PROBABILISTIC SPARSE VOLUMETRIC TRANSPORT FOR 3D RADAR NOWCASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Probabilistic 3D radar nowcasting forecasts how storms may move, develop vertically, and change in intensity. At high resolution, current methods reconstruct radar volumes from compressed representations such as 3D Gaussians and autoencoder latents, which can degrade high-intensity echoes. Diffusion and flow-matching models also repeat iterative sampling for every ensemble member, making high-resolution ensembles costly for time-critical nowcasting. We introduce Probabilistic Sparse Volumetric Transport (PSVT), which transports observed echoes along a small set of K 3D transport paths and estimates their growth and decay. Each path carries the full-resolution observed reflectivity through the native voxel space, so high-intensity echoes come directly from the observation rather than from reconstruction. The K path-wise predictions describe alternative evolutions of the same storm. Their deviations serve as K shared factors that define the spatiotemporal dependence of echo variations. PSVT models the local uncertainty of reflectivity at each voxel with a Gaussian distribution. Each member takes its variations from one linear combination of the K factors. This couples echo variations across space and time while avoiding iterative sampling and repeated network evaluations. On NEXRAD-3Y, PSVT improves single-forecast CSI at 40 dBZ by 114.0% over FlowCast and generates 16-member ensembles about $30\times$ faster.

1 INTRODUCTION

Volumetric radar nowcasting forecasts future 3D reflectivity volumes from recent radar observations. Compared with 2D radar sequences at specific altitudes, these volumes retain the complete echo structure across heights. Yet the same radar history can lead to multiple plausible futures that differ in how echoes move, grow, and decay. Probabilistic forecasting represents this uncertainty through an ensemble of complete future sequences. Such ensembles estimate the likelihood of high-impact events for severe convective weather warnings (Ravuri et al., 2021). At high resolution, probabilistic 3D nowcasting faces two key challenges.

The first challenge is predicting high-intensity echoes. Extreme precipitation involves convective-scale features at the kilometer scale, so its forecasts require high resolution (Gong et al., 2024). Extending 2D nowcasting models directly to 3D raises computational cost, which limits high-resolution scaling. To enable efficient high-resolution nowcasting, current methods forecast a compressed representation and reconstruct the full radar volume from it. STC-GS (Wang et al., 2025) renders the volume from groups of 3D Gaussians, and latent diffusion and flow-matching models decode it with an autoencoder (Gao et al., 2023; Perrone Ribeiro & Faganeli Pucer, 2026). However, autoencoder reconstruction skill declines at higher intensity thresholds (Gong et al., 2024), and deviations that barely change image appearance can still shift predicted precipitation (Schusterbauer et al., 2026). The high-intensity echoes most relevant to warnings are thus particularly difficult to reconstruct.

The second challenge is the cost of generating many complete futures. Nowcasting requires rapid updates as new radar volumes arrive every few minutes (Schusterbauer et al., 2026). Diffusion and flow-matching models evaluate the network for every ensemble member at each sampling step (Yu et al., 2024; Perrone Ribeiro & Faganeli Pucer, 2026). Each evaluation is expensive for high-resolution 3D radar volumes with multiple height levels. The total cost thus grows with the ensemble size and the number of sampling steps, which poses a challenge for time-critical nowcasting. Sam-

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.

pling each voxel independently from a local distribution is fast but breaks the correlated variations of a storm in space and time (Oskarsson et al., 2024; Zhdanov et al., 2026). Low-rank probabilistic forecasting models instead represent these correlations with a few shared factors, so each additional ensemble member requires only a linear combination of these factors (Salinas et al., 2019; Zheng & Sun, 2024). Fast probabilistic 3D nowcasting therefore requires shared factors that form coherent members from a single forward pass of the network.

To preserve high-intensity echoes, we introduce Probabilistic Sparse Volumetric Transport (PSVT), which forms future volumes by transporting observed echoes. Following motion-based video prediction, PSVT models the motion of echoes separately from their appearance (Hu et al., 2023; Zhong et al., 2023). Its backbone encodes all height levels of the radar history on a downsampled grid and predicts a small set of K 3D transport paths for several plausible motions. Each path carries the full-resolution observed reflectivity to future positions in the native voxel space (Jaderberg et al., 2015; Liu et al., 2017). Observed echoes therefore reach the forecast without Gaussian rendering and autoencoder decoding, so high-intensity echoes and fine-scale echo structures come directly from the observation at every height level. A shared reflectivity predictor then estimates local growth and decay (Zhang et al., 2023), so each path yields one complete echo evolution.

To generate ensembles rapidly, PSVT draws inspiration from copula models, which separate the dependency structure among variables from their marginal distributions (Drouin et al., 2022). Deviations of the K path-wise predictions serve as shared volumetric factors for the correlated variations of a storm. Each deviation corresponds to one alternative evolution of the same storm, so it varies coherently in space and time. These factors specify which variations occur jointly, while local Gaussian distributions determine how much reflectivity varies at each voxel and lead. Each member uses a single linear combination of these factors throughout the forecast, coupling its variations across lead times (Andrae et al., 2025). The factors and distribution parameters are computed in one forward pass. Additional members thus require only linear combinations and local mappings, without the repeated sampling of diffusion and flow-matching models.

Our contributions are threefold. First, PSVT predicts 3D echo motion, growth, and decay from downsampled history features and applies them directly to the full-resolution observed reflectivity, preserving high-intensity echoes. Second, we derive shared volumetric factors for the correlated variations of a storm from the path-wise predictions and pair them with local Gaussian distributions, so additional members require only linear combinations and local mappings. Third, on the NEXRAD-3Y dataset, PSVT improves single-forecast CSI at 40 dBZ by 114.0% over FlowCast and generates a 16-member ensemble about $30\times$ faster.

2 Related Work

Deterministic nowcasting. Deterministic nowcasting maps historical radar observations to a single future sequence by optimizing pointwise losses (Shi et al., 2015; Gao et al., 2022b). ConvLSTM (Shi et al., 2015) introduces convolutional recurrent transitions, and TrajGRU (Shi et al., 2017) learns location-dependent recurrent connections for spatially varying motion. PhyDNet (Le Guen & Thome, 2020) separates physical dynamics from complementary unknown factors. SimVP (Gao et al., 2022a) and TAU (Tan et al., 2023) predict all future frames in parallel with convolutional and attention-based modules, and Earthformer (Gao et al., 2022b) applies self-attention within space-time cuboids. Volumetric nowcasting further models radar structure across height levels. The 3D-ConvLSTM model of Sun et al. (2022) combines 3D convolutional feature extraction with recurrent prediction, and STC-GS (Wang et al., 2025) forecasts temporally coherent groups of 3D Gaussians with a memory-augmented state-space predictor.

Probabilistic nowcasting. Probabilistic 2D nowcasting represents multiple plausible radar futures while preserving realistic spatial structure and coherent temporal evolution within each forecast (Ravuri et al., 2021; Gao et al., 2023). DGMR (Ravuri et al., 2021) generates realistic precipitation sequences through conditional adversarial learning. NowcastNet (Zhang et al., 2023) combines a physics-based evolution network for advection and intensity change with a generative network that supplies fine-scale detail. PreDiff (Gao et al., 2023) performs latent diffusion and uses knowledge alignment to guide denoising toward domain-specific constraints. DiffCast (Yu et al., 2024) models global motion with a deterministic predictor and local stochastic variation with resid-

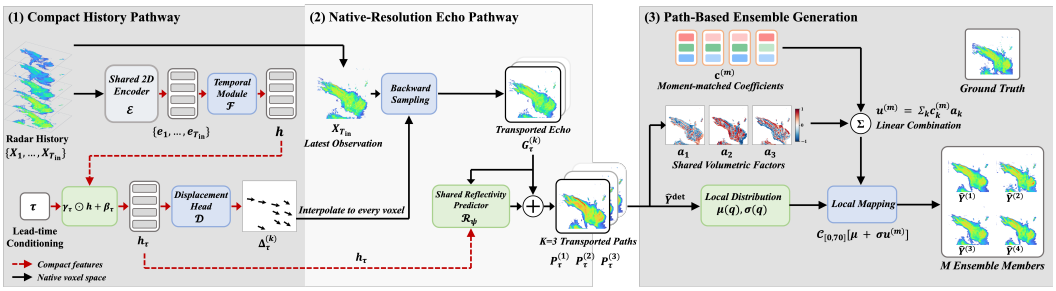

Figure 1: **Overview of PSVT.** The compact history pathway (1) predicts K displacement fields on a downsampled grid, the native-resolution echo pathway (2) transports the latest observation into K path-wise predictions, and path-based ensemble generation (3) forms M members from shared volumetric factors and local distributions. The example shows $K = 3$ transport paths and four of the M ensemble members.

ual diffusion. CasCast (Gong et al., 2024) cascades deterministic mesoscale prediction with latent diffusion for small-scale precipitation patterns. DuoCast (Wen et al., 2026) uses separate diffusion branches for low-frequency structure and high-frequency detail, and FlowCast (Perrone Ribeiro & Faganeli Pucer, 2026) learns a noise-to-data mapping in latent space through conditional flow matching.

3 METHOD

Given T_{in} observed radar volumes $X_{1:T_{\text{in}}}$, probabilistic 3D nowcasting models the conditional distribution of future reflectivity over the next T_{out} leads (Gao et al., 2023; Yu et al., 2024). Each observation $X_t \in \mathbb{R}^{Z\times H\times W}$ contains reflectivity in dBZ on Z height levels and an $H \times W$ horizontal grid. PSVT represents forecast uncertainty with M future sequences $\{\widehat{Y}_\tau^{(m)}\}_{\tau=1}^{T_{\text{out}}}$, $m = 1, \ldots, M$. Each member describes one possible echo evolution across the full radar volume and forecast horizon.

3.1 PSVT FRAMEWORK

Decoupling motion from appearance. PSVT models the motion of echoes separately from their appearance (Hu et al., 2023; Zhong et al., 2023). To maintain computational efficiency, its backbone encodes the radar history on a downsampled grid, while the latest observed volume $X_{T_{\text{in}}}$ stays at full radar resolution. Displacements predicted on the downsampled grid are interpolated to every voxel, and transport takes reflectivity from $X_{T_{\text{in}}}$ in the native voxel space (Jaderberg et al., 2015; Hu et al., 2023). Transported reflectivity thus comes from observed voxels rather than being reconstructed from compressed features, so high-intensity echoes and fine-scale echo structures remain on the native grid at every height level.

Precipitation evolution combines global moving trends with local growth and decay (Yu et al., 2024). The large-scale motion of precipitation is spatially coherent and highly predictable, so a downsampled grid retains the information needed to predict it (Gong et al., 2024). Local growth and decay are then predicted at full radar resolution from the history features and the transported volume (Sec. 3.2).

History encoding. A shared 2D encoder $\mathcal{E}$ takes each observed volume X_t with its Z height levels as input channels. It downsamples the horizontal grid and outputs features $e_t = \mathcal{E}(X_t)$ on an $H_f \times W_f$ grid. Treating height levels as channels allows the encoder to relate echoes across heights, while transport and reflectivity evolution are applied to the full 3D volume. A convolutional temporal module $\mathcal{F}$ maps the concatenated features $e_1, \ldots, e_{T_{\text{in}}}$ to a history feature h once per radar history. Its design follows the Temporal Attention Unit (TAU) (Tan et al., 2023). A small MLP predicts a channel-wise scale $\boldsymbol{\gamma}_\tau$ and offset $\boldsymbol{\beta}_\tau$ for each lead τ. The lead-conditioned history feature is $h_\tau = \boldsymbol{\gamma}_\tau \odot h + \boldsymbol{\beta}_\tau$.

Forecast pipeline. Figure 1 illustrates the complete forecast pipeline. With h_τ as input, sparse 3D transport and reflectivity evolution form the deterministic forecast and K path-wise predictions

(Sec. 3.2). A local Gaussian distribution determines how much reflectivity can vary at each voxel and lead (Sec. 3.3). Deviations of the K path-wise predictions from their weighted mean define K shared volumetric factors for the correlated variations of a storm (Sec. 3.4). PSVT combines these factors with the local Gaussian distributions to form M ensemble members (Sec. 3.5).

3.2 Sparse 3D Echo Transport

PSVT constructs K path-wise predictions of the full radar volume through learned 3D transport paths followed by reflectivity evolution. Echo motion varies across a storm and with height, so each path assigns a separate 3D displacement to every voxel and lead (Shi et al., 2017). The transport is sparse because PSVT predicts a small number K of paths, from which it later forms M ensemble members.

Sparse volumetric transport. For forecast lead τ, a displacement head $\mathcal{D}_{\theta_\Delta}$ maps h_τ to K displacement fields $\{\Delta_\tau^{(k)}\}_{k=1}^K$ on the downsampled grid. Interpolation extends them to every voxel, and their magnitude is bounded (Appendix C.1). The vector $\Delta_\tau^{(k)}(p) \in \mathbb{R}^3$ points from the source position in $X_{T_{\text{in}}}$ to the target voxel $p = (z, y, x)$ in grid units, and its vertical component shifts echoes across height levels. Because the same output channel of the head defines path k at every lead, the fields $\Delta_1^{(k)}, \ldots, \Delta_{T_{\text{out}}}^{(k)}$ form one transport path across forecast leads.

Transporting observed echoes. Each displacement field produces a full-resolution transported volume $G_\tau^{(k)}$ through differentiable backward sampling from $X_{T_{\text{in}}}$ (Jaderberg et al., 2015):

$$G_\tau^{(k)}(p) = X_{T_{\text{in}}}\left(p - \Delta_\tau^{(k)}(p)\right). \tag{1}$$

Trilinear interpolation evaluates reflectivity between grid points, so each transported value stays within the observed range of its neighboring voxels. Each lead takes reflectivity directly from $X_{T_{\text{in}}}$, so observed reflectivity passes through one differentiable interpolation per lead. NowcastNet instead advects fields step by step with nearest interpolation, since repeated bilinear interpolation blurs precipitation fields (Zhang et al., 2023). A path-weight head predicts weights $w_\tau^{(k)}(p)$ from h_τ, normalized by softmax over the K paths at each voxel. This gives the weighted transported volume $\bar{G}_\tau(p) = \sum_{k=1}^K w_\tau^{(k)}(p)\, G_\tau^{(k)}(p)$, and both $\bar{G}_\tau$ and each $G_\tau^{(k)}$ are retained for reflectivity prediction.

Reflectivity evolution. Transport determines future echo locations, while echo intensity can grow or decay during this motion (Zhang et al., 2023). The shared reflectivity predictor $\mathcal{R}_\psi$ takes h_τ and $G_\tau^{(k)}$, so growth and decay follow the echoes carried by path k. Its output is a reflectivity change in dBZ, bounded by a scaled $\tanh$ that keeps each voxel anchored to its transported value (Appendix C.1). Adding this change to $G_\tau^{(k)}$ gives the path-wise prediction $P_\tau^{(k)}$:

$$P_\tau^{(k)} = G_\tau^{(k)} + \mathcal{R}_\psi\left(h_\tau, G_\tau^{(k)}\right), \qquad k = 1, \ldots, K. \tag{2}$$

Removing this update causes the largest loss of 40 dBZ skill among the design ablations (Table 3). The deterministic forecast $\widehat{Y}_\tau^{\text{det}} = \mathcal{C}_{[0,70]}\left[\bar{G}_\tau + \mathcal{R}_\psi(h_\tau, \bar{G}_\tau)\right]$ applies the same update to $\bar{G}_\tau$ and clips it to $[0, 70]$ dBZ. It provides the reference for the local reflectivity distribution in Sec. 3.3.

3.3 Local Reflectivity Distribution

For a space-time coordinate $q = (\tau, p)$ with lead τ and voxel p, the deterministic forecast assigns one reflectivity value $\widehat{Y}^{\text{det}}(q) = \widehat{Y}_\tau^{\text{det}}(p)$. The predictability of precipitation differs between short-lived convective features and longer-lived mesoscale patterns (Gong et al., 2024; Zhang et al., 2023). PSVT therefore describes the reflectivity at each space-time coordinate with a local marginal distribution, which allows the forecast uncertainty to vary across the radar volume and forecast horizon. This distribution sets how much reflectivity can vary at q, while Sec. 3.4 models the dependency structure of these variations across space-time coordinates. Following Gaussian predictive distributions used to estimate uncertainty in data-driven weather prediction (Verma et al., 2024), PSVT approximates this distribution with a Gaussian censored to $[0, 70]$ dBZ:

$$\widetilde{Y}(q) \overset{d}{=} \mathcal{C}_{[0,70]}\left[\mu(q) + \sigma(q)U\right], \qquad U \sim \mathcal{N}(0, 1). \tag{3}$$

Here $\widetilde{Y}(q)$ follows the local reference distribution, $\mu(q)$ and $\sigma(q)$ are the mean and standard deviation of the uncensored Gaussian, and $\mathcal{C}_{[0,70]}$ is the clipping of Sec. 3.2. Censoring keeps the Gaussian density inside $(0, 70)$ dBZ and forms point masses at the lower and upper bounds. The mass at 0 dBZ collects the Gaussian tail below 0 dBZ and gives the probability that echoes are absent at the voxel. This allows the forecast to represent the echo-free background, which covers large parts of the radar volume.

The forecast uncertainty at each voxel depends on the echo structure around it, which we summarize with a feature vector $\mathbf{f}(q)$ computed from quantities already produced by the forecast. It includes the forecast reflectivity (high-intensity echoes), the magnitude of the predicted reflectivity change (evolving intensity), and the 3D gradient magnitude of $\bar{G}_\tau$ (sharp precipitation boundaries). Near these boundaries, a small shift in the predicted echo position can move the voxel between an echo and the echo-free background. The path-weighted standard deviation of the K path-wise predictions and the entropy of their path weights indicate uncertain echo motion. Two pointwise MLPs map $\mathbf{f}(q)$, together with lead and height embeddings, to the Gaussian parameters:

$$\mu(q) = \widehat{Y}^{\mathrm{det}}(q) + m_\phi\big(\mathbf{f}(q), \tau, z\big), \qquad \sigma(q) = s_\eta\big(\mathbf{f}(q), \tau, z\big). \tag{4}$$

The correction m_ϕ is bounded by $\pm\delta_{\max}$, so $\mu(q)$ retains the echo structure of the deterministic forecast. The standard deviation $\sigma(q)$ lies between fixed bounds $\sigma_{\min}$ and $\sigma_{\max}$. Appendix C.2 defines each entry of $\mathbf{f}(q)$ and gives the MLP architecture.

3.4 Volumetric Dependence Modelling

The local distributions in Sec. 3.3 set how much reflectivity can vary at each space-time coordinate. Each ensemble member should form a spatially coherent (Oskarsson et al., 2024) and temporally consistent (Andrae et al., 2025) radar sequence. Sampling each space-time coordinate independently from its local distribution breaks this coherence. PSVT instead derives an explicit dependency structure from how its K path-wise predictions deviate from their weighted mean, following low-rank probabilistic forecasting models (Salinas et al., 2019; Zheng & Sun, 2024). The dependence therefore follows the predicted echo evolutions.

Path-wise deviations. Each path-wise prediction $P^{(k)}$ carries the entire observed volume along the k-th path and adds local growth and decay (Eq. (2)). Its deviation from the path-weighted mean $\bar{P}$ represents one alternative evolution of the same storm across the whole radar volume and forecast horizon. For example, suppose the k-th path moves the storm farther east than the other paths. Its deviation is then positive on the eastern side of the storm and negative on the western side, at all space-time coordinates the storm occupies. At a space-time coordinate q, we collect these deviations in the weighted deviation vector $\mathbf{d}(q) \in \mathbb{R}^K$ with components

$$d_k(q) = \sqrt{w_k(q)}\,\big(P^{(k)}(q) - \bar{P}(q)\big), \qquad \bar{P}(q) = \sum_{k=1}^{K} w_k(q)\, P^{(k)}(q), \tag{5}$$

where $P^{(k)}(q) = P_\tau^{(k)}(p)$ is the unclipped path-wise prediction that keeps reflectivity values outside $[0, 70]$ dBZ, and $w_k(q) = w_\tau^{(k)}(p)$ is its path weight.

Shared volumetric factors. The magnitude $\|\mathbf{d}(q)\|_2$ is the path-weighted standard deviation of the K path-wise predictions, which already enters $\mathbf{f}(q)$ in Sec. 3.3. For nonzero deviations, PSVT therefore keeps only the direction $\mathbf{a}(q) = \mathbf{d}(q)/\|\mathbf{d}(q)\|_2$, so that $\sigma(q)$ sets the scale of the variations before censoring. Where $\mathbf{d}(q) = \mathbf{0}$, PSVT sets $\mathbf{a}(q) = \mathbf{0}$, and Sec. 3.5 draws an independent variation at q. This combines a shared low-rank component with independent local variations where the path factors vanish (Salinas et al., 2019). Each component $a_k(q)$ gives the response of coordinate q to the k-th alternative evolution. Collected over all space-time coordinates, the components $a_k(\cdot)$ form K shared volumetric factors. Each ensemble member scales each factor with one coefficient over the whole forecast (Sec. 3.5). Coordinates with aligned vectors $\mathbf{a}(q)$ therefore vary together, while coordinates with opposite vectors vary in opposite directions (Eq. (8)). In the example above, a positive coefficient on the k-th factor strengthens echoes on the eastern side of the storm and weakens those on the western side. A negative coefficient does the opposite. The storm therefore appears farther east in some members and farther west in others, and this spread across members represents uncertainty in the storm position.

3.5 Future Radar Ensemble Formation

PSVT forms M ensemble members by combining the local Gaussian distributions of Sec. 3.3, parameterized by $\mu(q)$ and $\sigma(q)$, with the K shared volumetric factors $a_1(\cdot), \ldots, a_K(\cdot)$ of Sec. 3.4. This separation of local intensity and dependence is inspired by Gaussian copula models (Salinas et al., 2019).

From local distributions to complete members. For member $m = 1, \ldots, M$, let $c_k^{(m)}$ denote its random coefficient for factor $a_k(\cdot)$, with $\mathbf{c}^{(m)} = (c_1^{(m)}, \ldots, c_K^{(m)})^\top$. PSVT draws $M/2$ independent vectors $\mathbf{c}$ from the K-dimensional standard normal distribution and adds their negatives $-\mathbf{c}$, so the coefficients come in symmetric pairs. It then whitens the resulting M vectors so that their sample covariance equals the identity matrix, which requires an even $M \geq 2K$ (Appendix C.3). Each member takes one of these vectors as $\mathbf{c}^{(m)}$ and keeps it the same at all leads, while $a_k(q)$, $\mu(q)$, and $\sigma(q)$ vary across space-time coordinates. Sharing member-specific randomness across lead times is also used in diffusion-based weather ensembles (Andrae et al., 2025). Where $\mathbf{a}(q) \neq \mathbf{0}$, the variation field $u^{(m)}(q)$ combines the factors with these coefficients, and member m departs from $\mu(q)$ by $\sigma(q)u^{(m)}(q)$ before censoring:

$$u^{(m)}(q) = \sum_{k=1}^{K} c_k^{(m)} a_k(q), \qquad \widehat{Y}^{(m)}(q) = \mathcal{C}_{[0,70]}\big[\mu(q) + \sigma(q)u^{(m)}(q)\big]. \tag{6}$$

Where $\mathbf{a}(q) = \mathbf{0}$, PSVT instead draws standard normal values independently across space-time coordinates and pairs each draw with its negative across members. The whitened coefficients provide a finite, moment-matched approximation to Gaussian sampling (Appendix C.3). Applying the local mapping in Eq. (6) at all space-time coordinates gives the future sequence $\{\widehat{Y}_\tau^{(m)}\}_{\tau=1}^{T_{\text{out}}}$ of member m. All members share the factors $a_k(\cdot)$ and the local parameters $\mu(q)$ and $\sigma(q)$, which PSVT computes once per radar history. Each additional member therefore requires only the linear combination and local mapping in Eq. (6), without further network evaluations.

Preserving local intensity and spread. The factors $a_k(\cdot)$ describe only which coordinates vary together, whereas $\mu(q)$ and $\sigma(q)$ are learned from observations and already set the intensity and uncertainty at each coordinate (Sec. 3.3). Member formation uses these same local parameters for every member. With equal member weight $1/M$, the paired and whitened coefficients satisfy the moment-matching conditions

$$\frac{1}{M}\sum_{m=1}^{M} \mathbf{c}^{(m)} = \mathbf{0}, \qquad \frac{1}{M}\sum_{m=1}^{M} \mathbf{c}^{(m)}\mathbf{c}^{(m)\top} = \mathbf{I}_K, \tag{7}$$

where $\mathbf{I}_K$ is the $K \times K$ identity matrix. Wherever $\mathbf{a}(q) \neq \mathbf{0}$, its unit norm and these conditions give $\frac{1}{M}\sum_m u^{(m)}(q) = 0$ and $\frac{1}{M}\sum_m u^{(m)}(q)^2 = 1$. At these coordinates, the uncensored ensemble has mean $\mu(q)$ and standard deviation $\sigma(q)$, using member weights $1/M$. Removing moment matching increases the continuous ranked probability score (CRPS) by 3.6% (Table 3).

Preserving spatiotemporal correlations. Moment matching preserves the correlations encoded by the shared factors before censoring. For coordinates with $\mathbf{a}(q) \neq \mathbf{0}$ and $\mathbf{a}(q') \neq \mathbf{0}$, the empirical correlation of the variation fields across the M members follows from Eqs. (6) and (7):

$$\frac{1}{M}\sum_{m=1}^{M} u^{(m)}(q)\, u^{(m)}(q') = \mathbf{a}(q)^\top \mathbf{a}(q'). \tag{8}$$

3.6 Training Objective

PSVT is trained end to end with a deterministic loss on the deterministic forecast $\widehat{Y}^{\text{det}}$ and a CRPS loss on the local parameters $\mu(q)$ and $\sigma(q)$. The deterministic loss combines a binary cross-entropy weighted toward high-intensity echoes (Shi et al., 2017) with a soft CSI loss (Ebert-Uphoff et al., 2021). We compute the CRPS of the local censored-Gaussian reference distribution in Eq. (3) in closed form (Gneiting et al., 2005; Jordan et al., 2019), so training requires no ensemble members.

Table 1: Forecast quality and computational cost on NEXRAD-3Y and TRRD. ConvGRU3D, PhyDNet3D, and SimVP3D use 128×128 inputs; all other methods use 512×512 inputs. Params reports total model parameters in millions.

Method	CRPS↓	SSIM↑	LPIPS↓	HSS↑ 20	HSS↑ 30	HSS↑ 40	CSI↑ 20	CSI↑ 30	CSI↑ 40	Params (M)↓	Time (s)↓
NEXRAD-3Y											
ConvGRU3D	0.995	0.852	0.275	0.488	0.225	0.043	0.340	0.133	0.022	4.594	0.860
PhyDNet3D	1.068	0.807	0.259	0.520	0.237	0.009	0.367	0.142	0.005	2.897	0.119
SimVP3D	1.099	0.813	0.257	0.585	0.353	0.063	0.430	0.225	0.033	2.663	0.406
STC-GS	1.098	0.848	**0.132**	0.424	0.259	0.128	0.283	0.155	0.070	11.622	2.842
FlowCast-Single	0.616	**0.923**	0.134	0.623	0.461	0.203	0.466	0.308	0.114	163.339	6.440
FlowCast-Ensemble	0.448	0.921	0.157	0.611	0.422	0.131	0.454	0.276	0.071	163.339	98.628
PSVT-Single	0.663	0.881	0.143	**0.642**	**0.559**	**0.389**	**0.486**	**0.397**	**0.244**	**0.524**	0.254
PSVT-Ensemble	**0.446**	0.872	0.147	0.618	0.508	0.352	0.460	0.349	0.216	**0.524**	3.301
TRRD											
FlowCast-Single	0.292	0.943	**0.136**	0.248	0.145	0.074	0.166	0.087	0.043	163.339	14.020
FlowCast-Ensemble	**0.234**	**0.944**	0.158	0.217	0.125	0.063	0.143	0.074	0.037	163.339	219.511
PSVT-Single	0.522	0.894	0.142	**0.299**	**0.222**	0.123	**0.206**	**0.141**	0.071	**0.524**	0.511
PSVT-Ensemble	0.239	0.931	0.140	0.272	0.198	**0.131**	0.182	0.122	**0.076**	**0.524**	6.833

This term fits the local marginals, while the shared paths supply the dependence structure. Through the features $\mathbf{f}(q)$, the CRPS loss also updates the backbone, the displacement head $\mathcal{D}_{\theta_\Delta}$, and the reflectivity predictor $\mathcal{R}_\psi$. Appendix B.2 gives the loss expressions and weights.

4 Experiments

Experimental settings. We evaluate PSVT on two 3D radar datasets with reflectivity on a 512×512 horizontal grid at 5-minute intervals and 36 vertical levels. NEXRAD-3Y is a three-year dataset constructed by STC-GS (Wang et al., 2025) from observations collected between 2020 and 2022, and we re-split it at the event level with events from 2022 as the test set. TRRD is a one-year dataset collected in 2025 by a weather radar in Singapore. On both datasets, each model observes five radar volumes and forecasts the next 20, covering a 100-minute horizon. Following CasCast (Gong et al., 2024), we also compare PSVT with a reproduced CasCast on the 2D SEVIR dataset (Veillette et al., 2020). Appendix B gives the network architectures and the training and testing details.

4.1 Main Results

High-intensity echoes. Table 1 compares PSVT with 3D nowcasting baselines following STC-GS (Wang et al., 2025). PSVT achieves the highest HSS and CSI at every threshold on both datasets. Its margin over FlowCast grows with the threshold, reaches 114.0% in single-forecast CSI at 40 dBZ on NEXRAD-3Y, and holds across lead times (Figure 2(e,f)). This trend is consistent with the echo source of PSVT: transport takes reflectivity directly from observed voxels, so high-intensity echoes bypass the compressed representations that can degrade them (Gong et al., 2024; Schusterbauer et al., 2026). The backbone only predicts how these echoes move, grow, and decay, using 0.3% of the parameters of FlowCast.

Ensemble efficiency. PSVT-Ensemble reaches CRPS comparable to FlowCast-Ensemble and generates 16 members about $30\times$ faster on NEXRAD-3Y and $32\times$ faster on TRRD. The gap persists up to 1024×1024 (Figure 2(a,b)). FlowCast evaluates its network at each of ten sampling steps for every member. PSVT computes the factors $a_k(\cdot)$ and the local parameters $\mu(q)$ and $\sigma(q)$ once, and each additional member requires only a linear combination and a local mapping (Sec. 3.5). The shared factors and local distributions thus form a competitive ensemble without iterative sampling.

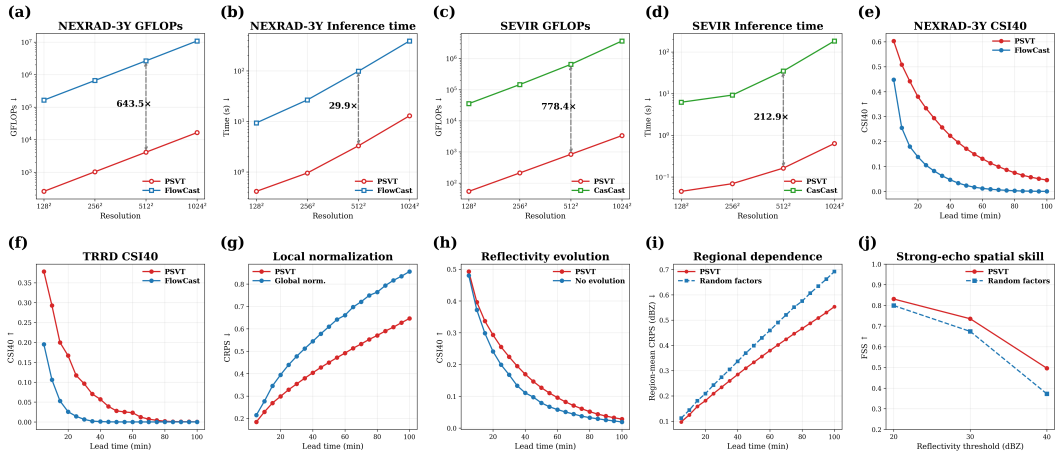

Figure 2: **Computational scaling, lead-time skill, and component analysis.** (a,b) GFLOPs and inference time versus resolution against FlowCast on NEXRAD-3Y and (c,d) against CasCast on SEVIR, with ratios annotated at 512^2. (e,f) Pool4 CSI at 40 dBZ over lead time on NEXRAD-3Y and TRRD. (g) CRPS with local and global normalization. (h) CSI at 40 dBZ with and without reflectivity evolution. (i,j) Regional CRPS and FSS with learned and random path factors.

Table 2: SEVIR comparison with deterministic and probabilistic models. We reproduce CasCast and take the other baseline results from Gong et al. (2024). PSVT uses $K = 4$ and 10 members, and its SSIM, HSS, and CSI are computed on the ensemble mean.

Method	CRPS↓	SSIM↑	HSS↑	CSI-M↑			CSI-181↑			CSI-219↑		
				POOL1	POOL4	POOL16	POOL1	POOL4	POOL16	POOL1	POOL4	POOL16
ConvLSTM	0.02640	0.7749	0.5232	0.4102	0.4163	0.4475	0.2453	0.2525	0.2977	0.1322	0.1380	0.1734
PredRNN	0.02710	0.7497	0.5192	0.4045	0.4161	0.4623	0.2416	0.2567	0.3214	0.1331	0.1447	0.1909
PhyDNet	0.02530	0.7649	0.5311	0.4198	0.4226	0.4410	0.2526	0.2532	0.2782	0.1362	0.1359	0.1526
SimVP	0.02590	**0.7772**	0.5280	0.4153	0.4226	0.4530	0.2532	0.2604	0.3000	0.1338	0.1394	0.1685
Earthformer	0.02510	0.7756	0.5411	0.4310	0.4319	0.4351	0.2622	0.2542	0.2562	0.1448	0.1409	0.1481
NowcastNet	0.02830	0.5696	0.5365	0.4152	0.4452	0.5024	0.2495	0.2935	0.3725	0.1422	0.1874	0.2700
LDM	0.02080	0.7495	0.4386	0.3465	0.3442	0.3520	0.1470	0.1391	0.1432	0.0671	0.0655	0.0717
PreDiff	0.02020	0.7648	0.4914	0.3875	0.3918	0.4157	0.2076	0.2069	0.2264	0.1032	0.1051	0.1213
CasCast (reprod.)	0.02049	0.7595	0.5687	0.4387	**0.4627**	**0.5207**	0.2868	**0.3169**	**0.3892**	0.1826	0.2107	**0.2805**
PSVT ($K = 4$)	**0.01824**	0.7648	**0.5752**	**0.4447**	0.4557	0.4792	**0.2965**	0.3158	0.3573	**0.1893**	**0.2146**	0.2715

Generalization to 2D nowcasting. On SEVIR, PSVT keeps the Earthformer backbone of CasCast and replaces its autoencoder and latent diffusion with sparse transport and ensemble formation (Secs. 3.2–3.5). The two methods reach comparable skill (Table 2): PSVT achieves lower CRPS and higher HSS and POOL1 CSI, whereas CasCast retains higher CSI-M at POOL4 and POOL16. Because CasCast runs 20 DDIM steps and decodes every member, PSVT generates ten members with 212.9× lower latency (Figure 2(c,d)).

Qualitative results. Appendix D visualizes forecasts on all three datasets. On NEXRAD-3Y and TRRD, FlowCast produces smooth fields and underestimates the high-intensity echoes, whereas PSVT keeps these echoes at their observed locations (Figures 3 and 4). Figure 5 shows how K pathwise predictions form ensemble members, and Figure 6 compares PSVT with CasCast on SEVIR.

4.2 Ablation Study

Core design components. **(i) Echo source.** PSVT transports the full-resolution observed volume so that high-intensity echoes remain on the native grid (Sec. 3.1). Replacing it with a coarse 128×128 source (Coarse echo source) reduces CSI at 40 dBZ by 20.8% and also lowers FAR, in-

Table 3: Ablations of PSVT on NEXRAD-3Y with $M = 16$ members.

Variant	CRPS↓	SSIM↑	HSS_{40} ↑	CSI↑ 20	CSI↑ 30	CSI↑ 40	FAR↓ 20	FAR↓ 30	FAR↓ 40
No intensity update	0.486	0.857	0.271	0.435	0.311	0.159	0.218	0.329	0.443
No echo conditioning	0.456	0.856	0.349	0.462	0.349	0.214	0.194	0.307	0.479
Uniform displacements	0.454	0.871	0.319	0.442	0.330	0.192	0.185	0.300	0.476
Global normalization	0.597	0.850	0.317	0.439	0.352	0.191	0.472	0.500	0.408
No moment matching	0.462	0.872	0.357	0.464	0.352	0.220	0.197	0.312	0.489
Coarse echo source	0.443	0.874	0.288	0.445	0.320	0.171	0.166	0.263	0.410
Full PSVT	0.446	0.872	0.352	0.460	0.349	0.216	0.180	0.291	0.455

Factors	CRPS↓ Voxel	CRPS↓ Region	CRPS↓ Δ_{20}	CRPS↓ Vert. Δ_3	$Occupancy_{40}$ CRPS↓ ($\times 10^3$) Region	$Occupancy_{40}$ CRPS↓ ($\times 10^3$) Δ_{20}	$Occupancy_{40}$ CRPS↓ ($\times 10^3$) Vert. Δ_3	FSS↑ 20	FSS↑ 30	FSS↑ 40
Path-derived factors	0.4462	**0.3371**	**0.3153**	**0.2288**	**1.111**	**1.162**	**0.729**	**0.8308**	**0.7353**	**0.4961**
Random normalized factors	0.4463	0.4095	0.3614	0.2613	1.425	1.334	0.909	0.7992	0.6740	0.3722

dicating more missed high-intensity echoes rather than more false alarms. **(ii) Displacement fields.** Echo motion varies across a storm and with height, so each path assigns a displacement to every voxel (Sec. 3.2). Sharing one displacement across all voxels for each path and lead (Uniform displacements) reduces CSI at 40 dBZ by 11.1%. **(iii) Reflectivity evolution.** Transport determines future echo locations, whereas $\mathcal{R}_\psi$ predicts their growth and decay along each path (Sec. 3.2). Removing $\mathcal{R}_\psi$ (No intensity update) reduces CSI at 40 dBZ by 26.4%, the largest reduction among the design ablations. Figure 2(h) shows that this variant has lower CSI at 40 dBZ at every lead time. Removing the transported echoes from the input of $\mathcal{R}_\psi$ (No echo conditioning) increases FAR at 40 dBZ by 5.3%, indicating that this conditioning anchors growth and decay to the transported echoes. Components (i)–(iii) thus underpin skill on high-intensity echoes.

(iv) Factor normalization. PSVT normalizes each nonzero deviation vector to unit length, separating path spread from the scale of uncensored member variations (Sec. 3.4). Normalizing with a single global scale (Global normalization) couples this magnitude to the path spread, which increases CRPS by 33.9% and FAR at 20 dBZ by 162.2%. Figure 2(g) shows that this variant has higher CRPS at every lead time. **(v) Moment matching.** Moment matching enforces zero mean and identity covariance on the finite set of shared coefficients (Sec. 3.5). Removing it (No moment matching) increases CRPS by 3.6% and FAR at 40 dBZ by 7.5%, despite a 1.9% gain in CSI at 40 dBZ. Components (iv) and (v) thus govern the distribution coverage of the ensemble. Appendix A.1 analyzes the number of transport paths and the backbone capacity.

Path-derived volumetric dependence. Replacing the factors $a_k(\cdot)$ with random normalized factors leaves voxel CRPS nearly unchanged. It increases regional, 20-minute-change, and vertical-difference CRPS by 21.5%, 14.6%, and 14.2% and reduces FSS at 40 dBZ by 25.0% (Table 3; Figure 2(i,j)). These gains hold on all 29 test events, indicating that the path-wise deviations capture spatial, vertical, and temporal dependence.

5 Conclusion

We presented PSVT, a probabilistic sparse volumetric transport framework for 3D radar nowcasting. Instead of reconstructing future radar volumes from compressed representations, PSVT transports the latest observed volume at full resolution along a small set of learned 3D transport paths. This design lets a lightweight backbone learn how echoes move, grow, and decay, while the transported volume preserves high-intensity echoes. Deviations of the path-wise predictions define shared volumetric factors, and local Gaussian distributions determine how much reflectivity varies at each voxel and lead. Additional ensemble members therefore require only linear combinations and local mappings, without iterative sampling. Experiments on NEXRAD-3Y and TRRD show higher skill on high-intensity echoes and faster ensemble generation than FlowCast. On SEVIR, the same ensemble formation can replace latent diffusion with comparable skill and lower inference cost.

AI USE STATEMENT

We used generative AI tools for the following tasks with recommended disclosure: summarizing and analyzing existing literature, identifying relevant literature, suggesting the structure of the paper, and editing the text for readability. These tools were further used to edit our PSVT code into an anonymized review package for the supplementary material. We checked every AI-identified reference against the original publication, reviewed all AI-assisted text. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Section 3 and Appendix C describe the model, including the transport bounds, the entries of the feature vector, and the construction of the ensemble coefficients. Appendix B gives the data splits, architecture details, training settings, evaluation metrics, and baseline implementations. NEXRAD-3Y (Wang et al., 2025) and SEVIR (Veillette et al., 2020) are publicly available, and the anonymized PSVT code for the NEXRAD-3Y experiments is provided in the supplementary material.

REFERENCES

Martin Andrae, Tomas Landelius, Joel Oskarsson, and Fredrik Lindsten. Continuous ensemble weather forecasting with diffusion models. In *International Conference on Learning Representations*, 2025.

Alexandre Drouin, Étienne Marcotte, and Nicolas Chapados. TACTiS: Transformer-attentional copulas for time series. In *International Conference on Machine Learning*, 2022.

Imme Ebert-Uphoff, Ryan Lagerquist, Kyle Hilburn, Yoonjin Lee, Katherine Haynes, Jason Stock, Christina Kumler, and Jebb Q. Stewart. CIRA guide to custom loss functions for neural networks in environmental sciences – version 1. *arXiv preprint arXiv:2106.09757*, 2021.

Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z. Li. SimVP: Simpler yet better video prediction. In *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.

Tilmann Gneiting, Adrian E. Raftery, Anton H. Westveld, III, and Tom Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133(5), 2005.

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 *International Conference on Machine Learning*, 2024.

Xiaotao Hu, Zhewei Huang, Ailin Huang, Jun Xu, and Shuchang Zhou. A dynamic multi-scale voxel flow network for video prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, 2015.

Alexander Jordan, Fabian Krüger, and Sebastian Lerch. Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, 90(12), 2019.

Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

Ziwei Liu, Raymond A. Yeh, Xiaoou Tang, Yiming Liu, and Aseem Agarwala. Video frame synthesis using deep voxel flow. In *IEEE International Conference on Computer Vision*, 2017.

Joel Oskarsson, Tomas Landelius, Marc Peter Deisenroth, and Fredrik Lindsten. Probabilistic weather forecasting with hierarchical graph neural networks. In *Advances in Neural Information Processing Systems*, 2024.

Bernardo Perrone Ribeiro and Jana Faganeli Pucer. FlowCast: Advancing precipitation nowcasting with conditional flow matching. In *International Conference on Learning Representations*, 2026.

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

David Salinas, Michael Bohlke-Schneider, Laurent Callot, Roberto Medico, and Jan Gasthaus. High-dimensional multivariate forecasting with low-rank Gaussian copula processes. In *Advances in Neural Information Processing Systems*, 2019.

Johannes Schusterbauer, Jannik Wiese, Nick Stracke, Timy Phan, and Björn Ommer. Probabilistic precipitation nowcasting with rectified flow transformers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

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.

Nengli Sun, Zeming Zhou, Qian Li, and Jinrui Jing. Three-dimensional gridded radar echo extrapolation for convective storm nowcasting based on 3D-ConvLSTM model. *Remote Sensing*, 14(17), 2022.

Cheng Tan, Zhangyang Gao, Lirong Wu, Yongjie Xu, Jun Xia, Siyuan Li, and Stan Z. Li. Temporal attention unit: Towards efficient spatiotemporal predictive learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

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*, 2020.

Yogesh Verma, Markus Heinonen, and Vikas Garg. ClimODE: Climate and weather forecasting with physics-informed neural ODEs. In *International Conference on Learning Representations*, 2024.

Ziye Wang, Yiran Qin, Lin Zeng, and Ruimao Zhang. High-dynamic radar sequence prediction for weather nowcasting using spatiotemporal coherent Gaussian representation. In *International Conference on Learning Representations*, 2025.

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 *AAAI Conference on Artificial Intelligence*, 2026.

Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

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 *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 (7970), 2023.

Maksim Zhdanov, Ana Lucic, Max Welling, and Jan-Willem van de Meent. (Sparse) attention to the details: Preserving spectral fidelity in ML-based weather forecasting models. In *International Conference on Machine Learning*, 2026.

Vincent Zhihao Zheng and Lijun Sun. Multivariate probabilistic time series forecasting with correlated errors. In *Advances in Neural Information Processing Systems*, 2024.

Yiqi Zhong, Luming Liang, Ilya Zharkov, and Ulrich Neumann. MMVP: Motion-matrix-based video prediction. In *IEEE/CVF International Conference on Computer Vision*, 2023.

A Additional Ablation Results

A.1 Number of Transport Paths and Backbone Capacity

Table 4: Number of transport paths and backbone capacity on NEXRAD-3Y with $M = 16$ members. The $K = 3$ and PSVT-Small rows repeat Full PSVT in Table 3.

				CSI↑			FAR↓		
Variant	CRPS↓	SSIM↑	HSS_{40} ↑	20	30	40	20	30	40
Number of transport paths									
$K = 1$	0.467	0.858	0.366	0.463	0.353	0.226	0.208	0.320	0.487
$K = 2$	0.451	0.874	0.358	0.460	0.350	0.220	0.188	0.302	0.472
$\mathbf{K = 3}$	0.446	0.872	0.352	0.460	0.349	0.216	0.180	0.291	0.455
$K = 4$	0.444	0.873	0.348	0.459	0.346	0.213	0.177	0.285	0.443
$K = 5$	0.444	0.871	0.343	0.459	0.344	0.210	0.176	0.283	0.439
Backbone capacity									
PSVT-Small (1×)	0.446	0.872	0.352	0.460	0.349	0.216	0.180	0.291	0.455
PSVT-Medium (10×)	0.426	0.883	0.372	0.476	0.360	0.231	0.169	0.275	0.436
PSVT-Large (20×)	0.424	0.884	0.375	0.477	0.362	0.233	0.167	0.276	0.439

Table 4 varies the number of transport paths and the backbone capacity.

Number of transport paths. With $K = 1$, PSVT degenerates into independent perturbations at each space-time coordinate (Sec. 3.4). This setting yields the highest CRPS and the lowest SSIM among all values of K. Increasing K to 3 reduces CRPS by 4.5% and FAR at 40 dBZ by 6.6%, with a 4.4% decrease in CSI at 40 dBZ. Further increasing K changes CRPS by less than 0.5%.

Backbone capacity. Scaling the backbone from 0.52M to 5.20M parameters (PSVT-Medium) reduces CRPS by 4.5% and improves CSI at 40 dBZ by 6.9%. Further scaling to 10.40M parameters (PSVT-Large) changes all metrics by at most 1.2%. The main results use PSVT-Small, which outperforms FlowCast in CSI at 40 dBZ with only 0.3% of its parameters. This result supports the lightweight backbone design in Sec. 3.1.

B Experimental Details

B.1 Datasets and Implementation Details

Datasets. We evaluate PSVT on two datasets of 3D radar reflectivity, NEXRAD-3Y and TRRD. Both datasets provide reflectivity in dBZ on a 512×512 horizontal grid at 5-minute intervals, and we use 36 vertical levels from each. NEXRAD-3Y is the three-year dataset constructed by STC-GS (Wang et al., 2025) from observations collected between 2020 and 2022. We re-split it at the event level: events from 2020 and 2021 form the training and validation sets, and events from 2022 form the test set. The training, validation, and test sets contain 786, 263, and 530 sequences from 47, 12, and 29 events, respectively. TRRD is a one-year dataset collected in 2025 by a weather radar in Singapore. Its 104,057 raw volumes contain 40 vertical levels from 0.5 to 20.0 km at 0.5-km spacing. We use the 36 levels from 1.5 to 19.0 km. For TRRD, the training, validation, and test sets contain 6,553, 499, and 60 sequences starting on 259, 19, and 60 distinct days, respectively. Following the experimental setting of CasCast (Gong et al., 2024) on SEVIR (Veillette et al., 2020), we also reproduce CasCast and compare it with PSVT on this 2D dataset.

Implementation details. *(i) Forecast setup.* We use five observed radar volumes to predict the next 20, covering a 100-minute forecast horizon. All probabilistic methods produce $M = 16$ members. Full evaluation details are given in Appendix B.3. *(ii) Architecture.* The encoder $\mathcal{E}$ uses ConvNeXt V2 blocks (Woo et al., 2023), and the temporal module $\mathcal{F}$ uses TAU blocks (Tan et al., 2023). The displacement head $\mathcal{D}_{\theta_\Delta}$ and the path-weight head process h_τ with the same ConvNeXt V2 blocks and use separate output layers for $\Delta_\tau^{(k)}$ and $w_\tau^{(k)}$. The reflectivity predictor $\mathcal{R}_\psi$ encodes

$G_\tau^{(k)}$ with a ConvNeXt V2 feature pyramid and fuses it with h_τ. *(iii) Bounds.* In Eqs. (1) and (2), $\tanh$ bounds the horizontal and vertical components of $\Delta_\tau^{(k)}(p)$ by 20 grid cells and 2 levels and $\mathcal{R}_\psi$ by 12 dBZ. In Eq. (4), $\delta_{\max} = 6$, $\sigma_{\min} = 0.02$, and $\sigma_{\max} = 35$ dBZ. Grid search on the validation set selects these bounds. *(iv) Training.* We train PSVT from scratch for 20 epochs with AdamW, batch size 1, and bfloat16 mixed precision for the backbone. Peak learning rates are 2×10^{-4}, 3×10^{-4}, and 10^{-3} for the backbone and the MLPs of $\mu(q)$ and $\sigma(q)$, respectively, with one-epoch linear warmup and cosine decay.

B.2 Training Details

PSVT is trained with

$$\mathcal{L} = \sum_{b \in \{20,30,40\}} \left(\mathcal{L}_{\mathrm{BCE}}^{(b)} + \lambda_{\mathrm{CSI}} \, \mathcal{L}_{\mathrm{CSI}}^{(b)} \right) + \lambda_{\mathrm{CRPS}} \, \mathcal{L}_{\mathrm{CRPS}}, \tag{9}$$

where b is a reflectivity threshold in dBZ, $\lambda_{\mathrm{CSI}} = 0.5$, and $\lambda_{\mathrm{CRPS}} = 20$. The threshold terms are computed on $\widehat{Y}^{\mathrm{det}}$ with the logit $(\widehat{Y}^{\mathrm{det}} - b)/T$ and a temperature $T = 4$ dBZ. $\mathcal{L}_{\mathrm{BCE}}^{(b)}$ is a binary cross-entropy with positive-class weights of 2, 5, and 10 for the three thresholds, and $\mathcal{L}_{\mathrm{CSI}}^{(b)}$ is one minus a soft CSI (Ebert-Uphoff et al., 2021). All forecast leads are weighted equally in both terms. $\mathcal{L}_{\mathrm{CRPS}}$ is the closed-form CRPS of the censored Gaussian in Eq. (3) on $[0, 70]$ dBZ (Jordan et al., 2019), divided by 70 dBZ. At each update, $\mathcal{L}_{\mathrm{CRPS}}$ is averaged over one 128×128 horizontal tile, and the 16 tiles of the grid are visited in a random order across updates. No ensemble members are therefore sampled during training. In addition to the settings in Appendix B.1, we apply a weight decay of 10^{-4} to the backbone only, decay the learning rate to 5% of its peak value, and clip the gradient norm at 5.0. The local-distribution MLPs and all losses are computed in float32.

B.3 Evaluation Details

We compute CSI, FAR, and HSS at 20, 30, and 40 dBZ from the counts of hits, false alarms, misses, and correct negatives, following their standard definitions (Gong et al., 2024). For ensembles, we evaluate the CRPS of the equally weighted empirical distribution formed by the members (Jordan et al., 2019) and average it over all space-time coordinates. For a single forecast, this score reduces to the absolute difference from the observation. For the path-factor ablation in Table 3, voxel CRPS is evaluated over the complete native 3D forecast. Region CRPS evaluates regional mean reflectivity; Δ_{20} evaluates its 20-minute change; and vertical Δ_3 evaluates differences between regions separated by three model levels. Per-event comparisons use pointwise paired 95% intervals.

B.4 Baseline Implementation Details

Adapting FlowCast to 3D nowcasting. We adapt the released FlowCast implementation to volumetric forecasting on NEXRAD-3Y, keeping its autoencoder and conditional flow-matching (CFM) objectives and two-stage training procedure (Perrone Ribeiro & Faganeli Pucer, 2026). *(i) Volumetric adaptation.* Each temporal frame is one complete radar volume, represented as a 36-channel 512×512 field. The autoencoder compresses each volume $8\times$ horizontally into a $64 \times 64 \times 144$ latent, encoding all height levels jointly. The five observed volumes therefore give $Z_{\mathrm{past}} \in \mathbb{R}^{5 \times 64 \times 64 \times 144}$, and the complete future sequence is represented by $Z_{\mathrm{future}} \in \mathbb{R}^{20 \times 64 \times 64 \times 144}$. The CFM model predicts Z_{future} jointly rather than sampling each forecast lead separately. We retain the released autoencoder widths $[128, 256, 512, 512]$ and the Cuboid Transformer U-Net with base width 192, depths $[4, 4]$, four attention heads, and axial Cuboid Attention. FlowCast is conditioned only on the radar history, without PSVT or baseline forecasts as additional inputs. *(ii) Training.* We train two CFM models with different seeds for 200 epochs with batch size 12 and report the average of the two runs. *(iii) Forecast products.* The radar history is encoded once and shared by all $M = 16$ members. For each member, one Gaussian latent covering all forecast leads and height levels is sampled and integrated with ten Euler steps. Each member is therefore one complete future radar sequence, and FlowCast requires 160 CFM evaluations per ensemble forecast. FlowCast-Single is the first prespecified member, without member selection. For FlowCast-Ensemble, CRPS is computed from all 16 members, and their voxelwise mean serves as the point forecast. HSS and CSI are computed on the point forecast after horizontal Pool4, and SSIM and LPIPS over all horizontal height slices.

C METHOD DETAILS

C.1 TRANSPORT DETAILS

The displacement head $\mathcal{D}_{\theta_\Delta}$ predicts the K displacement fields on the downsampled grid, and the fields are interpolated to every voxel of the radar volume. Each component of $\Delta_\tau^{(k)}(p)$ is bounded by a scaled $\tanh$, with at most 20 grid cells for the horizontal components and 2 height levels for the vertical component. Because every lead is sampled directly from $X_{T_{\text{in}}}$ in Eq. (1), these bounds apply to the total displacement from the latest observation rather than to the displacement between consecutive leads. Backward sampling is performed with trilinear interpolation, and positions outside the radar domain are assigned 0 dBZ. The reflectivity predictor $\mathcal{R}_\psi$ outputs $12\tanh(\cdot)$ dBZ, so each path-wise prediction $P_\tau^{(k)}(p)$ differs from its transported value $G_\tau^{(k)}(p)$ by at most 12 dBZ. The same bound applies to the update of $\bar{G}_\tau$ in the deterministic forecast $\widehat{Y}_\tau^{\text{det}}$.

C.2 LOCAL DISTRIBUTION DETAILS

Table 5 lists the seven entries of the feature vector $\mathbf{f}(q)$ at a space-time coordinate $q = (\tau, p)$. The gradient is computed with central differences along the height and horizontal axes, and $H(q)$ is the entropy of the path weights. The path support indicates whether the path-wise predictions deviate at q, which is equivalent to a nonzero $\mathbf{a}(q)$ (Sec. 3.4). For $K = 1$, the normalized path-weight entropy is defined as zero before mapping to $[-1, 1]$.

Table 5: Entries of the feature vector $\mathbf{f}(q)$.

Entry	Quantity	Scaling
Forecast reflectivity	$\widehat{Y}^{\text{det}}(q)$	divided by 70 dBZ, mapped to $[-1, 1]$
Reflectivity change	$\lvert\mathcal{R}_\psi(h_\tau, \bar{G}_\tau)(p)\rvert$	divided by 12 dBZ, mapped to $[-1, 1]$
Gradient magnitude	$\lVert\nabla\bar{G}_\tau(p)\rVert_2$	logarithmic, mapped to $[-1, 1]$
Path-weight entropy	$H(q)$	divided by $\log K$, mapped to $[-1, 1]$
Unit reflectivity	$\widehat{Y}^{\text{det}}(q)$	divided by 70 dBZ
Path spread	$\lVert\mathbf{d}(q)\rVert_2$	logarithmic
Path support	$\mathbf{d}(q) \neq \mathbf{0}$	binary

Two pointwise MLPs predict the location and the scale from $\mathbf{f}(q)$, an 8-dimensional lead embedding, and a 4-dimensional height embedding. Each MLP is a three-layer perceptron with input layer normalization, a hidden width of 32, and SiLU activations. The location correction is bounded to ± 6 dBZ by a scaled $\tanh$ and a sigmoid gate, and the scale output is mapped to $\sigma(q) \in [0.02, 35]$ dBZ by a sigmoid. The output layers are initialized so that $\mu(q) = \widehat{Y}^{\text{det}}(q)$ and $\sigma(q) = 2$ dBZ at the start of training.

C.3 ENSEMBLE FORMATION

We draw $M/2$ independent Gaussian coefficient vectors, add their negatives, and whiten the resulting set using equal weights $1/M$. This requires an even $M \geq 2K$, satisfied by $M = 16$ and $K = 3$. The censored Gaussian in Eq. (3) is the local reference distribution fitted by the analytic CRPS loss. The finite coefficients match the prescribed moments but are not exact Gaussian draws. Where the factors are nonzero, Eq. (7) gives the ensemble moments and correlations described in Sec. 3.5 before censoring; censoring need not preserve them.

In implementation, we treat $\lVert\mathbf{d}(q)\rVert_2 \leq 10^{-6}$ as zero path spread and set $\mathbf{a}(q) = \mathbf{0}$. At these coordinates, we draw $M/2$ standard normal values and add their negatives, independently across space-time coordinates and independently of the shared coefficients. Pairing gives the variations zero ensemble mean, but their ensemble second moment equals one only in expectation; no local sample-variance normalization is applied.

D QUALITATIVE RESULTS

D.1 COMPARISON WITH FLOWCAST ON NEXRAD-3Y AND TRRD

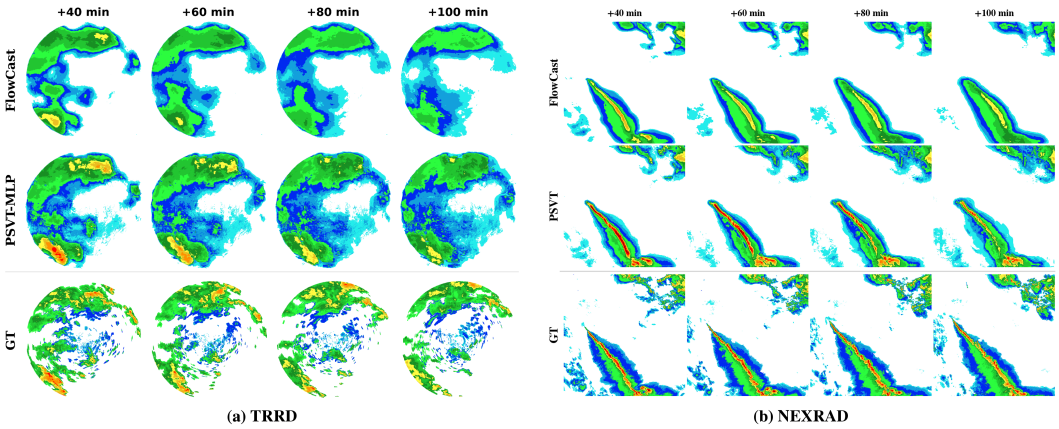

Figure 3: **Qualitative visualization of column-maximum reflectivity.**

Figure 3 qualitatively compares FlowCast and PSVT on NEXRAD-3Y and TRRD. FlowCast produces smooth fields and underestimates the high-intensity echoes, which disappear by +80 min on TRRD. PSVT keeps the high-intensity echoes at their observed locations up to +100 min, which is visually consistent with the ground truth. This is because PSVT transports the observed high-intensity echoes, whereas FlowCast decodes its forecasts from a compressed latent space. Appendix D.2 shows how PSVT forms ensemble members from the transport paths.

Figure 4 compares FlowCast and PSVT on four additional cases from each of NEXRAD-3Y and TRRD. In all eight cases, FlowCast produces smooth fields and underestimates the high-intensity echoes. PSVT keeps the high-intensity echoes at their observed locations, which is visually consistent with the ground truth. These cases are consistent with Table 1, where PSVT achieves higher CSI at 40 dBZ than FlowCast on both datasets. Figure 2(e,f) further shows that this advantage holds across lead times. As in Figure 3, this difference arises because PSVT transports the observed high-intensity echoes, whereas FlowCast decodes its forecasts from a compressed latent space.

D.2 TRANSPORT PATHS AND ENSEMBLE MEMBERS

Figure 5 shows how PSVT forms ensemble members from $K=3$ path-wise predictions. The three paths differ in how the high-intensity echoes move and change in intensity. A small set of paths thus represents several plausible futures. Each member preserves realistic spatial structure and differs from the others mainly around these echoes. This is because each member linearly combines the deviations of the K path-wise predictions, which define its dependency structure (Sec. 3.4).

D.3 COMPARISON WITH CASCAST ON SEVIR

Figure 6 shows qualitative comparisons between CasCast and PSVT on four SEVIR test sequences. PSVT better preserves the location, structure, and intensity of convective echoes as the forecast horizon increases. The advantage is particularly evident for the high-VIL cores in panels (a), (c), and (d), where PSVT remains more consistent with the ground truth while CasCast exhibits larger spatial and structural deviations. These results further illustrate the benefit of explicitly modeling echo transport and intensity evolution in PSVT.

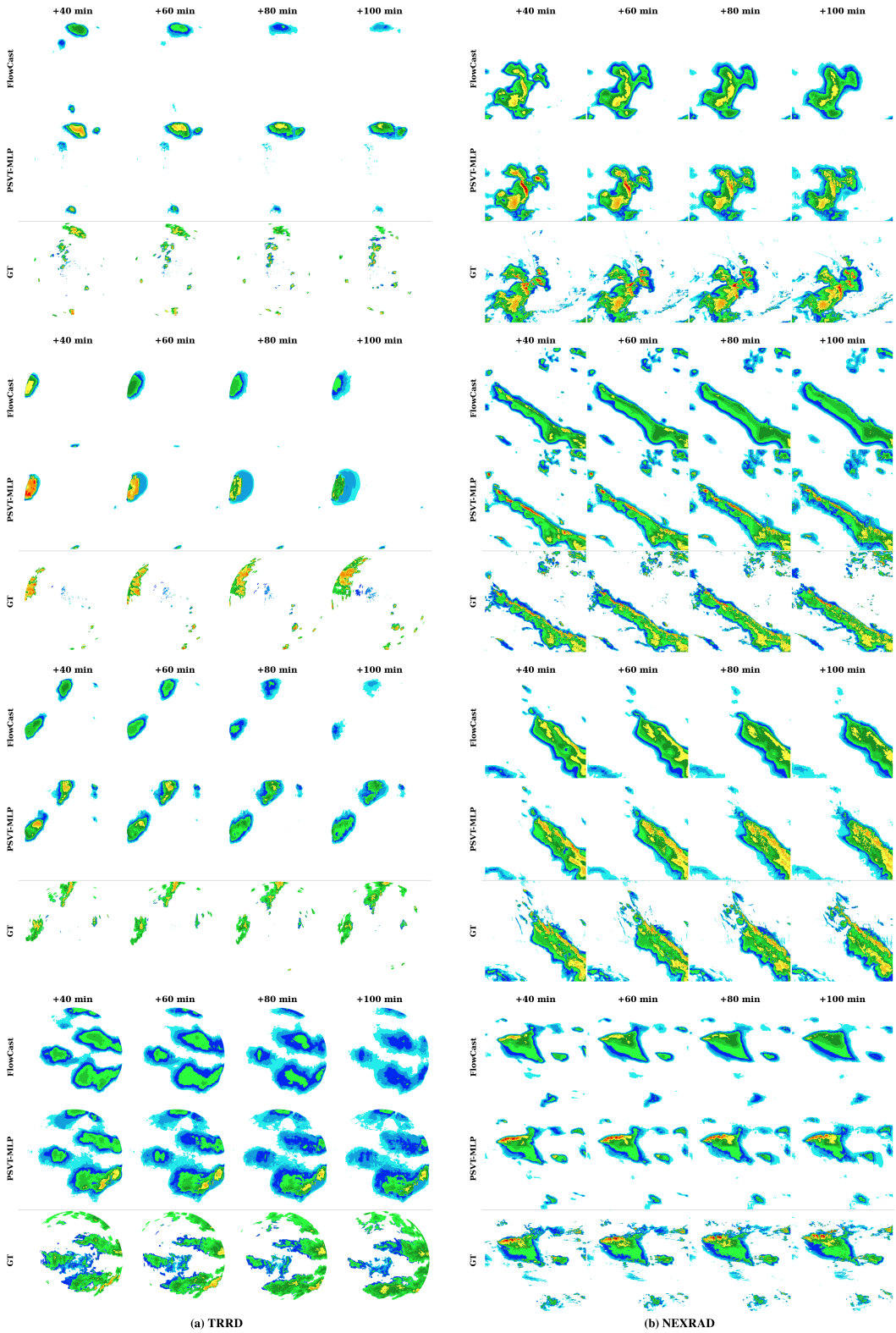

Figure 4: **Additional qualitative visualization of column-maximum reflectivity.** Four cases from NEXRAD-3Y and four cases from TRRD compare FlowCast, PSVT, and the ground truth. As in Figure 3, FlowCast produces smooth fields and underestimates the high-intensity echoes. PSVT keeps these echoes at their observed locations, which is visually consistent with the ground truth.

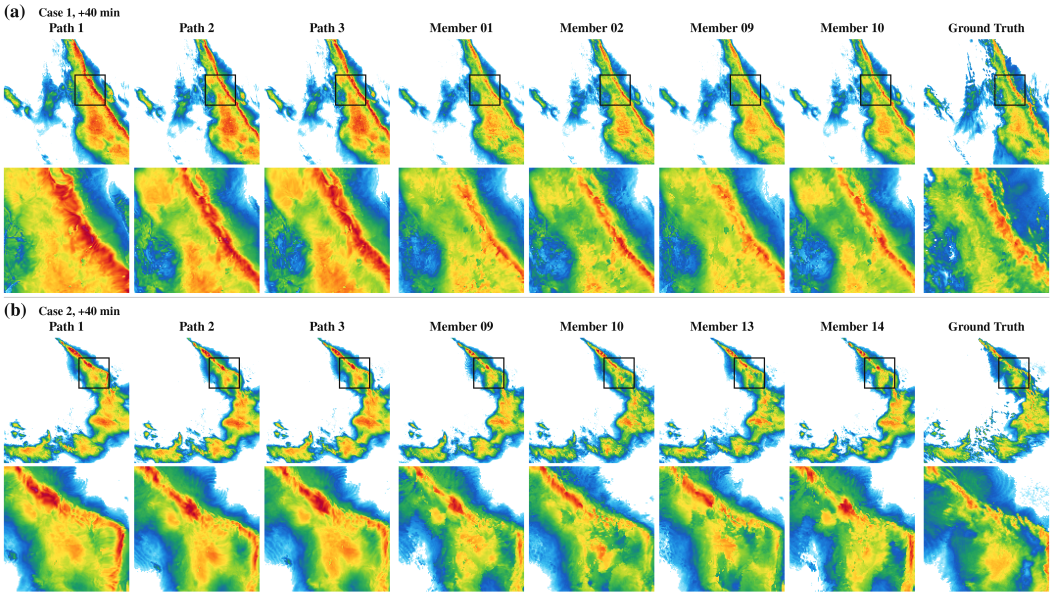

Figure 5: **Transport paths and ensemble members.** Two NEXRAD-3Y cases at +40 min, shown as column-maximum reflectivity. Columns show the $K=3$ path-wise predictions, four ensemble members, and the ground truth. The lower row of each case enlarges the boxed region.

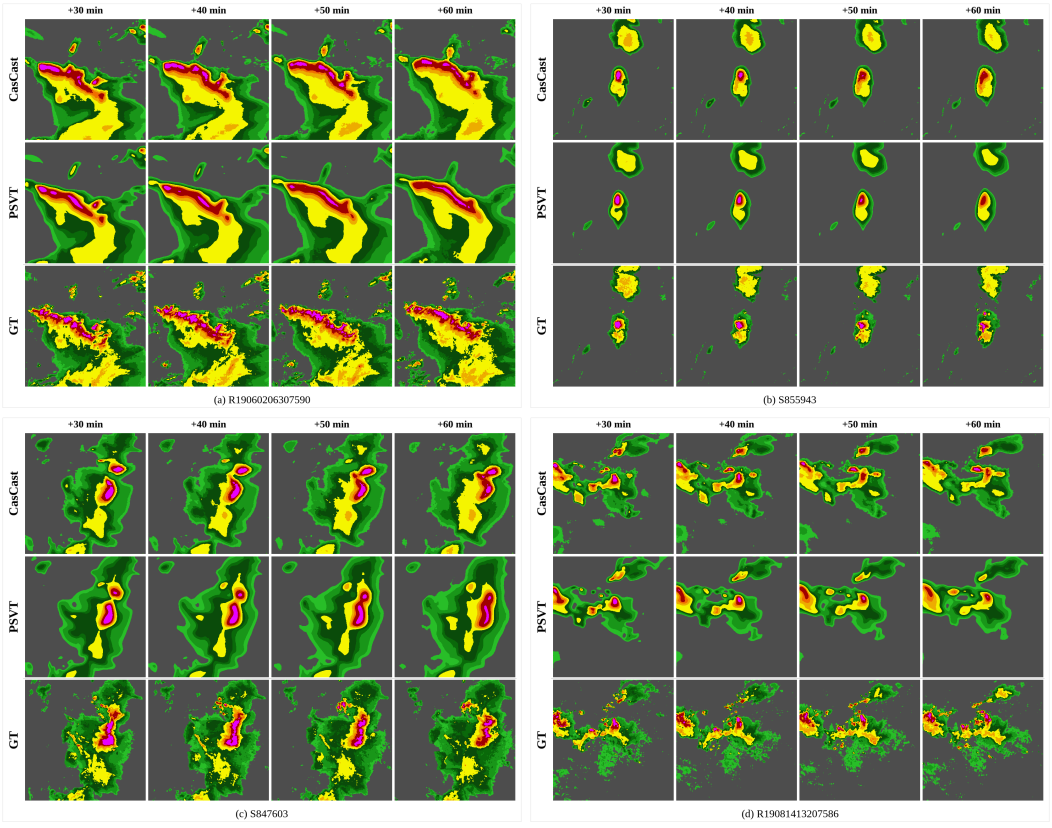

Figure 6: **Qualitative visualization of vertically integrated liquid (VIL) on SEVIR.** Four selected test cases are shown. Within each panel, rows correspond to CasCast, PSVT, and ground truth (GT), and columns correspond to +30, +40, +50, and +60 min. Both methods are represented by the mean of 10 ensemble members, with a shared color scale across all panels.

E LIMITATIONS

PSVT forms ensemble members from K shared volumetric factors and local Gaussian distributions. Where path spread is nonzero, the variation fields $u^{(m)}$ are restricted to the span of the K shared factors; the zero-spread fallback adds independent rather than structured variations. Individual members can thus contain fewer fine-scale details than samples from generative models. This is a trade-off for non-iterative ensemble formation. PSVT also has lower SSIM than FlowCast on both 3D datasets. Away from transported echoes, new precipitation relies on the bounded reflectivity update and local-distribution mapping.