IMALCAST: INTENSITY-AWARE MULTIMODAL ALIGNMENT LEARNING FOR HEAVY-RAINFALL NOWCASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate precipitation nowcasting, especially for extreme rainfall events, plays a vital role in supporting decision-making across multiple sectors of society. However, pointwise regression objectives commonly used in deep learning models are not well aligned with threshold-sensitive operational verification, particularly for high-intensity precipitation regions. Geostationary satellite observations offer complementary cloud information for precipitation forecasting, but differences in spatial and temporal resolution complicate their integration with radar data. To address these challenges, we propose IMALCast, an intensity-aware multimodal alignment learning framework for deterministic precipitation nowcasting. For objective alignment, IMALCast integrates weighted intensity regression with global and high-intensity structural losses motivated by threshold-based verification. For observation alignment, IMALCast uses dual-tower encoding and cross-modal fusion to preserve and integrate modality-specific radar and satellite information. Experiments on the nationwide China Radar–Satellite Dataset and the public SEVIR benchmark demonstrate strong forecasting performance. In particular, IMALCast achieves up to a 65.7% relative improvement in the regional-scale critical success index for heavy rainfall over the strongest baseline in our comparison.

1 INTRODUCTION

Precipitation nowcasting aims to predict the evolution of rainfall for the next few hours, and is critical for disaster early warning, urban flood management, power grid scheduling, and transportation safety (Ravuri et al., 2021; Zhang et al., 2023; Shi et al., 2017). Recent deep learning models have shown strong potential for learning spatiotemporal dynamics directly from observational sequences (Shi et al., 2015; Ayzel et al., 2020; Gao et al., 2022b; Espeholt et al., 2022). In this work, we consider this task using nationwide radar and geostationary satellite observations over mainland China, with a focus on forecasting heavy rainfall. A central challenge is that commonly used pointwise training objectives are not well aligned with threshold-based operational verification. We refer to this discrepancy as *objective misalignment*.

For heavy precipitation, the practical value of a nowcast depends on its ability to predict where and when hazardous rainfall will occur, as well as its intensity (Skinner et al., 2018; Ferro & Stephenson, 2011; Ebert et al., 2013). To reflect this priority, threshold-based verification evaluates predicted and observed regions exceeding physically meaningful rainfall levels, using categorical scores such as the Critical Success Index (CSI) (Schaefer, 1990; Wilks, 2011). Intuitively, CSI measures the intersection-over-union of these regions. However, under uncertain rainfall evolution, deterministic forecasts trained with pixel-wise losses such as MAE or MSE can become conservative, weakening or smoothing intense precipitation (Ravuri et al., 2021; Yan et al., 2024). As illustrated in Fig. 1, a forecast may attain a lower regression error while missing more heavy-rain regions and obtaining a lower CSI. Directly optimizing CSI appears to be a natural remedy, but its hard-threshold binary decisions provide zero gradients almost everywhere.

Existing objectives address different aspects of this problem. Intensity-weighted regression increases the penalty on errors at heavy-rain pixels while retaining a pointwise comparison (Shi et al., 2017). Fourier Amplitude and Correlation Loss (FACL) (Yan et al., 2024) combines spectral amplitude and correlation criteria to improve forecast sharpness and overall structure. However, reducing intensity errors or improving overall structure does not necessarily improve the overlap between predicted

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.

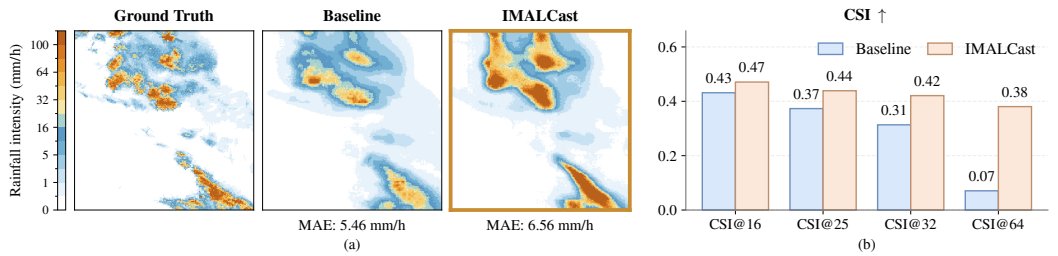

Figure 1: Intensity error and heavy-rain detection. (a) Observed and predicted precipitation fields with forecast MAE. (b) CSI across rainfall thresholds (mm/h). Scores use the displayed fields. The baseline has lower MAE but lower CSI than IMALCast, with a larger CSI gap at higher thresholds.

and observed heavy-rain regions. This motivates a differentiable training objective that explicitly promotes agreement in heavy-rain regions while retaining supervision of rainfall intensity.

Beyond improving the training objective, incorporating geostationary satellite observations can provide complementary cloud and convective context for radar-based heavy-rain forecasting (Sønderby et al., 2020; Espeholt et al., 2022; Andrychowicz et al., 2023). However, differences in spatial resolution, temporal sampling, and channel semantics complicate the integration of radar and satellite observations, giving rise to *observation misalignment* (Andrychowicz et al., 2023; Veillette et al., 2020). These differences motivate a multimodal design that preserves modality-specific information.

We therefore propose **IMALCast**, an **I**ntensity-aware **M**ultimodal **A**lignment **L**earning framework for precipitation nowcasting. To align training with threshold-based verification, we combine intensity-weighted regression with global and high-intensity structural losses, jointly supervising rainfall magnitudes, overall precipitation structure, and heavy-rain regions. To integrate complementary radar and satellite observations, IMALCast uses modality-specific patch embedding and dual-tower encoding to preserve their distinct information before cross-modal fusion.

In summary, our main contributions are as follows:

- We formulate an intensity-aware training objective motivated by a threshold-integrated analysis of verification. It jointly supervises rainfall intensity, full-field structure, and high-intensity exceedance patterns to improve heavy-rain detection.
- We develop IMALCast, a deterministic framework for high-resolution heavy-rainfall nowcasting using radar–satellite observations. It generates the complete forecast sequence in one forward pass, with structural losses introducing no additional inference cost.
- Experiments on the nationwide China Radar–Satellite Dataset (CRSD; $\sim$1 km, 6 min) and the public SEVIR benchmark demonstrate strong heavy-rain forecasting performance. The proposed objective further improves heavy-rain detection across forecasting backbones.

2 Related Work

2.1 Precipitation Nowcasting

Deterministic nowcasting models learn spatiotemporal dynamics through recurrent architectures (Shi et al., 2015; Wang et al., 2017; Shi et al., 2017; Le Guen & Thome, 2020), convolutional predictors (Ayzel et al., 2020; Gao et al., 2022a; Tan et al., 2025), and attention-based models (Gao et al., 2022b; Andrychowicz et al., 2023). Generative methods model multiple plausible futures to represent uncertainty and recover precipitation detail (Ravuri et al., 2021; Zhang et al., 2023; Gao et al., 2023; Yu et al., 2024). NowcastNet (Zhang et al., 2023) combines an evolution forecast with conditional generation, while CasCast (Gong et al., 2024) uses a deterministic forecast to guide latent diffusion. IMALCast improves deterministic forecasts through its training objective, which can be applied across forecasting backbones while retaining direct generation of the forecast sequence.

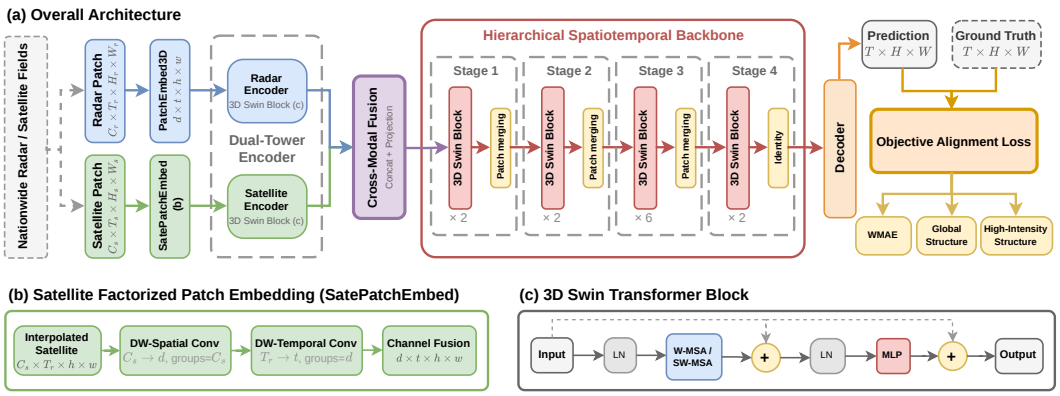

Figure 2: Overview of IMALCast. (a) Observation alignment embeds, encodes, and fuses radar–satellite patches from nationwide fields for hierarchical spatiotemporal prediction. Objective alignment combines weighted MAE with global and high-intensity structural losses during training. (b) Factorized satellite patch embedding. (c) 3D Swin Transformer block.

2.2 Training Objectives

MAE and MSE constrain pointwise intensity errors, and intensity-weighted variants assign greater importance to heavy-rain pixels (Shi et al., 2017). Structural and spectral objectives provide additional supervision on forecast patterns. FACL (Yan et al., 2024) regularizes Fourier amplitudes and uses a complementary correlation term to preserve structural agreement, improving forecast quality across architectures. Our objective shares the use of correlation-based structural criteria with FACL. It combines intensity-weighted regression with full-field structural alignment and explicit high-intensity structural supervision, giving heavy-rain regions their own alignment target.

2.3 Radar–Satellite Fusion

Radar observations describe precipitation intensity and local structures, while geostationary satellite imagery provides complementary cloud and convective context (Sadeghi et al., 2020). MetNet and its successors demonstrate the value of multi-source observations for forecasting (Sønderby et al., 2020; Espeholt et al., 2022; Andrychowicz et al., 2023). Differences in spatial resolution, temporal sampling, and channel semantics require careful integration of these observations (Veillette et al., 2020; Andrychowicz et al., 2023). IMALCast preserves modality-specific radar and satellite features through separate embedding and encoding before fusion.

3 Method

IMALCast combines intensity-aware objective alignment with modality-specific radar–satellite encoding and fusion for deterministic heavy-rain forecasting. We first define the forecasting setup, then describe the training objective and observation alignment architecture.

3.1 Problem Setup

We consider nationwide radar–satellite observations over mainland China in CRSD. Three hours of observations are used to forecast the next three hours at 6-minute intervals. We partition the observation fields into fixed-size patches for training and inference to enable high-resolution forecasting over this large domain (Sønderby et al., 2020; Andrychowicz et al., 2023).

Let $\mathbf{X}^r \in \mathbb{R}^{C_r \times T_r \times H_r \times W_r}$ and $\mathbf{X}^s \in \mathbb{R}^{C_s \times T_s \times H_s \times W_s}$ denote radar and satellite inputs, respectively. Here, C, T, H, and W denote channel, temporal, and spatial dimensions, with $C_r = 1$ for radar-derived rainfall intensity. The forecaster $\hat{\mathbf{Y}} = f(\mathbf{X}^r, \mathbf{X}^s)$ predicts a future sequence on the radar grid, with prediction and target $\hat{\mathbf{Y}}, \mathbf{Y} \in \mathbb{R}^{T_{\text{out}} \times H_r \times W_r}$. Figure 2 provides an overview of the framework.

3.2 INTENSITY-AWARE OBJECTIVE ALIGNMENT

Operational verification evaluates precipitation forecasts at multiple intensity thresholds, with particular importance placed on rare heavy-rain events. However, the non-differentiability of hard thresholding hinders direct optimization. We therefore examine the disagreement and overlap underlying threshold-based verification to guide joint supervision of rainfall intensities and heavy-rain regions. For a rainfall threshold η, define binary exceedance maps

$$B_i^\eta = \mathbb{I}(y_i > \eta), \qquad \hat{B}_i^\eta = \mathbb{I}(\hat{y}_i > \eta), \tag{1}$$

where $y_i, \hat{y}_i \geq 0$ denote the target and predicted precipitation values at spatiotemporal location i, and $\eta \geq 0$ is a threshold on the same scale. When $\sum_i B_i^\eta \hat{B}_i^\eta > 0$, the Critical Success Index (CSI) (Gao et al., 2022b; Schaefer, 1990) can be written as

$$\mathrm{CSI}_\eta = \left(1 + \frac{\sum_i (B_i^\eta - \hat{B}_i^\eta)^2}{\sum_i B_i^\eta \hat{B}_i^\eta}\right)^{-1}. \tag{2}$$

The ratio compares mismatched exceedance decisions with overlapping exceedances. For non-negative intensities, integrating exceedance disagreement over thresholds yields absolute intensity error. Following the threshold-integrated formulation in Appendix A, we construct an auxiliary objective that penalizes this error while rewarding exceedance overlap:

$$\mathcal{J} = \underbrace{\sum_i |y_i - \hat{y}_i|}_{\text{Intensity fidelity}} - \lambda \underbrace{\int_0^\infty \langle \mathbf{B}^\eta, \hat{\mathbf{B}}^\eta \rangle \, d\eta}_{\text{Exceedance overlap}}, \tag{3}$$

where $\mathbf{B}^\eta$ and $\hat{\mathbf{B}}^\eta$ are exceedance maps flattened over time and space, and $\lambda > 0$ is an auxiliary trade-off coefficient. Intensity fidelity constrains rainfall magnitudes, while exceedance overlap motivates structural supervision. We combine intensity-weighted regression with structural alignment of the full field and its high-intensity regions.

Intensity fidelity. We retain intensity fidelity through weighted MAE, assigning greater weight to higher-intensity rainfall to account for its rarity:

$$\mathcal{L}_{\mathrm{WMAE}} = \frac{1}{N} \sum_i w(y_i) \, |y_i - \hat{y}_i|, \qquad w(y_i) = 1 + \beta \sqrt{y_i}, \tag{4}$$

where N is the number of spatiotemporal locations and β controls the additional weight assigned to higher intensities. The square-root weighting increases sublinearly with intensity.

Global structural alignment. Motivated by the exceedance-overlap term, we use a global cosine similarity loss to compare predicted and observed precipitation distributions with reduced sensitivity to overall magnitude:

$$\mathcal{L}_{\mathrm{GS}} = 1 - \frac{\langle \mathbf{y}, \hat{\mathbf{y}} \rangle}{\|\mathbf{y}\|_2 \, \|\hat{\mathbf{y}}\|_2 + \varepsilon}, \tag{5}$$

where $\mathbf{y}$ and $\hat{\mathbf{y}}$ denote the target and prediction flattened over time and space, and ε is a small constant for numerical stability. This term emphasizes the relative distribution of precipitation across locations, while weighted MAE supplies the complementary magnitude constraint.

High-intensity structural alignment. Full-field similarity does not explicitly isolate regions exceeding a high rainfall threshold. We give these regions dedicated structural supervision by transforming both the target and prediction into soft exceedance maps:

$$T_\tau(x) = \sigma\big(\gamma(x - \tau)\big), \tag{6}$$

and define the high-intensity structural loss through cosine similarity between these maps:

$$\mathcal{L}_{\mathrm{HS}} = 1 - \frac{\langle T_\tau(\mathbf{y}), T_\tau(\hat{\mathbf{y}}) \rangle}{\|T_\tau(\mathbf{y})\|_2 \, \|T_\tau(\hat{\mathbf{y}})\|_2 + \varepsilon}, \tag{7}$$

where τ specifies the threshold on the same scale as y_i and $\hat{y}_i$, γ controls the transition sharpness, and σ denotes the sigmoid function. Applying the transform to both fields shifts the comparison from continuous rainfall magnitudes to soft exceedance patterns. Thus, high-intensity regions receive their own structural constraint in addition to intensity weighting. The global term continues to supervise the full field, while weighted MAE retains sensitivity to magnitude differences that the sigmoid suppresses well above the threshold.

In summary, the final training objective is

$$\mathcal{L}_{\text{full}} = \mathcal{L}_{\text{WMAE}} + \lambda_1 \mathcal{L}_{\text{GS}} + \lambda_2 \mathcal{L}_{\text{HS}}, \tag{8}$$

where λ_1 and λ_2 weight the structural terms. The objective depends only on predictions and targets, allowing transfer across forecasting backbones. Sections 4.3 and 4.4 evaluate its transferability and the complementary roles of the structural terms.

3.3 Observation Alignment Architecture

Figure 2 summarizes the observation alignment architecture described below.

Modality-specific patch embedding. Radar and satellite observations are cropped over corresponding geographic regions. Satellite inputs are resampled in space and time before embedding to accommodate their different sampling resolutions. For radar, a 3D convolutional patch embedding partitions the input into non-overlapping spatiotemporal patches and projects them into the latent space (Liu et al., 2022). For satellite, a factorized embedding applies separate spatial and temporal convolutions followed by channel mixing. These operations produce latent tensors of the same shape, $\mathbb{R}^{d\times t\times h\times w}$, while allowing each modality to use its own embedding parameters.

Dual-tower encoding and fusion. Separate towers of 3D Swin Transformer blocks (Liu et al., 2022; 2021) encode radar and satellite tokens, preserving features specific to rainfall and infrared observations. Their outputs are concatenated along the channel dimension and projected into a unified representation.

Hierarchical spatiotemporal modeling and decoder. The fused representation passes through a hierarchical spatiotemporal backbone with patch merging to enlarge the receptive field and capture multi-scale evolution. A lightweight decoder produces the target precipitation sequence on the radar grid. Preprocessing and architecture details appear in Appendices B.1 and B.2.

Training and inference. The forecaster is trained using equation 8. At inference, it produces the complete forecast sequence in one deterministic forward pass; the structural losses are used only during training. Training settings are provided in Appendix B.3.

4 Experiments

We evaluate high-intensity detection on CRSD and SEVIR, compare training objectives and backbone transfer, and analyze loss components, satellite contributions, detection–error trade-offs, and inference requirements.

4.1 Experimental Setup

CRSD. The China Radar–Satellite Dataset (CRSD) contains observations over mainland China during 2023–2025. Radar reflectivity, sampled at 0.01° resolution every 6 minutes, is converted to rainfall intensity (mm/h) using a Z–R relationship (Alfieri et al., 2010). Himawari-9 infrared Bands 08, 13, and 15 are sampled at 0.05° resolution every 10 minutes (Bessho et al., 2016). Three hours of history are used to forecast the next three hours: radar inputs and targets each contain 30 frames at 512×512, and satellite inputs contain 18 three-channel frames at 105×105. Training uses 401,565 samples from 2023–2024; validation and testing comprise 10,000 and 18,852 candidate samples from July–September 2025. Preprocessing is described in Appendix B.1.

SEVIR. SEVIR (Veillette et al., 2020) is a public storm-event dataset. We use its radar vertically integrated liquid (VIL) fields in the native-resolution forecasting setup of Earthformer (Gao et al.,

2022b): 13 input frames predict the next 12 frames at five-minute intervals on a 384×384 grid. Training, validation, and testing use the sample lists provided with CasCast (Gong et al., 2024). This radar-only benchmark evaluates forecasting in a different region and observation variable, complementing the radar–satellite setting on CRSD.

Baselines. CRSD baselines include seven deterministic models—ConvLSTM (Shi et al., 2015), TrajGRU (Shi et al., 2017), PhyDNet (Le Guen & Thome, 2020), RainNet (Ayzel et al., 2020), SimVP (Gao et al., 2022a), Earthformer (Gao et al., 2022b), and MetNet-3 (Andrychowicz et al., 2023)—and two generative models, NowcastNet (Zhang et al., 2023) and CasCast (Gong et al., 2024). The generative models' CRSD results score ensemble-mean forecasts with 8 and 10 members, respectively. SEVIR baseline results, including PredRNN (Wang et al., 2017), are taken from CasCast's native-resolution comparison. To assess objective design, we compare weighted mean absolute error (WMAE), Fourier Amplitude and Correlation Loss (FACL) (Yan et al., 2024), and the proposed objective with the IMALCast backbone and radar–satellite inputs fixed. Transfer experiments replace the training objective within three existing radar-only backbones.

Metrics. We report the Critical Success Index (CSI) and Heidke Skill Score (HSS) (Wilks, 2011) for threshold-based detection, central to precipitation warnings (Ma et al., 2022). CSI measures predicted–observed exceedance overlap; HSS measures skill relative to chance agreement. Rainfall thresholds are $\{0.1, 1, 5, 10, 16, 32, 64\}$ mm/h on CRSD, and VIL thresholds are $\{16, 74, 133, 160, 181, 219\}$ on SEVIR. CSI Avg and HSS Avg average these threshold sets. The frequency bias index (FBI) is the predicted-to-observed exceedance count ratio, ideally one. Probability of detection (POD) is the fraction of observed exceedances correctly predicted; false alarm ratio (FAR) is the fraction of predicted exceedances falsely identified. Definitions and calculation details appear in Appendix C.1.

We report CSI at three spatial scales. POOL1 denotes the native grid; POOL4/16 apply non-overlapping max pooling with kernel and stride 4/16, without padding, to the physical prediction and target fields before thresholding. HSS Avg uses the native grid. Scales and thresholds have different physical meanings across datasets, so comparisons are made within each dataset.

4.2 Main Results

CRSD. IMALCast achieves the highest native-grid CSI at both heavy-rain thresholds shown in Table 1. At 32 and 64 mm/h, IMALCast improves CSI over the strongest competing baseline, NowcastNet, from 0.147 to 0.156 and from 0.093 to 0.101, respectively. It also achieves the highest native-grid average CSI and HSS. With radar input alone, IMALCast achieves an average CSI of 0.289, compared with 0.251 for Earthformer, the strongest deterministic baseline on this metric, and 0.265 for NowcastNet; adding satellite observations raises it to 0.293.

The heavy-rain advantage becomes more pronounced at coarser evaluation scales. At 64 mm/h, relative CSI gains over NowcastNet increase from 8.6% on the native grid to 37.6% at POOL4 and 65.7% at POOL16, where CSI rises from 0.102 to 0.169. Max pooling evaluates whether intense precipitation occurs within a larger neighborhood, so the larger gains indicate improved regional heavy-rain detection. The radar-only variant also outperforms the reported baselines at both 32 and 64 mm/h across all three spatial scales, with satellite observations providing further gains in each setting. IMALCast's raw forecasts also outperform the distribution-corrected deterministic baselines in average CSI and HSS (Appendix E).

SEVIR. IMALCast achieves the highest reported native-grid scores on radar-only SEVIR, including average CSI of 0.4453 and HSS of 0.5759 (Table 2). Its advantage at the highest evaluated VIL threshold, 219, extends across all three scales: CSI is 0.1973 versus 0.1851 for CasCast at POOL1 and 0.3084 versus 0.2841 at POOL16. At this threshold, relative CSI improvements over CasCast range from 6.6% to 11.2% across the three spatial scales. Together with CRSD, these results demonstrate consistent improvements in high-intensity detection across the two benchmarks.

Inference requirements. IMALCast predicts the full sequence in one deterministic forward pass; its training-only structural losses add no parameters or operations at inference. For the reported CRSD point forecasts, NowcastNet averages 8 realizations, while CasCast averages 10, each requiring 20 DDIM denoising steps and decoding. IMALCast avoids these sampling stages (Appendix C.2).

Table 1: CRSD forecasting performance. R/S: radar/satellite; Avg: seven-threshold average; rainfall thresholds: mm/h. POOL1 is the native grid; POOL4/16 use max pooling by factors 4/16. HSS uses the native grid. Generative models use ensemble-mean forecasts. Bold marks the best in each column. Full native-grid results appear in Appendix D.1.

Model	HSS Avg↑	CSI Avg↑			CSI@32↑			CSI@64↑		
		POOL1	POOL4	POOL16	POOL1	POOL4	POOL16	POOL1	POOL4	POOL16
ConvLSTM	0.339	0.248	0.257	0.269	0.090	0.074	0.069	0.042	0.036	0.034
TrajGRU	0.291	0.216	0.225	0.232	0.046	0.041	0.040	0.018	0.020	0.022
PhyDNet	0.315	0.226	0.228	0.232	0.074	0.060	0.058	0.038	0.031	0.030
RainNet	0.320	0.232	0.243	0.264	0.088	0.077	0.078	0.050	0.047	0.048
SimVP	0.344	0.250	0.259	0.272	0.104	0.088	0.082	0.045	0.038	0.035
Earthformer	0.343	0.251	0.251	0.249	0.089	0.069	0.059	0.046	0.036	0.030
MetNet-3	0.339	0.247	0.250	0.257	0.087	0.072	0.068	0.047	0.042	0.041
NowcastNet	0.369	0.265	0.319	**0.363**	0.147	0.167	0.182	0.093	0.101	0.102
CasCast	0.343	0.247	0.278	0.318	0.122	0.127	0.145	0.064	0.063	0.065
IMALCast (R)	0.399	0.289	0.320	0.329	0.151	0.180	0.194	0.096	0.129	0.155
IMALCast (R+S)	**0.404**	**0.293**	**0.325**	0.335	**0.156**	**0.190**	**0.205**	**0.101**	**0.139**	**0.169**

Table 2: Radar-only SEVIR performance. CSI Avg averages the six VIL thresholds. Thresholds use VIL units. POOL1 is the native grid; POOL4/16 use max pooling by factors 4/16. HSS uses the native grid. Bold marks the best in each column.

Model	HSS Avg↑	CSI Avg↑			CSI@181↑			CSI@219↑		
		POOL1	POOL4	POOL16	POOL1	POOL4	POOL16	POOL1	POOL4	POOL16
ConvLSTM	0.5232	0.4102	0.4163	0.4475	0.2453	0.2525	0.2977	0.1322	0.1380	0.1734
PredRNN	0.5192	0.4045	0.4161	0.4623	0.2416	0.2567	0.3214	0.1331	0.1447	0.1909
PhyDNet	0.5311	0.4198	0.4226	0.4410	0.2526	0.2532	0.2782	0.1362	0.1359	0.1526
SimVP	0.5280	0.4153	0.4226	0.4530	0.2532	0.2604	0.3000	0.1338	0.1394	0.1685
Earthformer	0.5411	0.4310	0.4319	0.4351	0.2622	0.2542	0.2562	0.1448	0.1409	0.1481
NowcastNet	0.5365	0.4152	0.4452	0.5024	0.2495	0.2935	0.3725	0.1422	0.1874	0.2700
CasCast	0.5602	0.4401	0.4640	**0.5225**	0.2879	**0.3179**	**0.3900**	0.1851	0.2127	0.2841
IMALCast	**0.5759**	**0.4453**	**0.4675**	0.4952	**0.2968**	0.3158	0.3415	**0.1973**	**0.2366**	**0.3084**

4.3 Objective Comparison and Backbone Transfer

Comparison of training objectives. Within the same radar–satellite IMALCast backbone, the proposed objective improves heavy-rain detection relative to FACL (Table 3, upper block). The relative CSI gain is 2.1% at 10 mm/h and reaches 20.9% and 18.8% at 32 and 64 mm/h, respectively. Average CSI also improves from 0.286 to 0.293. The larger high-threshold gains support explicit supervision of high-intensity structure alongside the full field.

Transfer across backbones. The improvement extends to ConvLSTM, SimVP, and Earthformer, representing recurrent, convolutional, and Transformer architectures (Table 3, lower block). All three achieve higher average CSI. CSI@64 rises from 0.042–0.046 to 0.087–0.090, a relative gain of 89.1–114.3%. These consistent gains across the tested architectures show that the benefit of the objective extends beyond the IMALCast backbone.

4.4 Ablation Study

Objective components. The structural terms improve different intensity ranges (Table 4, upper block). Adding global alignment to WMAE raises average CSI from 0.255 to 0.276 and improves every displayed threshold. Adding only high-intensity alignment raises CSI@32 from 0.092 to 0.148 but reduces CSI at 1, 5, and 10 mm/h, leaving average CSI nearly unchanged at 0.254. Combining both achieves the highest average and heavy-rain CSI, supporting their complementary roles. Across training seeds, the full objective retains higher mean average CSI and CSI@32/64 than either single-term variant; these results and parameter sensitivity appear in Appendix D.2.

Satellite observations and fusion. Early fusion resamples satellite observations to the radar grid and temporal sampling, then concatenates both modalities along the channel dimension before shared

Table 3: Native-grid CSI on CRSD. Upper: objectives with the IMALCast backbone fixed. Lower: backbone transfer (original → proposed objective). Parentheses give relative gains from displayed scores over FACL (upper) or the original objective (lower). Bold marks the best scores.

Objective	CSI Avg↑	CSI@10↑	CSI@16↑	CSI@32↑	CSI@64↑
WMAE	0.255	0.196	0.141	0.092	0.049
FACL	0.286	0.238	0.181	0.129	0.085
Proposed	**0.293** (+2.4%)	**0.243** (+2.1%)	**0.198** (+9.4%)	**0.156** (+20.9%)	**0.101** (+18.8%)
Backbone	**CSI Avg↑**		**CSI@32↑**		**CSI@64↑**
ConvLSTM	0.248→**0.270** (+8.9%)		0.090→**0.149** (+65.6%)		0.042→**0.090** (+114.3%)
SimVP	0.250→**0.266** (+6.4%)		0.104→**0.147** (+41.3%)		0.045→**0.089** (+97.8%)
Earthformer	0.251→**0.271** (+8.0%)		0.089→**0.142** (+59.6%)		0.046→**0.087** (+89.1%)

Table 4: Ablation results for the objective alignment and observation alignment components. Upper: objective components with the radar–satellite backbone fixed. Lower: observation designs under the full objective. Parentheses give relative gains from displayed scores over WMAE (upper) or radar only (lower). Bold marks the best value within each block.

Variant	CSI Avg↑	CSI@1↑	CSI@5↑	CSI@10↑	CSI@16↑	CSI@32↑	CSI@64↑
Objective alignment							
WMAE	0.255	0.471	0.290	0.196	0.141	0.092	0.049
+ Global structural	0.276	**0.502**	0.318	0.224	0.165	0.115	0.078
+ High-intensity structural	0.254	0.449	0.258	0.193	0.166	0.148	0.083
Full objective	**0.293** (+14.9%)	0.496 (+5.3%)	**0.319** (+10.0%)	**0.243** (+24.0%)	**0.198** (+40.4%)	**0.156** (+69.6%)	**0.101** (+106.1%)
Observation alignment							
Radar only	0.289	0.487	0.313	0.238	0.193	0.151	0.096
Early fusion	0.267	0.475	0.297	0.221	0.175	0.137	0.090
Dual-tower fusion	**0.293** (+1.4%)	**0.496** (+1.8%)	**0.319** (+1.9%)	**0.243** (+2.1%)	**0.198** (+2.6%)	**0.156** (+3.3%)	**0.101** (+5.2%)

patch embedding and encoding. With the full objective fixed, dual-tower fusion outperforms early fusion and radar-only prediction in average CSI (Table 4, lower block). Early fusion underperforms radar alone at every displayed threshold, showing that satellite benefits depend on the integration design. Dual-tower fusion improves all displayed thresholds, with a larger relative gain of 5.2% at 64 mm/h. Appendix D.3 associates larger gains with growing precipitation systems.

4.5 Forecast Analysis

Lead-time performance. Forecast skill decreases with lead time for all models shown in Figure 3. IMALCast achieves higher CSI and HSS than the baselines shown at nearly all displayed lead times, maintaining a clear advantage even at the three-hour forecast horizon. Numerical comparisons with all seven deterministic baselines appear in Appendix D.1.

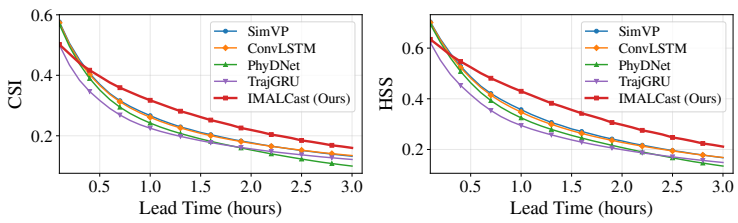

Figure 3: Native-grid CSI and HSS over the three-hour CRSD forecast horizon, averaged across seven rainfall thresholds.

Heavy-rain detection and frequency bias. We further examine how the heavy-rain detection gains relate to predicted exceedance frequency. Table 5 shows that IMALCast brings this frequency closer to observations at both thresholds than the compared baselines. At 64 mm/h, its POD is approximately three times that of the deterministic baselines, while its FBI remains below one, showing that improved detection is achieved without overpredicting the overall exceedance frequency. Although the predicted frequency is closer to observations, the higher POD comes with a higher FAR.

Table 5: Heavy-rain detection and frequency bias on CRSD, computed from aggregated native-grid contingency counts. Thresholds are in mm/h. Bold marks the best value in each row; FBI is best at 1.0. Calculation details appear in Appendix C.1.

Metric	ConvLSTM	PhyDNet	Earthformer	MetNet-3	NowcastNet	**IMALCast**
FBI@32	0.193	0.169	0.206	0.192	0.935	**1.035**
FBI@64	0.106	0.127	0.136	0.137	0.525	**0.831**
POD@64↑	0.059	0.057	0.066	0.067	0.159	**0.193**
FAR@64↓	**0.447**	0.554	0.516	0.512	0.697	0.767

Qualitative comparison. Figure 4 shows a CRSD event. At 1.5 and 2.0 hours, IMALCast retains a stronger upper-left high-intensity core than the baselines and intense rainfall along the lower-right band. The predicted cores remain smoother and more contiguous than observations, with shape and extent discrepancies indicating unresolved fine-scale structure. Appendix F provides additional cases.

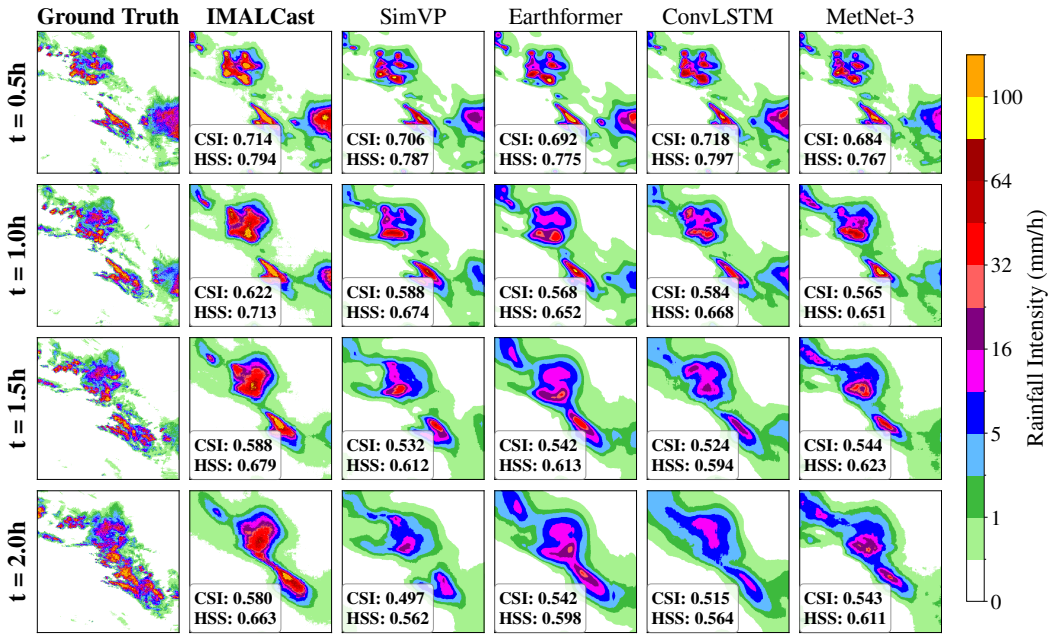

Figure 4: CRSD observations and forecasts (columns) at different lead times (rows). Colors show rainfall intensity (mm/h); annotations report framewise CSI and HSS at 1 mm/h.

5 Conclusion

We presented IMALCast, a deterministic nowcasting framework combining intensity-aware supervision with modality-specific radar–satellite encoding and fusion. Experiments on the nationwide China Radar–Satellite Dataset (∼1 km, 6 min) and the public SEVIR benchmark demonstrate strong heavy-rain forecasting performance. The results support complementary full-field and high-intensity supervision, with gains across forecasting backbones and further improvements from satellite observations at high thresholds. These findings highlight the value of combining intensity-aware supervision with complementary observations for heavy-rain detection.

AI USE STATEMENT

In this work, we used generative AI tools to assist with retrieving relevant literature and improving the language, clarity, and readability of the manuscript. We have reviewed all AI-assisted work. Specifically, we checked the suggested references against their original sources and reviewed the writing revisions to ensure that they accurately reflect our methods, experimental results, and intended claims. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The training objective is detailed in Section 3.2. Data preprocessing, architecture configurations, and training settings are provided in Appendices B.1, B.2, and B.3, respectively. Evaluation metrics are defined in Appendix C.1. The supplementary anonymous code includes the model implementation, loss functions, and a training example using synthetic inputs. The original CRSD observations cannot be redistributed due to data-use restrictions.

REFERENCES

L. Alfieri, P. Claps, and F. Laio. Time-dependent Z-R relationships for estimating rainfall fields from radar measurements. *Natural Hazards and Earth System Sciences*, 10:149–158, 2010.

Marcin Andrychowicz, Lasse Espeholt, Di Li, Samier Merchant, Alexander Merose, Fred Zyda, Shreya Agrawal, and Nal Kalchbrenner. Deep learning for day forecasts from sparse observations. *arXiv preprint arXiv:2306.06079*, 2023.

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.

Kotaro Bessho, Kenji Date, Masahiro Hayashi, Akio Ikeda, Takahito Imai, H. Inoue, Yukihiro Kumagai, Takuya Miyakawa, Hidehiko Murata, Tomoo Ohno, Arata Okuyama, Ryo Oyama, Yukio Sasaki, Yoshio Shimazu, Kazuki Shimoji, Yasuhiko Sumida, Masuo Suzuki, Hidetaka Taniguchi, Hiroaki Tsuchiyama, Daisaku Uesawa, Hironobu Yokota, and Ryo Yoshida. An introduction to Himawari-8/9–Japan's new-generation geostationary meteorological satellites. *Journal of the Meteorological Society of Japan*, 94:151–183, 2016.

E. Ebert, L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Göber, S. Joslyn, B. Brown, T. Fowler, and A. Watkins. Progress and challenges in forecast verification. *Meteorological Applications*, 20(2):130–139, 2013.

Lasse Espeholt, Shreya Agrawal, Casper Sønderby, Manoj Kumar, Jonathan Heek, Carla Bromberg, Cenk Gazen, Rob Carver, Marcin Andrychowicz, Jason Hickey, Aaron Bell, and Nal Kalchbrenner. Deep learning for twelve hour precipitation forecasts. *Nature Communications*, 13:5145, 2022.

Christopher A. T. Ferro and David B. Stephenson. Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, 26(5): 699–713, 2011.

Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z. Li. SimVP: Simpler yet better video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022a.

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

Zhihan Gao, Xingjian Shi, Boran Han, Hao Wang, Xiaoyong Jin, Danielle Maddix, Yi Zhu, Mu Li, and Yuyang Wang. PreDiff: Precipitation nowcasting with latent diffusion models. In *Advances in Neural Information Processing Systems*, 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, pp. 15809–15822, 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*, 2020.

Jingyuan Liu, Jianlin Su, Xingcheng Yao, Zhejun Jiang, Guokun Lai, Yulun Du, Yidao Qin, Weixin Xu, Enzhe Lu, Junjie Yan, Yanru Chen, Huabin Zheng, Yibo Liu, Shaowei Liu, Bohong Yin, Weiran He, Han Zhu, Yuzhi Wang, Jianzhou Wang, Mengnan Dong, Zheng Zhang, Yongsheng Kang, Hao Zhang, Xinran Xu, Yutao Zhang, Yuxin Wu, Xinyu Zhou, and Zhilin Yang. Muon is scalable for LLM training. *arXiv preprint arXiv:2502.16982*, 2025.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. Video Swin transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Zhifeng Ma, Hao Zhang, and Jie Liu. PrecipLSTM: A meteorological spatiotemporal LSTM for precipitation nowcasting. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–8, 2022.

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*, 597:672–677, 2021.

Mojtaba Sadeghi, Phu Nguyen, Kuolin Hsu, and Soroosh Sorooshian. Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information. *Environmental Modelling & Software*, 134:104856, 2020.

Joseph T Schaefer. The critical success index as an indicator of warning skill. *Weather and Forecasting*, 5(4):570–575, 1990.

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*, 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. In *Advances in Neural Information Processing Systems*, 2017.

Patrick S. Skinner, Dustan M. Wheatley, Kent H. Knopfmeier, Anthony E. Reinhart, Jessica J. Choate, Thomas A. Jones, Gerald J. Creager, David C. Dowell, Curtis R. Alexander, Therese T. Ladwig, et al. Object-based verification of a prototype warn-on-forecast system. *Weather and Forecasting*, 33(5):1225–1250, 2018.

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.

Cheng Tan, Zhangyang Gao, Siyuan Li, and Stan Z. Li. SimVPv2: Towards simple yet powerful spatiotemporal predictive learning. *IEEE Transactions on Multimedia*, 27:5170–5184, 2025.

Mark S. Veillette, Siddharth Samsi, and Christopher J. Mattioli. SEVIR: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. In *Advances in Neural Information Processing Systems*, 2020.

Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S. Yu. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In *Advances in Neural Information Processing Systems*, 2017.

Daniel S. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 100. Academic Press, 2011.

Chiu-Wai Yan, Shi Quan Foo, Van Hoan Trinh, Dit-Yan Yeung, Ka-Hing Wong, and Wai-Kin Wong. Fourier amplitude and correlation loss: Beyond using L2 loss for skillful precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

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*, 2024.

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: 526–532, 2023.

A THRESHOLD-BASED MOTIVATION FOR THE TRAINING OBJECTIVE

This section develops the mathematical motivation for the objective alignment loss in Section 3.2. We relate threshold-integrated disagreement to intensity error and express exceedance overlap, then explain how these relationships inform the intensity and structural supervision used in training.

For a rainfall threshold η, define binary exceedance maps

$$B_i^\eta = \mathbb{I}(y_i > \eta), \qquad \hat{B}_i^\eta = \mathbb{I}(\hat{y}_i > \eta), \tag{9}$$

where $y_i, \hat{y}_i \geq 0$ denote the target and predicted precipitation values at spatiotemporal location i, and $\eta \geq 0$ is a threshold on the same scale, following Section 3.2. The corresponding true positives, false negatives, and false positives are

$$\mathrm{TP}_\eta = \sum_i B_i^\eta \hat{B}_i^\eta, \tag{10}$$

$$\mathrm{FN}_\eta = \sum_i B_i^\eta (1 - \hat{B}_i^\eta), \tag{11}$$

$$\mathrm{FP}_\eta = \sum_i (1 - B_i^\eta) \hat{B}_i^\eta. \tag{12}$$

A representative threshold-based verification metric in precipitation forecasting is the Critical Success Index (CSI). When $\mathrm{TP}_\eta + \mathrm{FP}_\eta + \mathrm{FN}_\eta > 0$, the CSI at threshold η is

$$\mathrm{CSI}_\eta = \frac{\mathrm{TP}_\eta}{\mathrm{TP}_\eta + \mathrm{FP}_\eta + \mathrm{FN}_\eta}. \tag{13}$$

For $\mathrm{TP}_\eta > 0$, rearranging gives

$$\frac{1}{\mathrm{CSI}_\eta} - 1 = \frac{\mathrm{FP}_\eta + \mathrm{FN}_\eta}{\mathrm{TP}_\eta}. \tag{14}$$

Because $B_i^\eta, \hat{B}_i^\eta \in \{0, 1\}$,

$$(B_i^\eta - \hat{B}_i^\eta)^2 = \begin{cases} 1, & B_i^\eta \neq \hat{B}_i^\eta, \\ 0, & B_i^\eta = \hat{B}_i^\eta, \end{cases} \tag{15}$$

which exactly counts mismatched exceedance decisions. Therefore,

$$\mathrm{FP}_\eta + \mathrm{FN}_\eta = \sum_i (B_i^\eta - \hat{B}_i^\eta)^2. \tag{16}$$

Substituting equation 16 into equation 14 yields

$$\mathrm{CSI}_\eta = \frac{1}{\frac{\sum_i \left(B_i^\eta - \hat{B}_i^\eta\right)^2}{\sum_i B_i^\eta \hat{B}_i^\eta} + 1}. \tag{17}$$

Equation 17 shows that threshold-based skill depends on two coupled quantities: the disagreement between predicted and observed exceedance maps, and the overlap between the corresponding precipitation regions. These two factors motivate an auxiliary objective that penalizes disagreement and rewards overlap at a fixed threshold η:

$$\mathcal{J}_\eta = \sum_i (B_i^\eta - \hat{B}_i^\eta)^2 - \lambda \sum_i B_i^\eta \hat{B}_i^\eta, \tag{18}$$

where $\lambda > 0$ controls the trade-off between disagreement reduction and overlap encouragement. This weighted difference motivates a CSI-informed training objective. Its auxiliary coefficient λ is separate from the training weights λ_1 and λ_2 in equation 8.

To examine these two factors across intensity levels, we consider the threshold-integrated auxiliary objective

$$\mathcal{J} = \int_0^\infty \sum_i (B_i^\eta - \hat{B}_i^\eta)^2 \, d\eta - \lambda \int_0^\infty \sum_i B_i^\eta \hat{B}_i^\eta \, d\eta. \tag{19}$$

This continuous-threshold formulation connects binary exceedance decisions to intensity-space supervision. Below, we derive the integrated disagreement and overlap terms and relate them to the training losses.

From thresholded disagreement to intensity deviation. We first consider the disagreement term. Since the exceedance maps are binary, $(B_i^\eta - \hat{B}_i^\eta)^2 = |B_i^\eta - \hat{B}_i^\eta|$. For two non-negative rainfall intensities $a, b \geq 0$, define

$$D(a,b) = \int_0^\infty |\mathbb{I}(a > \eta) - \mathbb{I}(b > \eta)| \, d\eta. \tag{20}$$

Without loss of generality, assume $a \geq b$. Then

$$\mathbb{I}(a > \eta) - \mathbb{I}(b > \eta) = \begin{cases} 0, & 0 \leq \eta < b, \\ 1, & b \leq \eta < a, \\ 0, & \eta \geq a, \end{cases} \tag{21}$$

so

$$D(a,b) = \int_b^a 1 \, d\eta = a - b = |a - b|. \tag{22}$$

Hence,

$$\int_0^\infty |\mathbb{I}(a > \eta) - \mathbb{I}(b > \eta)| \, d\eta = |a - b|. \tag{23}$$

Applying this identity to $a = y_i$ and $b = \hat{y}_i$ and summing over locations yields the intensity-fidelity term in equation 3. To give higher-intensity rainfall greater emphasis, we introduce the target-dependent weight $w(y_i) = 1 + \beta\sqrt{y_i}$ and average the weighted errors over the N spatiotemporal locations, yielding $\mathcal{L}_{\mathrm{WMAE}}$ in Section 3.2. The weighting is a design choice motivated by the rarity of heavy rainfall.

From thresholded co-exceedance to structural alignment. We next consider the overlap term in equation 19. At each threshold, the sum of co-exceedance products is the inner product of the corresponding binary maps. Integrating over thresholds gives

$$\int_0^\infty \sum_i B_i^\eta \hat{B}_i^\eta \, d\eta = \int_0^\infty \langle \mathbf{B}^\eta, \hat{\mathbf{B}}^\eta \rangle \, d\eta, \tag{24}$$

where $\mathbf{B}^\eta$ and $\hat{\mathbf{B}}^\eta$ denote the exceedance maps flattened over time and space. Together with equation 23, this expression yields the auxiliary objective in equation 3. The inner-product form motivates comparing the spatial and temporal distribution of predicted and observed precipitation through similarity measures. The cosine losses in Section 3.2 are similarity-based choices informed by this overlap structure.

Specifically, $\mathcal{L}_{\mathrm{GS}}$ compares the full fields $\mathbf{y}$ and $\hat{\mathbf{y}}$, flattened over time and space. It measures structural agreement with reduced sensitivity to overall magnitude, while $\mathcal{L}_{\mathrm{WMAE}}$ retains intensity supervision. To emphasize heavy-rain regions, $\mathcal{L}_{\mathrm{HS}}$ compares the soft exceedance maps $T_\tau(\mathbf{y})$ and $T_\tau(\hat{\mathbf{y}})$ defined in equation 6, where τ is a threshold on the same scale as y_i and $\hat{y}_i$, and γ controls transition sharpness. The two structural terms supervise the full field and its high-intensity regions, respectively.

Taken together, the disagreement identity and overlap formulation motivate complementary supervision of rainfall magnitudes and precipitation structure. The final objective $\mathcal{L}_{\mathrm{full}}$ in equation 8 implements this motivation through weighted intensity regression, global structural agreement, and high-intensity structural agreement, with λ_1 and λ_2 controlling the contributions of the two structural losses.

B Datasets and Implementation Details

B.1 CRSD Dataset Details

B.1.1 Data Sources

CRSD contains multisource observations collected during 2023–2025 over mainland China, including radar-derived precipitation and geostationary satellite observations.

The radar data are obtained from nationwide weather-radar composite reflectivity products and converted to rainfall intensity (mm/h) via the Z–R relationship (Alfieri et al., 2010). The radar data have 0.01° spatial resolution (approximately 1 km) and 6-minute temporal resolution.

The satellite data are acquired from the Himawari-9 geostationary meteorological satellite (Bessho et al., 2016), with 0.05° spatial resolution and 10-minute temporal resolution. We use three infrared channels, namely Bands 08, 13, and 15.

B.1.2 Radar Sequence Construction

We first assemble consecutive radar frames into full-frame sequences of length 30, corresponding to 3 hours of observations at 6-minute intervals. When the missing-frame ratio of a candidate sequence is below 10%, missing radar frames are filled by nearest-neighbor interpolation along the temporal dimension; otherwise, the sequence is discarded.

All radar sequences are then mapped to a unified geographic domain. To obtain a standardized full-frame representation, the valid radar domain is padded to a fixed canvas of size 4200×6200, where invalid regions are filled with -1.

B.1.3 Patch Generation and Radar–Satellite Alignment

To enable scalable training and inference, we construct local patch-based samples from the full-frame radar sequences. A sliding window of size 512×512 is used to crop local radar patches. During training, the sliding-window stride is 32 pixels, and an additional random offset in the range $[0, 32]$ pixels is applied to increase sample diversity. During validation and testing, the stride is set to 64 pixels and no random offset is used. Each radar sample is therefore represented as a tensor of shape $30 \times 512 \times 512$.

Satellite samples are aligned to radar samples in both space and time. For each radar crop, we use its starting spatial location as the reference and extract a satellite crop that fully covers the radar patch. Because the satellite resolution (0.05°) is lower than the radar resolution (0.01°), each radar crop of size 512×512 corresponds to a satellite crop of size 105×105, which covers the entire radar region with a small spatial margin. The temporal dimension of the satellite sample contains 18 frames over 3 hours at 10-minute intervals. Using the selected three infrared channels, each satellite sample is represented as a tensor of shape $18 \times 3 \times 105 \times 105$.

The final multimodal sample therefore contains a radar tensor of shape $30 \times 512 \times 512$ and a satellite tensor of shape $18 \times 3 \times 105 \times 105$, together with the future radar-derived precipitation target sequence on the radar grid.

B.1.4 Normalization and Importance Sampling

Radar-derived rainfall intensity exhibits a highly long-tailed distribution. Following prior precipitation forecasting work (Andrychowicz et al., 2023), we apply the tanh-log transformation

$$x_{\mathrm{norm}} = \tanh\left(\frac{\log(1+x)}{4}\right), \tag{25}$$

where x denotes rainfall intensity in mm/h and x_{norm} is the normalized value. The corresponding inverse transform is

$$x = \exp(4\ \mathrm{arctanh}(x_{\mathrm{norm}})) - 1. \tag{26}$$

Under this normalization, the operating threshold $\tau = 0.75$ in the high-intensity structural loss corresponds to approximately 48 mm/h in the original rainfall scale. Satellite observations are normalized independently for each channel using min-max scaling.

To improve training efficiency, we further adopt an importance sampling strategy that prioritizes precipitation-bearing regions (Ravuri et al., 2021). For each candidate radar crop, an acceptance probability is computed according to the rainfall intensity statistics within the crop. The candidate is then accepted or rejected by rejection sampling based on this probability, which increases the proportion of informative precipitation samples during training.

B.2 Architecture Details

Table 6 provides the full configuration of IMALCast. The radar input ($1\times30\times512\times512$) is tokenized by a 3D convolution with kernel and stride $(3, 4, 4)$, while the satellite input ($3 \times 18 \times 105 \times 105$) is first interpolated to match the radar extent and then processed by a factorized patch embedding (depthwise spatial/temporal convolutions followed by $1\times1\times1$ channel mixing) with stride $(3, 1, 1)$. Both branches produce a $10 \times 128 \times 128$ token grid with 96 channels. In the notation of Fig. 2, this corresponds to latent feature tensors of shape $d \times t \times h \times w$, with $d = 96$, $t = 10$, and $h = w = 128$ in our implementation. After independent two-stage encoding without spatial downsampling, the two modalities are concatenated and linearly projected back to 96 channels. The fused representation then passes through a four-stage hierarchical mixer with patch merging after stages 1–3, progressively reducing spatial resolution while increasing channel dimension. A linear head with pixel shuffle recovers the target $30 \times 512 \times 512$ output. All stages use 3D shifted-window attention with window size $(2, 7, 7)$ and MLP ratio 4.

Table 6: IMALCast architecture configuration on CRSD. "PM" = patch merging ($2\times$ spatial downsampling, $2\times$ channel expansion). Channels lists input-to-output channel dimensions, with a single value indicating no change; decoder dimensions refer to the linear projection before rearrangement. Output Size denotes the temporal and spatial dimensions after each operation.

Component	Operation	Depth	Heads	Channels	Output Size
Patch Embedding					
Radar	Conv3D $(3, 4, 4)$	–	–	$1 \to 96$	$10 \times 128 \times 128$
Satellite	Factorized Conv, stride $(3, 1, 1)$	–	–	$3 \to 96$	$10 \times 128 \times 128$
Dual-Tower Encoder ($\times2$ independent towers, no downsampling)					
Stage 1	Swin-3D blocks	2	3	96	$10 \times 128 \times 128$
Stage 2	Swin-3D blocks	2	6	96	$10 \times 128 \times 128$
Fusion					
Combine	Concat + Linear	–	–	$192 \to 96$	$10 \times 128 \times 128$
Hierarchical Spatiotemporal Backbone					
Stage 1	Swin-3D blocks + PM	2	3	$96 \to 192$	$10 \times 64 \times 64$
Stage 2	Swin-3D blocks + PM	2	6	$192 \to 384$	$10 \times 32 \times 32$
Stage 3	Swin-3D blocks + PM	6	12	$384 \to 768$	$10 \times 16 \times 16$
Stage 4	Swin-3D blocks	2	24	768	$10 \times 16 \times 16$
Decoder					
Output	Linear + Rearrange	–	–	$768 \to 3072$	$30 \times 512 \times 512$

For SEVIR, the radar-only configuration uses embedding dimension 96, encoder depths $(2, 2)$, mixer depths $(2, 4, 2)$, and attention windows $(2, 4, 4)$, adapting the hierarchy to its geometry.

B.3 Training Settings

We use Muon (Liu et al., 2025) with a learning rate of 10^{-3} on both datasets. On CRSD, IMALCast is trained on NVIDIA A800 GPUs with 80 GB memory, using a batch size of 4 with bf16-mixed precision. Training takes approximately 4 hours per epoch. The objective parameters are $\beta = 19$, $\lambda_1 = 2$, $\lambda_2 = 1$, and $\gamma = 50$ on both datasets. The high-intensity threshold is $\tau = 0.75$ on CRSD and $\tau = 0.85$ on SEVIR.

Checkpoints are ranked by validation-average CSI. Early stopping monitors validation-average CSI for IMALCast and the baselines. The transfer experiments apply the proposed loss to radar-only ConvLSTM, SimVP, and Earthformer backbones.

IMALCast follows the same single-stage supervised training paradigm as the deterministic baselines, with intensity and structural supervision integrated into a combined objective. The generative baselines involve additional training stages: NowcastNet (Zhang et al., 2023) first trains an evolution network, followed by conditional adversarial training with a temporal discriminator; CasCast (Gong et al., 2024) separately trains a deterministic forecasting model and an autoencoder before training its conditional latent diffusion model.

C Evaluation Metrics and Protocols

C.1 Metric Definitions

At threshold η, predictions and targets are binarized to obtain true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). The corresponding scores are

$$\mathrm{CSI}_\eta = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{27}$$

$$\mathrm{HSS}_\eta = \frac{2(\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN})}{(\mathrm{TP} + \mathrm{FN})(\mathrm{FN} + \mathrm{TN}) + (\mathrm{TP} + \mathrm{FP})(\mathrm{FP} + \mathrm{TN})}. \tag{28}$$

MAE and RMSE measure mean absolute and root mean squared intensity errors in mm/h, respectively. Additional detection metrics are:

$$\mathrm{POD}_\eta = \frac{\mathrm{TP}_\eta}{\mathrm{TP}_\eta + \mathrm{FN}_\eta}, \quad \mathrm{FAR}_\eta = \frac{\mathrm{FP}_\eta}{\mathrm{TP}_\eta + \mathrm{FP}_\eta}, \tag{29}$$

where subscripts indicate the intensity threshold used to form the contingency counts.

To distinguish missed heavy rain from excessive predicted exceedance area, we compute the frequency bias index

$$\mathrm{FBI}_\eta = \frac{\mathrm{TP}_\eta + \mathrm{FP}_\eta}{\mathrm{TP}_\eta + \mathrm{FN}_\eta}. \tag{30}$$

Values below one indicate underforecasting of exceedance frequency, while values above one indicate overforecasting. FBI measures frequency rather than spatial localization and should be interpreted together with CSI, POD, and FAR.

C.2 Baseline Evaluation and Inference Procedures

Table 7 summarizes how the reported CRSD point forecasts are constructed. IMALCast maps the input observations to all 30 forecast frames in one forward pass. The global and high-intensity structural losses affect training only; for a fixed backbone, they introduce no additional learned parameters or operations for generating a forecast.

NowcastNet (Zhang et al., 2023) first computes an evolution forecast, then generates eight realizations conditioned on the observations and evolution forecast. Their mean is used for evaluation. CasCast (Gong et al., 2024) conditions latent diffusion on a coarse forecast. Each of its ten members undergoes 20 DDIM denoising steps with classifier-free guidance, followed by decoding; the decoded forecasts are then averaged. Coarse-forecast conditioning is shared across members, but denoising and decoding are repeated. For CRSD, CasCast is adapted to the radar-only 30-input/30-output task. IMALCast's deterministic formulation avoids ensemble generation and iterative denoising when only one forecast field is needed.

Table 7: Inference procedures for the reported CRSD point forecasts. Members are stochastic realizations averaged into one forecast field; DDIM steps are counted per member.

Model	Forecast construction	Members	DDIM steps
IMALCast	One forward pass for all target frames	None	None
NowcastNet	Evolution forecast, conditional generation, and averaging	8	None
CasCast	Coarse forecast, guided latent denoising, decoding, and averaging	10	20

D Additional Experimental Results

We first provide detailed forecasting results on CRSD, then examine the robustness of the training objective and the contribution of satellite observations.

D.1 Detailed Forecasting Results on CRSD

Full Native-Grid Results on CRSD. Table 8 reports the full native-grid CSI and HSS results to complement the multi-scale comparison in Table 1. Radar–satellite IMALCast achieves the highest CSI and HSS at thresholds from 1 to 64 mm/h, supporting its gains across the intensity range, including heavy rain.

Table 8: Per-threshold CSI and HSS on CRSD. R: radar; S: satellite. Thresholds are in mm/h; Avg includes all seven thresholds. Generative rows score ensemble-mean forecasts. Bold indicates the best result.

	CSI ↑								HSS ↑							
Model	0.1	1	5	10	16	32	64	Avg	0.1	1	5	10	16	32	64	Avg
ConvLSTM	0.521	0.456	0.297	0.195	0.134	0.090	0.042	0.248	0.579	0.575	0.437	0.313	0.229	0.161	0.079	0.339
TrajGRU	0.509	0.446	0.257	0.143	0.088	0.046	0.018	0.216	0.566	0.565	0.388	0.239	0.156	0.086	0.034	0.291
PhyDNet	0.493	0.423	0.252	0.175	0.124	0.074	0.038	0.226	0.571	0.550	0.382	0.285	0.212	0.135	0.071	0.315
RainNet	0.511	0.427	0.252	0.169	0.124	0.088	0.050	0.232	0.569	0.546	0.382	0.278	0.215	0.157	0.092	0.320
SimVP	0.525	0.449	0.282	0.198	0.148	0.104	0.045	0.250	0.585	0.569	0.419	0.318	0.250	0.183	0.084	0.344
Earthformer	0.538	0.466	0.290	0.193	0.135	0.089	0.046	0.251	0.602	0.589	0.429	0.309	0.230	0.159	0.085	0.343
MetNet-3	**0.545**	0.464	0.278	0.182	0.128	0.087	0.047	0.247	**0.613**	0.588	0.414	0.295	0.220	0.156	0.087	0.339
NowcastNet	0.533	0.427	0.266	0.213	0.179	0.147	0.093	0.265	0.605	0.541	0.393	0.335	0.293	0.250	0.166	0.369
CasCast	0.474	0.418	0.270	0.211	0.168	0.122	0.064	0.247	0.532	0.532	0.399	0.332	0.278	0.211	0.118	0.343
IMALCast (R)	0.541	0.487	0.313	0.238	0.193	0.151	0.096	0.289	0.609	0.613	0.457	0.371	0.314	0.256	0.172	0.399
IMALCast (R+S)	0.538	**0.496**	**0.319**	**0.243**	**0.198**	**0.156**	**0.101**	**0.293**	0.603	**0.621**	**0.464**	**0.378**	**0.322**	**0.263**	**0.180**	**0.404**

Lead-Time Performance. Table 9 reports CSI and HSS at six lead times. Its average summarizes these six displayed lead times. IMALCast achieves higher CSI and HSS than all seven deterministic baselines at each displayed lead time, with the advantage maintained at the three-hour forecast horizon.

Table 9: Per-frame CSI and HSS averaged over all rainfall thresholds at different forecast lead times. **Bold** indicates the best performance and underline indicates the second best.

	CSI ↑							HSS ↑						
Model	0.5h	1h	1.5h	2h	2.5h	3h	Avg	0.5h	1h	1.5h	2h	2.5h	3h	Avg
ConvLSTM	0.367	0.261	0.208	0.175	0.152	0.135	0.216	0.482	0.347	0.274	0.228	0.194	0.168	0.282
TrajGRU	0.317	0.225	0.184	0.157	0.137	0.122	0.190	0.416	0.295	0.238	0.199	0.171	0.149	0.245
PhyDNet	0.352	0.242	0.190	0.153	0.123	0.100	0.193	0.463	0.325	0.256	0.207	0.167	0.135	0.259
RainNet	0.363	0.245	0.187	0.149	0.125	0.109	0.196	0.480	0.327	0.246	0.191	0.156	0.132	0.255
SimVP	0.370	0.267	0.211	0.177	0.151	0.133	0.218	0.487	0.357	0.281	0.234	0.196	0.168	0.287
Earthformer	0.370	0.265	0.212	0.180	0.157	0.139	0.220	0.487	0.353	0.279	0.235	0.201	0.176	0.288
MetNet-3	0.363	0.259	0.206	0.174	0.153	0.137	0.215	0.479	0.345	0.271	0.228	0.197	0.175	0.282
IMALCast	**0.397**	**0.317**	**0.262**	**0.219**	**0.185**	**0.160**	**0.257**	**0.525**	**0.430**	**0.356**	**0.297**	**0.248**	**0.211**	**0.344**

D.2 Robustness of the Training Objective

Multi-Seed Objective Ablations. Table 10 summarizes objective ablations with the radar–satellite IMALCast backbone across three training runs with different random seeds. Results are reported as means and sample standard deviations. All three configurations retain the weighted-MAE term; Full includes both structural terms, while Global only and High-intensity only retain the corresponding structural term. The full objective achieves higher mean average CSI/HSS and CSI at 32 and 64 mm/h

Table 10: Multi-seed objective ablations for radar–satellite IMALCast on CRSD, reported as mean ± sample standard deviation across three training runs. The columns report average CSI/HSS and CSI at rainfall thresholds of 32 and 64 mm/h.

Configuration	CSI Avg	HSS Avg	CSI@32	CSI@64
Full	0.2928 ± 0.0002	0.4037 ± 0.0004	0.1558 ± 0.0005	0.0981 ± 0.0027
Global only	0.2746 ± 0.0046	0.3774 ± 0.0067	0.1154 ± 0.0109	0.0790 ± 0.0070
High-intensity only	0.2515 ± 0.0033	0.3488 ± 0.0049	0.1494 ± 0.0023	0.0750 ± 0.0071

than either variant with only one structural term, supporting the complementary roles of the two structural terms across training runs.

Parameter Sensitivity. Figure 5 shows the effect of the loss weights and embedding dimension. Performance is weakest when $\lambda_1 = 0$ and improves before gradually saturating as λ_1 increases, supporting the benefit of global alignment over the tested range. Removing the high-intensity term ($\lambda_2 = 0$) also reduces performance, while excessively large λ_2 weakens the threshold-averaged score. Performance varies less with β and d over their evaluated ranges.

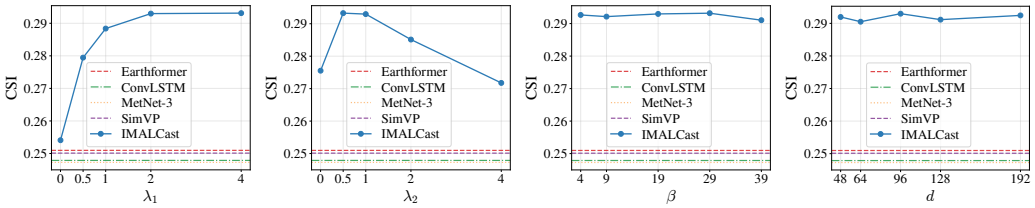

Figure 5: Sensitivity of $\mathrm{CSI}_{\mathrm{avg}}$ to the loss weights λ_1 and λ_2, the weighted-MAE coefficient β, and the embedding dimension d. Dashed lines indicate baseline performance.

D.3 Contribution of Satellite Observations

With the full objective fixed, adding satellite observations improves average CSI. The additional analyses show larger relative gains at high rainfall thresholds and larger average-CSI gains in growing precipitation systems. The relative CSI gain is largest at 64 mm/h. A separate pooled-contingency analysis at this threshold gives ΔCSI of 0.0055 with a 95% sample-level bootstrap interval of [0.0048, 0.0062] and $p < 0.001$ (Table 11). We estimate uncertainty using paired sample-level bootstrap resampling with replacement, applying identical sampled indices to both models and recomputing CSI gains from pooled contingency counts. The 95% confidence intervals are the 2.5th and 97.5th percentiles of the bootstrap gains.

Event groups are defined by changes in precipitation area. For each sample, we take the ratio of future to input precipitation coverage at 1 mm/h, adding 10^{-6} to both coverages before division. Ratios above 1.1, within [0.9,1.1], and below 0.9 define growth, steady, and decay cases, respectively. The average-CSI gain, computed from pooled counts within each group, is 0.0048 for growth cases, compared with 0.0033 for steady cases and 0.0032 for decay cases. These results show that satellite observations provide complementary information for heavy-rain detection, with larger average-CSI gains in the growth group.

Table 11: Satellite contribution on CRSD at the 64 mm/h threshold and across precipitation-area evolution groups. Gains are absolute CSI differences between radar–satellite and radar-only IMALCast, with the full objective fixed, computed from contingency counts pooled over the indicated samples. CI denotes the 95% paired sample-level bootstrap confidence interval; p denotes the proportion of bootstrap replicates with a gain at or below zero.

Sample group	Metric	ΔCSI	95% CI	p
All samples	CSI@64	+0.0055	[0.0048, 0.0062]	<0.001
Growth	CSI Avg	+0.0048	[0.0042, 0.0054]	<0.001
Steady	CSI Avg	+0.0033	[0.0024, 0.0041]	<0.001
Decay	CSI Avg	+0.0032	[0.0026, 0.0038]	<0.001

E Distribution-Correction Analysis

Post-processing can compensate for underestimated rainfall intensities by adjusting the forecast intensity distribution, but requires a reference distribution and can affect both detection and false alarms. We examine whether this additional correction enables the deterministic baselines to match the forecasting skill achieved directly by IMALCast.

We apply the same non-parametric distribution correction to all compared models, using the recent radar observations as the reference. For each sample independently, forecast values are ranked across

spatial and temporal locations and replaced by the correspondingly ranked values from that sample's preceding three-hour input sequence. This preserves forecast rank ordering while matching the empirical intensity distribution of past radar observations; future targets are not used in the correction. By imposing the past intensity distribution on future forecasts, this procedure introduces a strong prior assumption that the recent observed distribution remains representative of the forecast period.

E.1 Forecast Skill Before and After Correction

Table 12 compares IMALCast with the seven deterministic baselines before and after correction. All seven baselines improve in average CSI and HSS; for example, Earthformer's average CSI increases from 0.251 to 0.278. IMALCast achieves nearly the same average CSI and HSS before and after correction: average CSI remains at 0.293, while HSS changes from 0.404 to 0.407.

Even after correction, the baselines remain below uncorrected IMALCast in average CSI (at most 0.278 versus 0.293) and HSS (at most 0.389 versus 0.404). IMALCast therefore achieves higher average CSI and HSS directly, without requiring this additional post-processing step. Figure 6 summarizes performance after correction.

Table 12: Performance before and after distribution correction on CRSD over the full three-hour forecast horizon. CSI, POD, FAR, and HSS average the seven rainfall thresholds; MAE and RMSE are in mm/h. **Bold** and underline indicate the best and second-best results within each block, respectively.

	Raw						After Correction					
Model	CSI ↑	POD ↑	FAR ↓	HSS ↑	MAE ↓	RMSE ↓	CSI ↑	POD ↑	FAR ↓	HSS ↑	MAE ↓	RMSE ↓
ConvLSTM	0.248	0.374	0.456	0.339	0.679	**3.813**	0.277	0.424	0.589	0.389	0.797	4.676
TrajGRU	0.216	0.326	**0.416**	0.291	0.674	3.914	0.265	0.407	0.606	0.371	0.821	4.787
PhyDNet	0.226	0.307	0.481	0.315	0.678	4.035	0.242	0.376	0.635	0.341	0.865	4.995
RainNet	0.232	0.346	0.502	0.320	0.697	3.947	0.252	0.392	0.621	0.356	0.851	4.866
SimVP	0.250	0.375	0.483	0.344	0.690	3.846	0.270	0.416	0.598	0.379	0.814	4.729
Earthformer	0.251	0.362	0.459	0.343	0.654	3.833	0.278	0.424	0.589	0.389	0.796	4.686
MetNet-3	0.247	0.347	0.440	0.339	**0.645**	3.858	0.277	0.423	0.590	0.388	0.801	4.706
IMALCast	**0.293**	**0.459**	0.585	**0.404**	0.760	4.491	**0.293**	**0.440**	**0.573**	**0.407**	**0.772**	**4.674**

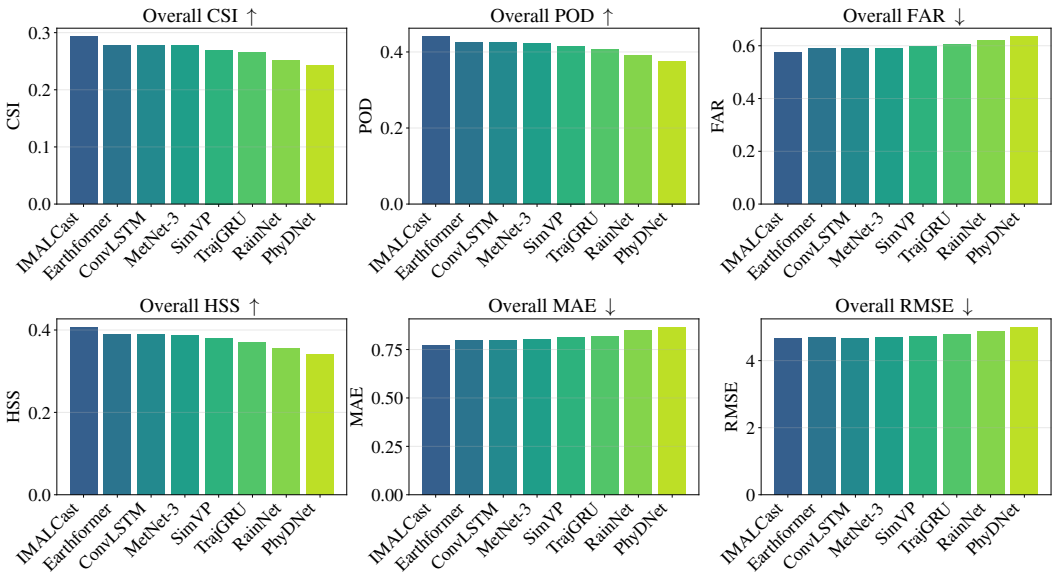

Figure 6: Overall evaluation metrics after distribution correction on CRSD over the full three-hour forecast horizon. CSI, POD, FAR, and HSS average the seven rainfall thresholds; MAE and RMSE measure intensity errors in mm/h.

E.2 Detection and False Alarms After Correction

The gains from correction are accompanied by changes in detection and false alarms. For all seven baselines, correction increases average POD but also average FAR (Table 12), showing that improved detection comes with additional false alarms. For IMALCast, both average POD and FAR decrease, while average CSI remains nearly unchanged.

Tables 13 and 14 provide the corresponding per-threshold results after correction. IMALCast retains the highest POD and lowest FAR among the compared models at every threshold, although the margins at the highest rainfall threshold are small. These results complement the main finding above: IMALCast's raw forecasts already achieve higher average CSI and HSS than the baselines obtain with this additional correction.

Table 13: Per-threshold POD after distribution correction on CRSD, computed over all 30 forecast frames. Thresholds are in mm/h. **Bold** and underline indicate the best and second-best results, respectively.

Model	0.1	1	5	10	16	32	64	Avg
ConvLSTM	0.687	0.622	0.477	0.394	0.330	0.268	0.194	0.424
TrajGRU	0.681	0.611	0.461	0.375	0.308	0.241	0.169	0.407
PhyDNet	0.661	0.597	0.422	0.333	0.263	0.200	0.157	0.376
RainNet	0.671	0.588	0.430	0.350	0.293	0.237	0.174	0.392
SimVP	0.682	0.607	0.459	0.381	0.322	0.265	0.193	0.416
Earthformer	0.695	0.626	0.478	0.394	0.328	0.263	0.187	0.424
MetNet-3	0.699	0.624	0.473	0.390	0.326	0.261	0.188	0.423
IMALCast (Ours)	**0.727**	**0.658**	**0.488**	**0.399**	**0.337**	**0.278**	**0.197**	**0.440**

Table 14: Per-threshold FAR after distribution correction on CRSD, computed over all 30 forecast frames. Thresholds are in mm/h. **Bold** and underline indicate the best and second-best results, respectively.

Model	0.1	1	5	10	16	32	64	Avg
ConvLSTM	0.313	0.383	0.535	0.622	0.688	0.753	0.828	0.589
TrajGRU	0.319	0.393	0.550	0.640	0.709	0.779	0.850	0.606
PhyDNet	0.338	0.408	0.589	0.681	0.752	0.817	0.860	0.635
RainNet	0.329	0.417	0.582	0.664	0.724	0.783	0.846	0.621
SimVP	0.319	0.399	0.553	0.635	0.696	0.756	0.828	0.598
Earthformer	0.305	0.379	0.535	0.622	0.690	0.758	0.834	0.589
MetNet-3	0.302	0.381	0.539	0.626	0.693	0.759	0.833	0.590
IMALCast (Ours)	**0.272**	**0.346**	**0.525**	**0.617**	**0.682**	**0.743**	**0.826**	**0.573**

F Additional Qualitative Results

Figure 7 shows an enlarged view of a heavy-rain band from the example in Figure 4. IMALCast retains a broader high-intensity region at one hour, while all three forecasts smooth the observed fine-scale structure. Figures 8–10 show three additional CRSD cases: an elongated rainband, a broad curved precipitation system, and localized precipitation cells. These examples illustrate how intensity retention and spatial smoothing vary across rainfall patterns and lead times; the panel scores also show that the relative performance depends on the case and metric.

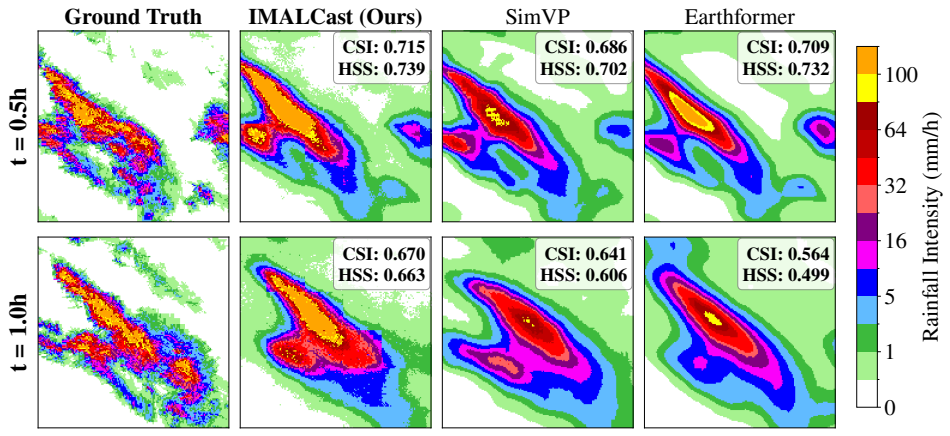

Figure 7: An enlarged heavy-rain band from the CRSD case in Figure 4. Rows show forecasts at 0.5 and 1.0 hours; columns show ground truth, IMALCast, SimVP, and Earthformer. Colors indicate rainfall intensity in mm/h, and annotations give CSI and HSS at 1 mm/h for each forecast panel.

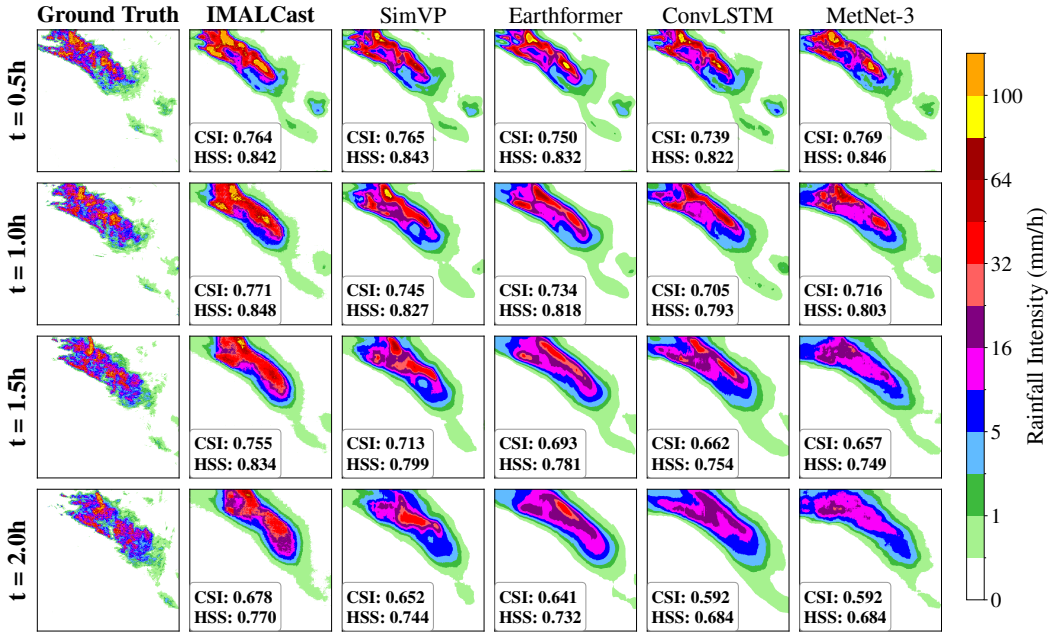

Figure 8: An additional CRSD case showing an elongated rainband. Rows indicate lead times of 0.5, 1.0, 1.5, and 2.0 hours. Columns show ground truth, IMALCast, SimVP, Earthformer, ConvLSTM, and MetNet-3. Colors indicate rainfall intensity in mm/h; panel annotations give CSI and HSS at 1 mm/h.

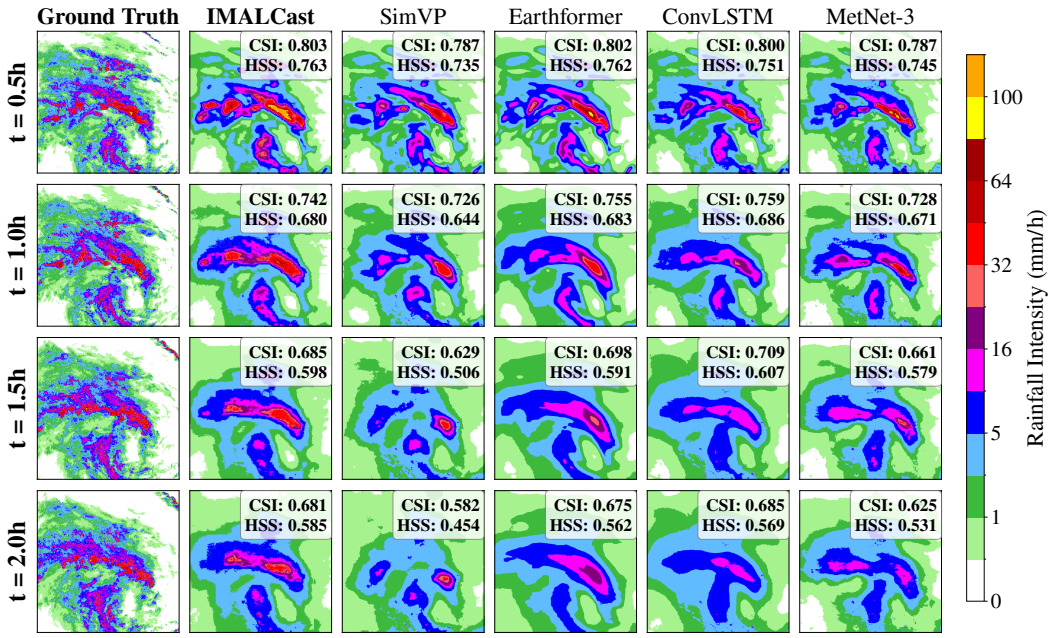

Figure 9: An additional CRSD case with a broad curved precipitation system. Rows indicate lead times of 0.5, 1.0, 1.5, and 2.0 hours. Columns show ground truth, IMALCast, SimVP, Earthformer, ConvLSTM, and MetNet-3. Colors indicate rainfall intensity in mm/h; panel annotations give CSI and HSS at 1 mm/h.

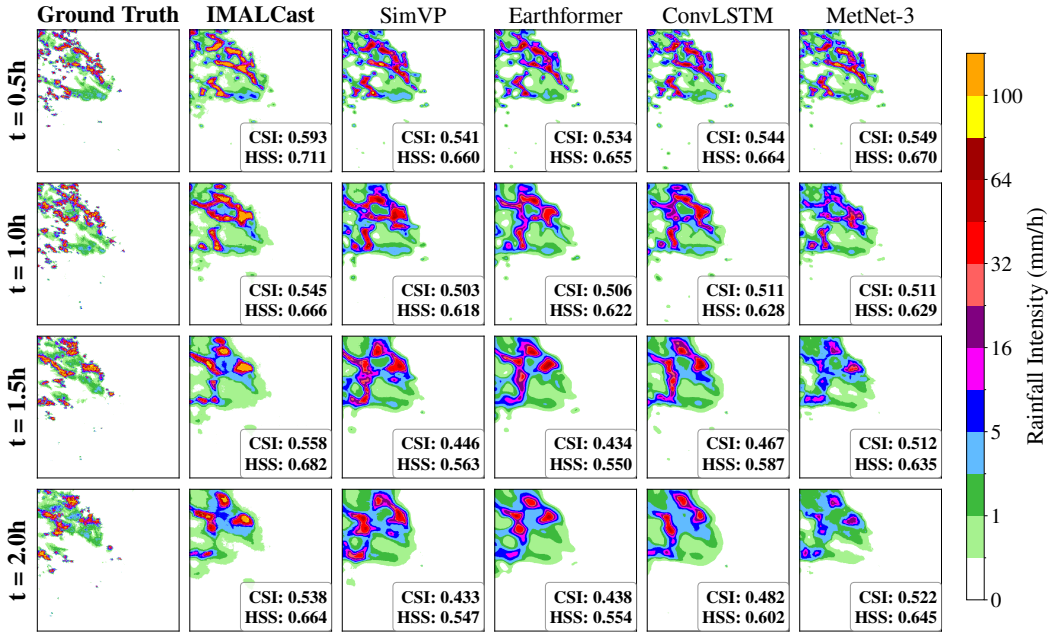

Figure 10: An additional CRSD case with localized precipitation cells. Rows indicate lead times of 0.5, 1.0, 1.5, and 2.0 hours. Columns show ground truth, IMALCast, SimVP, Earthformer, ConvLSTM, and MetNet-3. Colors indicate rainfall intensity in mm/h; panel annotations give CSI and HSS at 1 mm/h.