AIF-TVR: BELIEF-GUIDED FRAME ACQUISITION FOR TIME-VARYING 3D FLOW RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing a time-varying 3D flow sequence when only a few complete high-resolution (HR) frames can be acquired requires deciding which frames to observe and how to use them. We present AIF-TVR, an active-inference-inspired framework that couples sequential frame acquisition with conditional diffusion reconstruction. A sequence-conditioned temporal belief incorporates acquired evidence at queried time steps, while an observation predictor scores candidates through hypothetical belief updates. After the budget is exhausted, the acquired low-resolution (LR)–HR pairs supervise target-specific adaptation of the diffusion reconstructor. Across three flow datasets at $4\times$ spatial super-resolution, AIF-TVR achieved the best peak signal-to-noise ratio (PSNR), mean squared error (MSE), and mean absolute error (MAE) among the evaluated methods. Its PSNR gains over the strongest baseline on each dataset ranged from 0.56 to 2.87 dB. These results show that belief-guided acquisition can allocate a limited observation budget without accessing unqueried HR fields.

1 INTRODUCTION

Time-varying three-dimensional (3D) flow fields resolve fine-scale spatial structures and their temporal evolution, but storing, transferring, and processing every field at high resolution (HR) incurs substantial costs (Han & Wang, 2019; Gu et al., 2021). Prior work reconstructs flow from limited measurements (Dubois et al., 2022) and infers 3D turbulence from 2D observations (Yousif et al., 2023). Spatial and spatiotemporal super-resolution recover detail from low-resolution (LR) flow data (Fukami et al., 2019; 2021; Guo et al., 2020; Han & Wang, 2022). Volumetric generative networks, continuous space-time representations, and physics-informed models extend these reconstruction tools (Yu et al., 2022; Jiang et al., 2020; Bode et al., 2021). Learned flow maps, probabilistic flow-based models, and conditional diffusion offer other reconstruction routes (Sahoo et al., 2022; Shen & Shen, 2024; Shu et al., 2024; Guo et al., 2024; Li et al., 2025). Better reconstruction models, however, do not by themselves determine which scarce HR observations to acquire. CD-TVD adapts a conditional diffusion model using one HR frame selected according to LR entropy (Bi et al., 2025). We study the budgeted setting in which the complete LR sequence is available, but several HR frames can be acquired. Which frames should be selected sequentially so that their supervision benefits reconstruction of the entire sequence?

Related work establishes that scarce measurements can be chosen by expected information gain (Houlsby et al., 2011) or optimized for partial differential equation (PDE) inversion (Hemachandra et al., 2025). Data assimilation also updates state estimates from sparse observations (Singh et al., 2024; Si & Chen, 2025; Xiao et al., 2026). These objectives differ from selecting complete HR flow frames to supervise subsequent diffusion adaptation. Here a frame's value is its effect on full-sequence reconstruction, yet its HR content is unknown when it must be selected. The resulting error reduction cannot be evaluated directly. LR uncertainty is only a proxy because it specifies neither the prospective observation nor its induced belief update. Independently ranked frames may also duplicate temporal information. We therefore evaluate plausible outcomes from the LR sequence while accounting for the temporal locations of earlier acquisitions.

We address this problem with AIF-TVR, motivated by active inference's evaluation of actions through predicted outcomes and belief change (Friston et al., 2015; Parr & Friston, 2017; Parr et al., 2022). The framework separates belief-guided acquisition from observation-guided recon-

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.

struction. A temporal encoder uses the complete LR sequence to initialize factorized Gaussian beliefs over the unknown HR latent sequence. For each unobserved time step, an observation predictor estimates a distribution over its possible HR latent observation using only LR information. Samples from this distribution are incorporated through hypothetical analytic Gaussian variational-free-energy updates, allowing the candidate to be evaluated before its HR field is acquired. An expected-free-energy-inspired score combines predictive risk, expected belief change, acquisition cost, and temporal coverage to rank the candidates. After each acquisition, AIF-TVR updates the queried belief marginal and the selected-frame history before choosing the next time step. Once the budget is exhausted, the acquired LR–HR pairs provide target-specific supervision for adapting a conditional diffusion reconstructor, whose shared parameters are then used to recover every HR field in the sequence. The acquisition score is therefore an operational surrogate for observation value rather than a direct estimate of reconstruction-error reduction.

Our contributions are threefold. We formulate time-varying 3D flow reconstruction as budgeted HR frame acquisition followed by full-sequence reconstruction after target-specific adaptation. We develop an observation-predictive acquisition strategy that evaluates hypothetical Gaussian belief updates without accessing unqueried HR fields. Its score balances anticipated belief revision against predictive risk, acquisition cost, and temporal coverage. Across three flow datasets at $4\times$ spatial super-resolution, AIF-TVR achieves the best PSNR, MSE, and MAE among the compared reconstruction methods. Its PSNR improvements range from 0.56 to 2.87 dB over the strongest baseline on each dataset. Additional studies examine observation budget, adaptation duration, and auxiliary-loss weighting.

2 PRELIMINARIES

AIF-TVR connects two components with distinct roles. Conditional diffusion reconstructs an HR flow field from its LR condition, whereas active inference provides a belief-space principle for deciding which HR fields to acquire. We use t to index physical time and τ to index the diffusion process.

2.1 CONDITIONAL DIFFUSION FOR FLOW-FIELD RECONSTRUCTION

Let x_t be an HR flow field and $y_t = \mathcal{D}(x_t)$ its LR observation at physical time step t. Conditional diffusion learns to estimate the clean HR field from a noisy HR state while using y_t as a fixed condition (Saharia et al., 2023; Bi et al., 2025). During training, the forward process perturbs x_t at diffusion step τ:

$$x_{t,\tau} = \sqrt{\bar{\alpha}_\tau}\, x_t + \sqrt{1-\bar{\alpha}_\tau}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \tag{1}$$

where $\bar{\alpha}_\tau$ is the cumulative noise schedule. The LR observation is not corrupted and remains available to the clean-field predictor.

To reconstruct an HR field, the reverse process starts from Gaussian noise $x_{t,K}$ and repeatedly conditions on the same LR field y_t. Here K denotes the initial diffusion level in the sampling schedule. The predicted clean field determines the noise component in the next denoising diffusion implicit model (DDIM) step (Song et al., 2021). For deterministic sampling, the update is

$$\begin{aligned} \widehat{x}_t^{(\tau)} &= f_\theta(x_{t,\tau}, y_t, \tau), \\ \widehat{\epsilon}_{t,\tau} &= \frac{x_{t,\tau} - \sqrt{\bar{\alpha}_\tau}\, \widehat{x}_t^{(\tau)}}{\sqrt{1-\bar{\alpha}_\tau}}, \\ x_{t,\tau-1} &= \sqrt{\bar{\alpha}_{\tau-1}}\, \widehat{x}_t^{(\tau)} + \sqrt{1-\bar{\alpha}_{\tau-1}}\, \widehat{\epsilon}_{t,\tau}. \end{aligned} \tag{2}$$

Here $\tau - 1$ denotes the next level in the chosen sampling schedule, and the final state $x_{t,0}$ is the reconstructed HR field.

2.2 VARIATIONAL AND EXPECTED FREE ENERGY

Variational free energy (VFE) provides a principle for updating a belief after an observation is acquired (Parr & Friston, 2017; Parr et al., 2022). Let $b(s)$ be a prior belief over a latent state s, $p(o \mid s)$

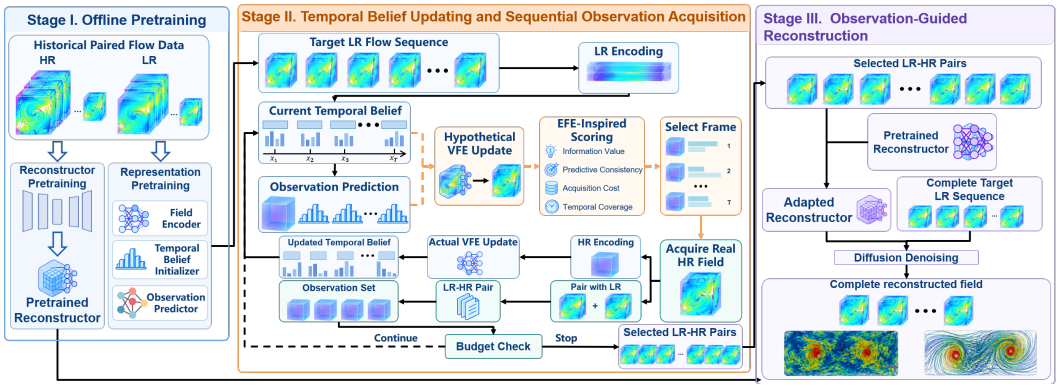


Figure 1: Overview of AIF-TVR. Stage I pretrains the conditional diffusion reconstructor, shared representation encoder, temporal belief initializer, and observation predictor. Stage II evaluates candidate frames through hypothetical Gaussian belief updates, acquires real HR fields, and updates the queried marginals and selected-frame history. Stage III adapts the reconstructor using the acquired LR–HR pairs and reconstructs the complete target sequence.

an observation model, and $q(s)$ a candidate posterior. For an observed outcome o, VFE is

$$\mathcal{F}(q; o) = \mathbb{E}_{q(s)}[-\log p(o \mid s)] + D_{\mathrm{KL}}(q(s) \,\|\, b(s)) \,. \tag{3}$$

Minimizing this objective balances observation fit against deviation from the prior belief. In our setting, the belief over the HR latent sequence is diagonal Gaussian, and an acquired latent follows an isotropic Gaussian observation model. These assumptions yield the closed-form Gaussian VFE update specified in Section 3.3. Because the belief factorizes across time, this update changes only the queried marginal. Temporal context enters through the sequence-conditioned initialization of the belief.

Action selection instead concerns observations that have not yet been acquired. A common expected-free-energy (EFE) decomposition evaluates a candidate policy π through its predicted outcomes and belief changes (Friston et al., 2015; Parr & Friston, 2017; Millidge et al., 2021; Parr et al., 2022):

$$G_{\mathrm{EFE}}(\pi) \simeq \mathbb{E}_{Q(o|\pi)}[-\log P^{\star}(o)] - \mathbb{E}_{Q(o|\pi)} D_{\mathrm{KL}}(Q(s \mid o, \pi) \,\|\, Q(s \mid \pi)) \,, \tag{4}$$

where $P^{\star}(o)$ represents preferred outcomes. The two terms express preference-related cost and anticipated information gain, respectively. In AIF-TVR, a candidate action selects an HR time step. We estimate its consequences using predicted latent observations and hypothetical Gaussian VFE updates. The acquisition score replaces the theoretical preference cost with a predictive-consistency KL and adds acquisition cost and temporal coverage. It is therefore an EFE-inspired, task-specific surrogate rather than an exact evaluation of Eq. 4.

3 Overview of AIF-TVR

Given the complete LR sequence and a limited HR acquisition budget, AIF-TVR maintains a sequence-wide belief over the unknown HR fields, as shown in Figure 1. The belief is initialized from the LR sequence and updated at each queried time step. Before acquisition, an observation predictor generates hypothetical HR latents for scoring expected belief revision, predictive consistency, cost, and temporal coverage. After the budget is exhausted, the acquired LR–HR pairs adapt a conditional diffusion reconstructor that is applied to every LR time step. The following sections define the task and the three processing stages.

3.1 Problem Formulation

At test time, the complete low-resolution (LR) flow sequence $\mathcal{Y}_{1:T} = \{y_t\}_{t=1}^{T}$ is available, but the high-resolution (HR) fields x_t are unknown. They satisfy $y_t = \mathcal{D}(x_t)$ for a spatial degradation

operator $\mathcal{D}$. We may acquire at most $B < T$ HR frames. After k acquisitions, S_k contains the selected time steps and $\mathcal{Q}_k = \{(y_t, x_t) : t \in S_k\}$ contains their LR–HR pairs. A policy π selects each new time step using $\mathcal{Y}_{1:T}$ and $\mathcal{Q}_k$, without access to unqueried HR fields.

The acquired pairs adapt a pretrained conditional reconstructor, giving $\theta_B^\pi = \mathrm{Adapt}(\theta_0, \mathcal{Q}_B^\pi)$ and $\widehat{x}_t^\pi = \mathcal{R}_{\theta_B^\pi}(y_t)$ at every time step. We seek a policy whose selected frames minimize the expected reconstruction error over the entire HR sequence:

$$\pi^\star = \arg\min_{\pi:\,|S_B^\pi| \le B} \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \mathcal{L}_{\mathrm{rec}}(\widehat{x}_t^\pi, x_t)\right]. \tag{5}$$

Here $\mathcal{L}_{\mathrm{rec}}$ measures discrepancy from the HR target, and the expectation covers target sequences and reconstruction randomness. This is the task objective, not a computable selection score: the policy cannot evaluate it before acquisition because unqueried HR fields are unknown.

3.2 Offline model preparation

Following the conditional diffusion and degradation-aware representation components of CD-TVD (Bi et al., 2025), we pretrain a reconstructor $\mathcal{R}_{\theta_0}$ and an encoder $\mathcal{E}$ on historical LR–HR sequences. The encoder maps the HR-grid-aligned LR field $\bar{y}_t = \mathcal{A}(y_t)$ and its HR counterpart to $c_t = \mathcal{E}(\bar{y}_t)$ and $z_t = \mathcal{E}(x_t)$ in $\mathbb{R}^{d_z}$. We denote their sequences by $c_{1:T}$ and $Z_{1:T}$. A temporal belief initializer and an observation predictor are also trained on these paired latents. Only the reconstructor is adapted on the target sequence, while the remaining components stay fixed. Training and information-access details are given in Appendix A.1.

3.3 Belief-guided budgeted acquisition

Gaussian temporal belief. The LR sequence describes coarse flow evolution but leaves the degraded spatial details unresolved. We therefore represent uncertainty over the unknown target HR latents by a belief over $Z_{1:T}$. After k observations, this belief is a factorized diagonal Gaussian distribution:

$$b_k(Z_{1:T}) = \prod_{t=1}^{T} \mathcal{N}\big(z_t \mid \mu_t^k, \mathrm{diag}(v_t^k)\big), \tag{6}$$

where $\mu_t^k \in \mathbb{R}^{d_z}$ and $v_t^k \in \mathbb{R}^{d_z}_{>0}$ denote the belief mean and variance at time step t. A temporal encoder predicts the initial parameters $\{\mu_t^0, v_t^0\}_{t=1}^T$ jointly from the complete condition sequence $c_{1:T}$. Thus, despite temporal factorization, each initial marginal incorporates sequence-level LR context rather than only its corresponding frame.

When time step $a \notin S_k$ is acquired, its encoded HR latent is $o_a = z_a$. We use $p(o_a \mid z_a) = \mathcal{N}(o_a \mid z_a, r\mathbf{I})$, with effective latent observation variance $r > 0$. This models uncertainty in the latent update rather than measurement noise in the HR reference, and avoids directly replacing the belief mean with o_a.

We incorporate the acquired evidence by minimizing variational free energy (VFE). For a candidate posterior $q(Z_{1:T})$, the update is defined by

$$\begin{aligned} \mathcal{F}_k(q \mid o_a) &= \mathbb{E}_{q(Z_{1:T})}[-\log p(o_a \mid z_a)] + D_{\mathrm{KL}}(q \,\|\, b_k), \\ b_{k+1} &= \arg\min_q \mathcal{F}_k(q \mid o_a). \end{aligned} \tag{7}$$

The two terms balance observation fit against deviation from the current belief.

Under these Gaussian assumptions, the VFE minimizer has a closed form. For latent coordinate $j \in \{1, \ldots, d_z\}$, the queried marginal becomes

$$\begin{aligned} v_{a,j}^{k+1} &= \big((v_{a,j}^k)^{-1} + r^{-1}\big)^{-1}, \\ \mu_{a,j}^{k+1} &= v_{a,j}^{k+1}\left(\frac{\mu_{a,j}^k}{v_{a,j}^k} + \frac{o_{a,j}}{r}\right). \end{aligned} \tag{8}$$

The mean combines prior and observation by precision, while $v_{a,j}^{k+1} < v_{a,j}^{k}$. Every unqueried marginal remains unchanged: $(\mu_t^{k+1}, v_t^{k+1}) = (\mu_t^k, v_t^k)$ for $t \neq a$. Temporal context enters through the LR-sequence-conditioned initialization, not through cross-time propagation of an acquired latent.

Observation-predictive selection. Marginal uncertainty alone does not specify which HR observation would be obtained. For each unobserved candidate $a \in \mathcal{A}_k = \{1, \dots, T\} \setminus S_k$, an observation predictor estimates $q_\phi(o_a \mid c_{1:T}) = \mathcal{N}(o_a \mid m_a, \mathrm{diag}(s_a))$ from the complete LR condition sequence. Here m_a and $s_a \succ 0$ are its predicted mean and diagonal variance. A hypothetical sample $\widetilde{o}_a \sim q_\phi(\cdot \mid c_{1:T})$ yields $\widetilde{b}_{k+1}^a = \mathcal{U}_{\mathrm{VFE}}(b_k, a, \widetilde{o}_a)$ through Eq. 8. The temporary belief does not change the actual acquisition state.

We score the expected, dimension-normalized belief change as

$$I_k(a) = \frac{1}{Td_z} \mathbb{E}_{\widetilde{o}_a \sim q_\phi} \left[D_{\mathrm{KL}} \left(\widetilde{b}_{k+1}^a \,\middle\|\, b_k \right) \right]. \tag{9}$$

A larger $I_k(a)$ means a larger anticipated belief revision, not a proven reduction in reconstruction error or an exact mutual information estimate. During selection, we substitute its M-sample estimate (Appendix A.2) into Eq. 11.

To penalize disagreement between the observation predictor and the current belief, we set $p_k^\star(o_a) = \mathcal{N}(o_a \mid \mu_a^k, \mathrm{diag}(v_a^k))$ and define predictive risk as

$$R_k(a) = \frac{1}{d_z} D_{\mathrm{KL}}(q_\phi(o_a \mid c_{1:T}) \,\|\, p_k^\star(o_a)). \tag{10}$$

A larger $R_k(a)$ indicates greater disagreement rather than a reconstruction loss or physical utility.

We also include nonnegative acquisition cost $C(a)$ and temporal coverage $H_k(a)$, the normalized distance to the nearest acquired time step (zero when $S_k = \varnothing$). Larger coverage discourages temporal concentration, with the exact expression given in Appendix A.2.

We group the acquisition criteria into a belief-based score $G_k^{\mathrm{bel}}(a) = \lambda_r R_k(a) - \lambda_i I_k(a)$ and a design score $G_k^{\mathrm{des}}(a) = \lambda_c C(a) - \lambda_h H_k(a)$, where λ_r, λ_i, λ_c, and λ_h are nonnegative weights. The next frame is selected by

$$G_k(a) = G_k^{\mathrm{bel}}(a) + G_k^{\mathrm{des}}(a), \qquad a_{k+1} = \arg\min_{a \in \mathcal{A}_k} G_k(a). \tag{11}$$

After selection, the real HR field updates its queried belief marginal and is retained for adaptation. The next round excludes that frame. Under the factorized belief and LR-only predictor, newly acquired values do not change I_k or R_k for unqueried candidates. Subsequent decisions instead reflect the updated candidate set and temporal coverage. Hypothetical observations are discarded after scoring.

3.4 Observation-guided adaptation and reconstruction

After all B frames are acquired, $\mathcal{Q}_B = \{(y_a, x_a) : a \in S_B\}$ supplies target-specific supervision to the pretrained conditional diffusion reconstructor. We fine-tune it from θ_0 with

$$\mathcal{L}_{\mathrm{adapt}}(\theta, \mathcal{Q}_B) = \mathcal{L}_{\mathrm{fit}}(\theta, \mathcal{Q}_B) + \beta \mathcal{L}_{\mathrm{deg}}(\theta, \mathcal{Q}_B). \tag{12}$$

Here $\mathcal{L}_{\mathrm{fit}}$ is the mean voxel-wise ℓ_1 loss on predicted clean HR fields. The degradation-aware representation loss $\mathcal{L}_{\mathrm{deg}}$ has weight $\beta \geq 0$. Both use only acquired LR–HR pairs, as detailed in Appendix A.3. After a fixed number of optimization steps, the adapted parameters θ_B are shared across all time steps: $\widehat{x}_t = \mathcal{R}_{\theta_B}(y_t)$ for $t = 1, \dots, T$. Even at acquired times, outputs come from the reconstructor rather than direct insertion of the HR observation. The belief determines which frames to measure, while real observations determine how the reconstructor is adapted.

4 Experiments

Experimental setup. Following the dataset and evaluation protocol of CD-TVD (Bi et al., 2025), we evaluate three time-varying 3D flow datasets and compare SRGAN (Ledig et al., 2017), SSR-VFD (Guo et al., 2020), and CD-TVD under a common $4\times$ spatial super-resolution protocol. Data

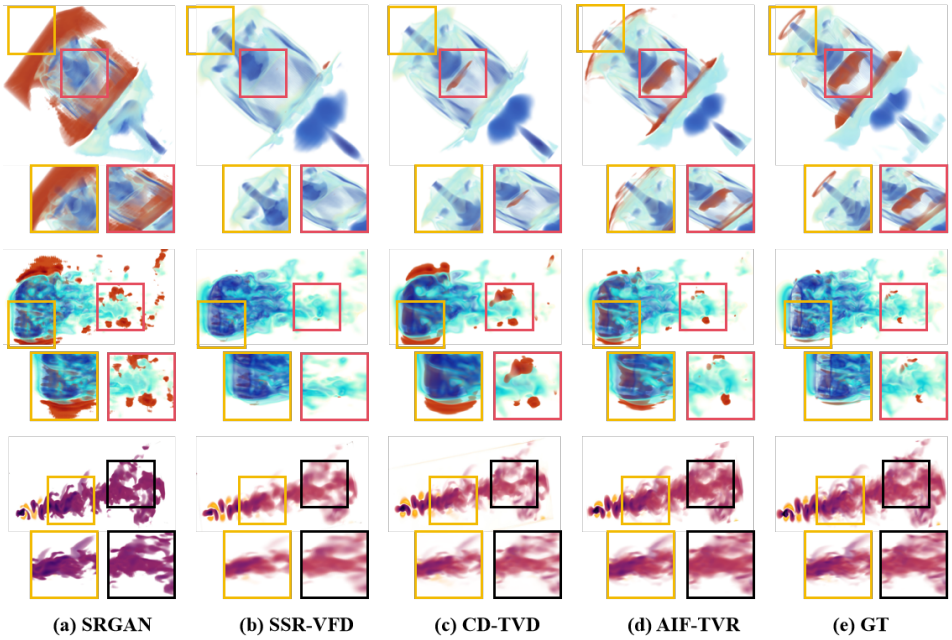


Figure 2: Volume-rendering comparison across the three evaluated datasets. Rows from top to bottom show Shock Interaction Vortex, Research Vessel Tangaroa, and Half Cylinder Ensemble. GT denotes the original HR ground-truth field.

partitions, LR generation, HR observation budgets, adaptation schedules, optimization settings, and rendering configuration are matched across methods. Complete dataset descriptions, baseline configurations, and fairness details are provided in Appendix B.1.

4.1 Comparison with existing methods

Visualization. Figure 2 shows that AIF-TVR more closely reproduces the reference flow structures across the three datasets. In Shock Interaction Vortex, the marked red structures retain their shape and continuity. In Research Vessel Tangaroa, vessel-adjacent boundaries and local color distributions are better resolved. In Half Cylinder Ensemble, the global pattern and local details more closely match the reference. These differences are consistent with target-specific HR supervision helping the conditional reconstructor recover details that are ambiguous in the LR input. The volume renderings illustrate structural differences, while the voxel-wise results below quantify reconstruction fidelity.

Quantitative comparison. Table 1 shows that AIF-TVR achieves the highest PSNR and the lowest MSE and MAE among the compared methods on all datasets. On Half Cylinder Ensemble, the target scenario at $Re = 6400$ differs from the low-Reynolds-number simulations used for pretraining. Relative to the strongest baseline, SSR-VFD, AIF-TVR improves PSNR by 2.87 dB and reduces MSE by 48.26%. Target-sequence HR supervision and shared-parameter adaptation provide a plausible explanation for improved reconstruction under this mismatch. On Shock Interaction Vortex, AIF-TVR exceeds CD-TVD by 1.77 dB and reduces MSE by 33.50%. This is consistent with acquired HR observations helping resolve details that remain ambiguous in the LR input. On Tangaroa, AIF-TVR improves PSNR by 0.56 dB over SSR-VFD and reduces MSE by 12.19%. Lower MAE across all datasets provides complementary evidence of reduced average absolute error. These comparisons assess the complete frameworks rather than isolate the contribution of the acquisition score.

Temporal comparison. Figure 3 examines whether the sequence-average gains over CD-TVD are concentrated at a few time steps. AIF-TVR has higher PSNR on all 40 Tangaroa test frames and all 51 ShockVortex test frames. The Tangaroa gap remains broadly stable. Although both ShockVortex

Table 1: Average PSNR, MSE, and MAE at $4\times$ spatial super-resolution. Higher PSNR and lower MSE or MAE indicate better reconstruction. The best value in each dataset and metric is bold.

Dataset	Method	PSNR (dB) ↑	MSE ↓	MAE ↓
Half Cylinder Ensemble	SRGAN	33.79	0.001670	0.02992
	SSR-VFD	36.96	0.000806	0.01496
	CD-TVD	36.36	0.000923	0.01696
	AIF-TVR	**39.83**	**0.000417**	**0.01373**
Shock Interaction Vortex	SRGAN	28.48	0.005674	0.05244
	SSR-VFD	31.61	0.002760	0.02994
	CD-TVD	38.28	0.000594	0.01545
	AIF-TVR	**40.05**	**0.000395**	**0.01245**
Research Vessel Tangaroa	SRGAN	38.26	0.0005968	0.01332
	SSR-VFD	43.87	0.0001641	0.00525
	CD-TVD	40.39	0.0003654	0.00831
	AIF-TVR	**44.43**	**0.0001441**	**0.00452**

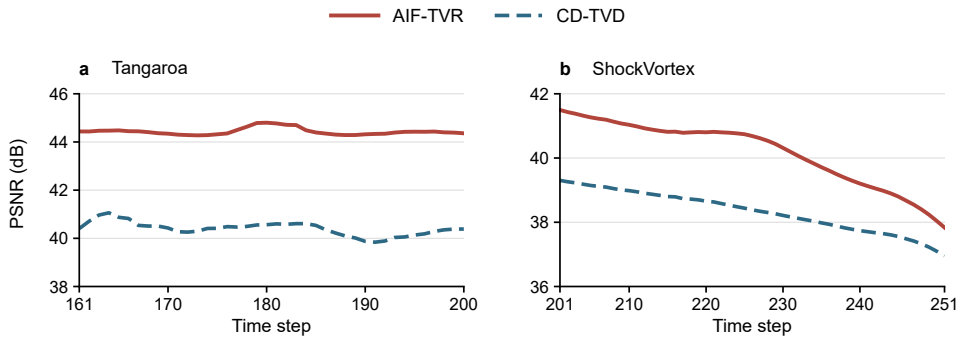


Figure 3: Frame-by-frame PSNR at $4\times$ spatial super-resolution on (a) Research Vessel Tangaroa (time steps 161–200) and (b) Shock Interaction Vortex (time steps 201–251). Curves use the corresponding comparison runs in Table 1. Higher is better, and the PSNR axis ranges differ between panels.

curves decline toward later frames, AIF-TVR retains its advantage throughout. The same adapted reconstructor generates every output, including frames without an acquired HR observation. Therefore, the persistent advantage is consistent with the acquired supervision transferring beyond the selected time steps, although it does not establish that the acquisition score is responsible for the gain. The Tangaroa curve compares AIF-TVR only with CD-TVD, whereas the smaller margin over SSR-VFD is established by Table 1.

4.2 Ablation Studies

Variant definitions. Table 2 lists five named variants and the budget variation M3-B8. M1 selects one frame by LR entropy and retains the contrastive adaptation term, so it is not an exact reproduction of pure-ℓ_1 CD-TVD adaptation. M2 uses observation-predictive EFE selection with ℓ_1 adaptation only. M3 is the full single-observation configuration, whereas M4 removes its expected-belief-revision term I_k. M3-B8 increases the budget from one to eight frames, and M5 disables the recorded belief-update switch at that budget. Under the factorized Gaussian model, this switch does not alter the exact belief-based scores of unqueried candidates. Thus, M3-B8 versus M5 does not establish cross-time propagation of acquired HR evidence.

Component effects. Figure 4 shows that adaptation behavior is strongly dataset dependent. Among the controlled variants, all settings reduce Tangaroa PSNR relative to the frozen SR checkpoint, with M2 producing the smallest decrease (-0.13 dB) and M4 the largest (-1.75 dB). Re6400 changes remain small (-0.04 to $+0.12$ dB), whereas every controlled ShockVortex variant improves the checkpoint, led by M2 at $+0.74$ dB. The headline $\mathrm{M3}^{\dagger}$ results are positive on all three datasets

Table 2: Core ablation variants. B is the HR frame budget, and Observation EFE is the observation-predictive acquisition score. The final column records the belief-update switch, not cross-time evidence propagation. $\mathrm{M3}^{\dagger}$ uses the separate framework-comparison result.

Variant	Frame selection	Adaptation loss	Update switch
M1	LR entropy, $B = 1$	ℓ_1 + contrastive, $\beta = 0.1$	Enabled
M2	Observation EFE, $B = 1$	ℓ_1 only	Enabled
$\mathrm{M3}^{\dagger}$	Observation EFE, $B = 1$	ℓ_1 + contrastive, $\beta = 0.1$	Enabled
M4	Observation EFE without I_k, $B = 1$	ℓ_1 + contrastive, $\beta = 0.1$	Enabled
M3-B8	Observation EFE, $B = 8$	ℓ_1 + contrastive, $\beta = 0.1$	Enabled
M5	Observation EFE, $B = 8$	ℓ_1 + contrastive, $\beta = 0.1$	Disabled

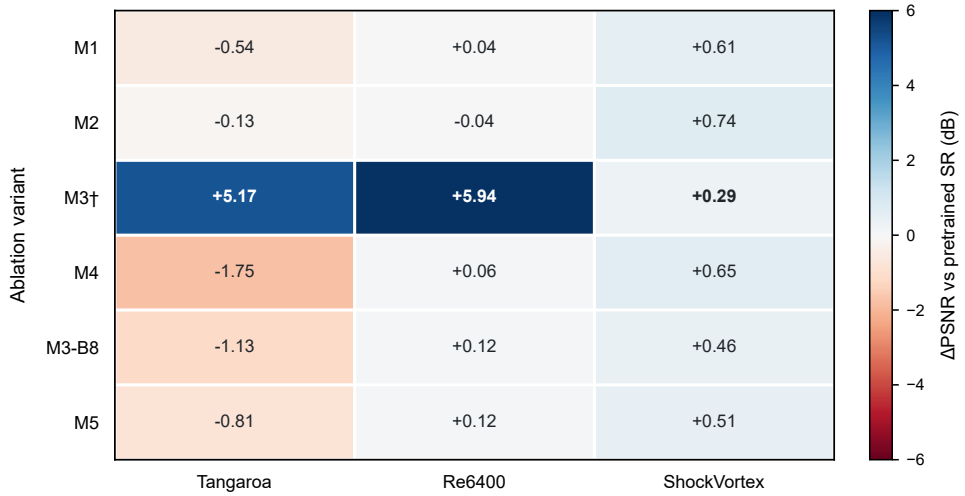


Figure 4: Ablation outcomes relative to the frozen pretrained SR model. Each cell reports the change in full-sequence PSNR after adaptation, with positive values indicating improvement. Rows denote the variants in Table 2, and columns denote the three datasets. $\mathrm{M3}^{\dagger}$ uses a separate framework-comparison run.

($+5.17$, $+5.94$, and $+0.29$ dB). However, their different run provenance prevents causal attribution to an individual module. Target-specific adaptation is therefore not uniformly beneficial, and small differences require repeated-seed validation. Complete metrics and protocol details are provided in Appendix C.

Hyperparameter sensitivity. Figure 5 reports a separate sensitivity protocol, whose absolute PSNR values are not directly comparable with Table 1. Budget sensitivity is non-monotonic, with eight frames performing best and no further gain at 16 frames. Increasing adaptation from 100 to 1,500 steps improves PSNR, whereas changing the auxiliary-loss weight yields a range of only 0.0019 dB. These observations characterize the tested operating points rather than establish a universally optimal budget or loss weight. Additional values and the frozen baseline are reported in Appendix C.2.

Physical consistency. Figure 6 tests whether reconstruction fidelity extends to flow structure, using the diagnostic definitions in Appendix C.3. AIF-TVR reduces median vorticity relative L_2 error from 0.6333 to 0.4117 on Tangaroa, from 0.4945 to 0.3914 on Re6400, and from 0.2432 to 0.2128 on ShockVortex. These changes correspond to reductions of 35.0%, 20.8%, and 12.5%, respectively. Energy preservation is less consistent, with normalized error decreasing by 76.2% on Tangaroa but increasing by 430.0% on Re6400 and 22.0% on ShockVortex. Thus, the diagnostic runs recover rotational structure more accurately, while improved PSNR does not guarantee preservation of global kinetic energy.

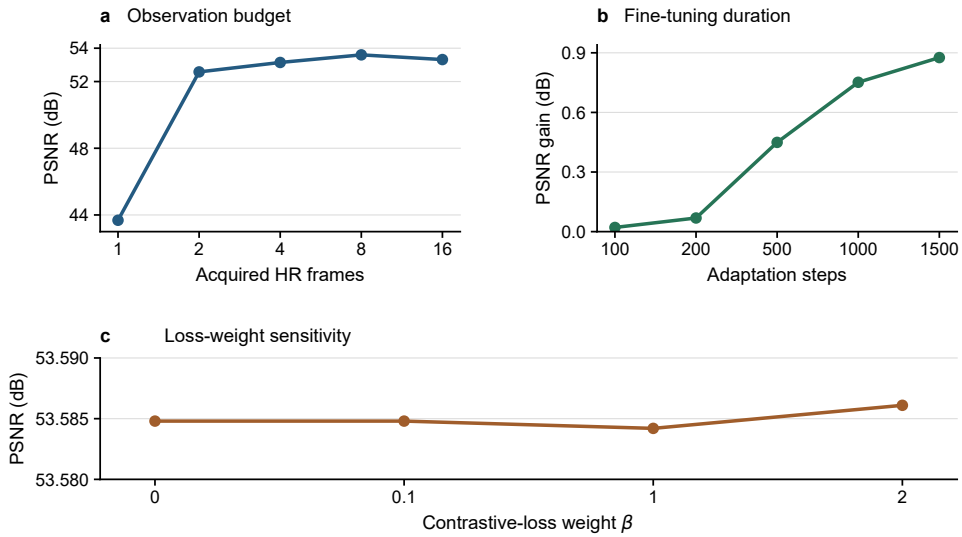


Figure 5: Sensitivity to (a) observation budget, (b) fine-tuning duration, and (c) contrastive-loss weight.

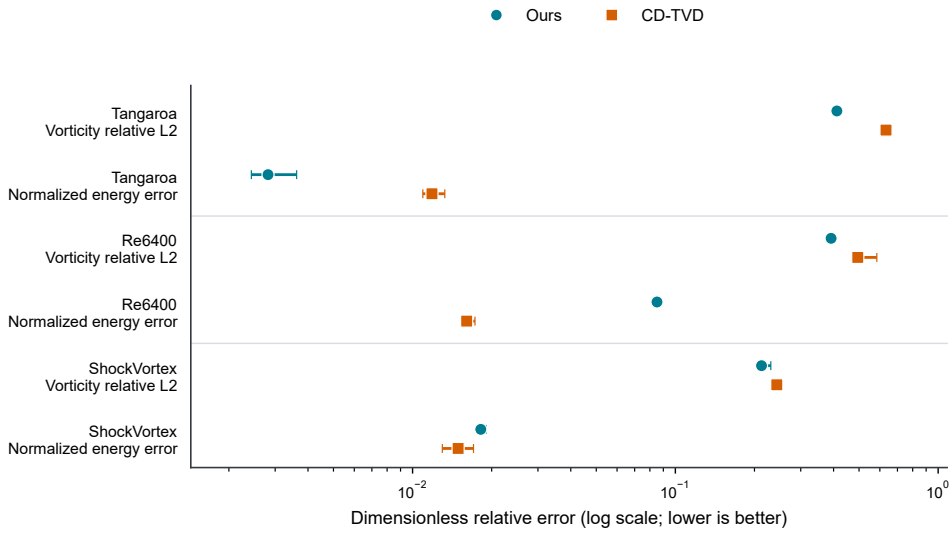


Figure 6: Physical-consistency errors for AIF-TVR and CD-TVD. Points show temporal medians and bars show interquartile ranges over 21 sampled test frames per dataset. Both metrics are dimensionless and lower is better.

5 Conclusion

We addressed full-sequence 3D flow reconstruction when the complete LR sequence is available but only a limited number of HR frames can be acquired. AIF-TVR evaluates hypothetical Gaussian belief updates to select frames and uses their LR–HR pairs to adapt a conditional diffusion reconstructor. It achieves the best reconstruction metrics among the compared methods on three datasets, although the current factorized belief does not propagate acquired values across time. The acquisition surrogate also does not guarantee reconstruction improvement, and physical diagnostics reveal remaining energy-preservation errors. Future work should test correlated temporal beliefs, physics-aware adaptation, and robustness across broader flow regimes and repeated random seeds.

REFERENCES

Chongke Bi, Xin Gao, Jiakang Deng, Guan Li, and Jun Han. CD-TVD: Contrastive diffusion for 3D super-resolution with scarce high-resolution time-varying data. *arXiv preprint arXiv:2508.08173*, 2025.

Mathis Bode, Michael Gauding, Zeyu Lian, Dominik Denker, Marco Davidovic, Konstantin Kleinheinz, Jenia Jitsev, and Heinz Pitsch. Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows. *Proceedings of the Combustion Institute*, 38(2):2617–2625, 2021. doi: 10.1016/j.proci.2020.06.022.

Pierre Dubois, Thomas Gomez, Laurent Planckaert, and Laurent Perret. Machine learning for fluid flow reconstruction from limited measurements. *Journal of Computational Physics*, 448:110733, 2022. doi: 10.1016/j.jcp.2021.110733.

Karl Friston, Francesco Rigoli, Dimitri Ognibene, Christoph Mathys, Thomas Fitzgerald, and Giovanni Pezzulo. Active inference and epistemic value. *Cognitive Neuroscience*, 6(4):187–214, 2015. doi: 10.1080/17588928.2015.1020053.

Kai Fukami, Koji Fukagata, and Kunihiko Taira. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics*, 870:106–120, 2019. doi: 10.1017/jfm.2019.238.

Kai Fukami, Koji Fukagata, and Kunihiko Taira. Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows. *Journal of Fluid Mechanics*, 909:A9, 2021. doi: 10.1017/jfm.2020.948.

Ping Gu, Jun Han, and Chaoli Wang. Reconstructing unsteady flow data from representative streamlines via diffusion and deep-learning-based denoising. *IEEE Computer Graphics and Applications*, 41(6):111–121, 2021. doi: 10.1109/MCG.2021.3089627.

L. Guo, S. Ye, J. Han, H. Zheng, H. Gao, D. Z. Chen, J.-X. Wang, and C. Wang. SSR-VFD: Spatial super-resolution for vector field data analysis and visualization. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 71–80, 2020. doi: 10.1109/PacificVis48177.2020.8737.

Yanan Guo, Xiaoqun Cao, Mengge Zhou, Hongze Leng, and Junqiang Song. Enhancing high-resolution reconstruction of flow fields using physics-informed diffusion model with probability flow sampling. *Physics of Fluids*, 36(11):115110, 2024. doi: 10.1063/5.0230525.

Jun Han and Chaoli Wang. Flow field reduction via reconstructing vector data from 3-d streamlines using deep learning. *IEEE Computer Graphics and Applications*, 39(4):54–67, 2019. doi: 10.1109/MCG.2018.2881523.

Jun Han and Chaoli Wang. TSR-VFD: Generating temporal super-resolution for unsteady vector field data. *Computers & Graphics*, 103:168–179, 2022. doi: 10.1016/j.cag.2022.02.001.

Apivich Hemachandra, Gregory Kang Ruey Lau, See-Kiong Ng, and Bryan Kian Hsiang Low. PIED: Physics-informed experimental design for inverse problems. In *International Conference on Learning Representations*, 2025.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Chiyu Max Jiang, Soheil Esmaeilzadeh, Kamyar Azizzadenesheli, Karthik Kashinath, Mustafa Mustafa, Hamdi A. Tchelepi, Philip Marcus, Prabhat, and Anima Anandkumar. MeshfreeFlowNet: A physics-constrained deep continuous space-time super-resolution framework. In *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–15, 2020. doi: 10.1109/SC41405.2020.00013.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 105–114, 2017.

Zeyu Li, Hongkun Dou, Shen Fang, Wang Han, Yue Deng, and Lijun Yang. Physics-aligned field reconstruction with diffusion bridge. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=D042vFwJAM.

Beren Millidge, Alexander Tschantz, and Christopher L. Buckley. Whence the expected free energy? *Neural Computation*, 33(2):447–482, 2021. doi: 10.1162/neco_a_01354.

Thomas Parr and Karl J Friston. Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface*, 14(136):20170376, 2017.

Thomas Parr, Giovanni Pezzulo, and Karl J Friston. *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press, 2022.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2023. doi: 10.1109/TPAMI.2022.3204461.

Saroj Sahoo, Yuzhe Lu, and Matthew Berger. Neural flow map reconstruction. *Computer Graphics Forum*, 41(3):391–402, 2022. doi: 10.1111/cgf.14549.

Jingyi Shen and Han-Wei Shen. PSRFlow: Probabilistic super resolution with flow-based models for scientific data. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):986–996, 2024. doi: 10.1109/TVCG.2023.3327171.

Dule Shu, Zijie Li, and Amir Barati Farimani. A physics-informed diffusion model for high-fidelity flow field reconstruction. *Journal of Computational Physics*, 510:111972, 2024. doi: 10.1016/j.jcp.2023.111972.

Phillip Si and Peng Chen. Latent-EnSF: A latent ensemble score filter for high-dimensional data assimilation with sparse observation data. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=urcEYsZOBz.

Ashutosh Singh, Ricardo Borsoi, Deniz Erdogmus, and Tales Imbiriba. Learning semilinear neural operators: A unified recursive framework for prediction and data assimilation. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ZMv6zKYYUs.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Pengpeng Xiao, Phillip Si, and Peng Chen. LD-EnSF: Synergizing latent dynamics with ensemble score filters for fast data assimilation with sparse observations. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=AWSVzzhbr7.

Mustafa Z. Yousif, Linqi Yu, Sergio Hoyas, Ricardo Vinuesa, and HeeChang Lim. A deep-learning approach for reconstructing 3D turbulent flows from 2D observation data. *Scientific Reports*, 13:2529, 2023. doi: 10.1038/s41598-023-29525-9.

Linqi Yu, Mustafa Z. Yousif, Meng Zhang, Sergio Hoyas, Ricardo Vinuesa, and Hee-Chang Lim. Three-dimensional ESRGAN for super-resolution reconstruction of turbulent flows with tricubic interpolation-based transfer learning. *Physics of Fluids*, 34(12):125126, 2022. doi: 10.1063/5.0129203.

A ADDITIONAL METHOD DETAILS

Algorithm 1 summarizes the complete procedure. Hypothetical observations are used only to score candidate frames. Only acquired HR fields enter the adaptation set.

Algorithm 1 AIF-TVR for budgeted time-varying flow reconstruction

Require: Historical LR–HR pairs $\mathcal{D}_{\text{hist}}$, target LR sequence $\mathcal{Y}_{1:T}$, budget B, sample count M, adaptation steps N_{adapt}
Ensure: Reconstructed HR sequence $\{\widehat{x}_t\}_{t=1}^{T}$

Stage 1: Offline preparation
Train $\mathcal{E}$, $\mathcal{R}_{\theta_0}$, the belief initializer, and q_ϕ on $\mathcal{D}_{\text{hist}}$
Compute $c_t = \mathcal{E}(\mathcal{A}(y_t))$ for all t
Initialize b_0 from $c_{1:T}$
Set $S_0 \leftarrow \varnothing$ and $\mathcal{Q}_0 \leftarrow \varnothing$
Stage 2: Budgeted HR frame acquisition
for $k = 0, \dots, B-1$ **do**
  Set $\mathcal{A}_k = \{1, \dots, T\} \setminus S_k$
  for $a \in \mathcal{A}_k$ **do**
    Predict $q_\phi(o_a \mid c_{1:T})$ and draw M hypothetical observations
    Compute $\widehat{I}_k(a)$ using Eq. 15
    Compute $R_k(a)$ using Eq. 10
    Compute $G_k(a)$ using Eq. 11
  end for
  Select $a_{k+1} = \arg\min_{a \in \mathcal{A}_k} G_k(a)$
  Discard the hypothetical observations
  Acquire the real HR field $x_{a_{k+1}}$
  Set $o_{a_{k+1}} = \mathcal{E}(x_{a_{k+1}})$
  Update only the selected marginal of b_k using Eq. 8
  Set $S_{k+1} = S_k \cup \{a_{k+1}\}$
  Set $\mathcal{Q}_{k+1} = \mathcal{Q}_k \cup \{(y_{a_{k+1}}, x_{a_{k+1}})\}$
end for
Stage 3: Adaptation and full-sequence reconstruction
Set $\theta \leftarrow \theta_0$ and keep $\mathcal{E}$ and the belief models fixed
for $n = 1, \dots, N_{\text{adapt}}$ **do**
  Update θ using $\mathcal{L}_{\text{adapt}}(\theta, \mathcal{Q}_B)$
end for
for $t = 1, \dots, T$ **do**
  Reconstruct $\widehat{x}_t = \mathcal{R}_\theta(y_t)$
end for
return $\{\widehat{x}_t\}_{t=1}^{T}$

A.1 OFFLINE MODEL PREPARATION

The interpolation operator $\mathcal{A}$ aligns each LR field to the HR grid before the shared encoder produces $c_t = \mathcal{E}(\mathcal{A}(y_t))$ and $z_t = \mathcal{E}(x_t)$. The temporal belief initializer predicts the mean and diagonal variance of each z_t jointly from $c_{1:T}$. The observation predictor estimates the distribution of a candidate HR latent. Both sequence-conditioned models are fitted on historical paired sequences by Gaussian negative log-likelihood. On a target sequence, they use the complete LR condition sequence, but an HR latent becomes available only after its frame is acquired. The encoder, initializer, and predictor remain fixed during acquisition, and only the conditional diffusion reconstructor is adapted after the budget is exhausted.

A.2 GAUSSIAN OBSERVATION AND ACQUISITION DETAILS

The real-observation likelihood used in the VFE update is

$$p(o_a \mid z_a) = \mathcal{N}(o_a \mid z_a, r\mathbf{I}), \qquad r > 0. \tag{13}$$

The observation predictor has diagonal Gaussian output

$$q_\phi(o_a \mid c_{1:T}) = \mathcal{N}(o_a \mid m_a, \operatorname{diag}(s_a)), \qquad s_a \succ 0. \tag{14}$$

Here $m_a \in \mathbb{R}^{d_z}$ and $s_a \in \mathbb{R}^{d_z}_{>0}$ are predicted from LR conditions alone. For each candidate, M hypothetical observations are sampled independently from this distribution. Writing $\widetilde{b}^{a,m}_{k+1} = \mathcal{U}_{\mathrm{VFE}}(b_k, a, \widetilde{o}^{(m)}_a)$, the Monte Carlo score used for Eq. 9 is

$$\widehat{I}_k(a) = \frac{1}{MTd_z} \sum_{m=1}^{M} D_{\mathrm{KL}}\Big(\widetilde{b}^{a,m}_{k+1} \,\Big\|\, b_k\Big), \qquad \widetilde{o}^{(m)}_a \sim q_\phi(\cdot \mid c_{1:T}). \tag{15}$$

Each temporary posterior starts from the same b_k and is discarded after candidate scoring. The preference density in Eq. 10 is exactly the current marginal,

$$p^\star_k(o_a) = \mathcal{N}\big(o_a \mid \mu^k_a, \mathrm{diag}(v^k_a)\big). \tag{16}$$

Temporal coverage is normalized by sequence length:

$$H_k(a) = \begin{cases} 0, & S_k = \varnothing, \\ \dfrac{\min_{s\in S_k} |a-s|}{\max(T-1,1)}, & S_k \neq \varnothing. \end{cases} \tag{17}$$

When all observations have equal cost, $C(a)$ is constant across candidates and does not affect their ranking.

A.3 Target-specific reconstruction adaptation

For acquired time a, sampled diffusion noise level τ, and Gaussian perturbation ϵ, let $\widehat{x}_{a,\tau}(\theta)$ denote the predicted clean HR field conditioned on y_a, with its dependence on ϵ suppressed. The fitting loss in Eq. 12 is

$$\mathcal{L}_{\mathrm{fit}}(\theta, \mathcal{Q}_B) = \mathbb{E}_{a\sim \mathrm{Unif}(S_B),\, \tau,\, \epsilon}\left[\frac{1}{N_a}\left\|\widehat{x}_{a,\tau}(\theta) - x_a\right\|_1\right]. \tag{18}$$

Here N_a is the number of scalar entries in the training field or patch, including its vector components. The auxiliary loss $\mathcal{L}_{\mathrm{deg}}$ compares reconstructed, HR, and degraded representations in the fixed degradation-aware space. It encourages identification of the paired HR representation, greater agreement with that representation than with the degraded input, and separation from residual degradation. The encoder stays fixed, so both losses update only the reconstructor and are averaged over acquired pairs and sampled perturbations. After a fixed number of optimization steps from θ_0, the resulting θ_B is used to reconstruct every LR frame. Hypothetical observations never enter the adaptation target set.

A.4 Theoretical analysis

We analyze the Gaussian belief update, acquisition score, and target-specific adaptation used in AIF-TVR. The results characterize inference and selection under the stated model assumptions. They do not establish an unconditional improvement in full-sequence reconstruction. The smoothness constant and estimation tolerances introduced below are analytical quantities rather than additional algorithmic hyperparameters.

Proposition 1: Optimality of the Gaussian VFE update. For $r > 0$ and strictly positive prior variances, the unique minimizer of Eq. 7 is the Bayesian posterior under the specified latent observation model. It changes only the queried marginal and strictly reduces its coordinate-wise variance.

Proof. Define $p_k(o_a) = \int p(o_a \mid z_a) b_k(Z_{1:T})\, dZ_{1:T}$ and $p_k(Z_{1:T} \mid o_a) = p(o_a \mid z_a) b_k(Z_{1:T}) / p_k(o_a)$. Substituting this posterior into the VFE objective gives

$$\mathcal{F}_k(q \mid o_a) = D_{\mathrm{KL}}(q \,\|\, p_k(Z_{1:T} \mid o_a)) - \log p_k(o_a). \tag{19}$$

The evidence term is independent of q, while the KL divergence is nonnegative and vanishes only when q equals the posterior. Therefore, minimizing VFE is equivalent to Bayesian conditioning under this model (Parr & Friston, 2017; Parr et al., 2022). Because the likelihood depends only on z_a, every other factor of b_k remains unchanged. Multiplying the queried Gaussian marginal by the Gaussian likelihood and completing the square yields Eq. 8. In particular, $v^{k+1}_{a,j} = v^k_{a,j} r / (v^k_{a,j} + r)$, which lies strictly between zero and $v^k_{a,j}$.

The posterior mean can equivalently be written as $\mu_{a,j}^{k+1} = \mu_{a,j}^{k} + \kappa_{a,j}(o_{a,j} - \mu_{a,j}^{k})$, where $\kappa_{a,j} = v_{a,j}^{k}/(v_{a,j}^{k} + r)$. An uncertain prior therefore gives the observation greater influence, whereas a concentrated prior produces a smaller correction. Here r represents effective uncertainty in the latent observation model and need not represent measurement noise in the ground-truth flow field.

Proposition 2: Closed-form expected belief revision. Fix a candidate a and acquisition step k. Abbreviate $v_j = v_{a,j}^{k}$, $\mu_j = \mu_{a,j}^{k}$, $m_j = m_{a,j}$, and $s_j = s_{a,j}$. Define $\delta_j = m_j - \mu_j$ and $\kappa_j = v_j/(v_j + r)$. The information term in Eq. 9 satisfies

$$I_k(a) = \frac{1}{2Td_z} \sum_{j=1}^{d_z} \left[\log\Big(1 + \frac{v_j}{r}\Big) - \kappa_j + \frac{\kappa_j^2}{v_j} \left(s_j + \delta_j^2\right) \right]. \tag{20}$$

Proof. Only the queried marginal contributes to the posterior-to-prior KL. Its hypothetical updated variance is $\widetilde{v}_j = v_j(1 - \kappa_j)$, and its mean displacement is $\widetilde{\mu}_j - \mu_j = \kappa_j(\widetilde{o}_{a,j} - \mu_j)$. Thus,

$$D_{\mathrm{KL}}\Big(\widetilde{b}_{k+1}^{a} \,\Big\|\, b_k\Big) = \frac{1}{2} \sum_{j=1}^{d_z} \left[\log \frac{v_j}{\widetilde{v}_j} + \frac{\widetilde{v}_j}{v_j} - 1 + \frac{(\widetilde{\mu}_j - \mu_j)^2}{v_j} \right]. \tag{21}$$

Taking expectation under the observation predictor gives $\mathbb{E}[(\widetilde{o}_{a,j} - \mu_j)^2] = s_j + \delta_j^2$. Substitution and division by Td_z yield Eq. 20.

Expected revision depends on predictive variance and predictor–belief disagreement, not solely on prior uncertainty. If the predictor matches the predictive distribution induced by the belief and likelihood, then $m_j = \mu_j$ and $s_j = v_j + r$. In this special case, $I_k(a) = (2Td_z)^{-1} \sum_j \log(1 + v_j/r)$, which is the dimension-normalized Gaussian mutual information under that joint model. Without this compatibility condition, $I_k(a)$ measures expected belief revision and is not generally mutual information.

Rationale for the acquisition objective. Large expected revision can reflect informative evidence or strong predictor–belief disagreement. Predictive risk counterbalances the latter through

$$R_k(a) = \frac{1}{2d_z} \sum_{j=1}^{d_z} \left[\log \frac{v_j}{s_j} + \frac{s_j + \delta_j^2}{v_j} - 1 \right]. \tag{22}$$

For fixed variances, the coefficient of δ_j^2 in $G_k^{\mathrm{bel}}(a)$ is $(\lambda_r - \lambda_i \kappa_j^2/T)/(2d_z v_j)$. Hence, $\lambda_r \geq \lambda_i/T$ is sufficient to ensure that increasing mean disagreement cannot lower the belief-based score. This property explains how risk can counterbalance disagreement-driven revision, but it does not establish that either distribution is calibrated.

The score follows the active-inference principle of evaluating predicted outcomes and epistemic value (Friston et al., 2015; Millidge et al., 2021), while making task-specific substitutions. Specifically, $D_{\mathrm{KL}}(q_\phi \| p_k^\star) = \mathbb{E}_{q_\phi}[-\log p_k^\star] - \mathcal{H}(q_\phi)$, where $\mathcal{H}$ denotes differential entropy. Because this entropy can vary across candidates, predictive risk is not generally equivalent to the expected preference cost in Eq. 4. Acquisition cost and temporal coverage encode additional design considerations. Equal costs do not affect ranking, while coverage discourages temporal concentration without guaranteeing nonredundant information. Thus, the score is EFE-inspired rather than an exact EFE decomposition.

Proposition 3: A sufficient condition for beneficial adaptation. For a fixed target sequence, define the full-sequence reconstruction risk as

$$\mathcal{L}_{\mathrm{all}}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\mathcal{L}_{\mathrm{rec}}(\mathcal{R}_\theta(y_t), x_t)]. \tag{23}$$

The expectation is over reconstruction randomness and is omitted for deterministic reconstruction. For this local analysis, assume that the risk is differentiable and its gradient is L-Lipschitz along the update segment, where $L > 0$ is an analytical smoothness constant. Let $S = S_B$ denote the final acquired set and consider $\theta^+ = \theta - \eta g_S$, where $\eta > 0$ is the step size and g_S is the actual optimizer

update direction obtained from $\mathcal{L}_{\text{adapt}}(\theta, \mathcal{Q}_B)$. Any preconditioning or momentum is included in g_S. Then

$$\mathcal{L}_{\text{all}}(\theta^+) \leq \mathcal{L}_{\text{all}}(\theta) - \eta \langle \nabla \mathcal{L}_{\text{all}}(\theta), g_S \rangle + \frac{L\eta^2}{2} \|g_S\|_2^2. \tag{24}$$

Consequently, the update strictly decreases this risk whenever $\langle \nabla \mathcal{L}_{\text{all}}(\theta), g_S \rangle > L\eta \|g_S\|_2^2/2$.

Proof. Let $h = -\eta g_S$ denote the parameter perturbation. The Lipschitz-gradient assumption gives $\mathcal{L}_{\text{all}}(\theta+h) \leq \mathcal{L}_{\text{all}}(\theta) + \langle \nabla \mathcal{L}_{\text{all}}(\theta), h \rangle + L\|h\|_2^2/2$. Substituting $h = -\eta g_S$ yields Eq. 24. The stated alignment condition makes its right-hand side strictly smaller than $\mathcal{L}_{\text{all}}(\theta)$.

This result identifies a local mechanism for transferring acquired supervision to unobserved frames through shared parameters. The update must sufficiently align with the full-sequence objective, and adding observations need not strengthen that alignment. The assumption concerns the reconstruction risk, not the smoothness of the ℓ_1 adaptation loss. Moreover, the full-sequence gradient depends on unqueried HR targets and is unavailable during acquisition. The proposition therefore does not establish that the acquisition surrogate selects aligned frames or guarantees improvement after arbitrary adaptation.

Proposition 4: Monte Carlo estimation and selection stability. Condition on the current acquisition state and fixed predictor. Define $Y_a = (Td_z)^{-1} D_{\text{KL}}(\widetilde{b}^a_{k+1} \| b_k)$ and $V_k(a) = \text{Var}(Y_a)$. All expectations, variances, and probabilities in this proposition are over hypothetical observation draws. With $M \geq 1$ independent draws per candidate,

$$\mathbb{E}[\widehat{I}_k(a)] = I_k(a), \qquad \text{Var}[\widehat{I}_k(a)] = \frac{V_k(a)}{M}. \tag{25}$$

Using the candidate-specific coordinate notation of Proposition 2,

$$V_k(a) = \frac{1}{2T^2 d_z^2} \sum_{j=1}^{d_z} \frac{\kappa_j^4}{v_j^2} \left(s_j^2 + 2 s_j \delta_j^2 \right). \tag{26}$$

For every estimation tolerance $\varepsilon > 0$, distinct from the diffusion perturbation ϵ,

$$\Pr\left(\max_{a \in \mathcal{A}_k} |\widehat{I}_k(a) - I_k(a)| \geq \varepsilon \right) \leq \frac{\sum_{a \in \mathcal{A}_k} V_k(a)}{M \varepsilon^2}. \tag{27}$$

Let $\widehat{G}_k$ replace only I_k by $\widehat{I}_k$, with the remaining score terms evaluated without sampling error. Suppose G_k has a unique minimizer $a^\star$, and define the score margin as $\Delta_k = \min_{a \neq a^\star} [G_k(a) - G_k(a^\star)]$. Uniform information-estimation errors of at most ε preserve the selected candidate whenever $\Delta_k > 2\lambda_i \varepsilon$.

Proof. Averaging independent copies of Y_a gives Eq. 25. Its random part is a sum of independent terms $\kappa_j^2 (\widetilde{o}_{a,j} - \mu_j)^2 / (2Td_z v_j)$. For a Gaussian variable with mean δ_j and variance s_j, the variance of its square is $2s_j^2 + 4s_j\delta_j^2$, giving Eq. 26. Chebyshev's inequality and a union bound yield Eq. 27, without requiring independence between candidates. Finally, $|\widehat{G}_k(a) - G_k(a)| \leq \lambda_i \varepsilon$. For any $a \neq a^\star$, this implies $\widehat{G}_k(a) - \widehat{G}_k(a^\star) \geq \Delta_k - 2\lambda_i \varepsilon > 0$, preserving the minimizer.

Increasing M reduces sampling variance, while nearly tied candidates require more accurate estimation to preserve their ordering. This guarantee concerns the exact score defined by the current predictor. It does not correct predictor misspecification or establish that the selected candidate maximizes reconstruction benefit.

Scope of sequential updating. Under the factorized belief and fixed LR-only predictor, an acquired HR value does not change the exact $I_k(a)$ or $R_k(a)$ of a remaining candidate. Subsequent selection changes through candidate exclusion and temporal coverage, although resampling can introduce Monte Carlo fluctuations. Real observations affect full-sequence reconstruction through later adaptation of shared parameters, not through cross-time Gaussian belief propagation. Accordingly, posterior variance contraction does not imply lower reconstruction error, and a larger observation budget does not guarantee higher PSNR.

B ADDITIONAL EXPERIMENTAL DETAILS

B.1 EXPERIMENTAL EXTENSIONS

Datasets and data splits. Following the dataset and evaluation protocol of CD-TVD (Bi et al., 2025), we evaluate all methods on three time-varying 3D flow datasets. Research Vessel Tangaroa contains an incompressible flow around a research vessel on a $300 \times 180 \times 120$ grid over 201 time steps. Half Cylinder Ensemble contains flow simulations around a half cylinder. We use $Re = 6400$ as the target scenario and simulations with $Re \in \{160, 320, 640\}$ for pre-training. Each pre-training simulation contains approximately 150 time steps, giving approximately 450 time steps in total. Shock Interaction Vortex models the interaction between shock waves and longitudinal vortices on a $160 \times 80 \times 80$ grid. For every dataset, we use the same variables, normalization procedure, and training/test partitions as CD-TVD. The Half Cylinder Ensemble pre-training data retain the original random 80%/20% split, while the $Re = 6400$ target case remains separate and no time step is reassigned across partitions.

Baselines and observation protocol. We compared AIF-TVR with CD-TVD (Bi et al., 2025), SRGAN (Ledig et al., 2017), and SSR-VFD (Guo et al., 2020). We retained the original configurations of the learning-based baselines: SRGAN uses five residual blocks and a perceptual loss adapted for 3D vector fields, whereas SSR-VFD uses magnitude and angular losses without its discriminator. Under the scarce-HR protocol, all learning-based methods receive the same HR observations. Under the single-frame protocol, all methods use the same time steps and selection rules.

Common settings and fairness protocol. For all methods, LR fields are generated from HR fields with the same trilinear downsampling operator, and the reconstruction scale is fixed to $4\times$. Within each comparison, we keep the variable normalization, data partitions, input and output resolutions, observation budget, and adaptation schedule fixed. Learning-based methods use Adam with learning rate 1×10^{-4}, while method-specific parameters follow the corresponding original formulations. Shared hyperparameters are not retuned for individual baselines or test samples, so the reconstruction or acquisition mechanism under study is the only changing factor. All experiments use the CD-TVD rendering configuration and a single NVIDIA A40 GPU.

C ADDITIONAL ABLATION AND DIAGNOSTIC RESULTS

This appendix supplements the core ablation summarized in Table 2 and Figure 4. Figures 7–9 report the absolute full-sequence MAE, PSNR, and MSE. We first clarify the controlled protocol and metric rankings, then provide additional sensitivity results and physical-diagnostic definitions.

C.1 COMPLETE METRICS FOR THE CORE ABLATION

Controlled protocol and provenance. All variants except $\mathrm{M3}^{\dagger}$ come from controlled reruns with seed 2026, a dataset-specific frozen SR checkpoint, candidate acquisition restricted to absolute time steps at or after 100, and 100 adaptation steps. Adaptation uses Adam with learning rate 10^{-5}, coefficients $(0.99, 0.99)$, no weight decay, and no gradient clipping. Unless varied in Table 2, the recorded acquisition settings use $M = 8$, $\lambda_i = 1$, $\lambda_r = 0.2$, $\lambda_c = 1$, $C(a) = 0$, and $\lambda_h = 0.001$. The recorded selector also enforces a minimum temporal gap of two steps. The $\mathrm{M3}^{\dagger}$ values instead come from separate framework-comparison runs. Differences involving that variant describe outcomes across runs and cannot isolate a single component. Moreover, the M5 update-switch comparison does not establish cross-time evidence propagation under the factorized Gaussian model in Appendix A.4.

PSNR. Figure 8 shows that $\mathrm{M3}^{\dagger}$ reaches the highest PSNR on Tangaroa (44.43 dB) and Re6400 (39.83 dB). On ShockVortex, M2 instead reaches 40.49 dB, compared with 40.05 dB for $\mathrm{M3}^{\dagger}$. Considering only the controlled reruns, M2 performs best on Tangaroa and ShockVortex, whereas M3-B8 has the highest Re6400 PSNR. The ranking thus varies across target datasets, and the unmatched $\mathrm{M3}^{\dagger}$ results should not be used to attribute gains to an individual design choice.

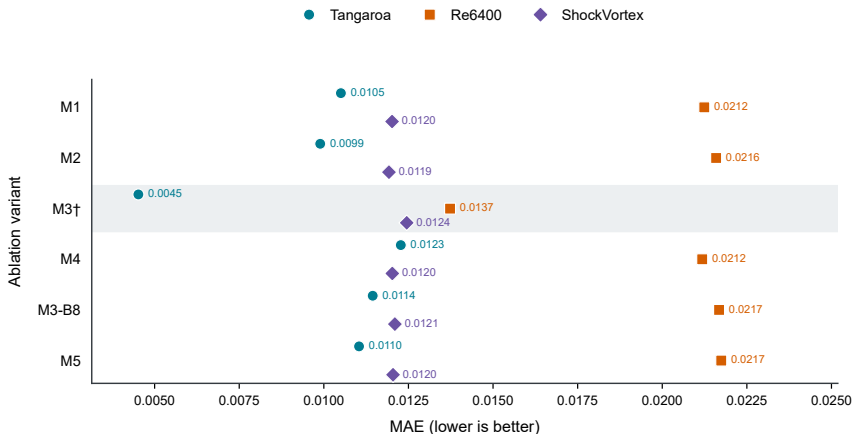


Figure 7: Full-sequence MAE for the core ablation variants. Lower is better. M3† uses the framework-comparison AIF-TVR result, and all other points use controlled reruns with seed 2026.

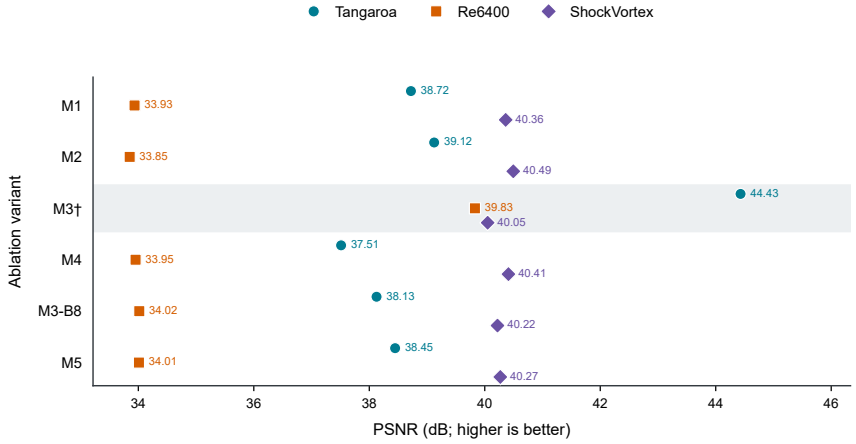


Figure 8: Full-sequence PSNR for the core ablation variants. Higher is better. M3† uses the framework-comparison AIF-TVR result, whereas the remaining points use controlled reruns. Each point is one run with seed 2026.

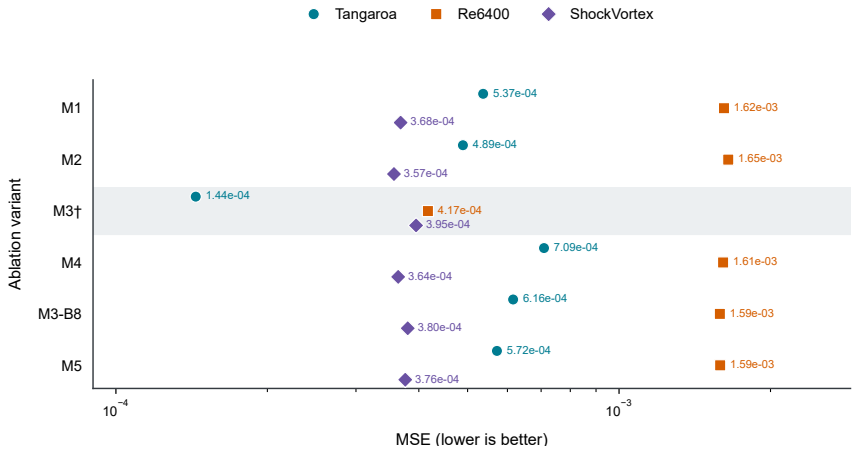


Figure 9: Full-sequence MSE for the core ablation variants on a logarithmic axis. Lower is better. The data provenance and single-seed limitation are the same as in Figure 8.

MSE. Figure 9 follows the PSNR ranking under the fixed-range evaluation. The lowest MSE belongs to $\mathrm{M3}^{\dagger}$ on Tangaroa (1.44×10^{-4}) and Re6400 (4.17×10^{-4}), while M2 has the lowest MSE on ShockVortex (3.57×10^{-4}). Because PSNR is computed from MSE, this agreement is expected and does not provide independent evidence for a component effect.

MAE. Figure 7 provides a complementary view of the absolute reconstruction error. $\mathrm{M3}^{\dagger}$ has the lowest MAE on Tangaroa (0.0045) and Re6400 (0.0137), whereas M2 has the lowest MAE on ShockVortex (0.0119). The controlled Re6400 variants do not have identical rankings across metrics. M3-B8 leads in PSNR and MSE, but M1 and M4 have lower displayed MAE values (0.0212 versus 0.0217). Thus, the preferred variant depends in part on the error measure. Since each setting has only one run, small differences between controlled variants require repeated-seed evaluation before a stable ranking or component-level conclusion can be established.

C.2 Additional hyperparameter sensitivity

Observation budget. On the common 130-frame evaluation set, the frozen reconstructor achieves 52.7943 dB. Budgets of 1, 2, 4, 8, and 16 acquired frames yield 43.6819, 52.5785, 53.1446, 53.6005, and 53.3189 dB, respectively. Eight frames perform best, but 16 frames provide no further improvement. The one-frame decrease demonstrates that HR supervision alone does not guarantee beneficial adaptation under this protocol.

Adaptation duration. At 100, 200, 500, 1,000, and 1,500 adaptation steps, gains over the corresponding frozen model are 0.0210, 0.0686, 0.4492, 0.7517, and 0.8760 dB. The 1,000-step setting reaches 53.5075 dB, while 1,500 steps give the largest tested gain. This trend need not extend to longer adaptation or other sequences.

Auxiliary-loss weight. With eight acquired frames, weights $\beta \in \{0, 0.1, 1, 2\}$ yield 53.5848, 53.5848, 53.5842, and 53.5861 dB. The total range is only 0.0019 dB. This small variation indicates limited sensitivity at these operating points, but does not establish a measurable benefit of the auxiliary term or justify extrapolation beyond the tested interval.

C.3 Physical-diagnostic definitions

Figure 6 uses separate diagnostic reconstruction runs. For each dataset, 21 test frames are selected by uniformly spacing indices over the evaluation interval and rounding to unique nearest indices, including both endpoints. Both methods share the HR reference and a grid subsampled with stride two along each axis. Derivatives use the corresponding physical spacing.

Let u_t and $\widehat{u}_t$ denote the reference and reconstructed velocity fields on this grid. Their vorticities are $\omega_t = \nabla \times u_t$ and $\widehat{\omega}_t = \nabla \times \widehat{u}_t$, computed with centered finite differences. We exclude one outer voxel layer from each boundary and denote the remaining grid by Ω_{int}. The relative vorticity error is

$$e_{\omega,t} = \frac{\|\widehat{\omega}_t - \omega_t\|_{2,\Omega_{\mathrm{int}}}}{\|\omega_t\|_{2,\Omega_{\mathrm{int}}}}, \tag{28}$$

where the norm includes all three vector components and all interior grid points. For the full subsampled grid Ω, define mean kinetic energy per unit mass as $E(u) = \frac{1}{2|\Omega|} \sum_{p \in \Omega} \|u(p)\|_2^2$. The normalized energy error is

$$e_{E,t} = \frac{|E(\widehat{u}_t) - E(u_t)|}{E(u_t)}. \tag{29}$$

Denominators are protected by the smallest positive normal floating-point value. Figure 6 reports temporal medians and interquartile ranges. These diagnostics measure reference agreement rather than satisfaction of every governing equation, so lower vorticity error can coexist with higher energy error.