YUSHI: UNIFIED PRECIPITATION ESTIMATION AND NOWCASTING FROM DIRECT OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Precipitation nowcasting is critical for hydrological forecasting and weather-related hazard warning, while most learning-based methods assume that spatially complete precipitation fields are already available as model inputs or prediction targets. In practice, such fields are usually derived from radar through quantitative precipitation estimation (QPE), interpolation, or multi-source merging rather than directly observed. In this paper, we propose **Yushi**, a unified framework for precipitation estimation and nowcasting directly from radar reflectivity and sparse automatic weather station (AWS) observations. Yushi first learns a gauge-constrained representation in which radar provides dense spatial structure while AWS measurements impose quantitative surface-precipitation constraints. Future precipitation states are then modeled in this latent space using causal conditional flow matching, and an intensity-aware balanced loss transfers precipitation-intensity information to latent flow optimization to emphasize heavy rainfall. Experiments over the Yangtze River Delta region show that Yushi consistently outperforms representative nowcasting baselines across various deterministic and spatially pooled metrics, with particularly clear improvements for intense precipitation. The results demonstrate the potential of directly learning and forecasting precipitation states from complementary primary observations without requiring a pre-constructed gridded precipitation analysis.

1 INTRODUCTION

Precipitation nowcasting is essential for early warning of floods, landslides, and other rainfall-induced hazards. Early learning-based methods primarily focused on extrapolating radar echoes to predict the short-term evolution of precipitating systems (Shi et al., 2015; 2017a; Wang et al., 2017; Gao et al., 2022b; 2023; Yu et al., 2024; Gong et al., 2024; Perrone Ribeiro & Faganeli Pucer, 2026), while recent studies have increasingly shifted toward forecasting radar-derived or gridded precipitation fields that are more directly relevant to hydrological applications (Ayzel et al., 2020; Sønderby et al., 2020; Ravuri et al., 2021; Espeholt et al., 2022; Zhang et al., 2023). This progression reflects a growing emphasis on predicting hydrologically meaningful quantities rather than only reproducing the visual evolution of radar echoes. In particular, precipitation-oriented forecasts are more directly connected to downstream applications such as runoff estimation, urban flood warning, and rainfall-threshold-based hazard assessment. However, these approaches typically assume that a spatially complete precipitation state is already available as the model input or prediction target.

In practice, such precipitation fields are often derived rather than directly observed. Radar provides dense and frequent measurements, but radar reflectivity is only indirectly related to surface precipitation and usually requires additional QPE procedures such as conversion, calibration, and gauge adjustment (Wilson & Brandes, 1979; Sauvageot, 1994; Krajewski et al., 2010; Berne & Krajewski, 2013; McKee & Binns, 2016). The relationship between radar reflectivity and surface rainfall can also vary across precipitation regimes and atmospheric conditions, making a universally accurate conversion difficult (Wilson & Brandes, 1979; Krajewski et al., 2010; Berne & Krajewski, 2013). Gridded precipitation analyses may further rely on interpolation or multi-source merging, introducing extra processing, latency, and dependence on region-specific operational products (McKee & Binns, 2016; Zhang et al., 2016; Ochoa-Rodriguez et al., 2019). As a result, the availability and quality of such analysis fields can vary substantially across regions, particularly where dense gauge networks or mature operational processing systems are unavailable. This motivates a more direct

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

question: *can precipitation be estimated and nowcast directly from primary observations, without requiring a pre-constructed gridded precipitation analysis?*

Among the primary observations available for precipitation monitoring, weather radar and AWS are two of the most important sources. Radar offers dense spatial coverage and captures the organization and motion of precipitation systems, but its measurements are only indirectly related to rainfall at the surface. In contrast, AWS gauges provide direct surface precipitation measurements, yet their spatial coverage is inherently sparse and irregular. Combining these two sources therefore offers a natural way to infer a spatially complete precipitation state directly from observations, with radar providing spatial structure and gauges supplying quantitative constraints (Berne & Krajewski, 2013; McKee & Binns, 2016; Ochoa-Rodriguez et al., 2019). Building on this complementarity, we formulate QPE and nowcasting as a unified observation-to-precipitation problem by learning a shared precipitation-aware latent state and modeling its future evolution.

Based on this formulation, **we propose Yushi, a unified framework for precipitation estimation and nowcasting from radar and sparse AWS observations**. Yushi first learns a gauge-constrained latent representation without requiring gauge interpolation or an external gridded precipitation analysis dataset. It then models future latent states using causal conditional flow matching (CFM; Lipman et al., 2023) and decodes them directly into precipitation fields. To better capture heavy rainfall, we further introduce precipitation-intensity-aware latent optimization, which assigns greater importance to latent flow errors associated with intense precipitation. Experiments on radar and AWS observations over the Yangtze River Delta demonstrate that Yushi consistently outperforms representative nowcasting baselines across event-based and spatially pooled metrics, with particularly strong improvements for heavy precipitation and substantially reduced under-forecasting. The advantage remains clear when 20% of AWS stations are completely withheld from both training and test-time inputs, indicating that the learned representation generalizes beyond previously observed gauge locations.

The contributions of this paper are as follows:

- We introduce a direct observation-to-precipitation framework that unifies QPE and nowcasting without relying on pre-generated gridded precipitation products.
- We propose gauge-constrained representation learning, where dense radar observations provide spatial structure and sparse AWS measurements impose quantitative constraints.
- We develop precipitation-intensity-aware latent flow matching to emphasize the dynamics of heavy precipitation and improve extreme-rainfall nowcasting.

2 Related Work

2.1 Radar Echo Extrapolation and Precipitation Nowcasting

Learning-based precipitation nowcasting has evolved from extrapolating radar echoes toward directly forecasting precipitation fields. Early approaches primarily modeled the spatiotemporal evolution of radar reflectivity using recurrent or convolutional architectures, such as ConvLSTM (Shi et al., 2015), TrajGRU (Shi et al., 2017b), PredRNN (Wang et al., 2017), and Earthformer (Gao et al., 2022b). More recent methods increasingly target precipitation quantities that are more directly relevant to hydrological applications. RainNet (Ayzel et al., 2020) predicts precipitation from radar observations, while DGMR (Ravuri et al., 2021), MetNet (Sønderby et al., 2020), and NowcastNet (Zhang et al., 2023) further demonstrate the potential of learning-based precipitation forecasting at high spatial and temporal resolutions. Meanwhile, diffusion- and flow-based generative approaches, including PreDiff (Gao et al., 2023), DiffCast (Yu et al., 2024), CasCast (Gong et al., 2024), and FlowCast (Perrone Ribeiro & Faganeli Pucer, 2026), have introduced increasingly expressive formulations for modeling the uncertainty and multiscale structure of future precipitation evolution.

This progression reflects a broader shift in the prediction target, from reproducing the evolution of radar echoes to forecasting precipitation fields that are more directly meaningful for hydrological applications. However, the distinction between these targets is not merely semantic. Radar reflectivity is an indirect measurement of surface precipitation, and converting radar observations into precipitation typically requires additional estimation and calibration procedures (Wilson & Brandes,

1979; Berne & Krajewski, 2013; Krajewski et al., 2010). Consequently, forecasting a radar-derived precipitation product is different from inferring precipitation directly from primary surface observations. Most existing precipitation-oriented nowcasting methods assume that a spatially complete precipitation field is already available for training and/or forecasting, leaving the construction of such a field as a separate problem. This separation motivates the direct observation-to-precipitation formulation of Yushi, which jointly learns precipitation states from primary observations and models their future evolution within a unified framework.

2.2 Radar-Gauge Fusion for Precipitation Estimation

Radar and rain gauges provide complementary information for precipitation estimation: radar offers dense spatial coverage, while gauges provide direct surface measurements. Their combination has been widely studied through bias adjustment, spatial interpolation, and radar-gauge merging to construct spatially complete precipitation fields (McKee & Binns, 2016; Ochoa-Rodriguez et al., 2019). Recent learning-based methods further integrate radar and station observations for precipitation forecasting. RN-Net (Zhang et al., 2021) uses interpolated station measurements together with radar observations, while RST-RainNet (Li et al., 2026) incorporates gauge-constrained QPE priors into radar-based forecasting. Although these approaches demonstrate the value of surface observations, they retain an explicit interpolation or QPE stage for constructing dense precipitation representations.

Generative data assimilation provides another way to incorporate sparse observations without explicit interpolation. For example, Manshausen et al. (2025) condition a learned generative prior on sparse weather-station observations to recover spatially coherent precipitation fields. However, the underlying dense prior is still learned from an existing atmospheric analysis. In contrast, Yushi learns a gauge-constrained precipitation representation directly from radar and sparse AWS observations, without requiring gauge interpolation or an external gridded precipitation analysis. Radar provides spatial structure while AWS measurements anchor the representation to surface precipitation, yielding a shared state for both precipitation estimation and subsequent nowcasting.

2.3 Generative Precipitation Nowcasting and Intensity-Aware Learning

Generative models have recently been adopted to better capture the stochasticity and multi-modality of future precipitation evolution. Latent diffusion approaches (Ho et al., 2020; Rombach et al., 2022) such as PreDiff (Gao et al., 2023), DiffCast (Yu et al., 2024) and CasCast (Gong et al., 2024) perform generation in compressed latent spaces, while FlowCast (Perrone Ribeiro & Faganeli Pucer, 2026) applies flow matching to latent representations for radar nowcasting. PixelFlowCast (Zhu et al., 2026) further explores flow-based generation without latent compression. These methods demonstrate the effectiveness of generative modeling for precipitation forecasting, but generally operate on predefined radar or precipitation representations. In contrast, Yushi learns a precipitation-aware latent state under sparse AWS supervision and performs flow matching directly in this representation, thereby coupling precipitation estimation with future state evolution.

Heavy precipitation presents an additional challenge because intense rainfall events are relatively rare but particularly important for hazard-oriented nowcasting. Existing approaches address this imbalance through intensity-weighted pixel losses (Shi et al., 2017b; Choi et al., 2025) or distribution-level objectives such as probability matching (Cao et al., 2025). These methods primarily reweight errors or constrain precipitation statistics in the output space. Other imbalance-aware approaches have considered latent representations, such as extreme-value objectives for discrete VQGAN tokens (Esser et al., 2021; Bi et al., 2023). Yushi instead transfers precipitation-intensity weights to corresponding spatial locations in a continuous latent field and uses them to weight local flow-matching errors, directly linking physical precipitation intensity with latent generative optimization.

3 Method

Yushi is a unified framework that directly maps primary observations to future precipitation fields, without relying on a pre-constructed gridded precipitation analysis. As illustrated in Figure 1, the framework consists of a gauge-constrained representation (GCR) learner for latent space modeling and a causal conditional flow matching (CCFM) model for forecasting. The GCR learner first trans-

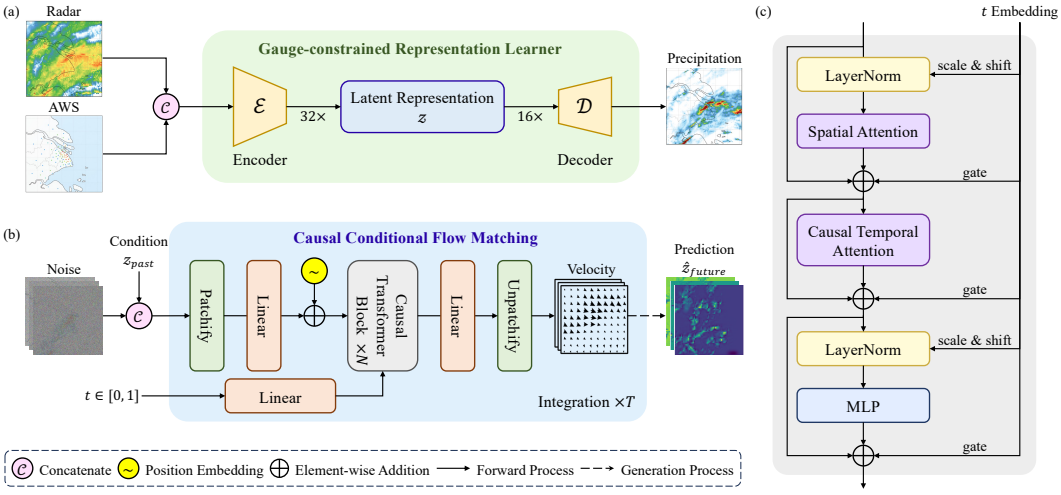

Figure 1: Overall framework of Yushi. **(a)** The gauge-constrained representation learner fuses radar echo intensity with AWS observations to construct a unified latent precipitation representation, and decodes it to a dense precipitation field. **(b)** The causal conditional flow matching model performs predictions in the latent space using causal Transformer blocks as its basic units. **(c)** Within each causal Transformer block, spatial attention and causal temporal attention are computed sequentially.

forms radar reflectivity and sparse AWS observations into a latent space that captures precipitation characteristics, providing a shared state that preserves both the spatial organization captured by radar and the quantitative surface precipitation information supplied by gauges. The CCFM model then learns the temporal evolution of these latent states and jointly generates future latent representations conditioned on the past observations. After iteratively predicting the velocity field of the trajectory and performing integration, the initial Gaussian noise gradually transforms into a state distribution that conforms to the characteristics of future precipitation. Finally, the predicted latent states are decoded by the GCR decoder into spatially complete precipitation fields. The framework is designed to unify precipitation estimation and nowcasting within the same latent representation, and a detailed description of each component is provided below.

3.1 Gauge-Constrained Representation Learning

The architecture of GCR learner adopts a hierarchical encoder-decoder design with residual blocks and self-attention modules to preserve both local details and large-scale spatial structure. Given a radar reflectivity field r in dBZ, sparse precipitation observations y, and the corresponding station availability mask m, the GCR learner encodes these observations into the parameters of a latent distribution and obtains a latent representation z through the reparameterization trick:

$$(\mu, \sigma) = \mathcal{E}(r, y, m), \quad z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{1}$$

where $\mathcal{E}$ denotes the encoder and μ and σ parameterize the latent distribution. The latent representation is then decoded into a spatially complete precipitation field:

$$\hat{x} = \mathcal{D}(z), \tag{2}$$

where $\mathcal{D}$ denotes the decoder. Importantly, radar reflectivity itself is not directly supervised in the precipitation space. Instead, we utilize the Z-R relationship to derive a dense precipitation field from radar reflectivity as background x_b and use it only as a training constraint. The GCR learner is trained with the following composite objective:

$$\mathcal{L}_{\text{GCR}} = \alpha \left\| \hat{x} - x_b \right\|_1 + \beta \left\| m \odot (\hat{x} - y) \right\|_1 + \mathcal{L}_{\text{SSIM}} + \mathcal{L}_{\text{grad}} + \eta \mathcal{L}_{\text{KL}}, \tag{3}$$

where the L1 term $\|\cdot\|_1$ is implemented as mean absolute errors (MAE) over the corresponding valid locations. The first two terms impose quantitative constraints on precipitation values, while the remaining terms provide regularization constraints on the spatial structure and latent representation. For detailed explanations and calculation formulas regarding the various losses, please refer to Appendix A.1. After training, the parameters of GCR learner are frozen and the encoder and decoder are used to construct the latent states and decode the precipitation fields for subsequent nowcasting.

3.2 CAUSAL CONDITIONAL FLOW MATCHING

Once the GCR learner provides precipitation-aware latent representations, we model their future evolution using CCFM. Given the past GCR sequence as the conditioning context, Yushi jointly generates the subsequent future latent frames in a single flow trajectory. During training, the target future latent sequence z_1 is obtained by encoding the corresponding future observations with the GCR encoder. A Gaussian noise sample ϵ is then interpolated with z_1 to construct an intermediate state along the probability path:

$$z_t = (1-t)\epsilon + tz_1, \quad u_t = z_1 - \epsilon, \quad t \in [0,1], \quad \epsilon \sim \mathcal{N}(0, I). \tag{4}$$

The velocity field v is trained to match the target velocity u_t by minimizing

$$\mathcal{L}_{\mathrm{CFM}} = \mathbb{E}_{t,z_1,\epsilon}\left[\|v(z_t, t \mid c) - u_t\|_2^2\right], \tag{5}$$

where c denotes the past GCR sequence, and the velocity-matching objective is implemented as mean squared errors (MSE) over the latent dimensions. At inference time, Gaussian noise is transformed into the future latent sequence by integrating the learned velocity field conditioned on the past representations.

We parameterize the velocity field with a causal spatiotemporal Transformer. The past representations serve as clean latent prompts and are concatenated with the future flow states along the temporal dimension. Each Transformer block consists of spatial attention and causal temporal attention, with the flow time embedding t injected through adaptive layer normalization (AdaLN), as illustrated in Figure 1c. The causal attention mask prevents each frame from attending to subsequent future frames:

$$\mathrm{CausalAttn}(Q, K, V) = \mathrm{Softmax}\left(\frac{QK^\top}{\sqrt{C}} + M\right)V, \tag{6}$$

where Q, K, and V denote the query, key, and value tensors, C is the channel dimension, and M is the causal mask. Thus, although the future frames are generated jointly within a single flow trajectory, the velocity prediction for each frame can only access the past context and preceding future frames, preventing information leakage from later frames. The resulting future latent states are finally decoded by the frozen GCR decoder into precipitation fields.

3.3 INTENSITY-WEIGHTED LATENT OPTIMIZATION

Precipitation exhibits a highly imbalanced intensity distribution, with weak precipitation occurring far more frequently than heavy precipitation. In standard MSE-based regression, this long-tailed distribution causes the optimization to be dominated by abundant weak-precipitation samples, leading to systematic underestimation of rare heavy precipitation (Cao et al., 2023; 2025). CFM differs from direct regression in that it supervises the velocity field along probability-flow trajectories rather than the generated latents directly. Nevertheless, the same imbalance can persist in velocity learning: gradient contributions from heavy-precipitation regions can be overwhelmed by those from weak-precipitation regions, resulting in insufficient transport toward intense precipitation states.

To alleviate this imbalance, we introduce an intensity-aware balanced (IAB) loss to optimize the flow matching objective, which transfers precipitation intensity information from the decoded precipitation field to the latent flow-matching loss:

$$\mathcal{L}_{\mathrm{IAB}} = \mathbb{E}_{t,z_1,\epsilon}\left[\frac{1}{N}\sum_{i=1}^{N} w_i^{\gamma}\left(v_i(z_t, t \mid c) - u_{t,i}\right)^2\right], \quad w = \mathrm{AvgPool}_{8\times 8}(1 + \mathcal{D}(z_1)). \tag{7}$$

where the spatial weight w is obtained by adding a constant offset of 1 to the precipitation field decoded from the GCR representation and applying 8×8 average pooling. This produces local precipitation-intensity weights at the spatial resolution of the latent features, thereby transferring intensity information from the physical space to the latent space. The exponent γ controls the strength of intensity-based reweighting. A larger γ increases the relative contribution of latent locations associated with intense precipitation, while preserving the standard velocity-matching objective for weak-precipitation regions. When $\gamma = 0$, the IAB loss degenerates into the standard CFM loss. The resulting IAB loss replaces the unweighted CFM objective during training, encouraging the learned velocity field to allocate greater capacity to the accurate transport of heavy-precipitation states.

4 EXPERIMENTS

4.1 DATASET AND IMPLEMENTATION DETAILS

We collect radar echo reflectivity and AWS observational data for the Yangtze River Delta region (Shanghai and its surrounding areas, covering a 461 × 461 km domain) spanning from January 2023 to August 2026. By constructing time-series samples using a 15-minute sliding window and filtering for continuous precipitation events, we obtain 38,447 samples with 5-minute observation intervals and 1 km spatial resolutions. Samples from January 2023 to December 2025 are used for model training, samples from January to April 2026 for validation, and samples from May to August 2026 for testing. During the pre-processing stage, we map the value from each station to the nearest grid point within a 500m radius, thereby gridding the AWS observations and achieving spatial alignment with the radar data. Corresponding station availability mask is then generated—assigning a value of 1 where a station is present and 0 otherwise—to allow grid points lacking stations to be excluded during loss calculation.

The GCR learner is trained using radar-AWS data pairs from a single time step as independent samples. The latent space dimensions are set to 58 × 58 with 4 channels. The CCFM utilizes a sequence of latent representations from past time steps as conditions to generate a sequence of representations for future time steps. The weighting coefficient γ for the IAB loss is set to 1 in default experiments. Both the GCR learner and the CCFM are trained for 50 epochs using the AdamW optimizer, with a batch size of 4 and a learning rate decaying from 1e-4 to 1e-5. During inference, we integrate the learned velocity field using a second-order multistep solver with $T = 20$ uniformly spaced flow-time steps.

To evaluate the forecasting performance of Yushi, we compare it with several representative nowcasting baselines, including ConvGRU (Shi et al., 2017b), SimVP (Gao et al., 2022a), PhyDNet (Guen & Thome, 2020), Earthformer (Gao et al., 2022b), and FlowCast (Perrone Ribeiro & Faganeli Pucer, 2026), under a unified experimental setting. For fair comparison, all models are trained and evaluated in the same computing environment using 4 NVIDIA H800 GPUs. They take data from the preceding 2 hours (i.e., 24 steps) as input to forecast precipitation for the subsequent 2 hours. Since the baseline models do not support sparse AWS observations as input, they are trained and evaluated directly on the background precipitation fields derived from radar reflectivity using the Z-R relation. This setting ensures that all methods directly predict precipitation fields rather than radar reflectivity, thereby maintaining a consistent prediction target across models. We do not adopt an alternative setting in which the baseline models first predict radar reflectivity and subsequently convert it to precipitation. Because the Z-R relation involves a nonlinear power-law transformation, even relatively small reflectivity errors in high-intensity regions can be substantially amplified after conversion, resulting in disproportionately large precipitation errors and degraded evaluation metrics (see Equation 8 and Figure 4 in the Appendix). Definitions and calculation methods for all evaluation metrics can be found in the Appendix A.2.

4.2 COMPARISON TO THE STATE OF THE ART

Table 1: Quantitative comparison of nowcasting under 5-minute precipitation thresholds of 1, 2 and 4mm. All metrics are computed over a 24-step forecast, and M means the average across all thresholds to evaluate general performance. For BIAS, values closer to 1 are better, while for the other metrics, higher values indicate better performance.

Model	CSI-M	HSS-M	BIAS-M	CSI			CSI-4POOL			CSI-16POOL		
				1mm	2mm	4mm	1mm	2mm	4mm	1mm	2mm	4mm
ConvGRU	0.0788	0.1434	0.2228	0.1071	0.0805	0.0489	0.1084	0.0806	0.0488	0.1339	0.0953	0.0580
SimVP	0.0648	0.1195	0.1632	0.0904	0.0631	0.0408	0.0910	0.0635	0.0411	0.1123	0.0784	0.0499
PhyDNet	0.0810	0.1466	0.2264	0.1168	0.0809	0.0453	0.1181	0.0812	0.0454	0.1438	0.0962	0.0536
Earthformer	0.0800	0.1450	0.2522	0.1114	0.0820	0.0466	0.1128	0.0824	0.0464	0.1392	0.1005	0.0565
FlowCast	0.0560	0.1031	0.3451	0.0765	0.0569	0.0345	0.0815	0.0605	0.0370	0.1582	0.1202	0.0754
Yushi	**0.1079**	**0.1921**	**0.7798**	**0.1610**	**0.1052**	**0.0575**	**0.1657**	**0.1109**	**0.0625**	**0.2492**	**0.1941**	**0.1269**

Table 1 shows that Yushi consistently outperforms all baselines across CSI, HSS, BIAS, and pooled CSI metrics. The gains are especially clear for heavier precipitation and under spatial pooling, in-

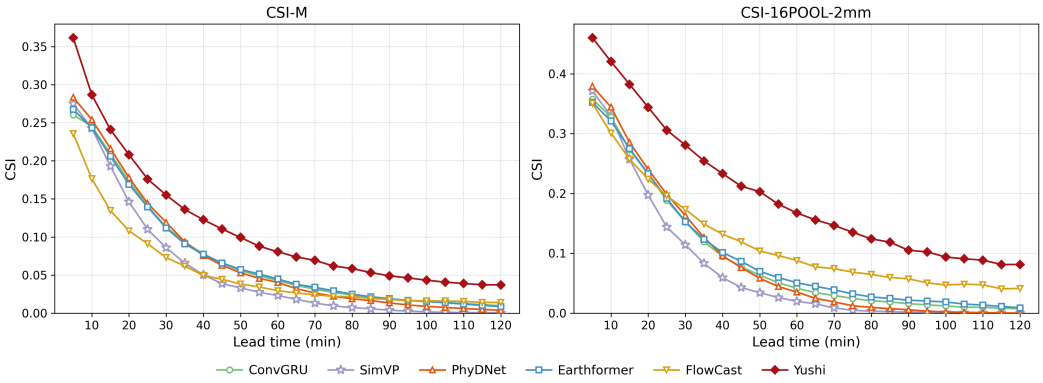

Figure 2: Frame-wise CSI-M and CSI-16POOL under 2mm threshold versus lead time.

dicating stronger skill in capturing intense precipitation within approximately correct spatial neighborhoods. The improvement in BIAS is particularly notable. BIAS measures the ratio between predicted and observed events, with 1 indicating an unbiased event frequency, values below 1 indicating under-forecasting, and values above 1 indicating over-forecasting. While all baselines exhibit substantial under-forecasting, Yushi achieves a BIAS-M much closer to 1. Together with its higher CSI and HSS, this suggests that Yushi reduces missed precipitation events without simply increasing false alarms. Figure 2 illustrates the temporal evolution of CSI-M and CSI-16POOL under 2mm threshold, showing that Yushi outperforms baseline models at all time steps. FlowCast is an ensemble forecasting method, which exhibits the lowest unpooled CSI when comparing under deterministic nowcasting. However, its pooled CSI outperforms other baseline models, particularly at larger aggregation scales. This contrast indicates that FlowCast preserves signals of heavy precipitation, even if the predicted location deviates slightly—errors that are difficult to avoid in convective nowcasting due to the inherent limits of moist convection predictability (Zhang et al., 2003; 2006). Spatial aggregation relaxes the requirement for pixel-level alignment, thereby better reflecting the model's ability to capture the occurrence and approximate location of heavy precipitation. The visualizations in Figure 3 further corroborate these results, and more visualizations can be found in the Appendix A.3.

Part of Yushi's improvement is attributable to its use of sparse AWS observations, whereas the baseline models rely only on radar-derived precipitation fields. Since AWS observations provide direct quantitative constraints on surface precipitation and are also used for evaluation, this gives Yushi additional information that is unavailable to the baselines. This setting is nevertheless operationally meaningful, as the objective is to exploit all available observations to improve forecast accuracy—which is the original design intent of the Yushi framework. To examine whether the gains are concentrated at station locations already seen by the model, we further conduct a station-holdout experiment in which 20% of the AWS stations are completely excluded from both training and test-time inputs and are used only for evaluation. As shown in Table 2, the baseline models are almost unaffected by this protocol because they do not use AWS observations as input. In contrast, Yushi experiences only a moderate performance decrease and still clearly outperforms all baselines across the main metrics. These results suggest that Yushi does not simply benefit from repeatedly observing the same gauge locations during training. Instead, the learned radar-AWS representation transfers useful precipitation information to locations whose gauges are never provided to the model, supporting the spatial generalization capability of the proposed methods.

In addition, we evaluate the QPE performance on the same 20% holdout stations. As shown in Table 3, Yushi consistently achieves lower MSE and higher correlation than both bilinear interpolation from AWS observations and the conventional Z-R conversion from radar reflectivity across all precipitation thresholds. The advantage remains clear for heavier precipitation, where estimation is typically more challenging. The results indicate that the GCR learner does not merely reproduce either the sparse gauge field or the radar-derived background precipitation. Instead, it effectively combines the complementary information from the two observation sources to produce better precipitation fields.

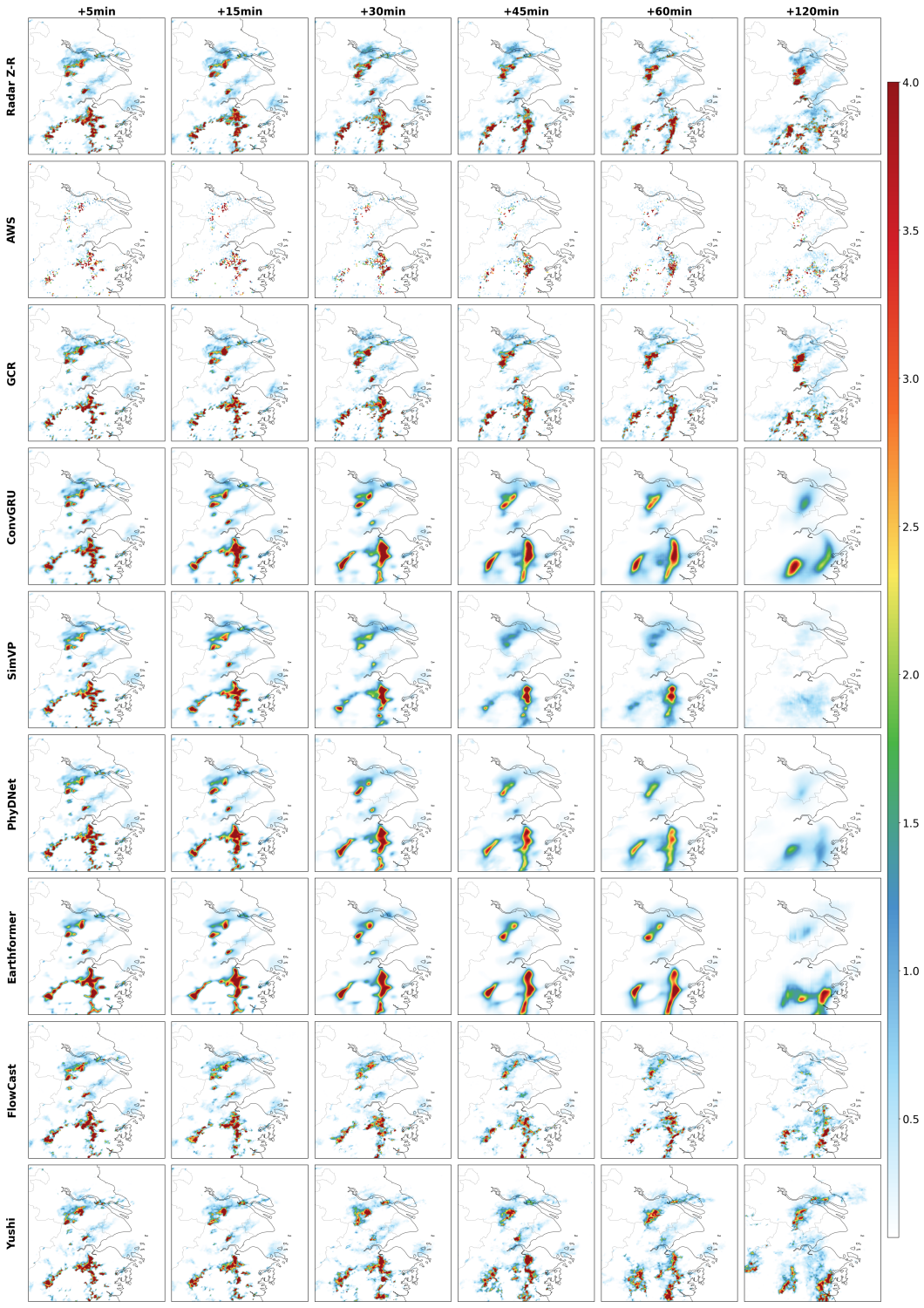

Figure 3: Visual comparison of precipitation nowcasting across different models. The first three rows show, respectively, the radar-derived background precipitation, sparse AWS observations, and GCR decoding output for the corresponding times. The remaining rows display the forecast results from the various models. The initialization date-time is set to 10:00 UTC on August 14, 2026.

Table 2: Quantitative comparison of precipitation nowcasting on the holdout stations.

Model	CSI-M	HSS-M	BIAS-M	CSI			CSI-4POOL			CSI-16POOL		
				1mm	2mm	4mm	1mm	2mm	4mm	1mm	2mm	4mm
ConvGRU	0.0784	0.1427	0.2199	0.1066	0.0793	0.0493	0.1069	0.0793	0.0492	0.1208	0.0874	0.0531
SimVP	0.0645	0.1191	0.1621	0.0903	0.0624	0.0410	0.0903	0.0624	0.0410	0.1006	0.0698	0.0445
PhyDNet	0.0809	0.1464	0.2241	0.1164	0.0803	0.0459	0.1168	0.0805	0.0459	0.1309	0.0882	0.0493
Earthformer	0.0797	0.1445	0.2497	0.1110	0.0811	0.0471	0.1113	0.0813	0.0472	0.1256	0.0913	0.0520
FlowCast	0.0565	0.1040	0.3499	0.0772	0.0572	0.0351	0.0783	0.0579	0.0358	0.1061	0.0794	0.0489
Yushi	**0.1035**	**0.1851**	**0.7648**	**0.1555**	**0.1000**	**0.0551**	**0.1570**	**0.1013**	**0.0563**	**0.1925**	**0.1347**	**0.0803**

Table 3: Comparison of different quantitative precipitation estimation methods on the holdout stations. Lower MSE and higher CORR values are better.

Method	MSE			CORR		
	1mm	2mm	4mm	1mm	2mm	4mm
Bilinear interpolation from AWS	2.4149	5.1681	12.8109	0.5540	0.4574	0.3489
Z-R conversion from radar reflectivity	2.9779	5.3865	10.2892	0.5720	0.5082	0.3805
Yushi (decoding from GCR)	**2.0254**	**3.8527**	**8.3188**	**0.6720**	**0.5959**	**0.4570**

4.3 Ablation Study

We conduct ablation studies on the two key modules and the designed loss function, as shown in Table 4. Specifically, Yushi w/o GCR replaces the GCR with compressed Z-R converted background precipitation, Yushi w/o CCFM replaces the CCFM with the FlowCast model, and Yushi w/o IAB employs the standard CFM loss with γ set to 0. The experiments also compare Yushi with other baseline models under the same settings from various perspectives, demonstrating that our method yields performance improvements of a general nature, rather than benefits specific only to our model. Furthermore, we investigate the role of γ by adjusting it to 1.5 and 0.5. As γ increases, the pooled CSI and BIAS values rise continuously—even leading to over-forecasting—indicating a more aggressive forecasting strategy for extreme precipitation. This also provides a controllable parameter to adjust the forecasting style according to different operational requirements.

Table 4: Ablation study for designed components and hyper-parameters in Yushi.

Model	CSI-M	HSS-M	BIAS-M	CSI			CSI-4POOL			CSI-16POOL		
				1mm	2mm	4mm	1mm	2mm	4mm	1mm	2mm	4mm
Yushi	**0.1079**	**0.1921**	0.7798	**0.1610**	**0.1052**	**0.0575**	**0.1657**	**0.1109**	**0.0625**	0.2492	0.1941	0.1269
Yushi w/o GCR	0.0774	0.1428	0.5766	0.1058	0.0770	0.0495	0.1117	0.0817	0.0532	0.1946	0.1484	0.1009
Yushi w/o CCFM	0.0643	0.1156	**0.8881**	0.0977	0.0607	0.0346	0.1021	0.0644	0.0375	0.1825	0.1338	0.0849
Yushi w/o IAB	0.1019	0.1826	0.4521	0.1513	0.0975	0.0568	0.1547	0.1020	0.0612	0.2247	0.1697	0.1137
Yushi $\gamma = 0.5$	0.1026	0.1837	0.5172	0.1526	0.0982	0.0569	0.1568	0.1033	0.0613	0.2362	0.1783	0.1174
Yushi $\gamma = 1.5$	0.1042	0.1863	1.1224	0.1552	0.1012	0.0564	0.1608	0.1074	0.0616	**0.2561**	**0.1987**	**0.1298**

5 Conclusion

We present Yushi, a unified framework for precipitation estimation and nowcasting directly from radar reflectivity and sparse AWS observations. Yushi learns a gauge-constrained latent representation, models its future evolution with causal conditional flow matching, and introduces intensity-aware latent optimization to better capture heavy precipitation. Experiments show consistent improvements over representative baselines, including reduced under-forecasting, stronger performance for intense precipitation, and good generalization to holdout stations. The learned representation also improves QPE over conventional Z-R conversion and gauge interpolation.

Several limitations remain. First, the current experiments are limited to the Yangtze River Delta region, and broader evaluation across different climates and observing systems is needed. Second, Yushi still depends on the availability and quality of AWS observations, motivating further study under sparser station networks. Third, although the forecasting model is generative, the present evaluation focuses mainly on deterministic skill. Future work should more systematically assess ensemble diversity, calibration, and predictive uncertainty.

AI USE STATEMENT

In this work, we used generative AI tools to provide feedback on the presentation of the methodology and experiments and to assist with the interpretation and presentation of experimental results. We also used generative AI tools for manuscript drafting and language editing, including improving the clarity, readability, and structure of the paper. The research methodology, model implementation, experiments, and reported results were developed, conducted, and verified by the authors. All AI-assisted suggestions and text were critically reviewed and revised by the authors. We take full responsibility for the final content of this work, including all claims, results, and conclusions.

REPRODUCIBILITY STATEMENT

To support reproducibility, codes will be available at `https://anonymous.4open.science/r/Yushi-1B03` after publication. Detailed descriptions of the dataset, data splits, and preprocessing procedures are provided in the experimental section of the main paper. The raw radar reflectivity and AWS observations used in this study can be obtained by contacting the corresponding author. These resources, together with the implementation and experimental details reported in the paper, are intended to facilitate reproduction of the proposed method and results.

REFERENCES

Georgy Ayzel, Tobias Scheffer, and Maik Heistermann. RainNet v1.0: A convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, 13(6):2631–2644, 2020.

Alexis Berne and Witold F Krajewski. Radar for hydrology: Unfulfilled promise or unrecognized potential? *Advances in Water Resources*, 51:357–366, 2013.

Haoran Bi, Maksym Kyryliuk, Zhiyi Wang, Cristian Meo, Yanbo Wang, Ruben Imhoff, Remko Uijlenhoet, and Justin Dauwels. Nowcasting of extreme precipitation using deep generative models. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023.

Yuan Cao, Danchen Zhang, Xin Zheng, Hongming Shan, and Junping Zhang. Mutual information based reweighting for precipitation nowcasting. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Yuan Cao, Shuaiyi Li, Jie Feng, Lei Chen, Hao Li, and Yijun Zhang. Enhancing machine learning models for nowcasting and short-term forecasting of precipitation with a novel probability-matching loss function. *Geophysical Research Letters*, 52(24):e2025GL119442, 2025.

Hyojeong Choi, Yongchan Kim, Soohyun Kim, and Dongkyun Kim. Unveiling the role of weighted loss functions in deep learning-based nowcasting of extreme rainfall events. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–22, 2025.

Lasse Espeholt, Shreya Agrawal, Casper Sønderby, Manoj Kumar, Jonathan Heek, Carla Bromberg, Cenk Gazen, Rob Carver, Marcin Andrychowicz, Jason Hickey, et al. Deep learning for twelve hour precipitation forecasts. *Nature communications*, 13(1):5145, 2022.

Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12873–12883, 2021.

Richard A. Fulton, Jay P. Breidenbach, Dong-Jun Seo, David A. Miller, and Timothy O'Bannon. The WSR-88D rainfall algorithm. *Weather and Forecasting*, 13(2):377–395, 1998.

Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z Li. Simvp: Simpler yet better video prediction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3160–3170. IEEE, 2022a.

Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Bernie Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems*, 35:25390–25403, 2022b.

Zhihan Gao, Xingjian Shi, Boran Han, Hao Wang, Xiaoyong Jin, Danielle Maddix, Yi Zhu, Mu Li, and Yuyang (Bernie) Wang. PreDiff: Precipitation nowcasting with latent diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 78621–78656, 2023.

Junchao Gong, Lei Bai, Peng Ye, Wanghan Xu, Na Liu, Jianhua Dai, Xiaokang Yang, and Wanli Ouyang. CasCast: Skillful high-resolution precipitation nowcasting via cascaded modelling. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 15809–15822. PMLR, 2024.

Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11474–11484, 2020.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Witold F Krajewski, Gabriele Villarini, and James A Smith. Radar-rainfall uncertainties: Where are we after thirty years of effort? *Bulletin of the American Meteorological Society*, 91(1):87–94, 2010.

Leijing Li, Jianzhu Li, Yongtao Wang, Chongju Shang, Zhixia Li, and Jin Liu. Deep learning-based precipitation nowcasting integrating radar echoes and rain gauge data. *Journal of Hydrology*, 679: 136210, 2026.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Peter Manshausen, Yair Cohen, Peter Harrington, Jaideep Pathak, Mike Pritchard, Piyush Garg, Morteza Mardani, Karthik Kashinath, Simon Byrne, and Noah Brenowitz. Generative data assimilation of sparse weather station observations at kilometer scales. *Journal of Advances in Modeling Earth Systems*, 17(10):e2024MS004505, 2025.

J. S. Marshall and W. M. Palmer. The distribution of raindrops with size. *Journal of Meteorology*, 5 (4):165–166, 1948.

Jack L. McKee and Andrew D. Binns. A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, 41(1–2):186–203, 2016.

Susana Ochoa-Rodriguez, L-P Wang, Patrick Willems, and Christian Onof. A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resources Research*, 55(8):6356–6391, 2019.

Bernardo Perrone Ribeiro and Jana Faganeli Pucer. Flowcast: Advancing precipitation nowcasting with conditional flow matching. In *International Conference on Learning Representations*, volume 2026, pp. 156863–156883, 2026.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. ieee, 2022.

Henri Sauvageot. Rainfall measurement by radar: A review. *Atmospheric Research*, 35(1):27–54, 1994.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-Chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 28, pp. 802–810, 2015.

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in neural information processing systems*, 30, 2017a.

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-Chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. In *Advances in Neural Information Processing Systems*, volume 30, 2017b.

Casper Kaae Sønderby, Lasse Espeholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner. Metnet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*, 2020.

Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S Yu. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Advances in neural information processing systems*, 30, 2017.

James W Wilson and Edward A Brandes. Radar measurement of rainfall—a summary. *Bulletin of the American Meteorological Society*, 60(9):1048–1060, 1979.

Demin Yu, Xutao Li, Yunming Ye, Baoquan Zhang, Chuyao Luo, Kuai Dai, Rui Wang, and Xunlai Chen. DiffCast: A unified framework via residual diffusion for precipitation nowcasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27758–27767, 2024.

Fuhan Zhang, Xiaodong Wang, Jiping Guan, Meihan Wu, and Lina Guo. RN-Net: A deep learning approach to 0–2 hour rainfall nowcasting based on radar and automatic weather station data. *Sensors*, 21(6):1981, 2021.

Fuqing Zhang, Chris Snyder, and Richard Rotunno. Effects of moist convection on mesoscale predictability. *Journal of the Atmospheric Sciences*, 60(17):1173–1185, 2003.

Fuqing Zhang, Andrew M. Odins, and John W. Nielsen-Gammon. Mesoscale predictability of an extreme warm-season precipitation event. *Weather and Forecasting*, 21(2):149–166, 2006.

Jian Zhang, Kenneth Howard, Carrie Langston, Brian Kaney, Youcun Qi, Lin Tang, Heather Grams, Yadong Wang, Stephen Cocks, Steven Martinaitis, et al. Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, 97(4):621–638, 2016.

Yuchen Zhang, Mingsheng Long, Kaiyuan Chen, Lanxiang Xing, Ronghua Jin, Michael I. Jordan, and Jianmin Wang. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619 (7970):526–532, 2023.

Yufeng Zhu, Chunlei Shi, Yongchao Feng, and Dan Niu. PixelFlowCast: Latent-free precipitation nowcasting via pixel mean flows. *arXiv preprint arXiv:2605.10046*, 2026.

A Appendix

A.1 Constraints for GCR Learning

Quantitative constraints. Because radar reflectivity is not directly a precipitation quantity, we first convert the input dBZ field into 5-minute accumulated precipitation (Fulton et al., 1998; Marshall & Palmer, 1948):

$$x_b = \left(\frac{10^{\mathrm{dBZ}/10}}{A}\right)^{1/b} \cdot \frac{1}{12}, \quad (A, b) = \begin{cases} (300, 1.4), & \text{convective}, \\ (200, 1.6), & \text{stratiform}. \end{cases} \tag{8}$$

Here, the rain type is determined according to the fraction of pixels exceeding 40 dBZ. The resulting x_b provides dense but potentially biased precipitation guidance derived from radar. To avoid treating the background precipitation derived from Z-R relationship as exact ground truth, we assign the AWS loss term a substantially larger weight, with $\alpha = 0.1$ and $\beta = 50$, so that the decoded precipitation field is strongly anchored to the direct surface observations. In this way, AWS observations correct the quantitative biases that may remain in the background precipitation while radar supplies dense spatial information between stations.

Regularization constraints. Sparse gauge supervision alone cannot fully determine the fine-scale spatial structure of precipitation. We therefore further constrain the decoded field using the background precipitation through structural similarity and gradient losses. Specifically, the SSIM term measures the structural similarity between the reconstruction and the radar QPE background. For each location p, the local means, variances, and covariance of the two fields are computed over 11×11 sliding windows (implemented via average pooling), yielding

$$\mathcal{L}_{\mathrm{SSIM}} = 1 - \frac{1}{|\Omega|} \sum_{p \in \Omega} \frac{(2\,\mu_{\hat{x}}(p)\,\mu_{x_b}(p) + C_1)\,(2\,\sigma_{\hat{x}x_b}(p) + C_2)}{\left(\mu_{\hat{x}}^2(p) + \mu_{x_b}^2(p) + C_1\right)\left(\sigma_{\hat{x}}^2(p) + \sigma_{x_b}^2(p) + C_2\right)}, \tag{9}$$

where μ and σ^2 denote the local mean and variance, $\sigma_{\hat{x}x_b}$ the local covariance, and $C_1 = 0.01^2$, $C_2 = 0.03^2$ are small constants stabilizing the ratio in degenerate regions. The gradient term

$$\mathcal{L}_{\mathrm{grad}} = \|\nabla_x \hat{x} - \nabla_x x_b\|_1 + \|\nabla_y \hat{x} - \nabla_y x_b\|_1 \tag{10}$$

penalizes discrepancies in the first-order horizontal and vertical finite differences, preserving the sharp edges and fine-grained textures of precipitation fields. In addition, we impose a weak KL regularization:

$$\mathcal{L}_{\mathrm{KL}} = D_{\mathrm{KL}}\left[q(z \mid r, y, m) \,\|\, \mathcal{N}(0, I)\right], \tag{11}$$

with $\eta = 10^{-7}$. The small coefficient makes this term a mild regularizer rather than a dominant training objective, allowing the latent representation to be primarily determined by the precipitation observations and their spatial constraints.

A.2 Evaluation Metrics

To comprehensively quantify forecast and estimation accuracy, we adopt a multi-level metric system covering continuous, categorical, and neighborhood-pooled evaluations.

Continuous metrics. MSE (mean squared error) and CORR (correlation coefficient) evaluate the overall numerical deviation and linear association between forecasts $\hat{x}$ and ground truth x:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2, \quad \mathrm{CORR} = \frac{\sum_{i=1}^{N}\left(\hat{x}_i - \bar{\hat{x}}\right)(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{N}\left(\hat{x}_i - \bar{\hat{x}}\right)^2}\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}, \tag{12}$$

where CORR approaching 1 and MSE approaching 0 indicate better performance.

Categorical metrics. For heavy and extreme precipitation evaluation, we compute contingency-based scores at precipitation thresholds. Let H, M, F, and C denote the numbers of hits, misses, false alarms, and correct negatives, respectively. The critical success index (CSI) and BIAS are defined as

$$\mathrm{CSI} = \frac{H}{H + M + F}, \quad \mathrm{BIAS} = \frac{H + F}{H + M}. \tag{13}$$

The Heidke skill score (HSS) measures the fractional improvement over random forecasts:

$$\mathrm{HSS} = \frac{2\,(H * C - F * M)}{(H + M)\,(M + C) + (H + F)\,(F + C)}. \tag{14}$$

CSI, BIAS and HSS approach 1 in the optimal case.

Pooled metrics. To tolerate small displacement errors of heavy-precipitation cores, we additionally report CSI on spatially pooled fields (CSI-4POOL and CSI-16POOL). Both the forecast and ground-truth fields are first aggregated with 4×4 and 16×16 max pooling, respectively, and CSI is then computed on the pooled fields under the same precipitation thresholds.

A.3 Supplementary Figures

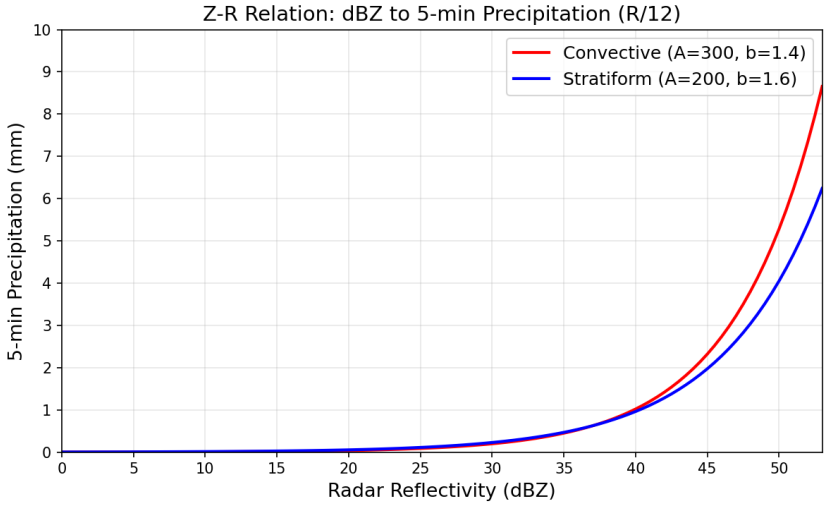

Figure 4: The mapping function curve corresponding to the Z-R relationship.

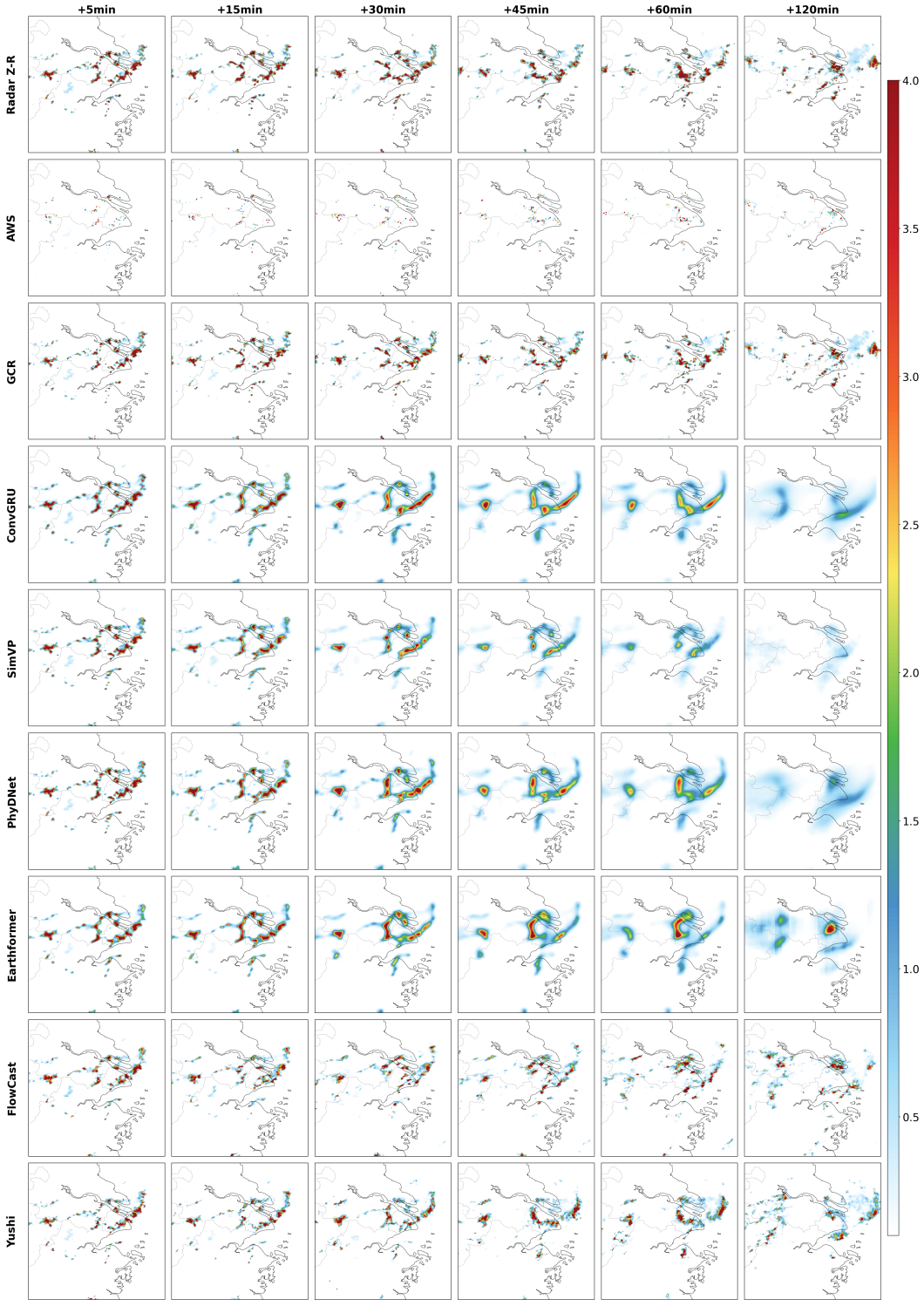

Figure 5: Visual comparison of precipitation nowcasting at 11:30 UTC on July 18, 2026.



Figure 6: Visual comparison of precipitation nowcasting at 09:50 UTC on July 21, 2026.

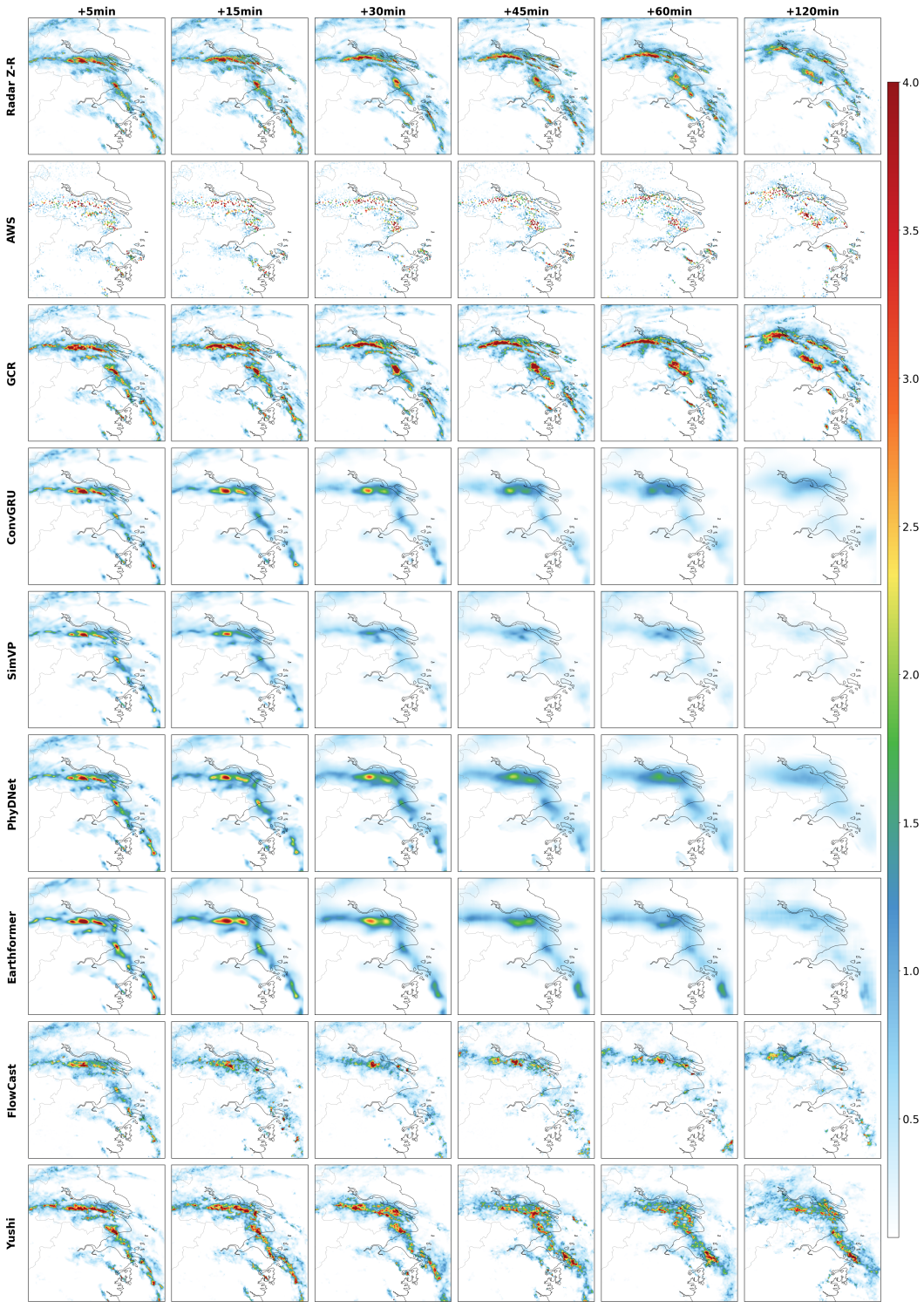

Figure 7: Visual comparison of precipitation nowcasting at 00:15 UTC on August 10, 2026.