TRANSPORT-GUIDED FACTORIZED EVOLUTION FOR PRECIPITATION NOWCASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Precipitation nowcasting predicts the spatial distribution and intensity of precipitation over the next few hours from recent radar observations. Motion-aware methods combine advective transport of existing echoes with intensity updates to model localized growth and decay. However, in recursive rollout, each forecast already incorporates predicted intensity updates and is then fed back as input to the next transport step. This forecast-field feedback carries earlier evolution errors into subsequent transport steps and can compound spatial distortion. We present **Transport-Guided Factorized Evolution (TFEvo)**, an end-to-end framework that breaks this forecast-field feedback by using learned displacement fields to warp the latest observed radar field directly to each forecast horizon. Conditioned on the transported prediction, TFEvo models residual evolution by factorizing the field into spatial coverage and intensity components. This factorization accounts for the zero-inflated sparsity of radar data by representing zero-valued background through the coverage component. These updates are learned with an asymmetric training objective applied to the evolved prediction to penalize the underestimation of extreme rainfall. Across three benchmarks, TFEvo outperforms competitive baselines on most forecasting metrics. Notably, at 384×384 resolution, it improves CSI-M by up to **4.5%** and reduces LPIPS by up to **40.6%** relative to a competitive high-resolution baseline, while using less than **1%** of the baseline's parameters. Code is available at `https://anonymous.4open.science/r/TFEvo-CECD`.

1 INTRODUCTION

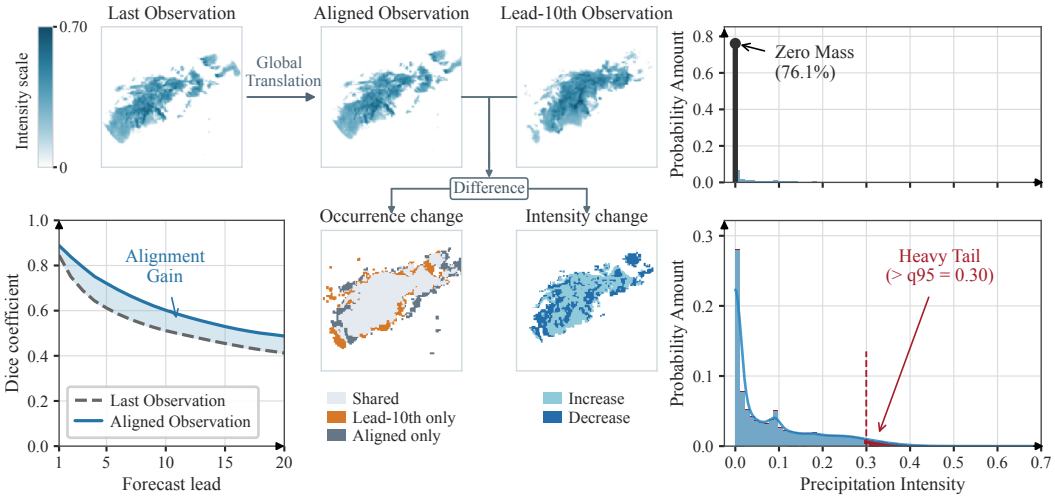
Precipitation nowcasting, the short-term prediction of precipitation within the next few hours, underpins weather-sensitive operations such as aviation planning, urban flood control, and disaster management (Ravuri et al., 2021; Espeholt et al., 2022). Recent advances in deep learning deliver accurate and efficient radar forecasts that sidestep the computational cost of numerical weather prediction (Lin et al., 2025; Wen et al., 2026). These forecasts must capture both the movement of existing echoes and changes in their coverage and intensity. To model both types of change explicitly, motion-aware methods with recursive rollout combine advection with predicted intensity updates at each step and then reuse the updated field as input to the next transport step(Zhang et al., 2023; Wu et al., 2026). However, this forecast-field feedback carries earlier evolution errors into subsequent transport steps, compounding spatial distortion.

To avoid this feedback, we consider transporting the latest observation directly to each forecast horizon and learning the remaining changes relative to the transported field. Figure 1 motivates the use of transported observations as spatial references. A coarse global translation improves overlap with future observations across forecast leads (Fig. 1a), indicating that recent observations retain alignable echo structure. Furthermore, learning these remaining changes requires accounting for the statistics of the target radar fields, which exhibit pronounced zero-inflation and relatively rare high-intensity echoes (Fig. 1b). These considerations motivate our central question:

> *How can we learn coverage and intensity changes relative to transported observations without forecast-field feedback, while accounting for zero-inflated radar targets?*

To answer this question, we introduce **TFEvo** (**T**ransport-Guided **F**actorized **Evo**lution), an end-to-end framework that learns residual evolution relative to transported observations. For the trans-

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.



(a) Diagnostic alignment and remaining changes (b) Zero inflation and long tail

Figure 1: **Empirical diagnostic of radar dynamics (MeteoNet). (a) Spatial alignment and remaining changes:** Global translation improves overlap (Dice) between the latest observed radar field and future observations, while coverage and intensity discrepancies remain after alignment. **(b) Target-value distribution:** Radar reflectivity contains a large zero-valued background and relatively rare high-intensity echoes.

port stage, an **Anchor-based Shared Transport (ASTrans)** module estimates displacement fields from radar history and warps the latest observed radar field directly to each forecast horizon, breaking forecast-field feedback. The resulting transported prediction captures macroscopic advection but cannot account for localized cell growth, decay, or initiation. To model this residual, a **Zero-inflation-aware Factorized Evolution (ZFEvo)** module then factorizes each transported prediction into spatial coverage and intensity components and learns their updates conditioned on the transported field, spatiotemporal context, and the learned motion representation. The coverage component represents zero-valued background, and the two updated factors jointly reconstruct the evolved prediction. To further counter the underestimation driven by rare extreme events, TFEvo incorporates an asymmetric training objective alongside joint supervision on both the transported and evolved predictions. Extensive experiments on SEVIR, MeteoNet, and Shanghai demonstrate that TFEvo outperforms competitive baselines on most evaluation metrics. With **0.29M** parameters, TFEvo uses less than **1%** of the parameters of exPreCast. At 384×384 resolution, it improves CSI-M by up to **4.5%** and reduces LPIPS by up to **40.6%** relative to exPreCast.

Our main contributions are summarized as follows:

- We propose TFEvo, an end-to-end nowcasting framework that breaks forecast-field feedback by transporting the latest observation directly to each forecast horizon and using the resulting field as an explicit reference for residual evolution.
- We develop ZFEvo, a factorized evolution module to learn coverage–intensity updates, combined with an asymmetric training objective to counter underestimation of extreme rainfall.
- Extensive experiments across three benchmarks show the complementary forecasting benefits of transport and evolution, while high-resolution comparisons demonstrate gains over a competitive high-resolution baseline with less than 1% of its parameters.

2 Related Work

Spatio-Temporal Transport in Precipitation Forecasting Deep learning approaches for precipitation nowcasting span deterministic architectures (Shi et al., 2015; Gao et al., 2022a;b; Song et al., 2026), probabilistic generators (Ravuri et al., 2021; Gao et al., 2023; Yu et al., 2024), and physics-informed models that decouple latent dynamics or spectral components (Le Guen & Thome, 2020;

Lin et al., 2025). Among these, motion-aware methods have become a dominant paradigm by separating spatial transport estimation from intensity updates (Zhang et al., 2023; Wu et al., 2026). However, these methods predominantly adopt a sequential rollout paradigm that recursively warps preceding forecasts, propagates local errors through repeated interpolation (Ayzel et al., 2019), and accumulates spatial distortion over extended horizons. In the video prediction, MMVP demonstrates that future frames can be directly referenced to the latest observation via motion correspondence (Zhong et al., 2023), providing independent evidence that observation anchoring is a viable alternative to recursive rollout. Following this line, TFEvo anchors transport exclusively to the latest observed radar field via accumulated displacement fields, while confining temporal recurrence to the motion representation so that no predicted radar field ever re-enters the transport pathway.

Zero-Inflated Sparsity and Long-Tailed Extreme Modeling Radar reflectivity fields exhibit severe zero-inflated sparsity alongside a heavy-tailed distribution of intense convective events (Scheuerer & Hamill, 2015; Wilson et al., 2022), where standard regression tends toward oversmoothed conditional means. Existing remedies operate at two levels. At the distributional level, parametric models such as hurdle distributions (Wilson et al., 2022), Tweedie processes (Gao et al., 2026a), and post-hoc tail calibration (Gao et al., 2026b) explicitly model the zero-inflated or heavy-tailed output distribution. At the objective level, balanced regression (Ren et al., 2022) and Fourier amplitude constraints (Yan et al., 2024; Song et al., 2026) re-weight or reshape the training loss to shift gradient focus toward underrepresented regimes. TFEvo takes a complementary route at the representation level, factorizing each transported prediction into spatial coverage and intensity components to structurally isolate the zero-valued background. This factorization is paired with an asymmetric quantile-based objective (Koenker & Hallock, 2001) that penalizes underestimation of extreme events.

3 PRELIMINARIES

Task Formulation Given L historical radar observations up to the forecast origin $t = 0$, precipitation nowcasting aims to predict the future sequence across T lead steps:

$$\widehat{\boldsymbol{X}}_{1:T} = \mathcal{F}_\theta(\boldsymbol{X}_{-L+1:0}), \tag{1}$$

where $\mathcal{F}_\theta$ denotes the forecasting model parameterized by θ, and $\widehat{\boldsymbol{X}}_{1:T}$ approximates the ground-truth sequence $\boldsymbol{X}_{1:T}$. Each radar field $\boldsymbol{X}_t \in [0,1]^{H\times W}$ represents the normalized precipitation intensity over a discretized spatial domain $\Omega \subset \mathbb{R}^2$.

Transport–Evolution Decomposition The dynamics of a continuous precipitation field $X(\boldsymbol{x}, \tau)$ are governed by the advection–evolution equation (Foresti et al., 2019):

$$\frac{\partial X}{\partial \tau} + \boldsymbol{v} \cdot \nabla X = r, \tag{2}$$

where $\boldsymbol{v}(\boldsymbol{x}, \tau) \in \mathbb{R}^2$ is the velocity field driving spatial transport, and $r(\boldsymbol{x}, \tau) \in \mathbb{R}$ is the Lagrangian source–sink rate governing localized growth and decay. Setting $r = 0$ yields pure advection, under which the field is conserved along flow trajectories, motivating the discrete decomposition

$$\boldsymbol{X}_t \approx \underbrace{\mathcal{W}(\boldsymbol{X}_0, \boldsymbol{d}_t)}_{\text{spatial transport}} + \underbrace{\boldsymbol{R}_t}_{\text{residual evolution}}, \tag{3}$$

where $\mathcal{W}$ denotes spatial warping by the cumulative displacement $\boldsymbol{d}_t$ from the forecast origin to lead t, and $\boldsymbol{R}_t$ captures the accumulated effect of growth and decay that transport alone cannot reproduce.

4 PROPOSED APPROACH

Overview. Guided by the decomposition in Eq. 3, TFEvo forecasts future precipitation through two cascading stages (Figure 2). **Anchor-based Shared Transport (ASTrans)** estimates the transport term $\mathcal{W}(\boldsymbol{X}_0, \boldsymbol{d}_t)$, and **Zero-inflation-aware Factorized Evolution (ZFEvo)** models the residual $\boldsymbol{R}_t$. To account for the zero-inflated sparsity of radar targets, ZFEvo models residual evolution

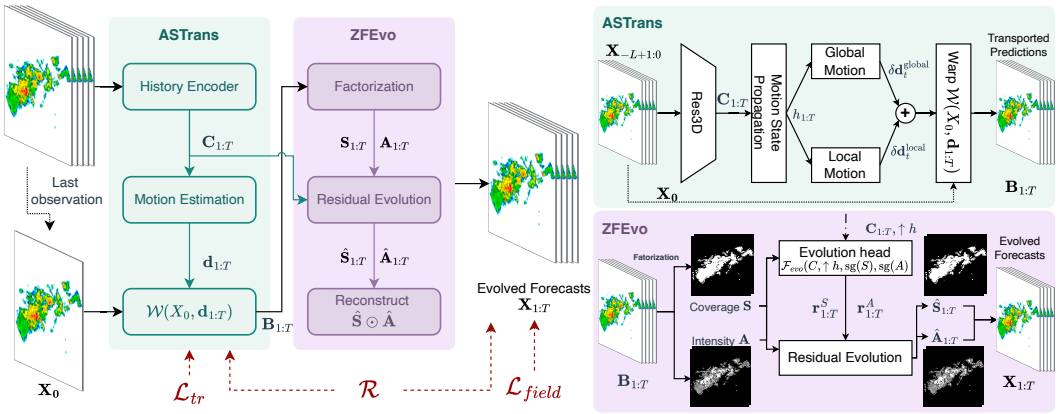


Figure 2: **Overview of TFEvo.** TFEvo decouples precipitation forecasting into anchor-based transport (**ASTrans**) and factorized residual evolution (**ZFEvo**). ASTrans accumulates displacement fields to warp the latest observation $\boldsymbol{X}_0$ into the transported prediction $\boldsymbol{B}_t$. ZFEvo factorizes $\boldsymbol{B}_t$ into spatial coverage and intensity components and predicts residual corrections under joint supervision.

through updates to the spatial coverage and intensity factors of the transport prior. Both stages are conditioned on a shared spatiotemporal context. A history encoder $\mathcal{F}_{\text{enc}}$ extracts

$$\boldsymbol{C}_{1:T} = \mathcal{F}_{\text{enc}}(\boldsymbol{X}_{-L+1:0},\ \boldsymbol{e}_{1:T}), \tag{4}$$

where $\boldsymbol{e}_t$ is a learnable embedding identifying the forecast lead t. The context $\boldsymbol{C}_t$ conditions a motion head $\mathcal{F}_{\text{mot}}$ within ASTrans and an evolution head $\mathcal{F}_{\text{evo}}$ within ZFEvo.

4.1 Anchor-based Shared Transport (ASTrans)

ASTrans maintains a recurrent motion state to track the temporal progression of advection, while warping strictly from the latest observation $\boldsymbol{X}_0$.

Motion State Propagation. Atmospheric advection evolves continuously across consecutive leads, so ASTrans tracks it with a hidden motion state $\boldsymbol{h}_t$. At lead t, context $\boldsymbol{C}_t$ produces a candidate state $\widetilde{\boldsymbol{h}}_t = \mathcal{F}_{\text{cand}}(\boldsymbol{C}_t)$ and an update gate $\boldsymbol{g}_t^{\text{mot}} = \sigma(\mathcal{F}_{\text{gate}}(\boldsymbol{C}_t))$, while a learned channel-wise decay balances retained motion against the current update:

$$\boldsymbol{\alpha} = \sigma(-\boldsymbol{\omega}), \qquad \boldsymbol{h}_t = \boldsymbol{\alpha} \odot \boldsymbol{h}_{t-1} + (1-\boldsymbol{\alpha}) \odot \boldsymbol{g}_t^{\text{mot}} \odot \widetilde{\boldsymbol{h}}_t, \tag{5}$$

where $\boldsymbol{h}_0 = \boldsymbol{0}$, $\boldsymbol{\omega}$ is a learnable parameter vector, $\sigma(\cdot)$ is the sigmoid function, and $\odot$ is elementwise multiplication. Channels with $\boldsymbol{\alpha}$ close to one preserve long-term advective persistence, while $\boldsymbol{g}_t^{\text{mot}}$ gates lead-specific updates.

Global-Local Motion Decoupling & Anchor Warping. From $\boldsymbol{h}_t$, the motion head $\mathcal{F}_{\text{mot}}$ predicts a spatially uniform global displacement $\delta\boldsymbol{d}_t^{\text{global}} \in \mathbb{R}^2$ and a pixel-wise local deformation field $\widetilde{\delta\boldsymbol{d}}_t^{\text{local}} \in \mathbb{R}^{2\times H\times W}$. To separate spatially uniform displacement from local deformation, the local field is zero-centered:

$$\delta\boldsymbol{d}_t^{\text{local}}(\boldsymbol{x}) = \widetilde{\delta\boldsymbol{d}}_t^{\text{local}}(\boldsymbol{x}) - \frac{1}{|\Omega|}\sum_{\boldsymbol{x}'\in\Omega} \widetilde{\delta\boldsymbol{d}}_t^{\text{local}}(\boldsymbol{x}'), \qquad \delta\boldsymbol{d}_t = \delta\boldsymbol{d}_t^{\text{global}} + \delta\boldsymbol{d}_t^{\text{local}}. \tag{6}$$

where Ω denotes the spatial grid. ASTrans accumulates displacement increments and warps the observation $\boldsymbol{X}_0$ in a single sampling step to obtain the transported prediction $\boldsymbol{B}_t$:

$$\boldsymbol{d}_t = \delta\boldsymbol{d}_t + \mathcal{W}(\boldsymbol{d}_{t-1}, \delta\boldsymbol{d}_t), \qquad \boldsymbol{B}_t = \mathcal{W}(\boldsymbol{X}_0, \boldsymbol{d}_t), \tag{7}$$

where $\boldsymbol{d}_0 = \boldsymbol{0}$, and $\mathcal{W}(\boldsymbol{X}, \boldsymbol{d})$ denotes the backward warping operator that samples field $\boldsymbol{X}$ at displaced positions $\boldsymbol{x} - \boldsymbol{d}(\boldsymbol{x})$ via differentiable spatial grid sampling (Jaderberg et al., 2015). Since every $\boldsymbol{B}_t$ is sampled from the observation $\boldsymbol{X}_0$, no predicted field re-enters the transport pathway.

4.2 ZERO-INFLATION-AWARE FACTORIZED EVOLUTION (ZFEVO)

ZFEvo models the residual $\boldsymbol{R}_t$ by factorizing the transported prediction $\boldsymbol{B}_t$ into spatial coverage and intensity factors, then predicting motion-guided residual corrections to the two factors.

Coverage–Intensity Factorization. To isolate precipitation coverage from cell intensity while preventing division singularities along dry–wet boundaries, ZFEvo uses a small threshold κ to factorize $\boldsymbol{B}_t$:

$$\boldsymbol{A}_t = \max(\boldsymbol{B}_t, \kappa), \qquad \boldsymbol{S}_t = \frac{\boldsymbol{B}_t}{\boldsymbol{A}_t}, \qquad \text{satisfying} \quad \boldsymbol{S}_t \odot \boldsymbol{A}_t \equiv \boldsymbol{B}_t. \tag{8}$$

disentangling spatial coverage $\boldsymbol{S}_t \in [0,1]$ from intensity $\boldsymbol{A}_t \in [\kappa, 1]$. This representation encodes zero-valued background through the coverage factor, while keeping the intensity factor bounded below by κ.

Motion-Guided Residual Evolution. The evolution head $\mathcal{F}_{\text{evo}}$ predicts residual adjustments conditioned on context $\boldsymbol{C}_t$, upsampled motion state $\uparrow \boldsymbol{h}_t$, and detached base factors: $(\boldsymbol{r}_t^A, \boldsymbol{r}_t^S) = \mathcal{F}_{\text{evo}}(\boldsymbol{C}_t, \uparrow \boldsymbol{h}_t, \text{sg}(\boldsymbol{S}_t), \text{sg}(\boldsymbol{A}_t))$, where $\text{sg}(\cdot)$ denotes the stop-gradient operator. The factors are updated and recombined to yield the forecast:

$$\widehat{\boldsymbol{A}}_t = \text{clip}(\boldsymbol{A}_t + \gamma \tanh \boldsymbol{r}_t^A), \quad \widehat{\boldsymbol{S}}_t = \boldsymbol{S}_t + (1 - \boldsymbol{S}_t) \odot \text{clip}(\tanh \boldsymbol{r}_t^S), \quad \widehat{\boldsymbol{X}}_t = \widehat{\boldsymbol{S}}_t \odot \widehat{\boldsymbol{A}}_t, \tag{9}$$

where $\text{clip}(\cdot)$ constrains values to $[0,1]$, and γ scales the maximum allowable intensity adjustment. $\widehat{\boldsymbol{A}}_t$ captures cell strengthening and dissipation. The coverage update is nonnegative and gated by $(1 - \boldsymbol{S}_t)$, so it increases coverage only where the transported coverage factor is below one. Decay is expressed through the intensity factor. When both residuals are inactive ($\boldsymbol{r}_t^A = \boldsymbol{0}$ and $\boldsymbol{r}_t^S \leq \boldsymbol{0}$), $\widehat{\boldsymbol{X}}_t$ recovers the transported prediction $\boldsymbol{B}_t$ exactly, so that localized evolution functions as a residual correction to the transported prediction.

4.3 TRAINING OBJECTIVES

TFEvo is trained end-to-end via a multi-task objective:

$$\mathcal{L} = \lambda \mathcal{L}_{\text{tr}} + (1 - \lambda)\mathcal{L}_{\text{field}} + \eta \mathcal{R}, \tag{10}$$

where $\lambda \in [0,1]$ balances intermediate transport supervision against final forecast fidelity, and η controls regularization strength.

The transported prediction is supervised via mean squared error: $\mathcal{L}_{\text{tr}} = \frac{1}{T}\sum_{t=1}^{T} \text{MSE}(\boldsymbol{B}_t, \boldsymbol{X}_t)$. To counteract the underestimation of extreme rainfall, the final forecast uses an asymmetric body–tail loss over each pixel:

$$\ell(\widehat{x}, x) = \begin{cases} (\widehat{x} - x)^2, & x < u, \\ \max\{q(x - \widehat{x}), (1 - q)(\widehat{x} - x)\}, & x \geq u, \end{cases} \tag{11}$$

where u separates moderate from heavy rainfall regimes, and $q \in (0.5, 1)$ imposes an asymmetric penalty that more severely punishes underestimation. The field loss $\mathcal{L}_{\text{field}}$ averages $\ell(\widehat{x}, x)$ across all samples, leads, and pixels. $\mathcal{R} = \mathcal{R}_{\text{motion}} + \mathcal{R}_{\text{evolution}}$ regularizes spatial smoothness and temporal consistency of displacement increments and factor updates. The complete inference procedure and further implementation details are provided in Appendix A.

5 EXPERIMENTS

We evaluate TFEvo in terms of forecast accuracy, structural and perceptual quality, and computational efficiency. We first compare forecasting performance and efficiency across standard benchmarks, then examine the roles of transport and factorized evolution through ablations, and finally visualize forecasts and intermediate fields in selected cases.

5.1 EXPERIMENTAL SETUP

We evaluate TFEvo across three real-world radar benchmarks spanning distinct meteorological scales (SEVIR, MeteoNet, and Shanghai) (Veillette et al., 2020; Larvor et al., 2020; Chen et al.,

Table 1: **Comparison with published baselines.** Bold and underline indicate the best and second-best results, respectively; – denotes an unreported value.

Method	SEVIR					MeteoNet					Shanghai				
	CSI-M↑	CSI-181↑	CSI-219↑	HSS↑	SSIM↑	CSI-M↑	CSI-24↑	CSI-32↑	HSS↑	SSIM↑	CSI-M↑	CSI-35↑	CSI-40↑	HSS↑	SSIM↑
ConvGRU	0.2903	0.0879	0.0350	0.3619	0.6100	0.3401	0.2990	0.1431	0.4667	0.7833	0.3612	0.3163	0.2062	0.4899	0.7796
MAU	0.3076	0.1071	0.0516	0.3863	0.6505	0.3233	0.2839	0.0997	0.4452	0.7897	0.3983	0.3621	0.2417	0.5346	0.7195
SimVP	0.3108	0.1106	0.0517	0.3924	0.6508	0.3351	0.3002	0.1130	0.4573	0.7804	0.3850	0.3549	0.2382	0.5194	0.7795
FourCastNet	0.2686	0.0717	0.0339	0.3355	0.5976	0.3027	0.2533	0.1085	0.4216	0.6450	0.3571	0.3108	0.2073	0.4868	0.5598
EarthFormer	0.2892	0.0844	0.0245	0.3665	0.6633	0.3205	0.2884	0.1237	0.4491	0.7772	0.3503	0.3178	0.1872	0.4844	0.7298
PhyDNet	0.3017	0.1040	0.0278	0.3812	0.6532	0.3384	0.3194	0.1366	0.4673	0.7823	0.3654	0.3236	0.2176	0.4957	0.7751
EarthFarseer	0.3004	0.0992	0.0413	0.3829	0.6327	0.3404	0.3170	0.1372	0.4726	0.7542	0.3926	0.3608	0.2343	0.5330	0.5405
NowCastNet	0.2791	0.0770	0.0351	0.3512	0.6839	0.3427	0.3206	0.1598	0.4751	0.7879	0.3953	0.3608	0.2450	0.5334	0.7902
DiffCast	0.3050	0.1300	0.0582	0.3996	0.6482	0.3512	0.3340	0.1808	0.4846	0.7887	0.4089	0.3740	0.2606	0.5476	0.7879
AlphaPre	0.3259	0.1332	0.0545	0.4110	**0.6884**	0.3824	0.3633	0.2002	0.5164	0.7968	0.4178	0.3854	0.2615	0.5534	0.7951
DuoCast	0.3375	–	–	0.4318	0.6827	0.3892	–	–	0.5297	0.7981	0.4252	–	–	0.5643	0.7788
MoCast	0.3114	–	–	0.3896	0.6776	0.3489	–	–	0.4738	0.8146	0.4156	–	–	0.5528	0.7812
TFEvo	**0.3578**	**0.2018**	**0.1163**	**0.4657**	0.6826	**0.4094**	**0.4005**	**0.2471**	**0.5512**	**0.8149**	**0.4340**	**0.3941**	**0.2926**	**0.5749**	**0.7964**

Table 2: **Detailed comparison and computational efficiency.** Params are reported in millions (M), and latency in milliseconds per forecast sequence (ms/seq). Bold and underline mark the best and second-best scores, respectively; – denotes an unreported value.

Method	Efficiency			SEVIR			MeteoNet			Shanghai		
	Params	GFLOPs*	Latency	CSI-219↑	POD-M↑	LPIPS↓	CSI-32↑	POD-M↑	LPIPS↓	CSI-40↑	POD-M↑	LPIPS↓
AlphaPre	89.026	3110.75	145.09	0.0800	0.4064	0.2848	0.2021	0.4789	0.2671	0.2771	0.5275	0.3236
exPreCast	32.004	26.54	17.07	0.0533	0.4460	0.3648	0.1668	0.4940	0.3101	0.2211	0.5950	0.3771
DuoCast	179.304	9625.14	2515.16	0.0945	0.4129	**0.2312**	–	–	–	–	–	–
TFEvo	0.287	65.90	34.98	**0.1163**	**0.5771**	0.2605	**0.2471**	**0.6581**	**0.1716**	**0.2926**	**0.6707**	**0.1843**

2020). Following the benchmark setting of Lin et al. (2025), all models take $L = 5$ historical radar frames to forecast the subsequent $T = 20$ frames at a spatial resolution of 128×128. We additionally evaluate high-resolution forecasting at 384×384. Because all methods follow this setting, published results can be compared directly (Lin et al., 2025; Wen et al., 2026; Wu et al., 2026). We compare TFEvo against 12 competitive baselines, among which AlphaPre, exPreCast, and DuoCast are additionally re-trained locally under the same protocol with their official implementations for operational evaluation. Forecast accuracy is measured by the Critical Success Index (CSI), Heidke Skill Score (HSS), and Probability of Detection (POD), reported as threshold-averaged CSI-M and POD-M and high-threshold CSI for severe rainfall. The perceptual quality is measured by SSIM and LPIPS. All locally evaluated models share the training protocol detailed in Appendix B.

5.2 Main Results

Table 3: **Forecasting performance at** 384×384**.** Params are in millions (M), and GFLOPs per forecast sequence. Bold marks the best score.

Method	Params	GFLOPs*	MeteoNet					Shanghai				
	(M)		CSI-M↑	CSI-32↑	POD-M↑	SSIM↑	LPIPS↓	CSI-M↑	CSI-40↑	POD-M↑	SSIM↑	LPIPS↓
exPreCast	32.00	189.97	0.3867	0.2144	0.5328	0.7920	0.3342	0.4059	0.2589	0.5743	0.7664	0.3016
TFEvo	0.29	593.08	**0.4041**	**0.2457**	**0.6393**	**0.8450**	**0.1986**	**0.4222**	**0.2847**	**0.6432**	**0.8482**	**0.2035**

Benchmark performance and efficiency. Table 1 and Table 2 summarize forecasting performance, perceptual quality, and computational costs against competitive baselines. From these results, we highlight three key findings: 1) ***Forecast accuracy and high-intensity detection.*** TFEvo achieves the highest reported CSI-M among the methods listed in Table 1 on all three benchmarks (up to +6.0% on SEVIR), while improving CSI-219 by 99.8% over DiffCast and POD-M by 42.0% over AlphaPre. 2) ***Forecast accuracy and perceptual quality.*** TFEvo reduces LPIPS by up to 43.0% relative to AlphaPre and achieves higher CSI-M in the main benchmark comparison. 3) ***Parameter count and inference latency.*** TFEvo delivers superior accuracy with only 0.29M parameters—requiring less than 1% of exPreCast (32.00M) and roughly 0.16% of DuoCast (179.30M). Furthermore, its inference latency (34.98 ms) is over 70× lower than that of DuoCast (2515.16 ms),

Table 4: **Component ablation.** Each + adds a component to the nearest preceding configuration with less indentation. Bold marks the complete TFEvo model.

Method	SEVIR					MeteoNet					Shanghai				
	CSI-M↑	CSI-219↑	POD-M↑	SSIM↑	LPIPS↓	CSI-M↑	CSI-32↑	POD-M↑	SSIM↑	LPIPS↓	CSI-M↑	CSI-40↑	POD-M↑	SSIM↑	LPIPS↓
Static anchor	0.2186	0.0367	0.3075	0.5988	0.1932	0.2404	0.1229	0.3700	0.7374	0.1341	0.2711	0.1608	0.4083	0.7324	0.1666
+ ZFEvo	0.3200	0.0744	0.4940	0.6646	0.2542	0.3576	0.2180	0.5281	0.7936	0.1748	0.3795	0.2563	0.5590	0.7764	0.1863
+ ASTrans	0.3413	0.1052	0.4684	0.6659	0.1813	0.3810	0.2194	0.5309	0.8003	0.1238	0.4024	0.2652	0.5736	0.7836	0.1566
+ Unfactorized Evo	0.3577	0.1178	0.5821	0.6734	0.2900	0.4109	0.2469	0.6626	0.8054	0.2219	0.4350	0.2912	0.6775	0.7694	0.2594
+ ZFEvo (TFEvo)	0.3578	0.1163	0.5771	0.6826	0.2605	0.4094	0.2471	0.6581	0.8149	0.1716	0.4340	0.2926	0.6707	0.7964	0.1843

indicating that persistent transport–evolution decoupling provides a more practical and scalable alternative than parameter scaling.

High-resolution forecasting. Table 3 evaluates model scalability on high-resolution 384×384 grids. Across both datasets, TFEvo outperforms exPreCast, improving CSI-M by up to 4.5%, increasing high-threshold CSI by up to 14.6% (CSI-32 on MeteoNet), and reducing perceptual distance (LPIPS) by up to 40.6%. TFEvo achieves these gains with only 0.29M parameters—less than 1% of exPreCast (32.00M). These gains come with higher computation per forecast sequence than exPreCast (593.08 vs. 189.97 GFLOPs).

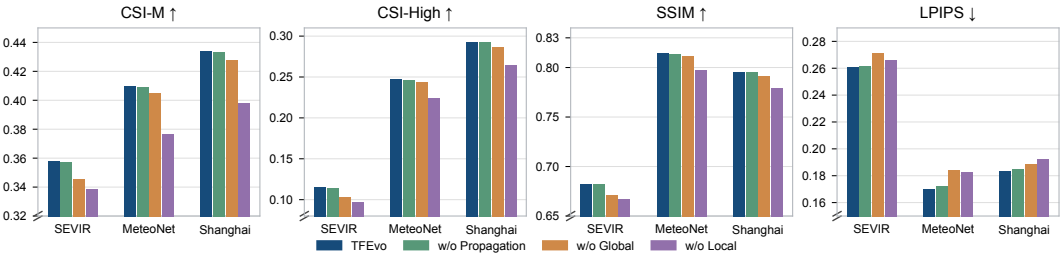


Figure 3: **Ablation on motion components.** Comparison of removing motion state propagation (*w/o Propagation*), global displacement (*w/o Global*), and local deformation (*w/o Local*).

5.3 Ablation Studies

Effect of transport and factorized evolution. Table 4 investigates the core modules of TFEvo. Starting from the static anchor, incorporating either ASTrans or ZFEvo individually yields notable improvements in CSI-M across all three datasets, and their joint formulation further improves CSI-M by up to 7.9% on Shanghai over transport alone. Comparing unfactorized evolution with ZFEvo highlights the necessity of decomposition: while both achieve comparable CSI-M, ZFEvo substantially improves structural and perceptual quality, reducing LPIPS by up to 28.9% on Shanghai (from 0.2594 to 0.1843) and raising SSIM from 0.7694 to 0.7964. This demonstrates that decomposing predictions into spatial coverage and intensity factors effectively prevents optimization bias caused by zero-inflated non-raining regions.

Effect of supervision balance. Fig. 4 investigates the sensitivity of TFEvo to the supervision weight $\lambda \in [0, 1]$, which balances intermediate transport supervision against final forecast supervision. As λ increases from 0 to 1, LPIPS improves monotonically by 27.6% (from 0.210 to 0.152), confirming that direct supervision on the transport prior yields physically realistic textures. However, omitting final field supervision ($\lambda = 1.0$) causes CSI-M to drop by 11.1% (from 0.434 to 0.386), because the model reduces to pure advection and fails to capture localized convective growth and decay. An equilibrium emerges across $\lambda \in [0.4, 0.6]$, where CSI-M peaks while LPIPS remains low, showing that dual-level supervision balances advective realism against extreme-rainfall sensitivity.

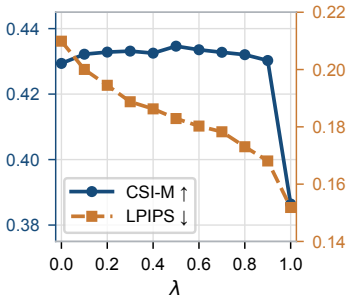


Figure 4: **Sensitivity to the supervision balance λ.** CSI-M and LPIPS of TFEvo on Shanghai as λ varies in $[0, 1]$.

Effect of motion components. Fig. 3 evaluates the three motion components in ASTrans by removing motion state propagation (*w/o Propagation*), global displacement (*w/o Global*), and local deformation (*w/o Local*). Removing local deformation causes the most severe performance drop, reducing CSI-M by up to 8.3% and CSI-High by up to 16.4% across datasets. This demonstrates that atmospheric transport involves strong localized deformations that cannot be modeled by uniform displacement alone. Removing global displacement also decreases CSI-M by up to 1.5% on MeteoNet and Shanghai, confirming its utility in capturing background motion. In contrast, removing motion state propagation results in marginal performance changes (under 1.0%), indicating that local deformation is the primary contributor to motion estimation.

5.4 Qualitative Analysis

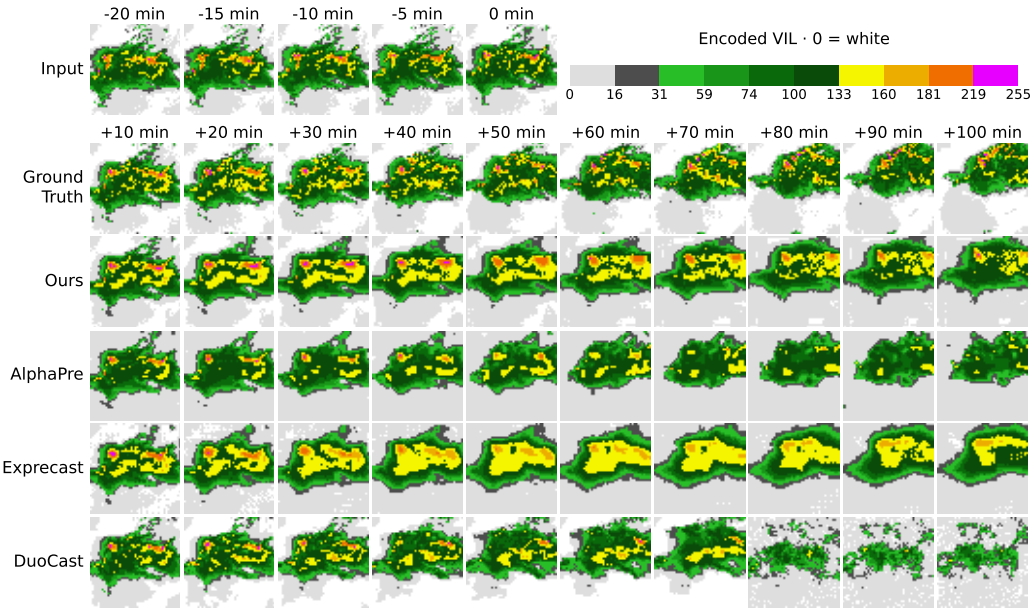


Figure 5: **Qualitative forecast comparison on a SEVIR convective event.** Observations, ground truth, and forecasts are displayed across a 100-minute horizon ($+10$ to $+100$ min) under identical VIL color mapping.

Long-horizon storm dynamics. Fig. 5 compares multi-lead predictions on a severe convective event from SEVIR across a 100-minute horizon. The target event features intense convective cores (VIL ≥ 219) embedded within an advecting frontal rainband. Across extended lead times ($+60$ to $+100$ min), baseline models exhibit distinct failure modes: AlphaPre progressively dissipates severe rainfall cores due to mean-field regression penalties, exPreCast over-smooths internal gradients into a diffused mass, and DuoCast suffers from spatial fragmentation and breaks the convective band unnaturally. In contrast, TFEvo consistently preserves the macroscopic continuity of the convective band while maintaining compact, high-intensity reflectivity cores up to 100 minutes. This demonstrates that directly anchoring transport to the latest observed radar field avoids recursive blur and error accumulation over long horizons.

Visualization of transport and residual evolution. Fig. 6 visualizes the internal states of TFEvo on a Shanghai event at leads 1, 10, and 20 by decomposing the evolved prediction $\widehat{\boldsymbol{X}}_t$ into the transported prediction $\boldsymbol{B}_t$ and the net innovation field $\widehat{\boldsymbol{X}}_t - \boldsymbol{B}_t$. ASTrans shifts and deforms the observed echo $\boldsymbol{X}_0$ to form the transported prediction $\boldsymbol{B}_t$. Because the target precipitation system undergoes natural dissipation at later leads, pure advection progressively overestimates intensities at leads 10 and 20. ZFEvo addresses this discrepancy by synthesizing spatially coherent negative innovations (red regions), which selectively attenuate the transported prediction to match the decaying ground truth.



Figure 6: **Visualization of transport and residual evolution on Shanghai.** Evolution is visualized at leads 1, 10, and 20. Radar fields share a normalized intensity scale $[0, 0.5]$, while the net innovation field $\widehat{\boldsymbol{X}}_t - \boldsymbol{B}_t$ uses a zero-centered signed scale $[-0.3, 0.3]$.

6 Conclusion, Limitations and Future Work

We presented TFEvo, an end-to-end framework that breaks forecast-field feedback by transporting the latest observed radar field directly to each forecast horizon. Conditioned on the transported prediction, factorized coverage and intensity updates model residual evolution, while an asymmetric objective penalizes underestimation of extreme rainfall. Across three benchmarks, TFEvo outperforms competitive baselines on most forecasting metrics with 0.29M parameters and achieves further gains over exPreCast at high resolution.

While TFEvo demonstrates strong empirical skill, several limitations outline avenues for future research. 1) As a deterministic model, TFEvo outputs a single trajectory, leaving the chaotic stochasticity of convective initiation unquantified. Extending our decoupled formulation to probabilistic forecasting via residual diffusion or flow matching represents a natural next step. 2) Continuous coordinate deformation across dense high-resolution grids incurs increased floating-point computation despite our compact parameter footprint. Future research could explore multi-scale motion representations or more efficient predictive backbones to further reduce computational overhead. 3) Anchoring solely to the latest observation assumes temporal motion coherence, which faces physical boundaries during abrupt topological transitions such as explosive squall-line genesis. Integrating multi-modal satellite data and 3D convective dynamics will be vital to strengthen physical robustness across broader meteorological regimes.

AI use statement

In this work, we used ChatGPT and Codex to assist with manuscript drafting and editing, literature search, code implementation and debugging, and visualization scripts. We have not used generative AI tools to generate the research ideas or technical contributions of this work, to produce experimental results or data, or to draft references or citations. We have reviewed all AI-assisted work: language produced with AI assistance was checked and revised by the authors, and AI-assisted code was verified and tested for correctness by the authors before it was used to produce the results reported in this paper. The authors take full responsibility for the content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

This work studies precipitation nowcasting using existing radar datasets. Improved forecasts may support weather-related decision-making, but missed severe events and false alarms can have adverse consequences. The reported benchmark results do not establish readiness for operational deploy-

ment. Use in warning systems would require further validation under local conditions, assessment of predictive uncertainty, and oversight by qualified meteorological practitioners.

Reproducibility statement

This statement summarizes where the information required to reproduce our results can be found. The model architecture, training objectives, and factorized inference procedure are described in Section 4 and Appendix A. Dataset protocols, preprocessing, evaluation thresholds, and forecast spans for SEVIR, MeteoNet, and Shanghai are specified in Appendix B. Optimization details are documented in Appendix B. All local baselines are re-trained with their official implementations under the shared experimental protocol described in Section 5, and extended per-dataset and high-resolution results are reported in Appendix C. Finally, We will release our code to facilitate reproducibility at the anonymous repository linked in the abstract.

References

Georgy Ayzel, Maik Heistermann, and Tanja Winterrath. Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1). *Geoscientific Model Development*, 12(4):1387–1402, 2019. doi: 10.5194/gmd-12-1387-2019.

Guangyao Chen, Yuan Cao, Leiming Ma, and Junping Zhang. A deep learning-based methodology for precipitation nowcasting with radar. *Earth and Space Science*, 7(2):e2019EA000812, 2020. doi: 10.1029/2019EA000812.

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(1):5145, 2022.

Loris Foresti, Ioannis V. Sideris, Daniele Nerini, Lea Beusch, and Urs Germann. Using a 10-year radar archive for nowcasting precipitation growth and decay: A probabilistic machine learning approach. *Weather and Forecasting*, 34(5):1547–1569, 2019. doi: 10.1175/WAF-D-18-0206.1.

Wentao Gao, Xiongren Chen, Xiaojing Du, Wenjun Yu, Andres Mauricio Cifuentes Bernal, and Ziqi Xu. Deep extreme transformer: Tackling zero-inflated time series for precipitation prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(45):38478–38486, 2026a. doi: 10.1609/aaai.v40i45.41189. URL https://ojs.aaai.org/index.php/AAAI/article/view/41189.

Wentao Gao, Xiaojing Du, Xiongren Chen, Yifan Guo, Andres Mauricio Cifuentes-Bernal, Renqiang Luo, and Ziqi Xu. Energy-efficient training-free zero-inflation correction for rainfall forecasting with time-series foundation models. In *Proceedings of the ACM Web Conference 2026*, pp. 8840–8850. Association for Computing Machinery, 2026b. doi: 10.1145/3774904.3792998. URL https://doi.org/10.1145/3774904.3792998.

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*, pp. 3170–3180, 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*, volume 35, 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*, volume 36, 2023.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, and koray kavukcuoglu. Spatial transformer networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Roger Koenker and Kevin F Hallock. Quantile regression. *Journal of economic perspectives*, 15(4): 143–156, 2001.

Gwennaëlle Larvor, Léa Berthomier, Vincent Chabot, Brice Le Pape, Bruno Pradel, and Lior Perez. MeteoNet, an open reference weather dataset by Météo-France. Météo-France, 2020.

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.

Kenghong Lin, Baoquan Zhang, Demin Yu, Wenzhi Feng, Shidong Chen, Feifan Gao, Xutao Li, and Yunming Ye. AlphaPre: Amplitude-phase disentanglement model for precipitation nowcasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17841–17850, 2025.

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.

Jiawei Ren, Mingyuan Zhang, Cunjun Yu, and Ziwei Liu. Balanced MSE for Imbalanced Visual Regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Ren_Balanced_MSE_for_Imbalanced_Visual_Regression_CVPR_2022_paper.html.

Michael Scheuerer and Thomas M. Hamill. Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Monthly Weather Review*, 143(11): 4578–4596, 2015. doi: 10.1175/MWR-D-15-0061.1. URL https://journals.ametsoc.org/abstract/journals/mwre/143/11/mwr-d-15-0061.1.xml.

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, 2015.

Changhoon Song, Teng Yuan Chang, and Youngjoon Hong. Extreme Weather Nowcasting via Local Precipitation Pattern Prediction. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=fDknsQhSgm.

Mark Veillette, Siddharth Samsi, and Chris Mattioli. SEVIR: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. In *Advances in Neural Information Processing Systems*, volume 33, pp. 22009–22019, 2020.

Penghui Wen, Mengwei He, Patrick Filippi, Na Zhao, Feng Zhang, Thomas Francis Bishop, Zhiyong Wang, and Kun Hu. DuoCast: Duo-probabilistic diffusion for precipitation nowcasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 39442–39450, 2026. doi: 10.1609/aaai.v40i46.41294.

Tyler Wilson, Andrew McDonald, Asadullah Hill Galib, Pang-Ning Tan, and Lifeng Luo. Beyond point prediction: Capturing zero-inflated and heavy-tailed spatiotemporal data with deep extreme mixture models. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2020–2028, 2022. doi: 10.1145/3534678.3539464.

Binqing Wu, Weiqi Chen, Shiyu Liu, Zongjiang Shang, Haiou Wang, Liang Sun, and Ling Chen. MoCast: Learning turbulent motions under physical guidance for precipitation nowcasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15950–15958, 2026. doi: 10.1609/aaai.v40i19.38628.

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*, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/b54532b0e57eb963b19e00583376cda3-Abstract-Conference.html.

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. doi: 10.1038/s41586-023-06184-4.

Yiqi Zhong, Luming Liang, Ilya Zharkov, and Ulrich Neumann. MMVP: Motion-matrix-based video prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4273–4283, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/papers/Zhong_MMVP_Motion-Matrix-Based_Video_Prediction_ICCV_2023_paper.pdf.

A Implementation and Evaluation Details

This appendix specifies the network, optimization, and evaluation used in the experiments.

A.1 Notation and tensor conventions

Symbol	Meaning
L, T	Observation history length and forecast length.
$\boldsymbol{X}_t$, $\widehat{\boldsymbol{X}}_t$	True and predicted fields in $[0,1]^{H\times W}$.
$\boldsymbol{C}_t$, $\boldsymbol{h}_t$	Lead context and recurrent motion state.
$\delta\boldsymbol{d}_t^{\mathrm{global}}$, $\delta\boldsymbol{d}_t^{\mathrm{local}}$	Global translation and local displacement increments with zero spatial mean.
$\delta\boldsymbol{d}_t$, $\boldsymbol{d}_t$	Incremental and cumulative displacement fields; horizontal and vertical sampling offsets at each location.
$\mathcal{W}$	Backward sampling operator with bilinear interpolation.
$\boldsymbol{B}_t$	Transport prior obtained by sampling $\boldsymbol{X}_0$.
$\boldsymbol{S}_t$, $\boldsymbol{A}_t$	Initial coverage and intensity factors.
$\widehat{\boldsymbol{S}}_t$, $\widehat{\boldsymbol{A}}_t$	Updated factors whose product is $\widehat{\boldsymbol{X}}_t$.
$\boldsymbol{r}_t^S$, $\boldsymbol{r}_t^A$	Coverage and intensity residuals.
$\boldsymbol{g}_t^{\mathrm{mot}}$	Motion update gate.
κ, γ	Small constant for continuous coverage and maximum intensity adjustment.
u, q	Target split and upper-regime asymmetry in the field loss.
λ, η	Supervision balance and common regularization weight.

Bold symbols denote grid arrays; scalar subscripts i index their entries. A superscript c denotes the coarse motion grid, and f the field grid. Layer-specific symbols and symbols used only in proofs are defined locally.

A.2 History and motion state

The history encoder has width 32: a 3D stem is followed by two residual blocks, each containing two $3\times3\times3$ convolutions with eight-group GroupNorm and SiLU. Averaging the encoded history over its L time steps gives a spatial representation. Spatial and lead projections combine this representation with learned 32-dimensional embeddings $\boldsymbol{e}_t$ to produce $\boldsymbol{C}_t$. The spatial projection is a 3×3 convolution followed by GroupNorm and SiLU; a 1×1 projection is applied after adding the lead embedding.

For motion estimation, $\boldsymbol{C}_t$ is resized to quarter resolution and concatenated with a separate lead embedding $\boldsymbol{\ell}_t$:

$$\boldsymbol{q}_t = \mathcal{F}_{\mathrm{in}}([\downarrow_4 \boldsymbol{C}_t, \boldsymbol{\ell}_t]), \qquad \widetilde{\boldsymbol{h}}_t = \tanh \mathcal{F}_{\mathrm{cand}}(\boldsymbol{q}_t), \qquad \boldsymbol{g}_t^{\mathrm{mot}} = \sigma\mathcal{F}_{\mathrm{gate}}(\boldsymbol{q}_t). \tag{12}$$

The input projection is a 1×1 convolution to 64 channels, followed by GroupNorm and SiLU. The candidate and gate use separate 3×3 convolutions. The 64-channel state has a learned decay per channel c,

$$\alpha_c = \sigma(-\omega_c), \qquad \boldsymbol{h}_t^{(c)} = \alpha_c \boldsymbol{h}_{t-1}^{(c)} + (1-\alpha_c)\boldsymbol{g}_t^{\mathrm{mot},(c)} \odot \widetilde{\boldsymbol{h}}_t^{(c)}, \quad \boldsymbol{h}_0 = \boldsymbol{0}. \tag{13}$$

The decay parameters are initialized to zero, giving $\alpha_c = 1/2$.

A.3 Global and local motion, composition, and sampling

A 3×3 convolution, eight-group GroupNorm, and SiLU project $\boldsymbol{h}_t$ to width-32 features $\boldsymbol{U}_t$. On the coarse motion grid, the two displacement components are

$$\begin{aligned} \delta\boldsymbol{d}_t^{\mathrm{global},c} &= \frac{b_g}{s}\tanh(W_g\,\mathrm{GAP}(\boldsymbol{U}_t)), \\ \boldsymbol{V}_t &= \tanh(W_l * \boldsymbol{U}_t), \\ \delta\boldsymbol{d}_t^{\mathrm{local},c} &= \frac{b_l}{s}\,\mathrm{Bound}_1(\boldsymbol{V}_t - \mathrm{mean}_{\boldsymbol{x}}\,\boldsymbol{V}_t). \end{aligned} \tag{14}$$

The global head is linear and the local head is a 3×3 convolution. Their weights are initialized from a zero-mean normal distribution with standard deviation 10^{-3}; their biases are zero. Here $s = 4$, $b_g = 4$ and $b_l = 1$; the bounds are specified in field-grid pixels. Bound_1 divides each component by the larger of one and its spatial maximum absolute value, preserving its zero spatial mean. The global translation is broadcast over the motion grid before adding the local component to form $\delta \boldsymbol{d}_t^c$. Composition precedes conversion to the field grid:

$$\begin{aligned} \boldsymbol{d}_t^c &= \delta \boldsymbol{d}_t^c + \mathcal{W}(\boldsymbol{d}_{t-1}^c, \delta \boldsymbol{d}_t^c), \qquad \boldsymbol{d}_0^c = \boldsymbol{0}, \\ \boldsymbol{d}_t^f &= \mathrm{ResizeFlow}_s(\boldsymbol{d}_t^c), \qquad \delta \boldsymbol{d}_t^f = \mathrm{ResizeFlow}_s(\delta \boldsymbol{d}_t^c), \\ \boldsymbol{B}_t &= \mathcal{W}(\boldsymbol{X}_0, \boldsymbol{d}_t^f). \end{aligned} \tag{15}$$

The first line corresponds to Eq. 7 in the main text. $\mathrm{ResizeFlow}_s$ upsamples bilinearly and multiplies both components by s. Displacement sampling and field sampling use bilinear sampling with zero padding. The field-grid increments $\delta \boldsymbol{d}_t^f$ are used in the regularizers below.

A.4 Factorized evolution and gradient paths

Eq. 8 uses $\kappa = 0.01$ and gives $(\boldsymbol{S}_t, \boldsymbol{A}_t) = (\boldsymbol{0}, \kappa)$ at a zero prior. The conditioner concatenates the 32-channel context, the resized 64-channel motion state, and two detached base factors (one channel each):

$$\boldsymbol{Q}_t = [\boldsymbol{C}_t, \uparrow \boldsymbol{h}_t, \mathrm{sg}(\boldsymbol{S}_t), \mathrm{sg}(\boldsymbol{A}_t)] \in \mathbb{R}^{98 \times H \times W}. \tag{16}$$

A width-32 2D stem and two residual blocks feed separate 1×1 output convolutions for $\boldsymbol{r}_t^S$ and $\boldsymbol{r}_t^A$. Both output layers are zero-initialized, and the same head is shared across leads. Eq. 9 uses $\gamma = 0.5$. The coverage gate uses clip, whose implemented derivative at the lower clipping boundary is one. The initial outputs reproduce $\boldsymbol{B}_t$ exactly.

Stop-gradient removes the conditioner-side paths through the initial factors. The reconstruction retains the gradient path from the final-field loss through the updated factors, their initial values, the transport prior, and the motion module. The evolution head sees only the current lead's conditioner; historical and preceding-lead information enter through $\boldsymbol{C}_t$ and $\boldsymbol{h}_t$.

A.5 Field objectives and regularization

The two field losses are defined in Section 4.3. The target split is $u = 0.2$, the asymmetry is $q = 10/11$, the supervision balance is $\lambda = 0.5$, and the common regularization weight is $\eta = 0.01$. For the same magnitude of error in the upper target regime, underestimation has $q/(1-q)$ times the slope of overestimation.

Define the mean absolute spatial first difference as TV_{xy}, and temporal differences by $D_t Z_t = Z_t - Z_{t-1}$ and $D_t^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2}$. The following means average over all valid batch, lead, component, and spatial entries. Spatial TV is the sum of the mean absolute horizontal and vertical differences. The regularization separates into a displacement-side group and a factor-side group. The displacement side acts on the field-grid increments $\delta \boldsymbol{d}_t^f$ and forms $\mathcal{R}_{\mathrm{motion}}$:

$$\begin{aligned} \mathcal{R}_{\mathrm{smooth}} &= \mathrm{TV}_{xy}(\delta \boldsymbol{d}^f), \\ \mathcal{R}_{\mathrm{accel}} &= \mathrm{mean}\,|D_t \delta \boldsymbol{d}_t^f| + \mathrm{mean}\,|D_t^2 \delta \boldsymbol{d}_t^f|^2, \end{aligned} \tag{17}$$

so that $\mathcal{R}_{\mathrm{motion}} = \mathcal{R}_{\mathrm{smooth}} + \mathcal{R}_{\mathrm{accel}}$. Temporal first and second differences use $t \geq 2$ and $t \geq 3$, respectively; this objective is defined for $T \geq 3$. The factor side acts on the effective factor changes $\boldsymbol{\Delta}_t = [\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t, \widehat{\boldsymbol{A}}_t - \boldsymbol{A}_t]$ and forms $\mathcal{R}_{\mathrm{evolution}}$:

$$\mathcal{R}_{\mathrm{evolution}} = \mathrm{mean}\,|\boldsymbol{\Delta}| + \mathrm{mean}\,|D_t \boldsymbol{\Delta}_t|. \tag{18}$$

Because $\boldsymbol{\Delta}_t$ aggregates the two update channels whose product expresses the growth and dissipation of rain cells, the evolution regularizer explicitly guides the growth–dissipation dynamics that pure advection cannot represent. The mean includes both correction channels with equal scale, and all terms in Eqs. 17 and 18 carry the common weight η in Eq. 10.

Inference pipeline. Algorithm 1 summarizes the complete inference procedure.

Algorithm 1 Inference Pipeline of TFEvo

Require: Historical radar sequence $\boldsymbol{X}_{-L+1:0}$, lead steps T.
Ensure: Future precipitation forecasts $\widehat{\boldsymbol{X}}_{1:T}$.

```
Extract spatiotemporal contexts: C_{1:T} ← F_enc(X_{-L+1:0}, e_{1:T})                ▷ Eq. 4
Initialize motion state h_0 ← 0, cumulative displacement d_0 ← 0
for t = 1 to T do
    ▷ ASTrans: Anchor-based Shared Transport
    Update recurrent state h_t via channel decay α                                   ▷ Eq. 5
    Predict displacement increment δd_t via motion head F_mot(h_t)                   ▷ Eq. 6
    Accumulate displacement: d_t ← δd_t + W(d_{t-1}, δd_t)                           ▷ Eq. 7
    Construct transported prediction: B_t ← W(X_0, d_t)
    ▷ ZFEvo: Zero-inflation-aware Factorized Evolution
    Factorize prior: A_t ← max(B_t, κ),   S_t ← B_t / A_t                            ▷ Eq. 8
    Predict residual controls (r_t^A, r_t^S) ← F_evo(C_t, ↑h_t, sg(S_t), sg(A_t))
    Compute refined factors Â_t, Ŝ_t and reconstruct: X̂_t ← Ŝ_t ⊙ Â_t               ▷ Eq. 9
end for
return X̂_{1:T}
```

B Additional Analyses

B.1 Data and preprocessing

We use $L = 5$ observations to predict $T = 20$ frames at 128×128. The frame interval is 5 minutes for SEVIR and MeteoNet and 6 minutes for Shanghai. Table 5 lists the forecast spans and evaluation thresholds.

Table 5: Dataset interfaces. Thresholds use the dataset-specific evaluation scale.

Dataset	Measurement	Forecast span	Thresholds
SEVIR	VIL	100 min	16, 74, 133, 160, 181, 219
MeteoNet	Reflectivity	100 min	12, 18, 24, 32
Shanghai	Radar echo	120 min	20, 30, 35, 40

B.2 Optimization and model selection

We train TFEvo with seed 42 and AdamW in single precision. The learning rate is fixed at 10^{-4}, with weight decay 10^{-4} and moments $(0.9, 0.95)$. Training runs for up to 150 epochs, with evaluation after each epoch and early stopping after 20 epochs without improvement in CSI-M. Exponential moving average weights use decay 0.99, and the checkpoint with the highest CSI-M is selected for evaluation. Architecture and objective constants are shared across datasets (Table 6).

The transported prediction receives MSE supervision, and the reconstructed forecast receives the body–tail loss; the latter supervises both residual controls. The supervision-balance study varies λ while keeping these loss definitions fixed.

Table 6: Shared architecture, objective, and optimizer settings.

Setting	Value
History/context/evolution width	32
Motion-state width; motion-grid reduction	64; $s = 4$
Global/local increment bound	$b_g = 4$, $b_l = 1$ field pixels
Coverage floor; intensity adjustment	$\kappa = 0.01$, $\gamma = 0.5$
Loss balance; target split	$\lambda = 0.5$, $u = 0.2$
Tail asymmetry; regularization	$q = 10/11$, $\eta = 0.01$
AdamW learning rate; weight decay	10^{-4}; 10^{-4}
AdamW moments; EMA decay	$(0.9, 0.95)$; 0.99

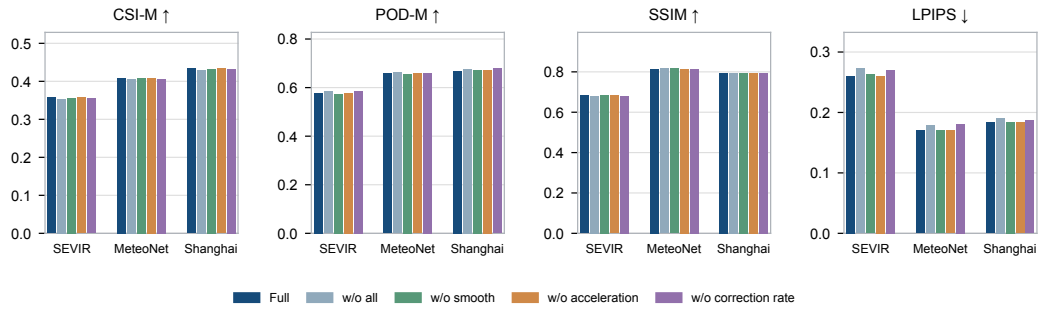


Figure 7: **Ablation of the regularization groups.** CSI-M, POD-M, SSIM, and LPIPS on SEVIR, MeteoNet, and Shanghai for the full model (*Full*), the model without any regularizer (*w/o all*), and models with the smoothness, acceleration, or correction group removed individually.

B.3 Regularization ablation

Fig. 7 reports CSI-M, POD-M, SSIM, and LPIPS when each regularizer group is removed on the three benchmarks. Removing the entire regularizer causes a clear drop in CSI-M and LPIPS, indicating a substantial overall effect of regularization. Among the groups, the correction regularizer shows the most consistent influence: by construction it shapes the factor updates that express growth and dissipation, the part of the forecast that pure advection cannot provide. The displacement-side groups (smoothness and acceleration) instead show more dataset-dependent behavior across the three benchmarks.

B.4 Performance across forecast leads

The main text reports aggregate metrics over the full $T = 20$ horizon; the per-lead curves below show how each advantage evolves with lead time. The static anchor repeats the latest observation unchanged at every lead and serves as the reference curve.

SEVIR. Figure 8 shows that TFEvo has the highest CSI-M from 20 minutes onward and the highest SSIM at the final 2 leads. The threshold-wise curves in Figure 9 show a faster loss of overlap at the highest thresholds than at the 2 lowest thresholds.

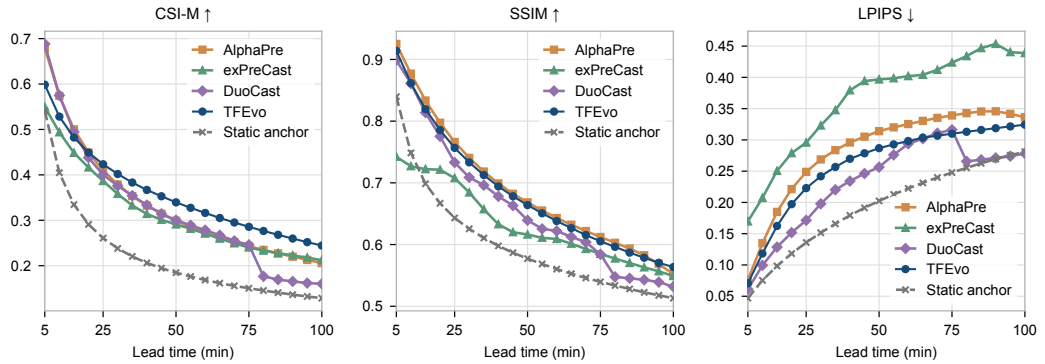


Figure 8: Forecast quality across lead times on SEVIR. Each panel identifies the methods shown.

MeteoNet. TFEvo exceeds the static anchor in CSI-M and SSIM at every lead, while the static anchor retains lower LPIPS (Figure 10). The largest relative CSI decline occurs at threshold 32 (Figure 11).

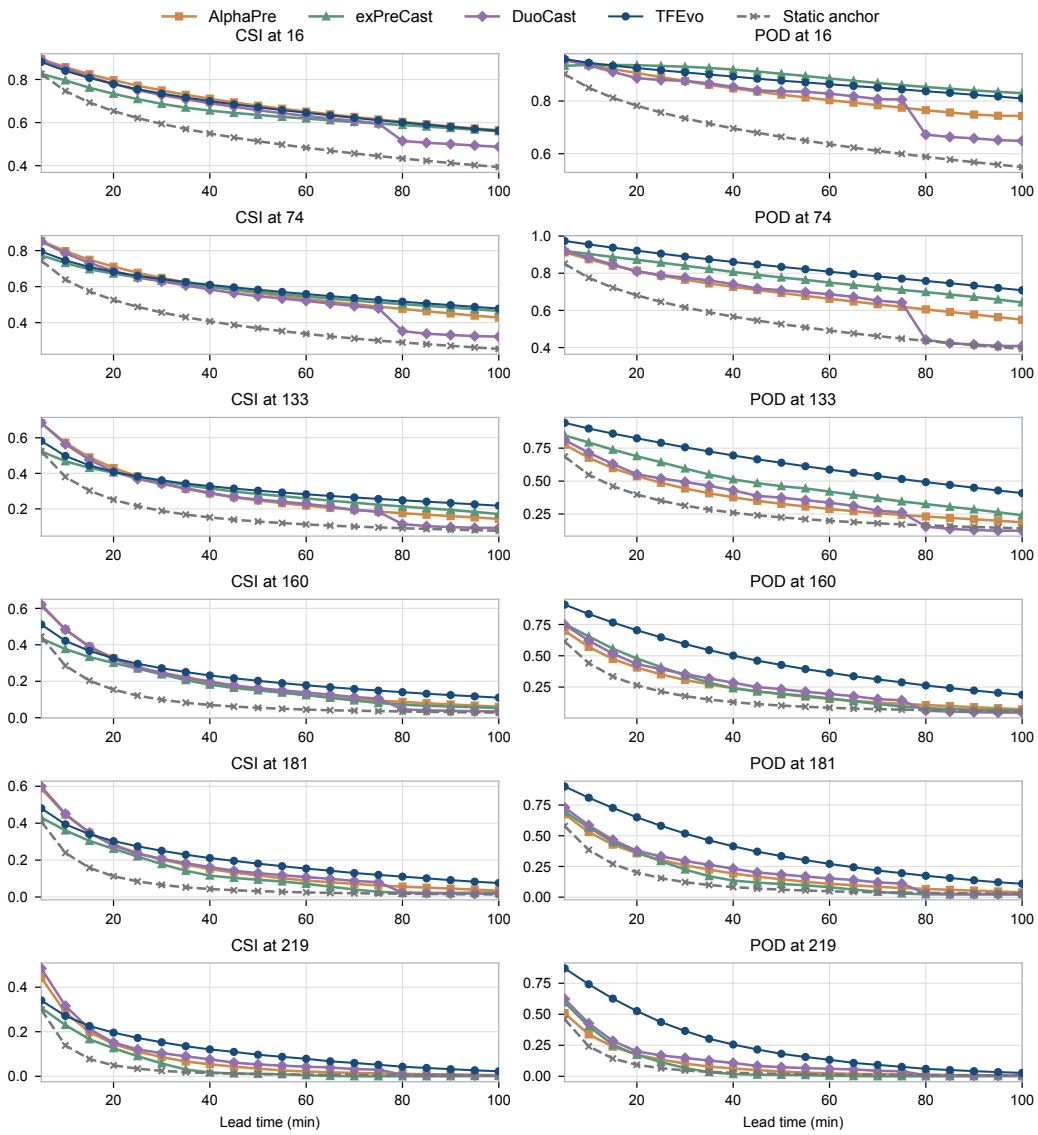


Figure 9: CSI and POD across forecast leads at each native threshold on SEVIR.

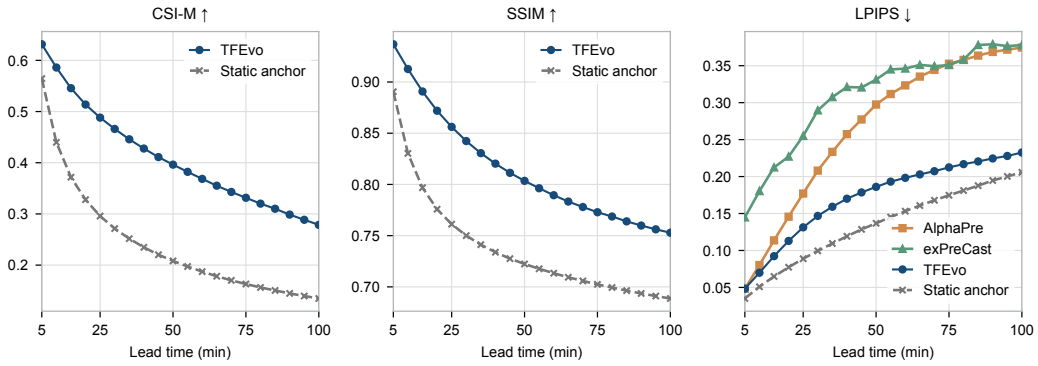


Figure 10: Forecast quality across lead times on MeteoNet. Each panel identifies the methods shown.

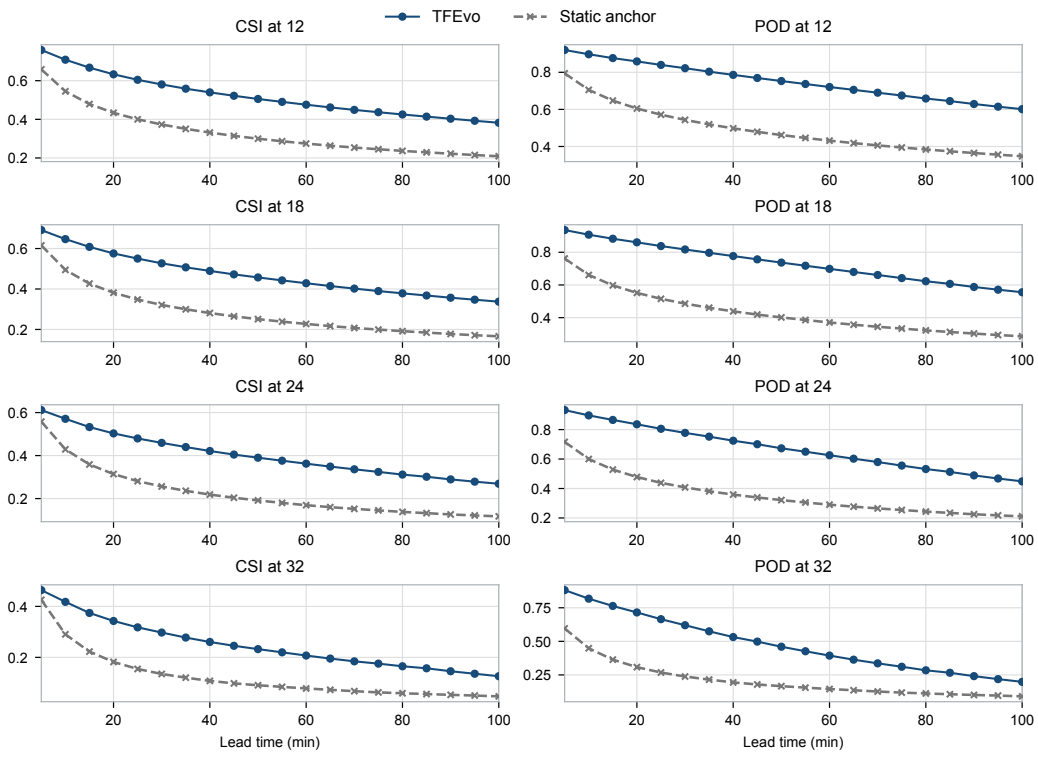


Figure 11: CSI and POD across forecast leads at each native threshold on MeteoNet.

Shanghai. TFEvo exceeds the static anchor in CSI-M and SSIM at all 20 leads and achieves lower LPIPS at the final 2 leads (Figure 12). The relative separation in CSI between the lowest and highest thresholds grows at late leads (Figure 13).

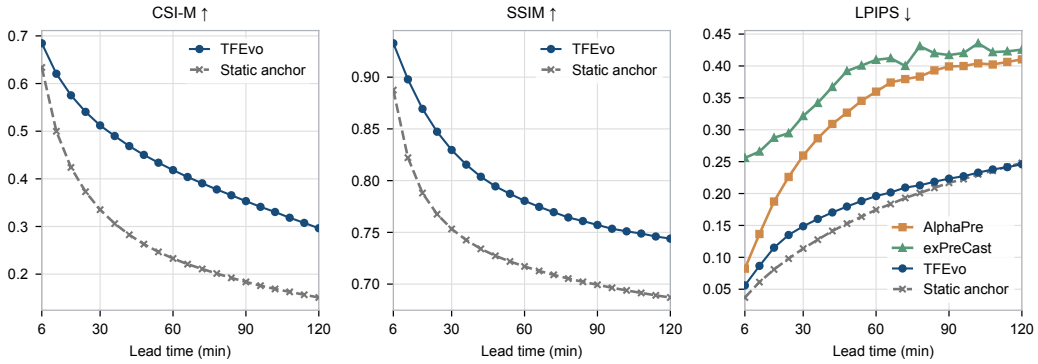


Figure 12: Forecast quality across lead times on Shanghai. Each panel identifies the methods shown.

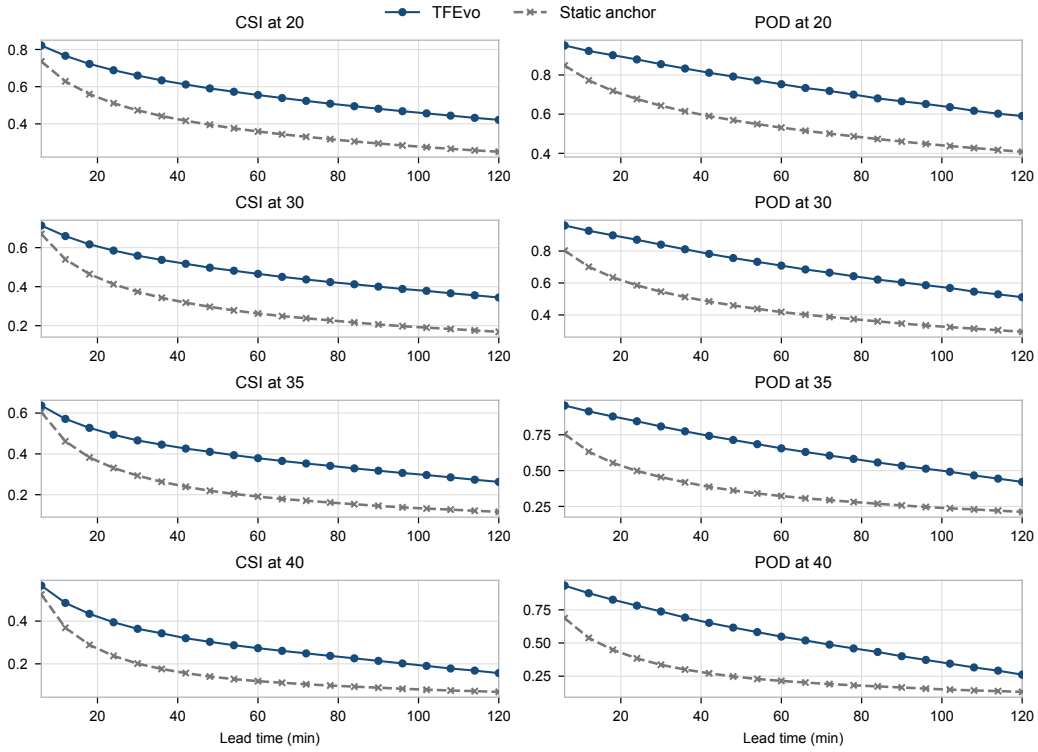


Figure 13: CSI and POD across forecast leads at each native threshold on Shanghai.

C Extended Results

This section expands the main-text results with complete per-dataset tables, per-lead high-resolution curves, and the remaining sensitivity studies. Metric definitions follow Section 5.1.

C.1 METEONET

Tables 7 and 8 provide the complete threshold comparison and additional aggregate metrics. TFEvo has lower MAE than both baselines, whereas AlphaPre has lower MSE.

Table 7: Threshold-wise detection performance on MeteoNet .

Model	12		18		24		32	
	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑
AlphaPre	0.5273	0.6677	0.4590	0.5532	0.3673	0.4378	0.2021	0.2569
exPreCast	0.4651	0.6519	0.4076	0.5549	0.3253	0.4627	0.1668	0.3065
TFEvo	0.5205	0.7501	0.4695	0.7326	0.4005	0.6714	0.2471	0.4783

Table 8: Additional detection and error metrics on MeteoNet.

Model	HSS-M↑	CSI-P4↑	MSE↓	MAE↓	RMSE↓
AlphaPre	0.5224	0.4014	12.7955	1.5812	3.5771
exPreCast	0.4744	0.3830	17.9510	1.6220	4.2369
TFEvo	0.5512	0.5185	17.5129	1.3053	4.1848

C.2 SHANGHAI

Tables 9 and 10 include threshold 35 and additional aggregate metrics. TFEvo improves HSS-M and CSI-P4 over both baselines, with lower MAE but higher MSE than AlphaPre.

Table 9: Threshold-wise detection performance on Shanghai.

Model	20		30		35		40	
	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑
AlphaPre	0.5615	0.6537	0.4728	0.5753	0.3877	0.5041	0.2771	0.3770
exPreCast	0.5216	0.6826	0.4290	0.6234	0.3350	0.5810	0.2211	0.4931
TFEvo	0.5696	0.7534	0.4795	0.7122	0.3941	0.6607	0.2926	0.5565

Table 10: Additional detection and error metrics on Shanghai.

Model	HSS-M↑	CSI-P4↑	MSE↓	MAE↓	RMSE↓
AlphaPre	0.5615	0.4486	28.0300	2.3475	5.2943
exPreCast	0.5148	0.4427	42.4166	2.6379	6.5128
TFEvo	0.5749	0.5307	37.5187	1.8464	6.1253

C.3 SEVIR

Tables 11 and 12 extend the comparison to all native thresholds and additional metrics. TFEvo achieves the highest HSS-M and CSI-P4, whereas AlphaPre attains lower MSE.

C.4 HIGH-RESOLUTION FORECASTING ACROSS LEADS

We extend the comparison in Table 3 to all $T = 20$ leads at 384×384, using $L = 5$ observations and the same ordered test samples and evaluation implementation. Figure 14 shows higher CSI-M and SSIM and lower LPIPS for TFEvo than exPreCast throughout both forecast windows.

Table 11: Threshold-wise detection performance on SEVIR.

Model	16		74		133		160		181		219	
	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑	CSI↑	POD↑
AlphaPre	0.6908	0.8292	0.5825	0.6999	0.2926	0.3698	0.2001	0.2443	0.1658	0.2018	0.0800	0.0936
exPreCast	0.6521	0.8929	0.5808	0.7761	0.2998	0.4854	0.1755	0.2586	0.1301	0.1779	0.0533	0.0849
DuoCast	0.6570	0.8117	0.5420	0.6704	0.2767	0.3792	0.1963	0.2646	0.1664	0.2242	0.0945	0.1271
TFEvo	0.6821	0.8775	0.5937	0.8322	0.3249	0.6432	0.2281	0.4600	0.2018	0.3847	0.1163	0.2651

Table 12: Additional detection and error metrics on SEVIR.

Model	HSS-M↑	CSI-P4↑	MSE↓	MAE↓	RMSE↓
AlphaPre	0.4304	0.3619	424.8598	8.6633	20.6121
exPreCast	0.4058	0.3547	453.8376	10.1251	21.3035
DuoCast	0.4155	0.3743	520.0758	9.7784	22.8052
TFEvo	0.4657	0.4413	500.8814	9.9564	22.3804

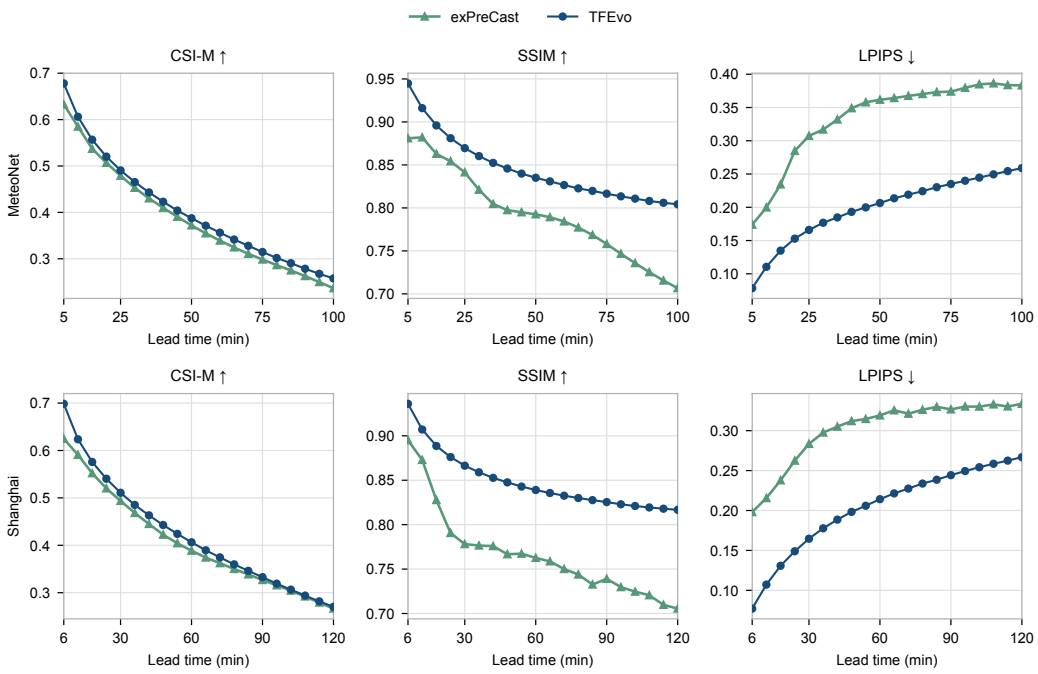


Figure 14: Forecast quality across lead times at 384×384. Rows correspond to the datasets indicated on the left.

Computation convention. GFLOPs in Table 3 are measured for a complete $L = 5$, $T = 20$ forecast with batch size 1 and FP32. Only supported convolution and matrix operations are counted, with one multiply-add counted as 2 operations. Grid sampling, interpolation, normalization, and activations are excluded.

C.5 SUPERVISION BALANCE

We vary λ to examine the balance between supervision on transported predictions and on the final field. Figure 4 visualizes the results, and Table 13 provides the corresponding detailed results for the Shanghai dataset. At $\lambda = 0$, the transported prediction receives no direct data loss; at $\lambda = 1$, the final forecast receives no direct data loss. Shared regularization remains active at both endpoints. Increasing λ lowers LPIPS, while the interior weights give similar CSI-M. The $\lambda = 1$ endpoint attains the lowest LPIPS but reduces CSI-M and SSIM, whereas the $\lambda = 0$ endpoint gives the highest LPIPS. Joint supervision retains both learning signals and balances perceptual similarity with event overlap.

Table 13: Sensitivity to the supervision weight λ on Shanghai.

	Detection				Image metrics				
Model	CSI-M↑	POD-M↑	HSS-M↑	CSI-P4↑	MSE↓	MAE↓	RMSE↓	SSIM↑	LPIPS↓
$\lambda = 0.0$	0.4294	0.6740	0.5700	0.5305	37.9543	1.9409	6.1607	0.7907	0.2099
$\lambda = 0.1$	0.4322	0.6749	0.5732	0.5330	37.6042	1.9032	6.1322	0.7935	0.2000
$\lambda = 0.2$	0.4328	0.6772	0.5740	0.5339	37.7242	1.8882	6.1420	0.7943	0.1945
$\lambda = 0.3$	0.4331	0.6745	0.5743	0.5326	37.6713	1.8746	6.1377	0.7949	0.1888
$\lambda = 0.4$	0.4325	0.6707	0.5735	0.5311	37.7744	1.8705	6.1461	0.7952	0.1863
$\lambda = 0.5$	0.4346	0.6710	0.5758	0.5317	37.5022	1.8521	6.1239	0.7960	0.1829
$\lambda = 0.6$	0.4335	0.6697	0.5748	0.5304	37.7976	1.8486	6.1480	0.7955	0.1803
$\lambda = 0.7$	0.4328	0.6666	0.5737	0.5293	37.9593	1.8399	6.1611	0.7956	0.1783
$\lambda = 0.8$	0.4320	0.6629	0.5730	0.5291	38.2056	1.8405	6.1811	0.7939	0.1731
$\lambda = 0.9$	0.4303	0.6524	0.5709	0.5287	38.3619	1.8332	6.1937	0.7919	0.1681
$\lambda = 1.0$	0.3864	0.4951	0.5210	0.4865	38.7614	1.7928	6.2259	0.7876	0.1518