kernel, and each pooling layer used a 2 × 2 × 2 window. The extracted features were flattened and processed by fully connected layers containing 1024, 256, and 64 units, followed by a one-unit regression output.

The comparison between 2DPoly-QPENet and 3DPoly-QPENet evaluates the combined effect of retaining the time-resolved radar sequence and applying three-dimensional convolution. It therefore provides the baseline for assessing the additional temporal, multiscale, and hybrid components introduced in the subsequent models.

2.5.2. 3DTime-QPENet and 3DEcho-QPENet

The 3DTime-QPENet receives a batch of complete radar sequences with dimensions $B \times 3 \times 10 \times 21 \times 21$, where B denotes batch size. Its three-stage three-dimensional convolutional pathway produces 16, 32, and 64 output channels in succession. Each stage consists of a $3 \times 3 \times 3$ convolution followed by batch normalization. Max pooling with a kernel and stride of $1 \times 2 \times 2$ follows the first two stages, yielding a feature tensor with dimensions $B \times 64 \times 10 \times 5 \times 5$.

Two representations are obtained from this tensor. Global average pooling across the temporal and horizontal dimensions produces a 64-dimensional global representation. In parallel, a ConvLSTM layer with 64 hidden channels and a 3×3 kernel processes the sequence of ten $64 \times 5 \times 5$ feature maps [23,24]. The spatially averaged hidden states are combined across the sequence using learned weights to produce a 64-dimensional temporal representation. The two representations are concatenated into a 128-dimensional vector and passed through fully connected layers with 64 and 32 units, followed by a one-unit layer that produces the hourly precipitation-rate estimate.

The 3DEcho-QPENet uses a two-level three-dimensional encoder–decoder architecture with skip connections [17,18]. It receives the complete $3 \times 10 \times 21 \times 21$ radar tensor. The two encoder stages produce 16- and 32-channel feature maps, respectively. Each stage contains two $3 \times 3 \times 3$ convolutional layers with batch normalization, followed by $2 \times 2 \times 2$ max pooling. The middle module further processes the encoded features using $3 \times 3 \times 3$ convolutions with 32 and 64 output channels.

The decoder contains two stages with 32 and 16 output channels. At each stage, an up-convolution operation increases the feature resolution, after which the upsampled features are concatenated with the corresponding encoder features through a skip connection. Two $3 \times 3 \times 3$ convolutional layers with batch normalization are then applied to the concatenated features. Following the second decoder stage, additional three-dimensional convolutions reduce the feature channels to 8 and then 1. The resulting feature map is processed by an additional upsampling layer and global average pooling, followed by a fully connected layer with 32 units and a one-unit hourly precipitation-rate output. This architecture combines information from different feature resolutions before regression.

2.5.3. 3DHybrid-QPENet

The 3DHybrid-QPENet receives a batch of radar sequences with dimensions $B \times 3 \times 10 \times 21 \times 21$. A two-stage encoder and a middle module form the shared feature-extraction pathway. The encoder stages contain two $3 \times 3 \times 3$ convolutions with batch normalization and produce 16 and 32 output channels, respectively. Each stage ends with max pooling using a kernel and stride of $1 \times 2 \times 2$. The middle module contains two $3 \times 3 \times 3$ convolutions with 64 output channels, producing a tensor with dimensions $B \times 64 \times 10 \times 5 \times 5$.

The encoder–decoder pathway contains two decoder stages with 32 and 16 output channels. Each stage upsamples the incoming feature map, concatenates it with the corresponding encoder output, and applies two $1 \times 3 \times 3$ convolutions with batch normalization. The output of the final decoder is spatially averaged at each time step and projected to 64 dimensions. It is then combined with the spatially averaged middle-module feature from the same time step. Averaging these combined features over time produces a 64-dimensional multiscale representation.

The middle-module tensor is also used to obtain global and temporal representations. Global average pooling across the temporal and horizontal dimensions produces a 64-dimensional global representation. In parallel, a ConvLSTM layer with 64 hidden channels and a 3×3 kernel processes the sequence of ten $64 \times 5 \times 5$ feature maps. As in 3DTime-

QPENet, the spatially averaged hidden states are combined across the sequence using learned weights to produce a 64-dimensional temporal representation.

The global, temporal, and multiscale representations are concatenated into a 192-dimensional vector. This vector is passed through fully connected layers with 64 and 32 units, followed by a one-unit layer that produces the hourly precipitation-rate estimate.

2.6. Locally Calibrated Composite Dual-Polarization QPE Estimator

A locally calibrated composite dual-polarization QPE estimator (Composite-DP) was used as the conventional reference for the complete independent test set and the two later event-based evaluations. To provide a polarimetric reference based on the same radar variables as the neural-network models, Composite-DP combines the $R(Z_H)$, $R(K_{DP})$, and $R(Z_H, Z_{DR})$ relations commonly used in dual-polarization radar QPE [4]. Let $Z = 10^{(ZH/10)}$ denote the linear radar reflectivity factor ($mm^6\ m^{-3}$). The three locally calibrated relations were

$$R(Z_H) = 0.21299983Z^{0.451011},$$

$$R(K_{DP}) = 35.834147{K_{DP}}^{0.379123},$$

$$R(Z_H, Z_{DR}) = 0.18882641Z^{0.472989}10^{-0.034979Z_{DR}},$$

where R is in mm h^{-1}, K_{DP} is in $^\circ$ km^{-1}, and Z_{DR} is in dB. At each radar gate, $R(K_{DP})$ was selected when $K_{DP} \geq 0.50^\circ$ km^{-1}; otherwise, $R(Z_H, Z_{DR})$ was used when $Z_H \geq 30$ dBZ and Z_{DR} was valid, and $R(Z_H)$ was used for the remaining valid echoes. Radar gates with $Z_H < 15$ dBZ were assigned zero rainfall, and the $R(Z_H, Z_{DR})$ relation was applied only when Z_{DR} was finite and $|Z_{DR}| \leq 5$ dB. The estimated rainfall rate was constrained to 0–200 mm h^{-1}. The use of multiple polarimetric relations with locally fitted coefficients is consistent with previous dual-polarization QPE studies in southern China, which showed that the estimator relationships depend on precipitation characteristics [19]. For the present dataset, the coefficients were fitted using the training subset by minimizing the mean squared error of the hourly estimates, whereas the K_{DP} and Z_H thresholds used to select among the relations were chosen on the validation set according to the overall RMSE. Composite-DP was applied separately to each of the ten radar scans, and the resulting scan-level estimates were averaged to obtain the hourly estimate.

2.7. Training Strategy and Implementation

To ensure consistent comparisons, all five neural-network models were trained using the same event-based data partition, radar variables, gauge-centered spatial neighborhood, precipitation target, normalization procedure, loss function, and evaluation code. The training, validation, and independent test subsets described in Section 2.4 were held fixed across all experiments. Apart from the architecture-specific input representations described in Section 2.5, no model-specific changes were made to the samples. Each network produced a scalar estimate of the gauge-observed hourly precipitation rate in mm h^{-1}.

A two-stage training strategy was adopted to reduce the influence of the long-tailed rainfall distribution. In the first stage, the training samples were grouped according to the five rainfall-rate intervals defined in Section 2.4 and undersampled to form a relatively balanced subset. This balanced pretraining increased the frequency with which the less common high-intensity samples contributed to the parameter updates. In the second stage, the pretrained weights were fine-tuned using the complete training subset with its original rainfall distribution. No resampling was applied to the validation or independent test subset.

The Balanced Mean Squared Error loss with a batch-based contrastive formulation (BMC) was used for model optimization [25]. For a mini-batch containing N samples, let $\hat{y}_i$ and y_i denote the estimated and observed hourly rainfall rates for the i-th sample, respectively. The pairwise score between the i-th estimate and the j-th observation was calculated as

$$s_{ij} = -\frac{\left(\hat{y}_i - y_j\right)^2}{2\sigma^2}, \tag{1}$$

where σ is a learnable noise-scale parameter that was jointly optimized with the network parameters. For each estimate $\hat{y}_i$, the index-matched observation y_i was treated as the positive target, whereas the other observations in the mini-batch served as contrasting targets. The BMC loss was then calculated as

$$\mathcal{L}_{\mathrm{BMC}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(s_{ii})}{\sum_{j=1}^{N} \exp\left(s_{ij}\right)}. \tag{2}$$

This formulation reduces the dominance of densely distributed low-rainfall samples during optimization while retaining the continuous nature of the regression target.

All models were optimized with Adam using a batch size of 512, an initial learning rate of 1×10^{-3}, and a weight decay of 1×10^{-5}. A ReduceLROnPlateau schedule halved the learning rate when the validation loss did not improve for three consecutive epochs, with a lower bound of 1×10^{-6}. Training stopped early if the validation loss did not improve for 10 consecutive epochs. The checkpoint with the lowest validation loss was retained for evaluation. 2DPoly-QPENet stopped after 37 epochs, with its best checkpoint at epoch 27; the best checkpoints for 3DPoly-QPENet, 3DTime-QPENet, 3DEcho-QPENet, and 3DHybrid-QPENet were obtained at epochs 33, 21, 33, and 42, respectively.

2.8. Evaluation Metrics

Model performance was evaluated using relative error (RE), mean absolute error (MAE), root-mean-square error (RMSE), Bias, and the Pearson correlation coefficient (CC). Let y_i and $\hat{y}_i$ denote the observed and estimated hourly rainfall rates for the i-th sample, respectively, and let n be the number of evaluated samples. The five metrics were calculated as follows:

$$\mathrm{RE} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{\sum_{i=1}^{n} y_i} \times 100\%, \tag{3}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i|, \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}, \tag{5}$$

$$\mathrm{Bias} = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i), \tag{6}$$

$$\mathrm{CC} = \frac{\sum_{i=1}^{n} (y_i - \overline{y})\left(\hat{y}_i - \overline{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}\sqrt{\sum_{i=1}^{n} \left(\hat{y}_i - \overline{\hat{y}}\right)^2}}, \tag{7}$$

where $\bar{y}$ and $\overline{\hat{y}}$ are the mean observed and estimated rainfall rates, respectively. RE measures the total absolute error relative to the total observed rainfall and is expressed as a percentage. MAE describes the mean magnitude of the estimation errors, whereas RMSE assigns greater weight to relatively large errors. Bias measures systematic deviation: positive values indicate overall overestimation, negative values indicate overall underestimation, and values closer to zero indicate better agreement. CC measures the linear correspondence between the estimates and observations, with values closer to 1 indicating stronger agreement.

The model target, rainfall-intensity thresholds, MAE, RMSE, and Bias were expressed in mm h^{-1}. RE was reported as a percentage, and CC was dimensionless. Lower RE, MAE, and RMSE values indicate better performance; the preferred Bias is closest to zero, and CC values closer to 1 indicate stronger linear agreement.

The metrics were calculated for the complete independent test set and separately for five rainfall-intensity classes: 0.1–5, 5–15, 15–20, 20–50, and >50 mm h^{-1}. Samples were assigned to these classes according to the gauge-observed rainfall rate rather than the model estimates. The stratified evaluation was used to examine how model errors varied with rainfall intensity, particularly for the less frequent heavy-rainfall samples. To further evaluate the RMSE differences across independent events, the six complete test events served as paired units in separate comparisons of 3DHybrid-QPENet with 3DTime-QPENet and 3DEcho-QPENet. Two-sided exact paired sign-flip tests assessed statistical significance. Paired bootstrap resampling of the six events (10,000 resamples) provided 95% confidence intervals for the mean RMSE differences, and the Holm method adjusted the two p values.

3. Results

3.1. Overall Performance on the Independent Test Set

Composite-DP provided the conventional dual-polarization reference on the independent test set. It yielded an RE of 43.652%, an MAE of 1.993 mm h^{-1}, an RMSE of 4.139 mm h^{-1}, a Bias of −0.860 mm h^{-1}, and a CC of 0.708. All five neural-network models had lower overall RE, MAE, RMSE, and Bias magnitude, together with higher CC, than Composite-DP.

The neural-network results first show the combined effect of retaining the time-resolved radar sequence and extending the convolutional operations to three dimensions. The 2D baseline, 2DPoly-QPENet, obtained an RE of 33.195%, an MAE of 1.524 mm h^{-1}, an RMSE of 3.307 mm h^{-1}, a Bias of −0.663 mm h^{-1}, and a CC of 0.832. When the temporally averaged 2D input was replaced by the complete time-resolved input and the convolutional operations were extended to three dimensions, 3DPoly-QPENet reduced RE, MAE, and RMSE to 27.516%, 1.321 mm h^{-1}, and 3.088 mm h^{-1}, respectively. Its Bias changed to −0.540 mm h^{-1}, and its CC increased to 0.860. Thus, the 3D baseline reduced both the overall error magnitude and the magnitude of the negative estimation bias relative to the 2D baseline.

The two structurally enhanced 3D models showed different performance characteristics. Compared with 3DPoly-QPENet, 3DTime-QPENet reduced RMSE from 3.088 to 3.081 mm h^{-1} and increased CC from 0.860 to 0.861. Its RE and MAE were 28.807% and 1.365 mm h^{-1}, respectively, whereas its Bias changed from −0.540 to −0.790 mm h^{-1}. By comparison, 3DEcho-QPENet reduced RE, MAE, and RMSE to 26.734%, 1.287 mm h^{-1}, and 3.035 mm h^{-1}, respectively, with a Bias of −0.578 mm h^{-1} and a CC of 0.863.

Among the evaluated methods, 3DHybrid-QPENet produced the best overall result for each evaluation metric. It achieved an RE of 26.331%, an MAE of 1.270 mm h^{-1}, an RMSE of 2.971 mm h^{-1}, a Bias of −0.537 mm h^{-1}, and a CC of 0.866. Relative to the 3D baseline, its MAE and RMSE decreased from 1.321 and 3.088 mm h^{-1} to 1.270 and 2.971 mm h^{-1}, respectively, while its Bias moved slightly closer to zero and its CC increased from 0.860 to 0.866. Compared with the 2D baseline, MAE and RMSE decreased from 1.524 and 3.307 mm h^{-1} to 1.270 and 2.971 mm h^{-1}, respectively, while Bias moved from −0.663 to −0.537 mm h^{-1} and CC increased from 0.832 to 0.866.

Across the six complete events in the test set, 3DHybrid-QPENet had a lower RMSE than both 3DTime-QPENet and 3DEcho-QPENet in every event. Taking the difference as 3DHybrid-QPENet minus the comparison model, the mean ΔRMSE was −0.113 mm h^{-1} relative to 3DTime-QPENet (95% event-bootstrap CI, [−0.164, −0.068]) and −0.045 mm h^{-1}

relative to 3DEcho-QPENet (95% CI, [−0.077, −0.018]). The exact paired p value was 0.031 for both comparisons; after Holm adjustment, both p values were 0.063, above the 0.050 threshold.

3.2. Performance Across Rainfall Intensities

The stratified results in Table 1 show that Composite-DP had higher RE, MAE, RMSE, and Bias magnitude than the four time-resolved 3D models in all five rainfall-intensity classes. The 2D baseline also yielded lower values for these four error measures in four of the five classes; the exception was the 5–15 mm h^{-1} class, where its values were marginally higher than those of Composite-DP. The transition from the 2D baseline to the 3D baseline reduced RE, MAE, RMSE, and the magnitude of Bias in all five rainfall-intensity classes. For light rainfall of 0.1–5 mm h^{-1}, 3DPoly-QPENet reduced MAE from 0.789 to 0.636 mm h^{-1} and RMSE from 1.476 to 1.342 mm h^{-1}. The absolute reductions in MAE and RMSE were largest in the >50 mm h^{-1} class: MAE decreased from 28.674 to 26.968 mm h^{-1}, RMSE decreased from 32.763 to 31.424 mm h^{-1}, and the magnitude of Bias decreased from 28.532 to 26.825 mm h^{-1}.

Table 1. Overall and rainfall-intensity-stratified performance of Composite-DP and the five neural-network models on the independent test set.

Model	Rainfall Class (mm h^{-1})	RE (%)	MAE (mm h^{-1})	RMSE (mm h^{-1})	Bias (mm h^{-1})	CC
Composite-DP	**Total**	**43.652**	**1.993**	**4.139**	**−0.860**	**0.708**
	0.1–5	41.803	1.265	2.103	0.721	
	5–15	51.236	3.716	4.710	−2.285	
	15–20	60.225	6.997	8.325	−6.109	
	20–50	79.461	14.431	16.715	−14.124	
	>50	86.705	41.957	44.937	−41.957	
2DPoly-QPENet	**Total**	**33.195**	**1.524**	**3.307**	**−0.663**	**0.832**
	0.1–5	29.261	0.789	1.476	0.106	
	5–15	53.115	3.759	4.753	−2.327	
	15–20	47.892	6.307	7.608	−4.617	
	20–50	46.166	9.823	12.144	−8.482	
	>50	63.931	28.674	32.763	−28.532	
3DPoly-QPENet	**Total**	**27.516**	**1.321**	**3.088**	**−0.540**	**0.860**
	0.1–5	23.843	0.636	1.342	−0.101	
	5–15	45.092	3.342	4.360	−2.179	
	15–20	45.068	6.071	7.389	−4.440	
	20–50	42.717	9.185	11.510	−7.619	
	>50	59.682	26.968	31.424	−26.825	
3DTime-QPENet	**Total**	**28.807**	**1.365**	**3.081**	**−0.790**	**0.861**
	0.1–5	24.585	0.658	1.328	−0.103	
	5–15	50.527	3.641	4.659	−2.127	
	15–20	45.661	6.143	7.624	−4.034	
	20–50	40.904	8.691	11.148	−6.190	
	>50	47.921	22.057	27.667	−21.280	
3DEcho-QPENet	**Total**	**26.734**	**1.287**	**3.035**	**−0.578**	**0.863**
	0.1–5	23.450	0.632	1.325	−0.071	
	5–15	42.379	3.210	4.268	−1.860	
	15–20	41.855	5.711	7.091	−3.912	
	20–50	40.956	8.879	11.265	−7.253	
	>50	58.428	26.568	30.885	−26.369	

Table 1. *Cont.*

Model	Rainfall Class (mm h^{-1})	RE (%)	MAE (mm h^{-1})	RMSE (mm h^{-1})	Bias (mm h^{-1})	CC
3DHybrid-QPENet	**Total**	**26.331**	**1.270**	**2.971**	**−0.537**	**0.866**
	0.1–5	23.258	0.629	1.270	−0.066	
	5–15	41.599	3.028	4.124	−1.835	
	15–20	41.317	5.660	7.005	−3.311	
	20–50	38.284	8.481	11.045	−5.563	
	>50	47.278	21.533	27.311	−20.031	

Note: Boldface identifies the overall metrics (Total) for each model; it does not denote the best-performing values.

For light rainfall, 3DTime-QPENet reduced RMSE slightly from 1.342 to 1.328 mm h^{-1} relative to 3DPoly-QPENet, while its RE and MAE were higher, at 24.585% and 0.658 mm h^{-1}, respectively. 3DEcho-QPENet obtained an RE of 23.450%, an MAE of 0.632 mm h^{-1}, and an RMSE of 1.325 mm h^{-1}. 3DHybrid-QPENet achieved the lowest errors in this class, with an RE of 23.258%, an MAE of 0.629 mm h^{-1}, and an RMSE of 1.270 mm h^{-1}. Composite-DP and the 2D baseline had positive Bias values of 0.721 and 0.106 mm h^{-1}, respectively, whereas the four 3D models had slightly negative Bias values. The Bias closest to zero was obtained by 3DHybrid-QPENet at −0.066 mm h^{-1}.

For the two intermediate rainfall classes, 3DHybrid-QPENet again produced the lowest errors. In the 5–15 mm h^{-1} class, it obtained an RE of 41.599%, an MAE of 3.028 mm h^{-1}, and an RMSE of 4.124 mm h^{-1}. In the 15–20 mm h^{-1} class, the corresponding values were 41.317%, 5.660 mm h^{-1}, and 7.005 mm h^{-1}. Relative to 3DPoly-QPENet, 3DTime-QPENet showed mixed changes: its RE, MAE, and RMSE were higher in both classes, whereas its Bias was closer to zero. 3DEcho-QPENet produced the next-lowest MAE and RMSE in both classes. All evaluated methods underestimated the observations in these two classes, and 3DHybrid-QPENet had the Bias closest to zero, at −1.835 and −3.311 mm h^{-1}, respectively.

For rainfall rates of 20–50 mm h^{-1}, 3DHybrid-QPENet achieved an RE of 38.284%, an MAE of 8.481 mm h^{-1}, an RMSE of 11.045 mm h^{-1}, and a Bias of −5.563 mm h^{-1}. These were the lowest values among the evaluated methods, followed by 3DTime-QPENet, with an MAE of 8.691 mm h^{-1}, an RMSE of 11.148 mm h^{-1}, and a Bias of −6.190 mm h^{-1}. In the >50 mm h^{-1} class, which contained 260 samples (0.18% of the independent test set), 3DHybrid-QPENet also obtained the lowest RE, MAE, and RMSE, at 47.278%, 21.533 mm h^{-1}, and 27.311 mm h^{-1}, respectively. Its Bias of −20.031 mm h^{-1} was closest to zero among the evaluated methods. 3DTime-QPENet ranked second in this class, with an MAE of 22.057 mm h^{-1} and an RMSE of 27.667 mm h^{-1}. By comparison, 3DEcho-QPENet produced an MAE of 26.568 mm h^{-1} and an RMSE of 30.885 mm h^{-1}.

The Bias results reveal a consistent intensity-dependent pattern. In the 0.1–5 mm h^{-1} class, Composite-DP and 2DPoly-QPENet overestimated rainfall, whereas the four 3D models slightly underestimated it. All evaluated methods underestimated rainfall in the four classes above 5 mm h^{-1}, and the magnitude of the negative Bias increased with rainfall intensity. Although 3DHybrid-QPENet had the smallest error magnitude and the Bias closest to zero in every rainfall-intensity class, its Bias of −20.031 mm h^{-1} indicates that underestimation of the highest rainfall intensities was not fully eliminated.

3.3. Error Distribution and Scatter-Density Analysis

Figure 8 presents the scatter-density distributions of the estimated and gauge-observed hourly rainfall rates for Composite-DP and the five neural-network models. In all panels, the highest sample densities occurred at observed rainfall rates below approximately 25 mm h^{-1}, reflecting the predominance of light-to-moderate rainfall samples in the inde-

pendent test set. Within this high-density range, the estimates were distributed on both sides of the 1:1 reference line, although the density ridge did not coincide completely with the reference line.

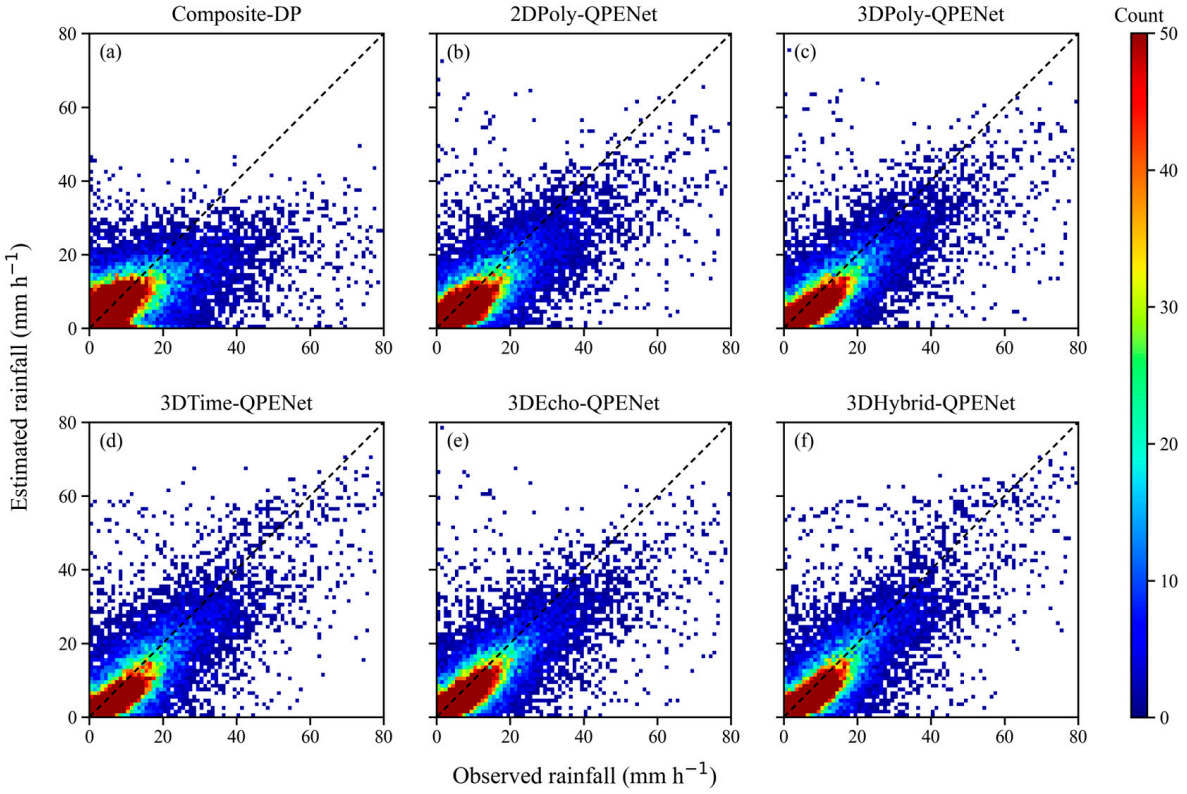


Figure 8. Scatter-density distributions of estimated versus gauge-observed hourly rainfall rates for (**a**) Composite-DP, (**b**) 2DPoly-QPENet, (**c**) 3DPoly-QPENet, (**d**) 3DTime-QPENet, (**e**) 3DEcho-QPENet, and (**f**) 3DHybrid-QPENet. The dashed line denotes the 1:1 reference line, and the shared color scale indicates the number of samples in each density bin.

As the observed rainfall rate increased, the scatter distributions became progressively broader, indicating larger estimation errors for heavier rainfall. For observed rainfall rates above approximately 50 mm h^{-1}, many estimates were located below the 1:1 line. This pattern is consistent with the negative Bias values reported for the higher rainfall-intensity classes in Table 1. A smaller number of samples were located above the reference line, showing that local overestimation also occurred, although underestimation was more prevalent at high rainfall intensities.

Differences among the six methods were most evident in the concentration and extent of the scatter distributions. Composite-DP showed a wider spread away from the 1:1 line than the neural-network models, particularly at moderate and high observed rainfall rates, consistent with its higher overall MAE and RMSE and lower CC in Table 1. Relative to 2DPoly-QPENet, the distributions produced by the 3D models were generally more concentrated around the 1:1 line within the main sample-density range. Among the 3D models, 3DHybrid-QPENet showed the closest overall correspondence between the density distribution and the reference line, consistent with its lowest overall MAE and RMSE and highest CC in Table 1. Nevertheless, its high-rainfall samples still showed considerable dispersion below the 1:1 line, confirming that the improved overall agreement did not eliminate underestimation at the upper end of the rainfall distribution.

3.4. Event-Based Case Comparison

Two selected precipitation events were used to examine the spatial performance of the models under different rainfall organizations: the event ending at 11:30 BJT on 18 August 2023 and the event ending at 10:30 BJT on 21 April 2024. Both events occurred outside the model-development period and were not used for neural-network training, validation, or checkpoint selection. They were selected because complete radar–gauge observations were available over the required 1 h period and because the observed rainfall fields had contrasting spatial structures. The first event occurred under a low-vortex/shear environment and was characterized by a relatively localized, irregularly shaped precipitation core over southeastern Guangzhou. The second event was associated with a squall line and exhibited an elongated rainband extending in the northeast–southwest direction. Figure 9 compares the gauge-observed hourly rainfall rates with estimates from Composite-DP and the five neural-network models. These case-specific results are used to compare the displayed spatial rainfall patterns and are treated as supplementary event-level evidence rather than as a statistically representative assessment of model generalization.

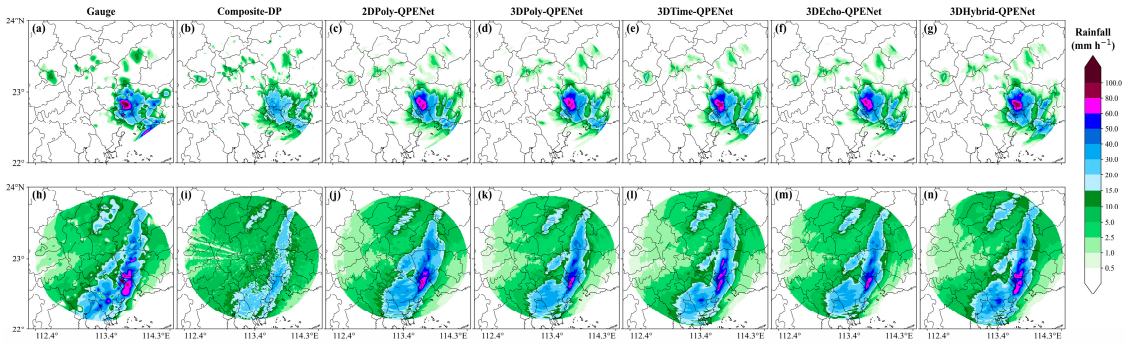


Figure 9. Spatial distributions of gauge-observed and estimated hourly rainfall rates for the events ending at 11:30 BJT on 18 August 2023 (top row) and 10:30 BJT on 21 April 2024 (bottom row). The columns from left to right show Gauge, Composite-DP, 2DPoly-QPENet, 3DPoly-QPENet, 3DTime-QPENet, 3DEcho-QPENet, and 3DHybrid-QPENet. Panels (**a**–**g**) correspond to the first event, and panels (**h**–**n**) correspond to the second event.

For the 18 August 2023 event, Composite-DP reproduced the approximate location of the principal rainfall area but underestimated the intensity and spatial extent of the heavy-rainfall core. The neural-network models provided a clearer representation of the observed core. 2DPoly-QPENet reproduced its general location but generated a relatively smoothed precipitation field. 3DPoly-QPENet produced a more distinct high-intensity center, while 3DTime-QPENet and 3DEcho-QPENet further modified the intensity and spatial organization of the core. Among the estimates, 3DHybrid-QPENet provided the closest overall representation of the observed core intensity and surrounding rainfall distribution, although local differences from the gauge observations remained.

For the 21 April 2024 event, Composite-DP captured the general orientation of the elongated rainband but underestimated rainfall intensity along much of the band, particularly near the southern high-intensity region. All five neural-network models reproduced the northeast–southwest rainband more clearly. Compared with 2DPoly-QPENet, the 3D models generated a more continuous representation of the moderate-to-heavy rainfall band. 3DHybrid-QPENet provided the closest correspondence with the observed position and intensity of the principal rainfall core while retaining the overall organization of the elongated rainband.

The event-based metrics in Table 2 support the spatial comparisons. For the 18 August 2023 event, Composite-DP obtained an RE of 40.896%, an MAE of 5.657 mm h^{-1}, an RMSE

of 10.535 mm h^{-1}, a Bias of −4.173 mm h^{-1}, and a CC of 0.872. All five neural-network models reduced RE, MAE, RMSE, and the absolute Bias while increasing CC. The transition from 2DPoly-QPENet to 3DPoly-QPENet reduced MAE from 5.320 to 5.064 mm h^{-1} and RMSE from 8.812 to 8.462 mm h^{-1}, while CC increased from 0.873 to 0.875. 3DTime-QPENet and 3DEcho-QPENet yielded MAE values of 4.595 and 4.264 mm h^{-1} and RMSE values of 7.862 and 7.098 mm h^{-1}, respectively. 3DHybrid-QPENet achieved the lowest RE, MAE, and RMSE, at 27.395%, 3.790 mm h^{-1}, and 6.306 mm h^{-1}, respectively. It also produced the Bias closest to zero (−0.491 mm h^{-1}) and the highest CC (0.929).

Table 2. Performance of Composite-DP and the five neural-network models for the two selected precipitation events.

Time (BJT)	Method	RE (%)	MAE (mm h^{-1})	RMSE (mm h^{-1})	Bias (mm h^{-1})	CC
18 August 2023 11:30	Composite-DP	40.896	5.657	10.535	−4.173	0.872
	2DPoly-QPENet	38.461	5.320	8.812	−2.918	0.873
	3DPoly-QPENet	36.612	5.064	8.462	−2.140	0.875
	3DTime-QPENet	33.216	4.595	7.862	−1.288	0.889
	3DEcho-QPENet	30.825	4.264	7.098	−1.818	0.916
	3DHybrid-QPENet	27.395	3.790	6.306	−0.491	0.929
21 April 2024 10:30	Composite-DP	42.075	4.921	8.362	−3.024	0.846
	2DPoly-QPENet	29.817	3.487	6.014	−1.974	0.893
	3DPoly-QPENet	28.903	3.380	5.615	−2.499	0.921
	3DTime-QPENet	27.852	3.257	5.353	−1.807	0.934
	3DEcho-QPENet	23.210	2.714	4.705	−1.903	0.945
	3DHybrid-QPENet	20.998	2.456	4.137	−0.667	0.951

For the 21 April 2024 event, Composite-DP obtained an RE of 42.075%, an MAE of 4.921 mm h^{-1}, an RMSE of 8.362 mm h^{-1}, a Bias of −3.024 mm h^{-1}, and a CC of 0.846. The corresponding MAE and RMSE decreased to 3.487 and 6.014 mm h^{-1} for 2DPoly-QPENet and to 3.380 and 5.615 mm h^{-1} for 3DPoly-QPENet. 3DTime-QPENet and 3DEcho-QPENet further reduced the errors, with MAE values of 3.257 and 2.714 mm h^{-1} and RMSE values of 5.353 and 4.705 mm h^{-1}, respectively. 3DHybrid-QPENet again produced the lowest RE (20.998%), MAE (2.456 mm h^{-1}), and RMSE (4.137 mm h^{-1}), together with the Bias closest to zero (−0.667 mm h^{-1}). Its CC was 0.951, compared with 0.945 for 3DEcho-QPENet. Thus, 3DHybrid-QPENet had the lowest error values and the Bias closest to zero in both selected events, consistent with the spatial distributions shown in Figure 9.

For complete-field QPE generation, all five neural-network models used a batch size of 512 and bfloat16 precision. Complete-field inference required 0.471–9.918 s, while the calculation from the radar sequence to the completed QPE field required 7.251–16.699 s. Model size and peak GPU memory ranged from 97,137 to 566,738 parameters and from 148.160 to 1281.688 MiB, respectively. Among the five models, 3DHybrid-QPENet had the largest computational demand, requiring 9.918 s for model inference and 16.699 s from the radar sequence to the QPE field, with a peak memory use of 1281.688 MiB.

4. Discussion

4.1. Contribution of the Time-Resolved Three-Dimensional Representation

The comparison between 2DPoly-QPENet and 3DPoly-QPENet reflects the combined effect of retaining the time-resolved radar sequence and extending the convolutional operations to three dimensions. The two models use the same dual-polarization radar variables and are evaluated under the same data partitioning and training protocol, but they represent the ten consecutive radar scans differently. The 2D model temporally averages these scans to

form a $C \times H \times W$ input, whereas the 3D model retains the complete sequence as a $C \times T \times H \times W$ input and performs convolution over the $T \times H \times W$ dimensions. Accordingly, the three convolutional dimensions comprise one temporal and two horizontal spatial dimensions and do not represent the vertical structure derived from multiple radar elevation angles.

As shown in Table 1, 3DPoly-QPENet reduced the overall RE from 33.195% to 27.516%, the MAE from 1.524 to 1.321 mm h^{-1}, and the RMSE from 3.307 to 3.088 mm h^{-1} compared with 2DPoly-QPENet. Its CC increased from 0.832 to 0.860, while the magnitude of its negative Bias decreased from 0.663 to 0.540 mm h^{-1}. Error reductions were also obtained in all five rainfall-intensity classes. For hourly rainfall rates exceeding 50 mm h^{-1}, the MAE and RMSE decreased by 1.706 and 1.339 mm h^{-1}, respectively. For the 0.1–5 mm h^{-1} class, the MAE and RMSE decreased by 0.153 and 0.134 mm h^{-1}, respectively. Although the absolute error reductions were larger in the highest rainfall class, considerable estimation errors remained for these less frequent, high-intensity samples.

These results are consistent with a difference in the information retained by the two input representations: the time-resolved sequence preserves scan-to-scan variation that is absent from the temporally averaged input. Nevertheless, the comparison simultaneously changes the input representation and the convolutional dimensionality. The observed differences therefore reflect their combined effect and should not be attributed exclusively to the use of three-dimensional convolution. In addition, 3DPoly-QPENet retained a Bias of −26.825 mm h^{-1} for rainfall rates above 50 mm h^{-1}, indicating that the 3D baseline did not resolve the underestimation of the highest rainfall intensities.

4.2. Complementary Effects of Temporal Modeling and Multiscale Echo Representation

The temporal and multiscale configurations represent the same ten-scan input in different ways. In 3DTime-QPENet, the complete sequence is first encoded by three-dimensional convolution. Horizontal pooling after the first two stages produces a sequence of ten $64 \times 5 \times 5$ encoded feature maps, which are passed sequentially to the ConvLSTM. The resulting hidden states are summarized across the sequence and combined with a global representation from the encoded tensor. In contrast, 3DEcho-QPENet uses an encoder–decoder pathway to reconstruct multiscale echo features. 3DHybrid-QPENet combines the ConvLSTM-based sequence representation and the global representation with the multiscale encoder–decoder features.

The performance of the temporal configuration varied with rainfall intensity. Compared with 3DPoly-QPENet, the overall changes obtained by 3DTime-QPENet were mixed: RMSE decreased only slightly from 3.088 to 3.081 mm h^{-1} and CC increased from 0.860 to 0.861, whereas RE and MAE were higher and Bias became more negative. A clearer advantage appeared in the two highest rainfall-intensity classes. For both 20–50 and >50 mm h^{-1}, 3DTime-QPENet had lower RE, MAE, RMSE, and absolute Bias than 3DPoly-QPENet. Above 50 mm h^{-1}, its RMSE decreased from 31.424 to 27.667 mm h^{-1}, while the magnitude of Bias decreased from 26.825 to 21.280 mm h^{-1}. The main advantage of the temporal configuration was therefore concentrated in higher-intensity rainfall rather than distributed uniformly across the test set.

The multiscale configuration showed a broader pattern of error reduction. Relative to 3DPoly-QPENet, 3DEcho-QPENet had lower RE, MAE, and RMSE in all five rainfall-intensity classes. Its overall RMSE decreased from 3.088 to 3.035 mm h^{-1} and its CC increased from 0.860 to 0.863, although its overall Bias became slightly more negative. This broader class-wise improvement complements the concentration of the temporal-modeling gains at higher rainfall intensities, while neither configuration eliminated the remaining underestimation of heavy rainfall.

3DHybrid-QPENet, which combines the two representations, produced lower overall errors than 3DTime-QPENet and 3DEcho-QPENet. Compared with 3DTime-QPENet, it reduced MAE from 1.365 to 1.270 mm h^{-1} and RMSE from 3.081 to 2.971 mm h^{-1}, while moving Bias closer to zero and increasing CC from 0.861 to 0.866. Its RMSE was also lower than those of both individual extensions in each of the five rainfall-intensity classes. Although the differences among the three-dimensional model variants were modest in the aggregate metrics, the paired analysis showed a consistent event-wise direction: 3DHybrid-QPENet had a lower RMSE than both 3DTime-QPENet and 3DEcho-QPENet in all six complete test events. For both comparisons, the event-bootstrap confidence intervals for the mean RMSE differences remained below zero, whereas the Holm-adjusted p value was 0.063. Given the limited number of independent event pairs, these results support a consistent direction within the available test events but do not establish statistically significant differences after adjustment. Overall, the temporal configuration showed its clearest advantage at higher rainfall intensities, whereas the multiscale configuration produced broader reductions across the rainfall classes. The joint result is consistent with the two representations providing complementary information, although this interpretation remains at the level of the complete model configurations rather than individual components.

4.3. Rainfall-Intensity Dependence and Residual Estimation Errors

The estimation errors varied systematically with the gauge-observed rainfall intensity. As expected for absolute-error metrics, MAE and RMSE increased as the magnitude of hourly rainfall increased. Bias provides additional information that is not evident from these metrics alone. In the 0.1–5 mm h^{-1} class, 2DPoly-QPENet had a small positive Bias of 0.106 mm h^{-1}, whereas the four 3D models had slightly negative Bias values ranging from −0.103 to −0.066 mm h^{-1}. All five neural-network models underestimated rainfall in the four classes above 5 mm h^{-1}, and the magnitude of the negative Bias increased with rainfall intensity. This intensity-dependent pattern was shared across the different model architectures.

The results of 3DHybrid-QPENet illustrate why the overall Bias should be interpreted together with the class-stratified values. Its overall Bias was −0.537 mm h^{-1}, while the class-specific values were −0.066, −1.835, −3.311, −5.563, and −20.031 mm h^{-1} as rainfall intensity increased. The overall value therefore does not represent the progressively stronger underestimation at higher rainfall intensities. Similarly, RE should be considered together with MAE, RMSE, and Bias because its value depends on the total observed rainfall within each class. The combined use of these metrics provides a more complete description of both relative error and error magnitude.

Although 3DHybrid-QPENet produced the lowest RE, MAE, and RMSE among the evaluated models in the >50 mm h^{-1} class, these metrics remained 47.278%, 21.533 mm h^{-1}, and 27.311 mm h^{-1}, respectively. Its Bias remained −20.031 mm h^{-1}, indicating systematic underestimation at the upper end of the rainfall distribution. Several factors may contribute to this residual error. First, high-intensity samples were much less frequent than light-to-moderate rainfall samples; the >50 mm h^{-1} class contained 3700 samples (0.12%) in the full retained dataset and 260 samples (0.18%) in the independent test set. Although class-balanced pretraining and the balanced mean squared error loss reduce the dominance of densely distributed low-rainfall samples, optimization is still governed by an aggregate training objective, and the limited number and diversity of extreme observations may not sufficiently constrain the upper tail of the regression. Second, localized high-intensity rainfall, particularly convective rainfall, can exhibit sharp spatial gradients, whereas the radar inputs represent finite sampling volumes over a spatial neighborhood and the target is a point-gauge accumulation. Spatial displacement and radar–gauge representativeness

differences can therefore produce large discrepancies near localized rainfall maxima. Third, rainfall microphysical variability may also contribute because changes in drop size distribution can alter the relationships between the polarimetric variables and surface rainfall. During intense rainfall, path attenuation can reduce the measured Z_H and affect Z_{DR}, while water on the radome may further attenuate or distort the received signals [4]. These microphysical and measurement effects may weaken or modify the radar signatures associated with the highest rainfall rates, thereby contributing to the observed negative Bias. Thus, 3DHybrid-QPENet reduced the error magnitude for high-intensity rainfall relative to the other evaluated models, but it did not eliminate systematic underestimation. Improving QPE above 50 mm h^{-1} remains a priority for future work.

4.4. Aggregate and Event-Based Comparisons

Composite-DP was included as a conventional dual-polarization reference for the complete independent test set and the two later events. As described in Section 2.6, it used the same three polarimetric variables and ten-scan temporal sampling as the neural-network models, but estimated hourly rainfall through locally calibrated empirical relations.

On the complete independent test set, Composite-DP yielded an overall RMSE of 4.139 mm h^{-1} and a CC of 0.708. All five neural-network models had lower overall RE, MAE, RMSE, and absolute Bias and higher CC. Across the rainfall-intensity classes, the four 3D models also had lower RE, MAE, and RMSE than Composite-DP. Among all neural-network results, the only exception to this class-wise improvement occurred for 2DPoly-QPENet in the 5–15 mm h^{-1} class, where its errors were slightly higher than those of Composite-DP. The difference was most pronounced above 50 mm h^{-1}, where 3DHybrid-QPENet reduced RMSE from 44.937 to 27.311 mm h^{-1} and the magnitude of Bias from 41.957 to 20.031 mm h^{-1}.

For the event at 11:30 on 18 August 2023, the relative performance of the methods was broadly consistent with the aggregate results. Composite-DP captured the general location of the observed rainfall but produced an RE of 40.896%, an MAE of 5.657 mm h^{-1}, an RMSE of 10.535 mm h^{-1}, and a Bias of −4.173 mm h^{-1}. All five neural-network models reduced these errors. Among them, 3DHybrid-QPENet obtained the lowest RE, MAE, and RMSE, with values of 27.395%, 3.790 mm h^{-1}, and 6.306 mm h^{-1}, respectively. It also increased CC from 0.872 for Composite-DP to 0.929 and reduced the magnitude of Bias from 4.173 to 0.491 mm h^{-1}. The hybrid model more closely represented the overall spatial rainfall pattern, although local discrepancies remained.

Similar behavior was observed for the elongated precipitation band at 10:30 on 21 April 2024. Composite-DP yielded an RE of 42.075%, an MAE of 4.921 mm h^{-1}, an RMSE of 8.362 mm h^{-1}, a Bias of −3.024 mm h^{-1}, and a CC of 0.846. The corresponding values for 3DHybrid-QPENet were 20.998%, 2.456 mm h^{-1}, 4.137 mm h^{-1}, −0.667 mm h^{-1}, and 0.951, respectively. The data-driven models generally reproduced the observed banded distribution more closely than Composite-DP, and the hybrid model achieved the lowest error values. Its Bias remained negative, but its magnitude decreased from 3.024 to 0.667 mm h^{-1}, indicating that underestimation was reduced rather than eliminated. As an average over the event, Bias does not describe localized positive and negative errors.

Across the two later events, the ordering among 3DTime-QPENet, 3DEcho-QPENet, and 3DHybrid-QPENet matched that of the complete test set, but the separation among them was larger. On the complete test set, the RMSE of 3DHybrid-QPENet was lower than those of 3DTime-QPENet and 3DEcho-QPENet by 0.110 and 0.064 mm h^{-1}, respectively, whereas the corresponding differences in the two event evaluations ranged from 0.568 to 1.556 mm h^{-1}. The complete-test metrics pool station–hour samples from six precipitation events across different rainfall intensities and gauge locations, with the lowest rainfall-

intensity class accounting for most samples. Local differences in intense rainfall cores or narrow rainbands therefore contribute less to the pooled errors than widespread low-intensity rainfall. The event maps retain the spatial organization of individual precipitation systems, allowing these localized differences to become more visible and, in the selected cases, to produce larger event-level metric gaps.

The available event sample nevertheless covers only six complete test events and two later cases and therefore represents a limited range of precipitation organizations. The magnitude and spatial form of the model differences may vary in other systems; a broader set of independent events will be needed to assess how consistently these patterns recur.

4.5. Limitations and Future Work

Several limitations define the scope of the present results. The models were developed and evaluated using observations from a single radar–gauge network in Guangzhou under a specific radar configuration and regional precipitation environment. The independent test set comprised six complete rainfall events, and two later independent events were examined separately. With only six independent events in the test set, between-event variability and small performance differences among the 3D models remain difficult to characterize. Event-based partitioning and the 24 h exclusion periods on either side of each validation and test event reduced temporal overlap between model development and evaluation. Nevertheless, the available events cover only a limited range of rainfall conditions and do not establish transferability to other radar systems, scanning strategies, climatic regions, or gauge-network configurations.

The radar inputs consist of ten consecutive single-elevation scans. The resulting $T \times H \times W$ convolutional dimensions describe temporal evolution and two-dimensional horizontal spatial structure but do not contain the vertical information available from multiple radar elevation angles. Accordingly, the present study evaluates time-resolved horizontal echo sequences and does not assess the contribution of the vertical structure of precipitation systems. Incorporating multi-elevation observations could extend the available information, but this would require consistent vertical coordinate construction, treatment of differences in sampling height and coverage, and a model architecture designed specifically for the additional elevation dimension.

The radar–gauge sampling configuration is another source of uncertainty. Excluding 59 of the 1353 available gauges (4.36%) in the persistently blocked azimuth sectors reduced coverage in those directions and may limit the representation of terrain-influenced rainfall within the affected sectors. The fixed gauge-centered window does not follow storm motion, so rapidly advecting echoes may enter or leave the sampled neighborhood during the 1 h accumulation period. Each input represents a radar neighborhood surrounding a gauge, whereas the target is a point-based one-hour gauge accumulation. This difference in spatial representativeness may be especially relevant for localized rainfall with strong spatial gradients. Although the gauge observations and radar data were quality controlled and temporally matched, the present evaluation does not distinguish errors caused by the model from those associated with measurement uncertainty and radar–gauge representativeness differences. Evaluation against denser gauge networks, independent precipitation products, or overlapping radar observations could help quantify the uncertainty associated with radar–gauge sampling and the excluded azimuth sectors. Future work could also compare the fixed gauge-centered window with a window shifted according to observed echo motion, particularly for rapidly moving precipitation systems.

The remaining underestimation above 50 mm h^{-1} is a major model-related limitation. Although 3DHybrid-QPENet reduced MAE and RMSE relative to the other evaluated models, its Bias remained −20.031 mm h^{-1}. Improving estimates in this part of the rainfall

distribution will require more high-intensity samples from a broader range of precipitation regimes, together with further evaluation of sampling strategies or loss functions that increase their influence during training. Because the model variants differ in more than one structural aspect, the observed performance differences cannot be assigned to individual components. Controlled comparisons under the same training settings, together with repeated training, would help separate component effects from run-to-run variability.

3DHybrid-QPENet contained 566,738 parameters. In the timing test, full-field inference required 9.918 s, while the total elapsed time from raw radar input to the QPE field was 16.699 s; peak memory use was 1281.688 MiB. The end-to-end time remained below the nominal 6 min interval between consecutive 0.5° scans, but was measured offline for a single radar sequence. Continuous processing would additionally depend on reliable data access, integration with existing processing procedures, handling of incomplete or interrupted inputs, and long-term stability, all of which remain to be evaluated.

5. Conclusions

This study investigated hourly QPE using ten consecutive single-elevation scans of Z_H, Z_{DR}, and K_{DP}, with particular attention to scan-to-scan evolution and multiscale horizontal echo structure. 3DTime-QPENet combined a globally pooled representation of the three-dimensional convolutional features with a ConvLSTM representation of the same feature sequence, while 3DHybrid-QPENet further incorporated multiscale features from an encoder–decoder pathway with skip connections. The model configurations were evaluated under a common training and testing protocol.

On the independent test set comprising six complete precipitation events from 2021, Composite-DP yielded an overall RMSE of 4.139 mm h^{-1} and a CC of 0.708. All five neural-network models had lower overall RE, MAE, RMSE, and absolute Bias and higher CC. The temporally averaged 2DPoly-QPENet had overall RE, MAE, and RMSE values of 33.195%, 1.524 mm h^{-1}, and 3.307 mm h^{-1}, whereas the corresponding values for 3DPoly-QPENet were 27.516%, 1.321 mm h^{-1}, and 3.088 mm h^{-1}, respectively. The CC was also higher for 3DPoly-QPENet (0.860 versus 0.832). This comparison reflects the combined changes in input representation and convolutional dimensionality. Among the evaluated configurations, 3DHybrid-QPENet had the lowest overall RE, MAE, and RMSE and the highest CC, with an RE of 26.331%, an MAE of 1.270 mm h^{-1}, an RMSE of 2.971 mm h^{-1}, a Bias of −0.537 mm h^{-1}, and a CC of 0.866.

Across the rainfall-intensity classes, all four time-resolved 3D models had lower RE, MAE, and RMSE than Composite-DP. The only class-wise exception among the neural-network results occurred for 2DPoly-QPENet in the 5–15 mm h^{-1} class, where its errors were slightly higher than those of Composite-DP. Among the neural-network configurations, 3DHybrid-QPENet obtained the lowest RE, MAE, and RMSE in every evaluated class. All five neural-network models underestimated rainfall in the four classes above 5 mm h^{-1}, and the magnitude of the negative Bias increased with rainfall intensity. For rainfall rates exceeding 50 mm h^{-1}, the hybrid model had an MAE of 21.533 mm h^{-1} and an RMSE of 27.311 mm h^{-1}, but its Bias remained −20.031 mm h^{-1}. At the event level, its RMSE was lower than those of 3DTime-QPENet and 3DEcho-QPENet in all six complete test events. With only six independent events, however, the statistical evidence remains limited.

The two later independent event evaluations provided spatially resolved comparisons under different precipitation organizations. The RMSE of 3DHybrid-QPENet was 6.306 mm h^{-1} compared with 10.535 mm h^{-1} for Composite-DP on 18 August 2023, and 4.137 mm h^{-1} compared with 8.362 mm h^{-1} on 21 April 2024. The hybrid model also reproduced the observed spatial rainfall patterns more closely in both selected cases. These cases provide supplementary evidence of model performance under contrasting spatial structures, but

their limited number does not support broader conclusions about cross-year or cross-region generalization.

Overall, the results indicate that retaining time-resolved horizontal radar information was associated with lower hourly QPE errors than temporal averaging, while combining temporal and multiscale representations produced the most consistent performance among the evaluated models. These conclusions remain limited to a single regional radar–gauge network, single-elevation inputs, a limited set of independent events, and the available high-intensity samples. Future work should evaluate broader multi-year and cross-radar datasets, incorporate multi-elevation observations where appropriate, increase the representation of extreme rainfall, and assess performance variability through controlled comparisons and repeated training experiments.

Author Contributions: Conceptualization and methodology, C.X., Y.Z. (Yi Zhang) and H.C.; data curation, software, formal analysis and visualization, C.X., L.L. and Y.C.; validation and resources, Y.Z. (Yonghua Zhang), L.L. and J.S.; writing—original draft preparation, C.X.; writing—review and editing, all authors; supervision, Y.Z. (Yi Zhang) and H.C.; project administration and funding acquisition, Y.Z. (Yi Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the GBA Academy of Meteorological Research through the Guangdong–Hong Kong–Macao Greater Bay Area Meteorological Science and Technology Collaborative Research Project, grant number GHMA2024Y05; the Key Laboratory of Intelligent Meteorological Observation Technology, China Meteorological Administration, through its Open Research Program, grant number ZNGC2024QN18; the Guangdong Basic and Applied Basic Research Foundation, grant number 2025A1515510029; the Liaoning Province Science and Technology Plan Joint Program (Natural Science Foundation—General Program), grant number 2025-MSLH-369; the Shanghai Typhoon Research Foundation, grant number TFJJ202606; the Guangzhou Municipal Science and Technology Project, grant number 2025B04J0018; and the Open Research Program of the China Meteorological Administration Radar Meteorology Key Laboratory, grant number 2026LRM-B06.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to operational data-management restrictions and the need to obtain authorization from the relevant data providers.

Acknowledgments: Generative AI tools ChatGPT (GPT-5.6, OpenAI) were used during the preparation and revision of this manuscript to assist with Chinese-to-English translation, language editing, and the organization of the manuscript and response letters. These tools were not used to generate or alter observational data or reported numerical results. All suggested wording and structural changes were reviewed against the source materials and revised by the authors. The authors take full responsibility for the scientific content and the final version of the manuscript.

Conflicts of Interest: Author Jian Sun is employed by HuaYun METSTAR Radar (Beijing) Company, Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Zhang, X.; Kang, A.; Song, Q.; Guo, Y.; Liu, Y.; Lei, X.; Wang, H. Characteristics and risk management of urban surface flooding in Guangzhou, China: Insights from 2022 ground monitoring. *J. Hydrol. Reg. Stud.* **2024**, *53*, 101831. [CrossRef]
2. Fu, X.; Xue, F.; Liu, Y.; Chen, F.; Yang, H. Evaluation of urban flood susceptibility under the influence of urbanization based on shared socioeconomic pathways. *Land* **2025**, *14*, 621. [CrossRef]
3. Marshall, J.S.; Palmer, W.M. The distribution of raindrops with size. *J. Meteorol.* **1948**, *5*, 165–166. [CrossRef]
4. Zhang, Z.; Zhao, Z.; Qi, Y.; Xiong, M. Dual-polarization radar quantitative precipitation estimation (QPE): Principles, operations, and challenges. *Remote Sens.* **2025**, *17*, 3619. [CrossRef]
5. Chen, X.; Huang, W.; Zhao, C.; Tian, Y. Rain detection from X-band marine radar images: A support vector machine-based approach. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2115–2123. [CrossRef]

6. Chen, X.; Huang, W. Identification of rain and low-backscatter regions in X-band marine radar images: An unsupervised approach. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4225–4236. [CrossRef]
7. Shin, K.; Song, J.J.; Bang, W.; Lee, G. Quantitative precipitation estimates using machine learning approaches with operational dual-polarization radar data. *Remote Sens.* **2021**, *13*, 694. [CrossRef]
8. Chen, H.; Chandrasekar, V.; Tan, H.; Cifelli, R. Rainfall estimation from ground radar and TRMM precipitation radar using hybrid deep neural networks. *Geophys. Res. Lett.* **2019**, *46*, 10669–10678. [CrossRef]
9. Li, W.; Chen, H.; Han, L. Polarimetric radar quantitative precipitation estimation using deep convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4102911. [CrossRef]
10. Yo, T.-S.; Su, S.-H.; Chu, J.-L.; Chang, C.-W.; Kuo, H.-C. A deep learning approach to radar-based QPE. *Earth Space Sci.* **2021**, *8*, e2020EA001340. [CrossRef]
11. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems 28*; Curran Associates, Inc.: Red Hook, NY, USA, 2015; pp. 802–810.
12. Tian, W.; Yi, L.; Liu, W.; Huang, W.; Ma, G.; Zhang, Y. Ground radar precipitation estimation with deep learning approaches in meteorological private cloud. *J. Cloud Comput.* **2020**, *9*, 22. [CrossRef]
13. Huangfu, J.; Hu, Z.; Zheng, J.; Zhu, Y.; Yin, X.; Zuo, Y. A study on polarization radar quantitative precipitation estimation using deep learning. *Acta Meteorol. Sin.* **2022**, *80*, 565–577. [CrossRef]
14. Peng, X.; Li, Q.; Jing, J. CNGAT: A graph neural network model for radar quantitative precipitation estimation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5106814. [CrossRef]
15. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning spatiotemporal features with 3D convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4489–4497. [CrossRef]
16. Chen, S.; Zhang, S.; Geng, H.; Chen, Y.; Zhang, C.; Min, J. Strong spatiotemporal radar echo nowcasting combining 3DCNN and bi-directional convolutional LSTM. *Atmosphere* **2020**, *11*, 569. [CrossRef]
17. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]
18. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*; Springer: Cham, Switzerland, 2016; Volume 9901, pp. 424–432. [CrossRef]
19. Zhang, Y.; Liu, L.; Bi, S.; Wu, Z.; Shen, P.; Ao, Z.; Chen, C.; Zhang, Y. Analysis of dual-polarimetric radar variables and quantitative precipitation estimators for landfall typhoons and squall lines based on disdrometer data in southern China. *Atmosphere* **2019**, *10*, 30. [CrossRef]
20. Gabella, M.; Notarpietro, R. Ground clutter characterization and elimination in mountainous terrain. In Proceedings of the Second European Conference on Radar Meteorology (ERAD 2002), Delft, The Netherlands, 18–23 November 2002; pp. 305–311.
21. Helmus, J.J.; Collis, S.M. The Python ARM Radar Toolkit (Py-ART), a library for working with weather radar data in the Python programming language. *J. Open Res. Softw.* **2016**, *4*, e25. [CrossRef]
22. Zhang, Y.; Bi, S.; Liu, L.; Chen, H.; Zhang, Y.; Shen, P.; Yang, F.; Wang, Y.; Zhang, Y.; Yao, S. Deep learning for polarimetric radar quantitative precipitation estimation during landfalling typhoons in South China. *Remote Sens.* **2021**, *13*, 3157. [CrossRef]
23. Graves, A. *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 385. [CrossRef]
24. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef] [PubMed]
25. Ren, J.; Zhang, M.; Yu, C.; Liu, Z. Balanced MSE for imbalanced visual regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 7926–7935. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.