PROJECTION, CONVECTION, AND RESPONSE CORRECTION FOR LONG-HORIZON NEURAL FLUID SIMULATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural network models accelerate fluid simulation through autoregressive rollouts that feed each forecast into the next prediction, but accumulated errors limit long-horizon accuracy. Complementing unrolling and pushforward training, we introduce Mass Projection (MP) and Scalar Consistency (SC) to couple incompressible velocity prediction with scalar transport. MP enforces exact discrete incompressibility, including on walled domains. SC evolves persistent subcell states with the current and projected velocities, retaining scalar structure within grid cells. Its closed-domain transport conserves scalar amount and preserves bounds independently of learned weights, while diffusion, sources and boundary exchange are included explicitly. Both components are differentiable, allowing scalar errors to train the velocity model. We further propose Response Correction (RC), a post-training method that learns how the trained model would continue from the ground-truth next state and uses the predicted response to correct its forecasts. Across three flows from The Well database, MP+RC lowers full-rollout error in 15 of 18 architecture–training settings. In matched one-step comparisons, MP+RC reduces full-rollout error by 89% on Rayleigh–Bénard convection, and MP+SC by 82% on shear flow.

1 INTRODUCTION

Computational fluid dynamics (CFD) predicts fluid motion, heat transfer, and mixing by numerically solving the governing equations. Repeating these solves over long trajectories is computationally demanding (Kochkov et al., 2021). Neural network models reduce this cost by learning to advance flow fields directly from simulation data (Li et al., 2021; Pfaff et al., 2021; Ohana et al., 2024).

An autoregressive rollout feeds each forecast into the next step, allowing errors to accumulate and alter subsequent dynamics (McCabe et al., 2023). Noise augmentation perturbs training inputs, unrolling differentiates through successive forecasts, and pushforward training supervises predictions from model-generated inputs (Sanchez-Gonzalez et al., 2020; Stachenfeld et al., 2022; Brandstetter et al., 2022; List et al., 2025). These strategies improve robustness to states encountered during rollout. Physical constraints and conservative solvers provide complementary structure through projected neural operators and learned velocity coupled to numerical transport (Xu et al., 2025; Horie & Mitsume, 2024; Agarwal et al., 2025). For incompressible flow, mass conservation requires zero velocity divergence, meaning balanced volume inflow and outflow. This constraint alone does not ensure that an independently predicted scalar, such as temperature or concentration, follows the forecast flow. Physical admissibility alone does not ensure an accurate trajectory. We therefore couple velocity and scalar transport within each transition and correct the forecast errors that remain.

We introduce Mass Projection (MP) and Scalar Consistency (SC) to couple the predicted velocity and its transport (Figure 1). MP differentiably enforces discrete incompressibility, including an exact Fourier–Chebyshev projection with zero normal wall velocity. SC retains persistent subcell states, which store scalar averages inside grid cells across rollout steps, and advances them with the current and projected velocities. These states distinguish scalar arrangements with equal grid averages but different subsequent fluxes. A circulation head, a small shared network, learns internal redistribution without changing coarse-boundary volume fluxes. Closed-domain transport preserves scalar amount

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

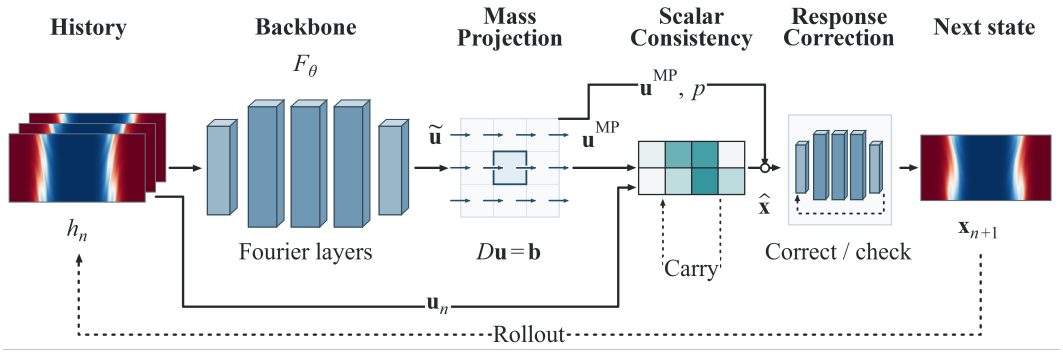


Figure 1: **Overview of the proposed framework.** The backbone F_θ, with weights θ, maps history h_n to a velocity proposal $\tilde{\mathbf{u}}$. MP enforces $D\mathbf{u} = \mathbf{b}$, where D is discrete divergence and $\mathbf{b}$ the constraint right-hand side, and returns projected velocity $\mathbf{u}^{\mathrm{MP}}$ and pressure p. SC transports persistent scalar states using the current velocity $\mathbf{u}_n = \mathbf{u}^n$ and the projected forecast. The backbone is trained through both components. RC refines the proposal $\hat{\mathbf{x}}$ into the retained rollout state $\mathbf{x}_{n+1}$. The Fourier layers illustrate a Fourier neural operator (FNO).

and bounds independently of learned weights, with diffusion and heated-wall exchange included explicitly. Joint adaptation lets scalar errors supervise the carrying velocity. Projection-aware pushforward (PA) restores supervision of velocity components removed by projection (Kasaura et al., 2023).

Response Correction (RC) addresses remaining errors through the frozen model's response, meaning its following prediction after a candidate enters the history. A companion network learns the response obtained by appending the recorded ground-truth next field during training. At inference, the response derivative with respect to the candidate guides a constrained forecast correction. A bounded trial is accepted only if it sufficiently reduces the response discrepancy. The model's weights remain fixed, and deployment requires no new observations.

Our contributions are as follows.

- **Coupled velocity and scalar evolution.** MP and SC combine exact discrete incompressibility on periodic and walled grids with differentiable, persistent conservative transport. PA restores supervision in discarded velocity directions.
- **Correction through the learned response.** RC learns a response target from training trajectories and uses the frozen model's input derivative to construct a bounded correction verified by a response trial.
- **Long-horizon improvements across training strategies.** On three flows from The Well (Ohana et al., 2024), MP+RC improves steps 13–30 in 17 of 18 architecture–training settings and full rollouts in 15, using PA for pushforward. MP and SC also extend to Transolver++ (Luo et al., 2025).

2 Related Work

Research on neural simulators has increasingly connected one-step prediction to behavior under repeated rollout. Unrolling differentiates through successive predictions, allowing errors at later steps to supervise earlier transitions. List et al. (2025) separate this temporal-gradient effect from the change in training inputs caused by feeding forecasts back into the model. Pushforward training uses the latter mechanism through a detached rollout prefix followed by a supervised transition (Brandstetter et al., 2022). Architecture also influences error growth. McCabe et al. (2023) use spectral normalization and filtering to control sensitivity and aliasing in autoregressive neural operators. PDE-Refiner learns successive denoising updates to recover non-dominant frequency information important to long rollouts (Lippe et al., 2023). Our approach combines recurrent training with physical constraints on each predicted transition and uses the trained dynamics to refine forecasts.

In fluid simulation, incompressibility provides a direct constraint on each predicted velocity. Pressure projection separates the velocity update from enforcement of this constraint through a pressure solve (Chorin, 1968). Jiang et al. (2020) incorporate a differentiable spectral projection into a neural network, allowing end-to-end training through a divergence-free output. Kim (2026) examines how constraint enforcement affects recurrent prediction, studying exact periodic projection and boundary-preserving correction. MP brings this structure into autoregressive simulation with an exact Fourier–Chebyshev projection that enforces zero normal velocity at walls. Projection also removes proposal directions from the output and hence from the projected-velocity training signal. In recurrent training, this connects the physical admissibility of an output with the directions in which the projected-velocity loss can train the proposal. The learning problem therefore depends on how projected and unprojected states enter the rollout history. PA addresses this learning effect through raw-proposal feedback and discarded-velocity supervision, adapting pre-projection feedback and constraint-violation penalties from action-constrained learning (Kasaura et al., 2023).

Zhuang et al. (2021) learn local discretization coefficients from the scalar and velocity fields, using the resulting face values within conservative finite-volume updates. Chen et al. (2023) learn semi-Lagrangian remapping coefficients from cell averages and characteristic displacements, with a constraint layer that enforces mass conservation. In the finite-volume formulation, the same face flux enters neighboring cells with opposite signs. In the semi-Lagrangian formulation, the remapping weights distribute the amount carried by each source cell conservatively. These constructions show how learned transport can retain a discrete conservation law independently of the accuracy of its coefficients. The connection to predicted velocity is explicit in mantle-convection models, where a network predicts mass-conserving flow and a finite-volume solver uses that flow to advect and diffuse temperature (Agarwal et al., 2025). FluxNet learns cumulative transport amounts between cells, with capacity-constrained heads that enforce lower or upper scalar bounds (Lan et al., 2026). SC couples learned transport to the forecast velocity and retains a subcell distribution across rollout steps. Retaining this distribution lets later fluxes depend on scalar arrangements that share the same coarse average. The carrier remains tied to the projected velocity, so learned redistribution evolves with the mass-conserving flow throughout a rollout. Its internal circulation changes local redistribution while the transport construction conserves scalar amount on closed domains and prevents new extrema. Joint training lets scalar errors supervise both the transport parameters and the velocity that carries the scalar.

Learned correction provides a way to improve forecasts while retaining a trained model. ARC-STAR combines a global corrector for broad prediction errors with local refinement of the remaining residual (Li et al., 2026a). PnP-Corrector addresses coupled forecasting systems by training a correction agent around frozen simulation models (Wu et al., 2026). These approaches place the additional learning in corrections to the predicted state, allowing the existing forecaster to remain fixed.

Frozen models can also supply signals that guide refinement during generation. TANGO evaluates a candidate video continuation through a frozen look-ahead critic and adapts model weights to improve the statistics of the predicted noise (Karageorgiou et al., 2026). Cycle-World uses a frozen reverse predictor to refine generated latents through cycle consistency (Su et al., 2026). Both use the response of a learned model to assess a proposed continuation. For fluid forecasts, RC uses the forward model response following the ground-truth next field as its correction target. A companion learns this target from training trajectories, and the frozen model guides a bounded, trial-checked correction that moves the forecast response toward it. The target is specific to the model and the current history, so correction follows the dynamics of the model that will continue the rollout. Differentiating its response identifies how changing the proposed field affects that continuation.

3 Physical Constraints and Response Correction

Let $\mathbf{x}_n$ collect the field channels stored at frame n, initialized from observations and subsequently supplied by forecasts. The recurrent input h_n contains the history $(\mathbf{x}_{n-H_{\mathrm{ctx}}+1}, \ldots, \mathbf{x}_n)$ of length H_{ctx} and, when enabled, SC's persistent subcell state. MP and SC form the next prediction, which RC refines before it enters the history. Superscript T denotes transpose, I the identity matrix and $\|\cdot\|_2$ the Euclidean norm.

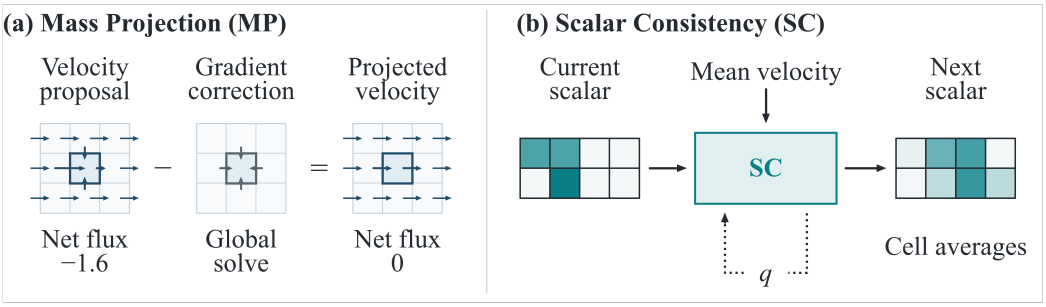


Figure 2: **Mass Projection and Scalar Consistency.** (a) MP removes net flux from the highlighted control volume through a global gradient correction. (b) SC advances the persistent transport state q with the mean of the current and projected velocities. The illustrated cell-average readout applies to periodic grids. Convection uses interpolation to its observation nodes.

3.1 Mass Projection

MP projects the physical velocity proposal $\widetilde{\mathbf{u}}$ onto admissible velocities $\mathbf{u}$ satisfying $D\mathbf{u} = \mathbf{b}$ (Figure 2). Here D discretizes divergence, the net outward volume flux per unit volume, and incompressibility sets $\mathbf{b} = 0$. Write $\mathbf{u} = \mathbf{u}_b + R_f\mathbf{z}$, where $\mathbf{z}$ contains free velocity coordinates, $\mathbf{u}_b$ satisfies the boundary data and the columns of R_f form an orthonormal basis for velocity changes that leave prescribed boundary values fixed. The free proposal is $\mathbf{a} = R_f^\mathsf{T}(\widetilde{\mathbf{u}} - \mathbf{u}_b)$, with restricted divergence $D_f = DR_f$ and right-hand side $\mathbf{b}_f = \mathbf{b} - D\mathbf{u}_b$. The condition $\mathbf{b}_f \in \operatorname{range} D_f$ ensures that an admissible velocity exists. A fixed symmetric positive-definite matrix M weights free velocity changes $\mathbf{w}$ through $\|\mathbf{w}\|_M^2 = \mathbf{w}^\mathsf{T} M\mathbf{w}$. With $L = D_f M^{-1} D_f^\mathsf{T}$, MP returns the closest admissible proposal,

$$\begin{aligned} \Pi(\mathbf{a}) &= \operatorname*{arg\,min}_{D_f\boldsymbol{\xi}=\mathbf{b}_f} \tfrac{1}{2}\|\boldsymbol{\xi} - \mathbf{a}\|_M^2 = \mathbf{a} - M^{-1}D_f^\mathsf{T} L^+(D_f\mathbf{a} - \mathbf{b}_f), \\ \mathbf{u}^{\mathrm{MP}} &= \mathbf{u}_b + R_f\Pi(\mathbf{a}). \end{aligned} \tag{1}$$

Here L^+ is the Moore–Penrose pseudoinverse, a generalized inverse that remains defined when redundant constraints make L singular (Penrose, 1955). A residual head predicts pressure p in the dataset convention from the backbone proposal, projected velocity, signed correction $\mathbf{u}^{\mathrm{MP}} - \widetilde{\mathbf{u}}$ and current state. For fixed geometry and boundary data, the projection branch maps output gradients $\mathbf{g}_{\mathrm{out}}$ to input gradients $\mathbf{g}_{\mathrm{in}}$ through

$$P = I - M^{-1}D_f^\mathsf{T} L^+ D_f, \qquad \mathbf{g}_{\mathrm{in}} = R_f P^\mathsf{T} R_f^\mathsf{T}\mathbf{g}_{\mathrm{out}}. \tag{2}$$

Here P differentiates Π with respect to the free proposal, and $\mathbf{g}_{\mathrm{in}}$ is the projection branch's gradient with respect to $\widetilde{\mathbf{u}}$. The identities $P^2 = P$, $D_f P = 0$ and $P^\mathsf{T} M = MP$ establish an M-orthogonal projection onto divergence-free variations. Periodic domains use spectral Leray projection (Jiang et al., 2020; Xu et al., 2025; Li et al., 2026b). On Rayleigh–Bénard, Fourier–Chebyshev projection uses Fourier modes in the periodic direction and Chebyshev polynomials in the wall-normal direction (Kleiser & Schumann, 1980). It enforces discrete incompressibility and zero normal wall velocity exactly (Appendix B.1).

Projection-aware pushforward. Projection-aware pushforward (PA) adapts pre-projection feedback and constraint penalties from action-constrained learning (Kasaura et al., 2023). During the detached prefix, early rollout steps are evaluated without backpropagating through them. Raw velocity proposals enter the backbone history while MP supplies the physical state. At the supervised step, a unit-weight discarded-velocity penalty restores supervision removed by projection. For the free-coordinate velocity target $\mathbf{y}$,

$$\begin{aligned} \mathcal{L}_{\mathrm{PA}}(\mathbf{a}, \mathbf{y}) &= \|\Pi(\mathbf{a}) - \mathbf{y}\|_M^2 + \|\mathbf{a} - \Pi(\mathbf{a})\|_M^2 \\ &= \|\mathbf{a} - \Pi(\mathbf{y})\|_M^2 + \|\mathbf{y} - \Pi(\mathbf{y})\|_M^2. \end{aligned} \tag{3}$$

With both terms in the same M-metric, this equals regression onto $\Pi(\mathbf{y})$ plus a proposal-independent constant. It restores discarded-direction supervision without changing the per-sample optimal deployed velocity. One-step and unrolled objectives are unchanged.

3.2 Scalar Consistency

SC retains subcell averages q_s^n, where $\mathcal{S}_i$ is the set of four subcells s in coarse cell i. Each stores a spatial average of a transported field, such as tracer concentration or buoyancy, and is initialized once per rollout or independent training window. The finer scalar grid follows the multiresolution approach of Ostilla-Mónico et al. (2015). Persistence distinguishes scalar arrangements with identical grid averages but different subsequent fluxes. The midpoint carrier $\overline{\mathbf{u}} = (\mathbf{u}^n + \mathbf{u}^{\mathrm{MP},n+1})/2$ averages the current physical velocity and the projected forecast. For subcell volume $V_s > 0$, face neighbors $r \in \mathcal{N}(s)$ and scalar flux rate $\mathcal{F}_{sr}^{(k)}$ directed from s to r, internal Euler stage k is

$$\begin{aligned} q_s^{(k+1)} &= q_s^{(k)} - \frac{\Delta\tau_k}{V_s} \sum_{r\in\mathcal{N}(s)} \mathcal{F}_{sr}^{(k)}, \qquad \mathcal{F}_{sr}^{(k)} = -\mathcal{F}_{rs}^{(k)}, \\ c_i^{n+1} &= \mathcal{O}_i(q^{n+1}). \end{aligned} \tag{4}$$

Here $\Delta\tau_k$ is the stage duration, q^{n+1} collects the subcell values after the complete update, and c_i^{n+1} is the returned scalar observation. Readout $\mathcal{O}$ averages periodic subcells or interpolates to native convection samples. The readout enters the backbone, while the subcell state persists.

A pointwise circulation head is a small network shared across coarse cells. It changes internal redistribution without altering coarse-boundary volume fluxes. Antisymmetric scalar fluxes conserve $\sum_s V_s q_s$ on closed domains. Balanced volume fluxes and adaptive Courant–Friedrichs–Lewy (CFL) numbers at most 0.8 limit the fraction of subcell volume transported per stage. With upwind scalar fluxes, each Euler update is therefore a convex combination, with nonnegative weights summing to one. Second-order strong-stability-preserving Runge–Kutta (SSPRK2) stages and positive conservative row remapping preserve the total and scalar bounds for any learned weights (Appendix B.3).

Physical diffusion and local sources complete the update over the saved-frame interval Δt. SC uses tracer diffusivity $D_c = (\mathrm{Re}\,\mathrm{Sc})^{-1}$ on shear flow and thermal diffusivity $\kappa = (\mathrm{Ra}\,\mathrm{Pr})^{-1/2}$ on convection, where Re, Sc, Ra and Pr are the Reynolds, Schmidt, Rayleigh and Prandtl numbers. Periodic diffusion conserves the scalar total, while heated-wall diffusion accounts for exchange and preserves bounds set by the incoming field and plate values. Active-matter orientation sources obey separate realizability constraints. SC uses trajectory diffusivities unavailable to the backbone and replaces the transported-field outputs. Joint adaptation passes their loss through transport to the forecast velocity and circulation head. The learned head controls internal circulation, while the conservative flux construction controls how scalar amounts move between subcells. This separates the accuracy of learned redistribution from the transport invariants. On nonuniform convection grids, conservation refers to the volume-weighted persistent state, whose interpolated observations need not have a conserved unweighted sum.

3.3 Response Correction

Let F be the frozen deterministic differentiable model with any enabled MP and SC components. For $h = h_n$, $\widehat{\mathbf{x}} = F(h)$ predicts the ground-truth state $\mathbf{x}$ at frame $n+1$ (Figure 3). A response is one further model forecast after a candidate $\boldsymbol{\xi}$ enters the history. Using fixed per-channel normalization, define the response and innovation,

$$\begin{aligned} R_h(\boldsymbol{\xi}) &= F(\mathrm{commit}(h, \boldsymbol{\xi})), \qquad \mathbf{a}^\star = R_h(\mathbf{x}), \\ \mathbf{r} &= \mathbf{a}_\phi(h) - R_h(\widehat{\mathbf{x}}). \end{aligned} \tag{5}$$

The operation commit shifts the history and appends the candidate. With SC, it also advances transport memory from the same preceding state for every candidate. A companion F_ϕ with trainable parameters ϕ, initialized from F, learns the target response $\mathbf{a}_\phi(h) \approx \mathbf{a}^\star$ from frozen-model responses to ground-truth next states. Training takes one GPU-hour on model-generated histories, including histories recollected under correction (Appendix C).

One vector–Jacobian product, evaluated without forming the full Jacobian matrix, maps the innovation to the correction direction $\delta\mathbf{x}$, and $\alpha\delta\mathbf{x}$ is the applied increment. The Jacobian $J_{R_h}(\widehat{\mathbf{x}})$ differ-

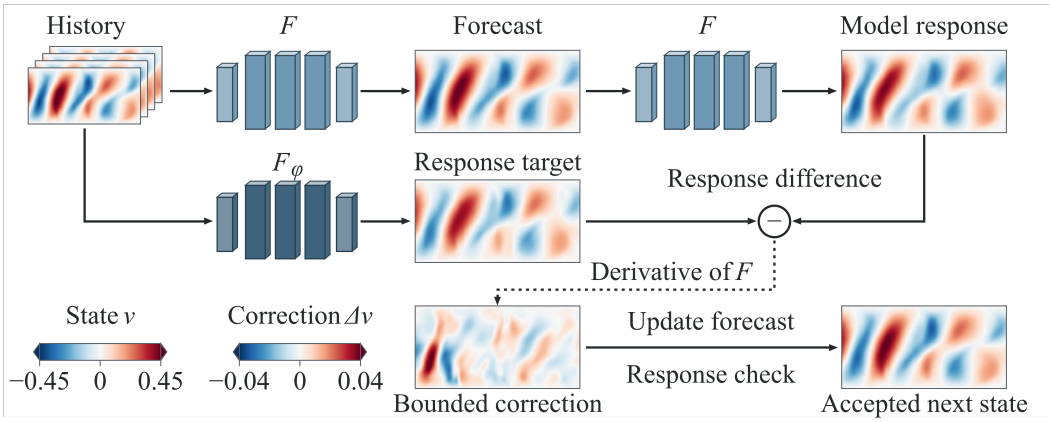


Figure 3: **Response Correction.** A companion initialized from the frozen model F learns the response obtained after appending the ground-truth next state. This response target guides correction through F. The companion F_ϕ has trainable weights ϕ. The displayed vertical velocity $v = u_y$ and its bounded correction Δv use separate color scales.

entiates the response with respect to the candidate while holding the preceding recurrent input fixed. The map P_c enforces physical increment constraints before returning to normalized coordinates. Velocity increments are divergence-free, other corrected channels have zero-mean increments, and prescribed coordinates and SC-replaced channels receive zero updates. For step length $\alpha > 0$ and Euclidean norms in normalized coordinates, RC accepts a finite trial and response only if

$$\begin{aligned} \delta\mathbf{x} &= P_c J_{R_h}(\widehat{\mathbf{x}})^{\mathsf{T}}\mathbf{r}, \qquad \mathbf{x}_\alpha = \widehat{\mathbf{x}} + \alpha\,\delta\mathbf{x}, \\ \|R_h(\mathbf{x}_\alpha) - \mathbf{a}_\phi(h)\|^2 &\le \|\mathbf{r}\|^2 - \tfrac{\alpha}{2}\|\delta\mathbf{x}\|^2. \end{aligned} \tag{6}$$

The Jacobian maps a candidate perturbation into a change in the following forecast. Its transpose transfers the response discrepancy to input directions that influence that forecast, and P_c restricts the resulting direction to admissible increments. The target $\mathbf{a}^\star$ includes the frozen model's response error, so an exact target makes response misfit vanish at $\mathbf{x}$ even for a biased model. Every accepted nonzero update reduces this misfit. Different states can nevertheless have the same response. The positive-margin condition in Eq. (20) is sufficient for field-error reduction and accounts for companion error, response nonlinearity, and the finite correction size.

Each channel's physical root-mean-square (RMS) correction is capped at 5% of the larger current-field or predicted-increment RMS. This fixed bound gives mean response-linearization error below 7% on the probed shear and active-matter validation rollouts. RC starts after the first transition at a step selected on training or validation data (Appendix C.3).

4 Experiments

We evaluate Rayleigh–Bénard convection, shear flow, and active matter from The Well (Ohana et al., 2024), using 35, 13, and 26 test trajectories over 194, 194, and 76 forecast steps. The Fourier neural operator (FNO) (Li et al., 2021) and CNextU-Net, a U-Net with ConvNeXt blocks (Ohana et al., 2024; Liu et al., 2022), receive four input frames. One-step training uses observed inputs, unrolling backpropagates through five forecast steps, and curriculum pushforward supervises after a detached prefix of increasing length (Brandstetter et al., 2022). Baseline and MP models are trained under a 12 GPU-hour budget with matched optimization settings, and their last checkpoints are evaluated. Appendix A specifies the datasets, hardware, and evaluation.

We measure prediction accuracy using variance-scaled root-mean-square error (VRMSE). For a predicted field $\widehat{y}$ and reference y at one frame,

$$\mathrm{VRMSE}(\widehat{y}, y) = \sqrt{\frac{\langle|\widehat{y} - y|^2\rangle}{\langle|y - \langle y\rangle|^2\rangle + \epsilon_{\mathrm{num}}}}, \qquad \epsilon_{\mathrm{num}} = 10^{-7}, \tag{7}$$

where $\langle\cdot\rangle$ denotes the spatial mean. For an individual field and frame, VRMSE above 1 indicates greater mean-square error than predicting the reference field's spatial mean. Scores average over fields, scored steps, and test trajectories, with standard error (SE) across trajectories. Scores above 10 or nonfinite are displayed as > 10, and mean percentage reductions use pairs with both scores at most 10. The fixed MP+RC comparison covers 18 settings and uses PA for pushforward. RC settings are selected on training or validation data. Best-variant summaries select components separately on the test split and describe the evaluated configurations.

4.1 LONG-HORIZON PREDICTION

Table 1: Long-rollout variance-scaled root-mean-square error (VRMSE) and improvement over the matching Baseline. Full-rollout values are mean ± standard error (SE). A plus sign adds a component. N/A means RC stays off, bnd. means the proposed mean is at most 10 while the Baseline exceeds 10, and — means both exceed 10. Bold/underline mark the best/second-best within each model, rollout method, dataset and error window, including the Baseline.

		Rayleigh–Bénard		Shear flow				Active matter	
Rollout method	Components	Full	Improv.	13–30	Improv.	Full	Improv.	Full	Improv.
Model **FNO**									
One-step	Baseline	8.62±0.84		> 10		> 10		1.53±0.44	
	+MP	1.54±0.18	82%	> 10	—	> 10	—	1.50±0.42	2%
	+RC	4.16±0.47	52%	N/A	N/A	N/A	N/A	1.38±0.37	10%
	+MP+RC	**0.98±0.05**	**89%**	**0.179**	bnd.	> 10	—	**1.37±0.37**	**10%**
Unrolling	Baseline	4.78±0.64		4.196		> 10		1.43±0.31	
	+MP	3.22±0.67	33%	1.561	63%	> 10	—	1.40±0.29	2%
	+RC	1.35±0.14	72%	N/A	N/A	N/A	N/A	1.40±0.30	2%
	+MP+RC	**0.90±0.07**	**81%**	**1.100**	**74%**	> 10	—	**1.37±0.27**	**4%**
Pushforward	Baseline	3.98±0.57		0.991		> 10		1.16±0.13	
	+MP	5.52±0.86	−39%	0.361	64%	> 10	—	1.15±0.14	1%
	+RC	1.85±0.23	54%	0.235	76%	> 10	—	1.12±0.12	4%
	+MP+RC	3.34±0.43	16%	**0.227**	**77%**	> 10	—	1.12±0.12	4%
	+MP(PA)	2.01±0.29	50%	1.233	−25%	> 10	—	1.13±0.11	3%
	+MP(PA)+RC	**1.51±0.21**	**62%**	0.979	1%	> 10	—	**1.08±0.09**	**7%**
Model **CNextU-Net**									
One-step	Baseline	3.85±0.57		0.162		5.05±2.88		1.37±0.26	
	+MP	3.04±0.52	21%	0.144	11%	4.28±2.13	15%	1.46±0.30	−7%
	+RC	1.12±0.10	71%	0.203	−25%	2.45±1.00	51%	1.29±0.23	6%
	+MP+RC	**0.87±0.10**	**77%**	**0.142**	**12%**	**1.37±0.48**	**73%**	**1.20±0.24**	**12%**
Unrolling	Baseline	2.82±0.51		0.468		> 10		1.39±0.28	
	+MP	2.76±0.53	2%	**0.231**	**51%**	**2.51±1.18**	bnd.	1.43±0.30	−3%
	+RC	1.88±0.31	33%	0.452	3%	> 10	—	1.28±0.24	8%
	+MP+RC	**1.68±0.32**	**40%**	0.235	50%	2.55±1.21	bnd.	**1.27±0.26**	**9%**
Pushforward	Baseline	2.84±0.59		0.103		0.71±0.42		1.43±0.29	
	+MP	2.58±0.50	9%	**0.098**	**5%**	1.20±0.62	−70%	1.39±0.26	3%
	+RC	2.23±0.46	21%	0.109	−6%	0.76±0.47	−8%	1.28±0.24	11%
	+MP+RC	**1.79±0.33**	**37%**	0.121	−17%	0.87±0.52	−24%	1.29±0.24	10%
	+MP(PA)	2.80±0.55	2%	0.110	−7%	0.83±0.54	−17%	1.38±0.29	3%
	+MP(PA)+RC	2.02±0.36	29%	0.137	−33%	**0.69±0.36**	**2%**	**1.27±0.25**	**12%**

MP+RC lowers steps 13–30 error in 17 of 18 settings and full-rollout error in 15. The remaining three full-rollout pairs both exceed 10. Test-selected variants improve steps 13–30 in all 18 settings (Tables 1–2). Against matching Baselines, one-step FNO with MP+RC reduces convection full-rollout VRMSE from 8.620 to 0.977 (89%), and one-step CNextU-Net with MP+SC reduces shear-flow error from 5.047 to 0.933 (82%). Among these configurations, the best convection full-rollout mean is 69% below the best Baseline across models and training strategies (Appendix D.1).

MP+RC extends the horizon before test-mean VRMSE first exceeds 1 in 14 of 18 settings. Test-selected variants extend it in 15, including from 77 to 125 steps for one-step CNextU-Net with MP+SC on shear flow (Table 3(a)).

Table 2: Shear-flow VRMSE (mean ± SE). Bold and underline mark the best and second-best per model and error window across rollout methods and component combinations. Transport-cap handling is given in Appendix B.3.

	+MP		+MP+SC		+MP+SC+RC	
Rollout method	13–30	Full	13–30	Full	13–30	Full
Model **FNO**						
One-step	> 10	> 10	> 10	> 10	N/A	N/A
Unrolling	1.561 ± 0.763	> 10	0.388 ± 0.147	> 10	0.231 ± 0.092	> 10
Pushforward	0.361 ± 0.207	> 10	$\underline{0.112 \pm 0.040}$	> 10	$\mathbf{0.097 \pm 0.030}$	> 10
Pushforward (PA)	1.233 ± 0.823	> 10	0.120 ± 0.042	> 10	0.116 ± 0.039	> 10
Model **CNextU-Net**						
One-step	0.144 ± 0.039	4.283 ± 2.129	0.124 ± 0.035	0.933 ± 0.490	N/A	N/A
Unrolling	0.231 ± 0.104	2.513 ± 1.179	0.214 ± 0.095	3.279 ± 1.714	N/A	N/A
Pushforward	$\underline{0.098 \pm 0.032}$	1.196 ± 0.624	$\mathbf{0.093 \pm 0.030}$	1.150 ± 0.616	N/A	N/A
Pushforward (PA)	0.110 ± 0.037	$\underline{0.826 \pm 0.535}$	0.102 ± 0.034	$\mathbf{0.821 \pm 0.518}$	N/A	N/A

Table 3: **Forecast horizons and Transolver++.** (a) Forecast horizon for the Baseline → the best component combination from Tables 1 and 2, selected on the test split. The horizon counts steps before the test-mean VRMSE first exceeds 1. Bold marks increases. (b) Full-rollout mean VRMSE on Rayleigh–Bénard. Bold marks the best in each row. Parentheses give improvement over +MP.

(a) Forecast horizon

Rollout method	Rayleigh–Bénard	Shear flow	Active matter
Model FNO			
One-step	2 → **9**	14 → **39**	8 → 8
Unrolling	0 → **56**	20 → **31**	9 → **11**
Pushforward	0 → **16**	25 → **47**	16 → **19**
Model CNextU-Net			
One-step	9 → **27**	77 → **125**	12 → 11
Unrolling	15 → **19**	28 → **69**	10 → **11**
Pushforward	17 → **19**	151 → **152**	11 → 11

(b) Transolver++ on convection

Rollout method	Baseline	+MP	+MP+SC
One-step	> 10	7.59	**7.50** (1%)
Unrolling	> 10	> 10	**6.02**
Pushforward	9.61	5.04	**2.91** (42%)
Pushforward (PA)		5.18	**2.52** (51%)

4.2 Ablations and controls

MP lowers step-1 error in 23 of 24 models by 21% on average, but this gain need not persist. For convection FNO with standard pushforward, MP raises full-rollout error from 3.98 to 5.52. PA reduces it to 2.01, supporting the need to address projection's effect on training. A separate one-step CNextU-Net control shows that training through MP improves on projection applied only during rollout. Steps 13–30 error is 4.731 with MP training, 7.279 with rollout-only projection and 8.749 for the raw model. Training through MP improves on rollout-only projection in 28 of 35 trajectories (Appendix E.1).

SC lowers steps 13–30 error in seven of eight shear-flow MP models by 38% over bounded pairs. Adding RC improves all three deployed SC models in this window, reaching a 90% reduction from the Baseline for standard-pushforward FNO. In Figure 4, both SC variants have tracer-only VRMSE at most 0.007. RC lowers the step-30 all-field error from 0.557 to 0.130 through velocity and pressure corrections while SC retains the tracer. SC also lowers full-rollout error in all eight convection models, although earlier-window gains are mixed. Its orientation-source extension worsens all eight active-matter models, showing that conservative transport alone does not ensure accurate coupled-field dynamics (Appendix E.2).

Across the 43 RC deployments in the main configurations, RC lowers steps 13–30 and full-rollout errors in 36 and 32 cases, averaging 20% and 26% over bounded pairs. For unrolled convection FNO with MP, full-rollout VRMSE falls from 3.220 to 0.901, with improvements on all 35 test trajectories (Figure 6). At the same step bound, RC also outperforms direct state correction over steps 13–30 on all four tested models. These controls support the response-guided procedure, with direction and acceptance effects assessed together. A second control keeps the response target and compares the Jacobian-based rule with half-innovation, which uses the innovation without an input

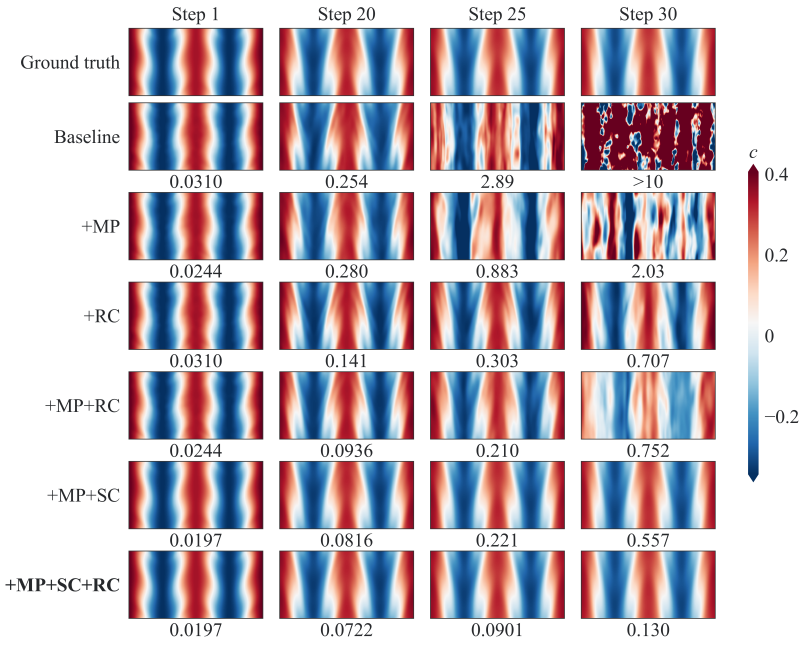


Figure 4: **Shear-flow tracer c rollouts.** Standard-pushforward FNO on one selected test trajectory. Panels share the truth color scale, and numbers give all-field VRMSE. Appendix D.2 specifies the selection and tracer errors.

Jacobian. Among 42 deployments with both rule scores available, the training-selected rule gives lower steps 13–30 error than the alternative in 32 and lower full-rollout error in 28 (Appendix E.3).

Compared with the corresponding models trained for one additional hour, SC lowers steps 13–30 error in seven of eight shear-flow models, and RC does so in 32 of 40 comparisons. Compared with models trained for two additional hours, SC+RC lowers this error in nine of twelve deployments, including all three shear-flow models. Relative to the same 12-hour model, RC reduces steps 13–30 error by 20% on average over bounded pairs, compared with 2% for one more hour of ordinary training. The matched controls therefore show that average gain of RC exceeds that of one additional hour of ordinary training (Appendix E.4).

On convection, MP and SC also improve the attention-based Transolver++ (Luo et al., 2025). SC lowers full-rollout error in all four MP configurations, by 42% under standard pushforward and 51% under PA (Table 3(b)). These scores include MP-only continuation after the transport cap (Appendix B.3). RC is not applied because this implementation resamples its slice assignment at each forward pass, whereas RC requires a fixed response map.

5 Conclusion

MP enforces exact discrete incompressibility on periodic and walled domains. SC transports persistent scalar states with the forecast velocity. Its closed-domain transport conserves scalar amount and preserves bounds independently of learned weights. Joint differentiation lets scalar errors train the carrying velocity, while PA restores supervision removed by projection. RC uses the frozen model's response to correct forecasts without updating its weights at inference. Each bounded correction must sufficiently reduce the discrepancy from the learned response target. Across three flows, MP+RC lowers full-rollout error in 15 of 18 architecture–training settings, using PA for pushforward. Matched controls support gains beyond additional training. Accurate source coupling and correction cost remain challenges. Together, these results show how physical constraints within each transition and correction through the response of a trained model improve long-horizon fluid prediction across architectures and training strategies.

AI USE STATEMENT

In this work, we used generative AI tools to help execute experiments, identify and review related literature, and draft and revise parts of the manuscript. The authors reviewed and revised all AI-assisted material, checked numerical results against experimental records, and verified citations and their associated claims against the original sources. The authors take responsibility for the methodology, results, and final content of this work.

REPRODUCIBILITY STATEMENT

The appendices specify projection and its derivative, the projection-aware objective, scalar transport, response correction, dataset slices and evaluation. Training budgets, hardware and checkpoint selection appear in Appendix A, transport and adaptation in Appendix B.3, and companion training and starting-step selection in Appendix C. Detailed rollout results and cumulative error curves appear in Appendix D.2, with controlled comparisons in Appendix E. Code and detailed configurations will be released upon acceptance.

REFERENCES

Siddhant Agarwal, Ali Can Bekar, Christian Hüttig, David S. Greenberg, and Nicola Tosi. Physics-based machine learning for mantle convection simulations. *Physics of Fluids*, 37(8):086624, 2025. doi: 10.1063/5.0281832. URL https://arxiv.org/abs/2505.16041.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=vSix3HPYKSU.

Yongsheng Chen, Wei Guo, and Xinghui Zhong. A learned conservative semi-Lagrangian finite volume scheme for transport simulations. *Journal of Computational Physics*, 490:112329, 2023. doi: 10.1016/j.jcp.2023.112329. URL https://doi.org/10.1016/j.jcp.2023.112329.

Alexandre J. Chorin. Numerical solution of the Navier–Stokes equations. *Mathematics of Computation*, 22(104):745–762, 1968. doi: 10.1090/S0025-5718-1968-0242392-2. URL https://math.berkeley.edu/~chorin/chorin68.pdf.

Masanobu Horie and Naoto Mitsume. Graph neural PDE solvers with conservation and similarity-equivariance. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 18785–18814. PMLR, 2024. URL https://proceedings.mlr.press/v235/horie24a.html.

Chiyu Max Jiang, Karthik Kashinath, Prabhat, and Philip Marcus. Enforcing hard physical constraints in CNNs through differentiable PDE layer. In *ICLR Workshop on Integration of Deep Neural Models and Differential Equations*, 2020. URL https://cfd.me.berkeley.edu/wp-content/uploads/2022/02/enforcing_physical_constraints.pdf.

Dimitrios Karageorgiou, Symeon Papadopoulos, Ioannis Kompatsiaris, and Efstratios Gavves. Test-time noise guided adaptation for realistic autoregressive video generation. *arXiv preprint arXiv:2607.15849*, 2026. doi: 10.48550/arXiv.2607.15849. URL https://arxiv.org/abs/2607.15849.

Kazumi Kasaura, Shuwa Miura, Tadashi Kozuno, Ryo Yonetani, Kenta Hoshino, and Yohei Hosoe. Benchmarking actor-critic deep reinforcement learning algorithms for robotics control with action constraints. *IEEE Robotics and Automation Letters*, 8(8):4449–4456, 2023. doi: 10.1109/LRA.2023.3284378. URL https://arxiv.org/abs/2304.08743.

Bum Jun Kim. Exactness matters for physical rule enforcement. *arXiv preprint arXiv:2605.08285*, 2026. URL https://arxiv.org/abs/2605.08285.

L. Kleiser and U. Schumann. Treatment of incompressibility and boundary conditions in 3-D numerical spectral simulations of plane channel flows. In E. H. Hirschel (ed.), *Proceedings of the Third GAMM Conference on Numerical Methods in Fluid Mechanics*, volume 2

of *Notes on Numerical Fluid Mechanics*, pp. 165–173. Vieweg+Teubner Verlag, 1980. doi: 10.1007/978-3-322-86146-7_17. URL https://link.springer.com/chapter/10.1007/978-3-322-86146-7_17.

Dmitrii Kochkov, Jamie A. Smith, Ayya Alieva, Qing Wang, Michael P. Brenner, and Stephan Hoyer. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21):e2101784118, 2021. doi: 10.1073/pnas.2101784118. URL https://www.pnas.org/doi/10.1073/pnas.2101784118.

Zishuo Lan, Junjie Li, Lei Wang, and Jincheng Wang. FluxNet: Learning capacity-constrained local transport operators for conservative and bounded PDE surrogates. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.01941.

Chengze Li, Lingwei Wei, Li Sun, Hongbo Lv, Jie Yang, Hanrong Zhang, Kening Zheng, Wei-Chieh Huang, Enze Ma, and Philip S. Yu. ARC-STAR: Auditable Post-Hoc Correction for PDE Foundation Models. *arXiv preprint arXiv:2605.22222*, 2026a. URL https://arxiv.org/abs/2605.22222.

Xigui Li, Hongwei Zhang, Ruoxi Jiang, Deshu Chen, Chensen Lin, Limei Han, Yuan Qi, Xin Guo, and Yuan Cheng. Project and generate: Divergence-free neural operators for incompressible flows. *arXiv preprint arXiv:2603.24500*, 2026b. URL https://arxiv.org/abs/2603.24500.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Phillip Lippe, Bastiaan S. Veeling, Paris Perdikaris, Richard E. Turner, and Johannes Brandstetter. PDE-Refiner: Achieving accurate long rollouts with neural PDE solvers. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/d529b943af3dba734f8a7d49efcb6d09-Abstract-Conference.html.

Bjoern List, Li-Wei Chen, Kartik Bali, and Nils Thuerey. Differentiability in unrolled training of neural physics simulators on transient dynamics. *Computer Methods in Applied Mechanics and Engineering*, 433:117441, 2025. doi: 10.1016/j.cma.2024.117441. URL https://www.sciencedirect.com/science/article/pii/S0045782524006960.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. URL https://arxiv.org/abs/2201.03545.

Huakun Luo, Haixu Wu, Hang Zhou, Lanxiang Xing, Yichen Di, Jianmin Wang, and Mingsheng Long. Transolver++: An Accurate Neural Solver for PDEs on Million-Scale Geometries. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 41432–41449. PMLR, 2025. URL https://proceedings.mlr.press/v267/luo25o.html.

Suryanarayana Maddu, Scott Weady, and Michael J. Shelley. Learning fast, accurate, and stable closures of a kinetic theory of an active fluid. *Journal of Computational Physics*, 504:112869, 2024. URL https://arxiv.org/abs/2308.06675.

Michael McCabe, Peter Harrington, Shashank Subramanian, and Jed Brown. Towards stability of autoregressive neural operators. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=RFfUUtKYOG.

Ruben Ohana, Michael McCabe, Lucas Meyer, Rudy Morel, Fruzsina J. Agocs, Miguel Beneitez, Marsha Berger, Blakesley Burkhart, Stuart B. Dalziel, Drummond B. Fielding, et al. The Well: A large-scale collection of diverse physics simulations for machine learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 44989–45037, 2024. doi: 10.52202/079017-1430.

Rodolfo Ostilla-Mónico, Yantao Yang, Erwin P. van der Poel, Detlef Lohse, and Roberto Verzicco. A multiple-resolution strategy for direct numerical simulation of scalar turbulence. *Journal of Computational Physics*, 301:308–321, 2015. doi: 10.1016/j.jcp.2015.08.031. URL https://arxiv.org/abs/1502.01874.

R. Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413, 1955. doi: 10.1017/S0305004100030401. URL https://doi.org/10.1017/S0305004100030401.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.03409.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8459–8468. PMLR, 2020. URL https://proceedings.mlr.press/v119/sanchez-gonzalez20a.html.

Kimberly Stachenfeld, Drummond B. Fielding, Dmitrii Kochkov, Miles Cranmer, Tobias Pfaff, Jonathan Godwin, Can Cui, Shirley Ho, Peter Battaglia, and Alvaro Sanchez-Gonzalez. Learned coarse models for efficient turbulence simulation. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2112.15275.

Zihan Su, Teng Hu, Jiangning Zhang, Ruiyan Wang, Ran Yi, Lizhuang Ma, and Dacheng Tao. Cycle-World: Mitigating Error Accumulation in Long-term Video World Models via Reverse-Prediction Cycle Consistency. *arXiv preprint arXiv:2607.11836*, 2026. URL https://arxiv.org/abs/2607.11836.

Hao Wu, Fan Xu, Yuxu Lu, Penghao Zhao, Fan Zhang, Hao Jia, Yuxuan Liang, Ruijian Gou, Qingsong Wen, Xian Wu, Xiaomeng Huang, and Yuan Gao. PnP-Corrector: A Universal Correction Framework for Coupled Spatiotemporal Forecasting. *arXiv preprint arXiv:2605.08935*, 2026. URL https://arxiv.org/abs/2605.08935.

Qingsong Xu, Jonathan L. Bamber, Nils Thuerey, Niklas Boers, Paul Bates, Gustau Camps-Valls, Yilei Shi, and Xiao Xiang Zhu. Physically consistent and uncertainty-aware learning of spatiotemporal dynamics. *arXiv preprint arXiv:2510.21023*, 2025. URL https://arxiv.org/abs/2510.21023.

Jiawei Zhuang, Dmitrii Kochkov, Yohai Bar-Sinai, Michael P. Brenner, and Stephan Hoyer. Learned discretizations for passive scalar advection in a two-dimensional turbulent flow. *Physical Review Fluids*, 6:064605, 2021. doi: 10.1103/PhysRevFluids.6.064605. URL https://doi.org/10.1103/PhysRevFluids.6.064605.

A EXPERIMENTAL SETTINGS, EVALUATION AND RESOURCES

Physical systems. Rayleigh–Bénard convection uses a 512×128 grid that is periodic in the horizontal direction and bounded by walls at the bottom and top. Its fields are buoyancy, pressure and velocity. Training uses $\mathrm{Ra} = 10^6$ and $\mathrm{Pr} \in \{1, 0.5\}$, with 80 training and 10 validation trajectories. The test split covers all seven Prandtl numbers at $\mathrm{Ra} = 10^6$, with 35 trajectories and 194 predicted steps. Periodic shear flow uses a 256×512 grid on $[0, 1) \times [-1, 1)$, with tracer, pressure and velocity fields. Trajectories repeated across the release's splits are removed. The shear-flow training slice has $\mathrm{Re} = 10^4$ and $\mathrm{Sc} \in \{1, 0.1\}$, with 50 training and 9 validation trajectories. The test split contains 13 trajectories covering six Schmidt numbers, with 194 predicted steps. Active matter uses the complete dataset (175 training, 24 validation and 26 test trajectories) on a 256×256 periodic grid. Concentration, two velocity components, four orientation-tensor and four strain-rate-tensor components form 11 channels, evaluated for 76 steps. All three systems are from The Well (Ohana et al., 2024), with frame intervals 0.25, 0.1 and 0.25, and all models receive four input frames.

Training budget and checkpoint. Following the training-budget scheme of The Well (Ohana et al., 2024), we allocate 12 GPU-hours to each Baseline and MP configuration, excluding validation. Training uses cosine learning-rate decay after linear warm-up. Evaluation uses the last checkpoint.

Hardware and numerical precision. Each model is trained on one NVIDIA GeForce RTX 5090 (32 GB) in float32 with PyTorch 2.8.0, CUDA 12.8 and cuDNN 9.10.2. Evaluation runs with TensorFloat-32 (TF32) disabled.

Optimization and batches. FNO and CNextU-Net use AdamW optimizer with weight decay 10^{-4} and the peak learning rates in Table 4, decaying to 10^{-6}. The first 1,000 updates accumulate per-channel normalization statistics from training batches without updating model weights. One-step and pushforward use batch size eight, and unrolling uses the largest batch size that fits. Baseline, MP and projection-aware variants use the same batch size within each comparison.

Table 4: Peak learning rates, unrolling batch sizes and epochs completed within the training protocol. Epoch counts follow Baseline/MP for one-step and unrolling, and Baseline/MP/PA for pushforward.

Dataset	Model	Peak rate	Unrolling batch	One-step	Unrolling	Pushforward
Rayleigh–Bénard	FNO	10^{-4}	8	209/187	45/37	134/121/120
	CNextU-Net	5×10^{-4}	6	198/178	57/48	121/112/110
Shear flow	FNO	10^{-3}	5	177/158	57/50	112/105/104
	CNextU-Net	5×10^{-4}	3	145/132	56/52	86/81/80
Active matter	FNO	10^{-3}	8	230/218	51/50	154/150/149
	CNextU-Net	5×10^{-3}	7	200/184	49/48	127/124/122

Metrics and table conventions. We use the VRMSE definition in Eq. (7), with equal weights over spatial grid samples, including the Rayleigh–Bénard Chebyshev rows. The full rollout includes every predicted step. Means above 10, including nonfinite values, are reported as > 10. Improv. is relative to the Baseline of the same model and training strategy. An unsigned percentage denotes lower error, $-$ denotes higher error, and 0% denotes a change below 0.5%. The labels bnd. and — indicate that only the row stays bounded or that both exceed 10. Averaged percentage reductions use pairs in which both means are at or below 10. Trajectory counts compare individual scores before the display threshold is applied. Rankings and percentages use unrounded values, and ties share their bold or underline mark. N/A means RC found no starting step and the model runs unchanged. CNU abbreviates CNextU-Net and PA denotes projection-aware pushforward. The column "RC starts" gives the first step at which correction is enabled, which is at least 2. Superscripts V and H mark validation selection and half-innovation.

Forecast horizons count the steps before per-step test-mean VRMSE first exceeds 1, treating nonfinite errors as infinite. A zero horizon means that the first prediction crosses the threshold. Baseline+RC is eligible for a horizon gain only if correction changes a forecast before the crossing. Fixed MP+RC counts use PA for pushforward. Best-variant counts and best-per-dataset summaries

select the lowest test error separately in each window among the configurations in Tables 7–9. A bounded pair has both means at most 10. This reporting threshold does not itself establish numerical divergence.

Resources. Under the same training budget, MP completes 82–98% as many epochs as the matching Baseline. SC adaptation and RC companion training each receive a budget of one GPU-hour. Across the measured configurations, RC increases batched evaluation time by 42–67% relative to the same model without correction.

B Mass Projection and Scalar Consistency

B.1 Projection and its derivative

Each grid type has its own implementation of the projection in Eq. (1) and its derivative in Eq. (2). The projection layer has no trainable weights. Periodic grids use the established Fourier Leray projection (Jiang et al., 2020; Xu et al., 2025; Li et al., 2026b), with the benchmark's zero- and Nyquist-mode conventions. On active matter, whose velocity follows incompressible Stokes flow (Maddu et al., 2024), it projects the two velocity channels. On walled Rayleigh–Bénard, Fourier modes represent variation along the periodic horizontal direction, while Chebyshev polynomials represent variation in the bounded wall-normal direction (Kleiser & Schumann, 1980). The Fourier transform separates horizontal wavenumbers, allowing Eq. (1) to be applied as a dense projection at each wavenumber with Chebyshev differentiation in the wall-normal coordinate. The projected velocity satisfies discrete incompressibility and zero normal velocity at the physical wall endpoints. This projection does not prescribe tangential wall velocity. Both spectral projections use the identity metric in their transform coordinates, which need not induce an unweighted Euclidean metric on physical grid values. Exactness refers to the specified discrete operators, up to floating-point roundoff. The displayed derivation uses real coordinates, while complex spectral representations use Hermitian adjoints. MP's pressure readout adds 12,769 parameters on convection and shear flow. Active matter has no pressure channel, so MP adds no parameters there.

Eq. (1) minimizes $\frac{1}{2}\|\mathbf{z}-\mathbf{a}\|_M^2$ over the nonempty set $\mathcal{C} = \{\mathbf{z} : D_f\mathbf{z} = \mathbf{b}_f\}$, where M is fixed, symmetric and positive definite and $\langle \mathbf{w}_1, \mathbf{w}_2\rangle_M = \mathbf{w}_1^\mathsf{T} M \mathbf{w}_2$. The columns of R_f span velocity variations satisfying homogeneous boundary conditions. With Lagrange multiplier $\boldsymbol{\lambda}$, stationarity and feasibility give $M(\mathbf{z}-\mathbf{a}) + D_f^\mathsf{T}\boldsymbol{\lambda} = 0$ and $D_f\mathbf{z} = \mathbf{b}_f$, hence $L\boldsymbol{\lambda} = D_f\mathbf{a} - \mathbf{b}_f$. Since $\mathrm{range}(L) = \mathrm{range}(D_f)$, this equation is consistent. Since L is symmetric positive semidefinite, its Moore–Penrose pseudoinverse L^+ inverts positive eigenvalues and maps zero-eigenvalue directions to zero (Penrose, 1955). It remains defined when redundant constraints make L singular and agrees with L^{-1} when L is invertible. For this consistent system, $L^+(D_f\mathbf{a} - \mathbf{b}_f)$ is the multiplier with smallest Euclidean norm. It lies in $\mathrm{range}(L)$, and every multiplier solution gives the same velocity because $\ker L = \ker D_f^\mathsf{T}$. Strict convexity makes $\Pi(\mathbf{a})$ the unique minimizer, even with redundant constraints and multiple null modes. The multiplier enforces the constraint, while physical pressure comes from the readout.

The linear part P satisfies $P^2 = P$, $D_fP = 0$ and $P^\mathsf{T}M = MP$. Thus P is the M-orthogonal projector onto $\ker D_f$. For an admissible target $\mathbf{y}$, $\Pi(\mathbf{a}) - \mathbf{y}$ lies in this kernel and $\mathbf{a} - \Pi(\mathbf{a})$ is orthogonal to it, yielding

$$\|\mathbf{a}-\mathbf{y}\|_M^2 = \|\Pi(\mathbf{a})-\mathbf{y}\|_M^2 + \|\mathbf{a}-\Pi(\mathbf{a})\|_M^2. \tag{8}$$

The same orthogonality makes Π nonexpansive in the M-norm. For a finite solver $\mathcal{S}$ with input $\boldsymbol{\xi} = D_f\mathbf{a} - \mathbf{b}_f$, let $J_\mathcal{S}$ be its input Jacobian and $\mathbf{g}_f$ the upstream free-coordinate gradient of a scalar loss $\mathcal{L}$. Reverse mode differentiates the executed map to give $\nabla_\mathbf{a}\mathcal{L} = \mathbf{g}_f - D_f^\mathsf{T} J_\mathcal{S}(\boldsymbol{\xi})^\mathsf{T} D_f M^{-1}\mathbf{g}_f$. When $J_\mathcal{S}(\boldsymbol{\xi}) = L^+$, this gives the free-coordinate gradient $P^\mathsf{T}\mathbf{g}_f$ used in Eq. (2). Direct inputs to the pressure head contribute additional gradient paths.

B.2 Projection-aware objective

The discarded-velocity penalty supervises directions removed by the projection. To derive Eq. (3), write $\overline{\mathbf{y}} = \Pi(\mathbf{y})$. The decomposition $\Pi(\mathbf{a}) - \mathbf{y} = [\Pi(\mathbf{a}) - \overline{\mathbf{y}}] + [\overline{\mathbf{y}} - \mathbf{y}]$ is M-orthogonal, as is

$\mathbf{a}-\overline{\mathbf{y}} = [\Pi(\mathbf{a})-\overline{\mathbf{y}}]+[\mathbf{a}-\Pi(\mathbf{a})]$. Subtracting their squared norms gives the unit-weight identity. The projected velocity loss alone is unchanged by displacements in $\ker P$. Adding the penalty yields, with the usual factor $1/2$,

$$\nabla_{\mathbf{a}} \left[\tfrac{1}{2}\|\Pi(\mathbf{a}) - \mathbf{y}\|_M^2 + \tfrac{1}{2}\|\mathbf{a} - \Pi(\mathbf{a})\|_M^2\right] = M(\mathbf{a} - \overline{\mathbf{y}}). \tag{9}$$

Both terms use the projection metric M and the same physical velocity scaling. The pointwise minimizer is $\overline{\mathbf{y}}$, which the deployed projection leaves unchanged.

B.3 Scalar transport and joint adaptation

SC stores subcell averages q_s^n on four subcells per coarse cell, with positive volumes V_s. The state is initialized once at the start of each rollout or independent training window and then carried from step to step. The carrier is $\overline{\mathbf{u}} = (\mathbf{u}^n + \mathbf{u}^{\mathrm{MP},n+1})/2$, with gradients through both inputs. These are physical-state velocities, including when PA feeds raw proposals to the backbone.

Observation and initialization. On periodic scalar grids, the observation map in Eq. (4) is the volume-weighted average

$$\mathcal{O}_i(q) = \frac{\sum_{s\in\mathcal{S}_i} V_s q_s}{\sum_{s\in\mathcal{S}_i} V_s}. \tag{10}$$

For native Rayleigh–Bénard point samples, initialization averages a piecewise-bilinear interpolant of the samples and prescribed plate values over each subcell. Readout interpolates subcell-center values, together with plate values near the walls, to the observation nodes using nonnegative weights that sum to one. These weights preserve bounds relative to the subcell and plate inputs. Conservation and wall-exchange budgets are measured on the persistent state as $\sum_s V_s q_s$, rather than an unweighted sum of point samples. Readout does not reinitialize this state.

Learned circulation. The shared pointwise head $g_{\theta_{\mathrm{circ}}}$, with learned parameters θ_{circ}, uses features constructed from the carrier velocity and its spatial derivatives at coarse cell i, given by

$$\mathbf{f}_i = \left[\frac{\overline{\mathbf{u}}}{U_{\mathrm{ref}}}, \frac{L_{\mathrm{ref}}}{U_{\mathrm{ref}}}\,\mathrm{vec}(\nabla\overline{\mathbf{u}})\right]_i, \qquad a_{\theta_{\mathrm{circ}},i} = g_{\theta_{\mathrm{circ}}}(\mathbf{f}_i). \tag{11}$$

Here $\mathbf{f}_i$ is the input feature vector, $a_{\theta_{\mathrm{circ}},i}$ the scalar head output and vec stacks the four velocity-gradient entries. The constants $U_{\mathrm{ref}}, L_{\mathrm{ref}} > 0$ are fixed velocity and length scales. The head uses 1×1 convolutions with widths $6 \to 32 \to 1$, a sigmoid linear unit (SiLU) activation and a zero-initialized final layer, giving 257 parameters. With $V_{\min,i} = \min_{s\in\mathcal{S}_i} V_s$, its internal potential is

$$\begin{aligned} \delta\psi_i &= \beta_{\mathrm{circ}} V_{\min,i}\left(\sqrt{\|\nabla\overline{\mathbf{u}}_i\|_F^2 + \epsilon_{\mathrm{circ}}^2} - \epsilon_{\mathrm{circ}}\right)\tanh(a_{\theta_{\mathrm{circ}},i}), \\ \beta_{\mathrm{circ}} &= 0.1, \qquad \epsilon_{\mathrm{circ}} = 10^{-6} U_{\mathrm{ref}}/L_{\mathrm{ref}}. \end{aligned} \tag{12}$$

The subscript F denotes the Frobenius norm, β_{circ} sets the circulation amplitude and ϵ_{circ} smooths the gradient magnitude. The potential is placed at the interior vertex of each 2×2 subcell block, with zero correction at its boundary vertices. Its discrete curl changes internal volume fluxes while preserving volume balance and every coarse-boundary volume flux. The correction is zero at initialization and for a spatially uniform carrier. Scalar transfers combine these volume fluxes with the transported subcell values.

Transport stages and physical terms. Let $\mathcal{U}_{sr} = -\mathcal{U}_{rs}$ be the signed residual volume flux rate and $a_{sr} = \max\{\mathcal{U}_{sr}, 0\}$ its outgoing part. For Euler-stage duration $\Delta\tau$, the upwind residual kernel uses

$$\mathcal{F}_{sr} = a_{sr} q_s - a_{rs} q_r, \qquad \sum_{r\in\mathcal{N}(s)} a_{sr} = \sum_{r\in\mathcal{N}(s)} a_{rs}, \qquad \frac{\Delta\tau}{V_s}\sum_{r\in\mathcal{N}(s)} a_{sr} \le 0.8. \tag{13}$$

Antisymmetric scalar transfers conserve $\sum_s V_s q_s$. Balance and the CFL bound make each Euler update a convex combination. If $T_{\Delta\tau}$ denotes this Euler kernel, a second-order strong-stability-preserving Runge–Kutta (SSPRK2) substep takes $q^{(1)} = T_{\Delta\tau}(q)$ and $q^+ = \frac{1}{2}q + \frac{1}{2}T_{\Delta\tau}(q^{(1)})$. The carrier remains fixed over the frame interval. Transport extracts periodic row-mean advection from

the base carrier and adds the learned curl to the residual flux. Each frame interval applies a residual half-interval, one full positive conservative row remap and a second residual half-interval. The adaptive number of stages is set by the combined residual flux, including the learned circulation. The row remap and residual stages preserve the volume-weighted amount, constants and scalar bounds for any learned weights.

For coupled laws, q collects all carried components at every subcell. Over the saved-frame interval Δt, let $\mathcal{A}^{\Delta t}_{\theta_{\mathrm{circ}}}$ denote this advection operator, $\mathcal{D}^{\Delta t}$ diffusion and $\mathcal{R}^{\Delta t/2}$ a local source half-step. The complete frame update is

$$q^{n+1} = \big(\mathcal{R}^{\Delta t/2} \circ \mathcal{D}^{\Delta t} \circ \mathcal{A}^{\Delta t}_{\theta_{\mathrm{circ}}}(\overline{\mathbf{u}}) \circ \mathcal{R}^{\Delta t/2}\big)(q^n). \tag{14}$$

Absent source or diffusion terms use identity operators. The convexity argument applies to transport and does not assert scalar interval preservation by general sources.

SC uses each trajectory's physical coefficients, with $D_c = (\mathrm{Re}\,\mathrm{Sc})^{-1}$ for the shear-flow tracer and $\kappa = (\mathrm{Ra}\,\mathrm{Pr})^{-1/2}$ for buoyancy. Here $\mathrm{Re}, \mathrm{Sc}, \mathrm{Ra}, \mathrm{Pr}$ denote the Reynolds, Schmidt, Rayleigh and Prandtl numbers. SC uses recorded frame intervals on convection and nominal intervals on shear flow and active matter. Implicit finite-volume diffusion conserves the weighted amount on periodic domains. With prescribed plates, it accounts for wall exchange and preserves bounds set by the incoming field and plate values.

Active-matter orientation law. The active-matter extension transports concentration c and the raw second orientation moment D with identical spatial weights and translational diffusivity d_T. The tensor D differs from MP's divergence matrix D. Tensor indices i, j, k, l range over the two Cartesian directions x, y. Write $G_{ij} = \partial_j u_i$ for the physical velocity gradient, $E = (G + G^{\mathsf{T}})/2$ for strain and $\mathsf{T} = E + 2\zeta\mathsf{D}$. With physical time t, spatial gradient ∇ and Laplacian Δ, the adapter evolves

$$\begin{aligned} (\partial_t + \mathbf{u}\cdot\nabla)c &= d_T \Delta c, \\ (\partial_t + \mathbf{u}\cdot\nabla)\mathsf{D} &= d_T\Delta\mathsf{D} + G\mathsf{D} + \mathsf{D}G^{\mathsf{T}} - 2\widehat{\mathsf{C}}_\vartheta + 4\zeta\mathsf{D}^2 - 4d_R(\mathsf{D} - cI/2), \\ (\widehat{\mathsf{C}}_\vartheta)_{ij} &= \sum_{k,l} (\widehat{\mathsf{S}}_\vartheta)_{ijkl}\mathsf{T}_{kl}. \end{aligned} \tag{15}$$

Here ζ is the trajectory's alignment parameter, d_R the rotational diffusivity, and $d_T = d_R = 0.05$. The tensor $\widehat{\mathsf{S}}_\vartheta$ is the fourth-moment closure with learned parameters ϑ, and $\widehat{\mathsf{C}}_\vartheta$ is its contraction with T. Shared transport and diffusion preserve $c \geq 0$, $\mathsf{D} = \mathsf{D}^{\mathsf{T}} \succeq 0$ and $\mathrm{tr}\,\mathsf{D} = c$. Here tr denotes trace and $\succeq 0$ positive semidefiniteness. Off-diagonal components may be negative. The flow remains a neural prediction constrained by MP, without an additional active-stress solve.

For $c > 0$, the closure uses the complex orientation and normalized strain

$$z = \frac{\mathsf{D}_{xx} - \mathsf{D}_{yy} + 2\mathrm{i}\mathsf{D}_{xy}}{c}, \qquad \eta = \frac{E_{xx} - E_{yy} + 2\mathrm{i}E_{xy}}{\sqrt{(E_{xx} - E_{yy})^2 + 4E_{xy}^2 + (\zeta c)^2 + \epsilon_\omega^2}}, \tag{16}$$

Here $\mathrm{i}^2 = -1$, and $\epsilon_\omega > 0$ is a fixed inverse-time scale. For complex values, $\Re$ and $\Im$ denote real and imaginary parts, an overbar denotes conjugation and $|\cdot|$ denotes modulus. A pointwise $8 \to 32 \to 3$ SiLU head with a zero-initialized final layer receives the invariant features c/c_{ref}, $|z|^2$, $|\eta|^2$, $\Re(z\bar{\eta})$, $\Im(z\bar{\eta})^2$, $\zeta c T_{\mathrm{ref}}$, $d_R T_{\mathrm{ref}}$ and $\alpha_{\mathrm{act}} T_{\mathrm{ref}}$. The fixed positive scales c_{ref} and T_{ref} normalize concentration and time, and α_{act} is the trajectory's dipole-strength parameter. Applying $\tanh$ to its outputs gives ℓ_1, ℓ_2, ℓ_3 and

$$v = \ell_1 z^2 + \ell_2 z\eta + \ell_3 \eta^2, \qquad w = z^2 + (1 - |z|^2)\frac{v}{\sqrt{1 + |v|^2}}, \qquad |w - z^2| \leq 1 - |z|^2. \tag{17}$$

Here v is an auxiliary complex value and w the normalized fourth angular moment. This admissible parameterization defines five independent tensor components, with the parameter subscript ϑ suppressed

$$\begin{aligned} \widehat{\mathsf{S}}_{xxxx} &= \tfrac{c}{8}(3 + 4\Re z + \Re w), & \widehat{\mathsf{S}}_{yyyy} &= \tfrac{c}{8}(3 - 4\Re z + \Re w), \\ \widehat{\mathsf{S}}_{xxyy} &= \tfrac{c}{8}(1 - \Re w), & \widehat{\mathsf{S}}_{xxxy} &= \tfrac{c}{8}(2\Im z + \Im w), \\ \widehat{\mathsf{S}}_{xyyy} &= \tfrac{c}{8}(2\Im z - \Im w). \end{aligned} \tag{18}$$

Index permutations share the same value, giving $\operatorname{tr}\widehat{\mathsf{C}}_\vartheta = \sum_{ij} \mathsf{D}_{ij}\mathsf{T}_{ij}$. For $c = 0$, admissibility gives $\mathsf{D} = \widehat{\mathsf{S}}_\vartheta = 0$.

Each source half-interval uses eight equal substeps, holding c and the midpoint carrier gradient $G = \nabla\overline{\mathbf{u}}$ fixed while recomputing the closure at every source stage. A substep brackets projected SSPRK2 deformation with exact rigid-rotation and rotational-relaxation half-steps. For the deformation source $\mathcal{J}_\vartheta(\mathsf{D}) = E\mathsf{D} + \mathsf{D}E - 2\widehat{\mathsf{C}}_\vartheta + 4\zeta\mathsf{D}^2$, each Euler stage maps a symmetric trial moment B to $\mathcal{Q}_c(B)$ using its trace-free part H and magnitude r_H

$$H = B - \tfrac{1}{2}\operatorname{tr}(B)I, \qquad r_H = \sqrt{H_{xx}^2 + H_{xy}^2}, \qquad \mathcal{Q}_c(B) = \tfrac{c}{2}I + \begin{cases} H, & r_H \le c/2, \\ \dfrac{c}{2r_H}H, & r_H > c/2. \end{cases} \tag{19}$$

If needed, the trial is first symmetrized using $\operatorname{sym}(B) = (B + B^{\mathsf{T}})/2$. The map returns zero when $c = 0$. Projection and SSPRK2 averaging preserve the joint moment constraints. The scalar maximum principle applies to transport, whereas orientation sources change individual moment components. The source uses the base carrier gradient without the learned circulation, and the strain readout is $\operatorname{sym}(\nabla\mathbf{u}^{\mathrm{MP},n+1})$.

Joint adaptation. SC is attached to the last checkpoint of a trained MP model. The backbone and enabled SC heads are jointly adapted using the existing training strategy and loss, allowing scalar errors to train the forecast velocity through SC. AdamW starts the backbone at 0.05 times its peak rate and the SC head at that peak, with 5% linear warm-up followed by cosine decay to 5% of the starting rates and no gradient clipping. One-step adaptation uses batch size eight, with 1,000 initial updates to collect normalization statistics. Pushforward adaptation resumes the curriculum where training ended and samples prefix depths 0–4 from the start with batch size eight. Unrolling uses batches of 8/5 on convection, 4/3 on shear flow and 5/4 on active matter (FNO/CNextU-Net). The training half-step cap is 64 (128 for active-matter FNO unrolling). All 24 adapted models are evaluated.

Transport cap and continuation. The evaluation cap is 128 transport substeps per residual half-interval. For SC without RC, exceeding it disables SC for the remaining trajectory and continues with MP alone. On shear-flow FNO, this occurs in 13/13 one-step trajectories at steps 16–168, 13/13 unrolled trajectories at steps 86–158, 6/13 standard-pushforward trajectories at steps 89–180 and 8/13 PA trajectories at steps 50–192. Three one-step trajectories switch at steps 16–19, so their steps 13–30 scores include MP-only continuation. On Transolver++, continuation occurs in 35/35 standard-pushforward, 32/35 unrolled and 1/35 PA trajectories, with no switches under one-step training. These scores evaluate the combined transport and continuation policy.

For SC+RC, exceeding the cap stops the trajectory and remaining steps are scored as failures. This affects 3/13 shear-flow FNO standard-pushforward, 6/13 PA and 10/13 unrolled trajectories. The earliest stop is at step 42, so steps 1–30 precede stopping. Their full-rollout means are reported as > 10.

C Response Correction

RC predicts how the frozen model would respond after the ground-truth next state $\mathbf{x}$ is appended to the history and uses this predicted response to refine $\widehat{\mathbf{x}}$. The quantities R_h, $\mathbf{a}^\star$, $\mathbf{a}_\phi(h)$ and $\mathbf{r}$ are defined in Eq. (5).

C.1 Companion training

The companion uses the same architecture and normalizers as F. Its output is decoded as $\mathbf{a}_\phi(h) = \mathbf{x}_n + 2(F_\phi(h) - \mathbf{x}_n)$ and fitted to $\mathbf{a}^\star$, holding F fixed while training ϕ. Because F_ϕ is initialized from F, training starts from $\mathbf{a}_\phi(h) = 2\widehat{\mathbf{x}} - \mathbf{x}_n$. Companion training runs for one GPU-hour. Histories come from model rollouts started on training trajectories. Of the sampled rollout steps, 80% lie between 1 and 30 and 20% span the whole rollout. In the second half of the hour, half of the batches are collected again under the current correction. The companion minimizes squared response error, normalized by the uncorrected forecast's squared response error with a batch-mean

stabilizer weighted by 0.1. The normalized loss is capped at 100. AdamW updates the companion alone with zero weight decay and gradient clipping at 1.0. Its peak rate is 0.05 times the model's, with 200 warm-up updates and wall-clock cosine decay to 5% of this peak. Sixteen training trajectories (12 on shear flow) are held out of the updates. The retained checkpoint gives the largest RC gain on these trajectories over steps 6–12. These trajectories also support starting-step selection (Appendix C.3). For SC models, the companion shares the frozen SC component of F, and only its remaining parameters are trained.

C.2 Constructing and accepting a correction

RC constructs the direction in Eq. (6) using $P_c = S^{-1}P_{\mathrm{adm}}S$. Here S is the positive diagonal matrix of fixed channel standard deviations, and P_{adm} is the linear map that enforces the admissible-increment constraints of Section 3.3 in physical units. Because channel scaling need not commute with P_{adm}, P_c is generally nonsymmetric. The finite trial check therefore determines whether the proposed direction decreases response misfit.

For a nonzero correction direction, RC takes the largest shared step $0 \le \alpha \le \|\mathbf{r}\|_2^2/(2\|\delta\mathbf{x}\|_2^2)$ satisfying all channel bounds. In physical units, each channel's RMS correction is capped by both the RMS response discrepancy and 5% of the larger of the current field RMS and predicted increment RMS.

A zero direction or step length leaves the forecast unchanged. Otherwise the trial is accepted exactly when $\mathbf{x}_\alpha$ and $R_h(\mathbf{x}_\alpha)$ are finite and the sufficient-decrease test in Eq. (6) holds. An accepted trial lowers the squared response discrepancy by at least $\alpha\|\delta\mathbf{x}\|_2^2/2$.

All candidate responses use the same preceding history and, for SC models, the same preceding persistent transport state. The retained candidate determines the next history and transport state. Its response, when already evaluated, is reused as the next proposal. The half-innovation control starts from $\widehat{\mathbf{x}} + \frac{1}{2}P_c\mathbf{r}$, uses the same step bounds and finite response-trial check, and requires no input Jacobian.

Step-bound evidence. For a candidate perturbation $\Delta\mathbf{x}$, write $J = J_{R_h}(\widehat{\mathbf{x}})$ and $\Delta R = R_h(\widehat{\mathbf{x}} + \Delta\mathbf{x}) - R_h(\widehat{\mathbf{x}})$. These are the response Jacobian and actual response change used in the linearization check. The step bound is fixed at 5% before any evaluation. On the probed validation rollouts of shear flow and active matter, the first-order response model has mean relative error below 7% up to this step bound (Table C.1). On the two Rayleigh–Bénard models of Table 13, the actual-to-predicted misfit-decrease ratio is 1.00 up to 20% and 1.01 at 50%. A validation search over 5%, 20% and no bound on the four models of Table 13 favors 5% on all three whose full validation rollout stays bounded.

Table C.1: **Validation of the RC step bound.** Mean relative linearization error, $\|\Delta R - J\Delta\mathbf{x}\|_2/\|\Delta R\|_2$, over steps 6–30 of the first four validation trajectories per model. Both are +MP models trained with standard pushforward. Bold marks the fixed 5% step bound.

Step size	1%	2%	5%	10%	20%	50%
Shear flow, FNO	1.7%	3.0%	**6.9%**	12.5%	21.6%	45.1%
Active matter, FNO	3.6%	4.3%	**6.4%**	10.2%	17.4%	38.0%

C.3 Choosing the first corrected step

RC starts at a step chosen on held-out training trajectories. Candidate starting steps $k \in \{1, \dots, 30\}$ are screened for positive median relative one-step gain on the held-out training trajectories. Candidate k first permits correction at step $\max\{2, k\}$, leaving the first transition unchanged. The procedure tests the Jacobian-based correction at eligible starting steps in increasing order, then the half-innovation control. It retains the first candidate whose mean error over steps 6–12 is at most 1.02 times the model's and whose average of the steps 6–12 and 13–30 mean errors is below the model's. The candidate must also preserve or improve the median failure horizon and bounded-rollout count. In each scored window, its prediction-to-truth energy ratio must be at least

$\min\{\text{model ratio}, 1\} - 0.01$. If no candidate is retained, eligible models proceed to validation selection, while MP+SC models keep RC off. Validation compares both corrections at candidate starting steps $\{1, 4, 8, 13, 20\}$ and selects the candidate with the lowest error ratio to the model, averaged over steps 6–12, steps 13–30 and the full rollout. RC remains off unless this average is below one.

C.4 Field-error reduction

A correction reduces field error when its alignment with the true forecast error outweighs the effect of its finite step size. Let $\mathbf{e} = \mathbf{x} - \widehat{\mathbf{x}}$ and $J = J_{R_h}(\widehat{\mathbf{x}})$. Define the mapped error $\mathbf{e}_R = JP_c^{\mathsf{T}}\mathbf{e}$ and the mismatch $\boldsymbol{\eta} = \mathbf{r} - \mathbf{e}_R$. This mismatch combines companion prediction error, nonlinear response effects and the effect of the admissible-direction map.

For $\delta\mathbf{x} = P_c J^{\mathsf{T}}\mathbf{r}$, we have $\langle \mathbf{e}, \delta\mathbf{x}\rangle = \langle \mathbf{e}_R, \mathbf{r}\rangle$. Expanding the squared field error and applying Cauchy–Schwarz shows that, for $\alpha > 0$, the corrected forecast has strictly lower field error whenever

$$\|\mathbf{e}_R\|_2^2 > \|\mathbf{e}_R\|_2\|\boldsymbol{\eta}\|_2 + \frac{\alpha}{2}\|\delta\mathbf{x}\|_2^2. \tag{20}$$

The condition connects field-error reduction to response accuracy and step size. At inference, RC evaluates the response-decrease test in Eq. (6).

D Complete results

D.1 Results across models and training strategies

Table 5 separates fixed compositions from test-selected variants. Counts use unrounded means, so the 0.47% steps 13–30 reduction for active-matter CNextU-Net with unrolling counts as lower despite displaying as 0%. Pairs in which both means exceed 10 do not count as improved. Table 6 compares the best evaluated Baseline and proposed configuration separately for each dataset and window. Its paired sign tests describe these test-selected comparisons and do not account for configuration selection.

Table 5: Models (out of 18) with lower mean VRMSE than their own Baseline. The two +MP+RC rows differ only on pushforward-trained models, which use projection-aware or standard pushforward. The best evaluated variant is selected on the test split, separately for each window.

Composition	Step 1	6–12	13–30	Full
+MP+RC, projection-aware pushforward	17/18	12/18	17/18	15/18
+MP+RC, standard pushforward	17/18	12/18	15/18	14/18
Best evaluated variant (test-selected)	17/18	14/18	18/18	15/18

D.2 Every evaluated configuration

Tables 7–9 give Baseline, MP and RC results, with SC for shear flow. Figures 5, 7 and 8 include SC on all three systems. At step t, each curve averages trajectory VRMSE over steps $1, \ldots, t$ and then across test trajectories, with bands giving one SE across these trajectory averages. Figure 6 shows paired RC gains.

Bands show $\pm$ one SE. Vertical axes are logarithmic with panel-specific limits. Long dashes denote projection-aware pushforward. Triangles mark the first step at which a curve exceeds 10, and crosses mark the last defined point before a nonfinite or stopped rollout.

Shear-flow visualization. Figure 4 uses the only test trajectory with Schmidt number 0.1 (Reynolds number 10^4) and an FNO trained with standard pushforward. RC starts at steps 4 (+RC), 10 (+MP+RC) and 13 (+MP+SC+RC). Among test trajectories whose MP+RC error over steps 13–30 is below 1, this trajectory has the largest MP+RC gain over MP. Steps 1, 20, 25 and 30 correspond to simulation times 0.5, 2.4, 2.9 and 3.4. Each panel shows the full periodic domain, rotated $90°$ clockwise, with a shared color scale fixed by the truth over these steps. The numbers in the panels give this trajectory's all-field VRMSE. Both SC rows have tracer-only VRMSE at most 0.007 at the displayed steps. Their step-30 all-field scores of 0.557 and 0.130 are dominated by pressure and velocity errors.

Table 6: Best test VRMSE per dataset and window among the configurations in Tables 7–9, selected separately for Baseline and proposed configurations. Full rollout reports mean ± SE. Bold marks the best mean in each row. Lower counts trajectories with lower error than the selected Baseline. The p column gives the two-sided paired sign-test p-value without adjustment for configuration selection. A dash in the p column indicates identical predictions.

Window	Baseline	Proposed	Configuration	Improv.	Lower	p
Rayleigh–Bénard						
Step 1	0.078	**0.058**	CNU / One-step / +MP	26%	35/35	5.8×10^{-11}
6–12	0.333	**0.246**	FNO / Pushforward / +MP+RC	26%	31/35	3.5×10^{-6}
13–30	2.770	**0.418**	FNO / Unrolling / +MP+RC	85%	35/35	5.8×10^{-11}
Full	2.815±0.506	**0.871±0.099**	CNU / One-step / +MP+RC	69%	35/35	5.8×10^{-11}
Shear flow						
Step 1	0.0045	**0.0039**	CNU / One-step / +MP+SC	12%	13/13	2.4×10^{-4}
6–12	0.035	**0.034**	FNO / Pushforward / +MP+SC	2%	7/13	1
13–30	0.103	**0.093**	CNU / Pushforward / +MP+SC	10%	8/13	0.58
Full	0.705±0.416	**0.692±0.361**	CNU / Pushforward / +MP(PA)+RC	2%	4/13	0.27
Active matter						
Step 1	**0.090**	**0.090**	CNU / One-step / +RC	0%	0/26	—
6–12	0.511	**0.501**	FNO / Pushforward / +RC	2%	15/26	0.56
13–30	1.121	**1.003**	FNO / Pushforward / +MP(PA)+RC	10%	17/26	0.17
Full	1.162±0.133	**1.082±0.094**	FNO / Pushforward / +MP(PA)+RC	7%	21/26	2.5×10^{-3}

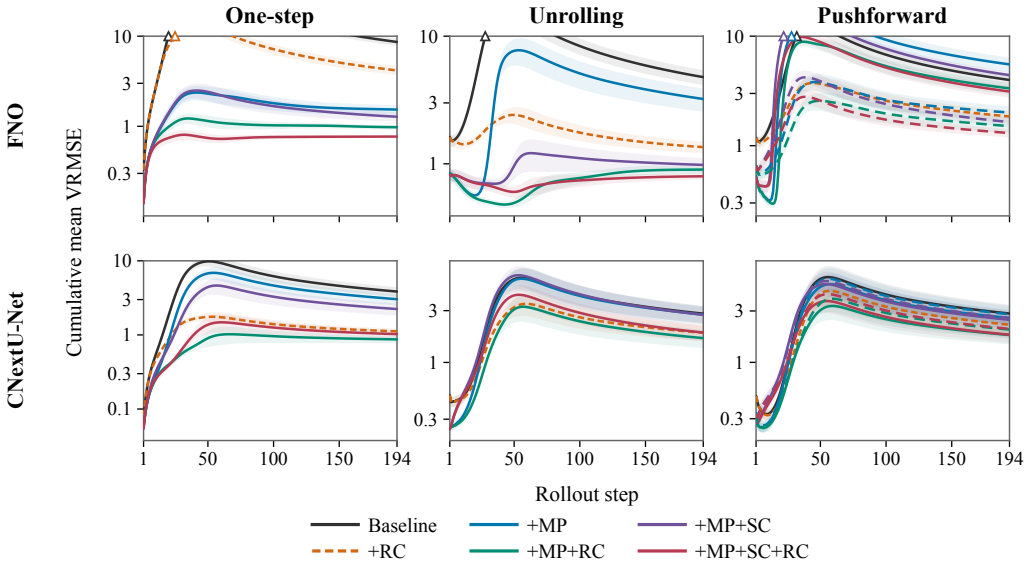


Figure 5: Cumulative mean VRMSE on Rayleigh–Bénard convection over 194 forecast steps (35 test trajectories).

Table 7: Test VRMSE on Rayleigh–Bénard. Full rollout reports mean ± SE. Improvements are relative to the matching Baseline. Within each model, rollout method and error window, bold and underline mark the best and second-best results, including the Baseline. The matching improvements carry the same emphasis. Rankings use unrounded means. Other conventions follow Appendix A.

Rollout method Components	Step 1	Improv.	6–12	Improv.	13–30	Improv.	Full ± SE	Improv.	RC starts
Model FNO									
One-step / Baseline	0.329		5.010		> 10		8.620±0.843		
+MP	**0.142**	**57%**	1.128	77%	2.939	bnd.	1.535±0.183	82%	
+RC	0.329	0%	5.010	0%	> 10	—	4.164±0.471	52%	13
+MP+RC	**0.142**	**57%**	**0.984**	**80%**	**1.484**	bnd.	**0.977±0.053**	**89%**	7
Unrolling / Baseline	1.620		2.488		> 10		4.779±0.638		
+MP	**0.843**	**48%**	0.538	78%	1.419	bnd.	3.220±0.666	33%	
+RC	1.620	0%	1.374	45%	2.256	bnd.	1.352±0.137	72%	2
+MP+RC	**0.843**	**48%**	**0.529**	**79%**	**0.418**	bnd.	**0.901±0.069**	**81%**	2
Pushforward / Baseline	1.176		1.909		> 10		3.982±0.568		
+MP	**0.527**	**55%**	0.272	86%	> 10	—	5.524±0.855	−39%	
+RC	1.176	0%	1.347	29%	3.933	bnd.	1.848±0.231	54%	2
+MP+RC	**0.527**	**55%**	**0.246**	**87%**	> 10	—	3.336±0.428	16%	2
+MP(PA)	0.662	44%	0.735	62%	3.818	bnd.	2.008±0.286	50%	
+MP(PA)+RC	0.662	44%	0.633	67%	**2.284**	bnd.	**1.509±0.207**	**62%**	2
Model CNextU-Net									
One-step / Baseline	0.078		0.953		8.749		3.850±0.574		
+MP	**0.058**	**26%**	0.453	52%	4.731	46%	3.036±0.522	21%	
+RC	0.078	0%	0.713	25%	2.024	77%	1.117±0.101	71%	2
+MP+RC	**0.058**	**26%**	**0.380**	**60%**	**0.739**	**92%**	**0.871±0.099**	**77%**	6
Unrolling / Baseline	0.498		0.550		3.563		2.815±0.506		
+MP	**0.253**	**49%**	0.386	30%	3.099	13%	2.756±0.533	2%	
+RC	0.498	0%	0.508	8%	2.116	41%	1.875±0.312	33%	3
+MP+RC	**0.253**	**49%**	**0.340**	**38%**	**1.662**	**53%**	**1.675±0.321**	**40%**	2
Pushforward / Baseline	0.488		0.333		2.770		2.839±0.591		
+MP	**0.303**	**38%**	0.313	6%	2.582	7%	2.583±0.500	9%	
+RC	0.488	0%	0.312	6%	1.877	32%	2.234±0.457	21%	3
+MP+RC	**0.303**	**38%**	**0.284**	**15%**	**1.604**	**42%**	**1.795±0.326**	**37%**	4
+MP(PA)	0.345	29%	0.312	7%	2.571	7%	2.796±0.545	2%	
+MP(PA)+RC	0.345	29%	0.287	14%	1.686	39%	2.017±0.364	29%	2

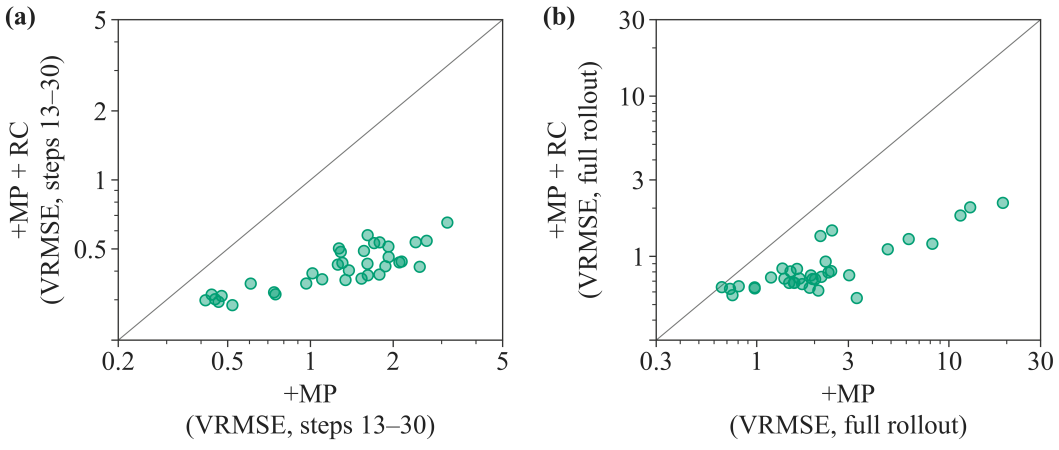


Figure 6: Paired errors for FNO+MP trained with unrolling on Rayleigh–Bénard. RC lowers error on all 35 test trajectories. Mean VRMSE falls from 1.419 to 0.418 over steps 13–30 and from 3.220 to 0.901 over all 194 steps. Each marker represents one trajectory, and the diagonal denotes equal errors. Axes are logarithmic with matched limits. RC starts at step 2 and leaves step 1 unchanged.

Table 8: Test VRMSE on shear flow. Ranking and reporting follow Table 7. Transport-cap handling is given in Appendix B.3. SC results for Rayleigh–Bénard and active matter appear in Appendix E.2.

Rollout method Components	Step 1	Improv.	6–12	Improv.	13–30	Improv.	Full ± SE	Improv.	RC starts
Model FNO									
One-step / Baseline	0.0091		0.070		> 10		> 10		
+MP	0.0071	22%	0.112	−61%	> 10	—	> 10	—	
+RC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
+MP+RC	0.0071	22%	**0.060**	**15%**	**0.179**	bnd.	> 10	—	6
+MP+SC	**0.0067**	**27%**	0.064	9%	> 10	—	> 10	—	
+MP+SC+RC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Unrolling / Baseline	0.033		0.063		4.196		> 10		
+MP	0.027	18%	0.048	23%	1.561	63%	> 10	—	
+RC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
+MP+RC	0.027	18%	0.049	21%	1.100	74%	> 10	—	8
+MP+SC	**0.024**	**26%**	**0.043**	**31%**	0.388	91%	> 10	—	
+MP+SC+RC	**0.024**	**26%**	0.044	30%	**0.231**	**95%**	> 10	—	9
Pushforward / Baseline	0.029		0.035		0.991		> 10		
+MP	0.025	13%	0.036	−3%	0.361	64%	> 10	—	
+RC	0.029	0%	0.043	−23%	0.235	76%	> 10	—	4^V
+MP+RC	0.025	13%	0.038	−8%	0.227	77%	> 10	—	10
+MP(PA)	0.025	13%	0.036	−5%	1.233	−25%	> 10	—	
+MP(PA)+RC	0.025	13%	0.036	−5%	0.979	1%	> 10	—	14
+MP+SC	0.023	20%	**0.034**	**2%**	0.112	89%	> 10	—	
+MP+SC+RC	0.023	20%	**0.034**	**2%**	**0.097**	**90%**	> 10	—	13
+MP(PA)+SC	**0.023**	**22%**	0.035	0%	0.120	88%	> 10	—	
+MP(PA)+SC+RC	**0.023**	**22%**	0.035	0%	0.116	88%	> 10	—	24
Model CNextU-Net									
One-step / Baseline	0.0045		0.050		0.162		5.047 ± 2.885		
+MP	0.0040	10%	**0.042**	**16%**	0.144	11%	4.283 ± 2.129	15%	
+RC	0.0045	0%	0.050	0%	0.203	−25%	2.450 ± 1.004	51%	20^V
+MP+RC	0.0040	10%	**0.042**	**16%**	0.142	12%	1.369 ± 0.479	73%	26
+MP+SC	**0.0039**	**12%**	0.046	8%	**0.124**	**23%**	$\mathbf{0.933 \pm 0.490}$	**82%**	
+MP+SC+RC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Unrolling / Baseline	0.041		**0.102**		0.468		> 10		
+MP	0.038	7%	0.114	−12%	0.231	51%	$\mathbf{2.513 \pm 1.179}$	bnd.	
+RC	0.041	0%	**0.102**	**0%**	0.452	3%	> 10	—	23
+MP+RC	0.038	7%	0.118	−15%	0.235	50%	2.552 ± 1.213	bnd.	6^H
+MP+SC	**0.036**	**12%**	0.106	−3%	**0.214**	**54%**	3.279 ± 1.714	bnd.	
+MP+SC+RC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Pushforward / Baseline	0.031		0.048		0.103		0.705 ± 0.416		
+MP	0.025	21%	0.045	6%	0.098	5%	1.196 ± 0.624	−70%	
+RC	0.031	0%	0.048	0%	0.109	−6%	0.761 ± 0.474	−8%	20^V
+MP+RC	0.025	21%	0.051	−7%	0.121	−17%	0.875 ± 0.518	−24%	8^V
+MP(PA)	0.027	15%	0.050	−4%	0.110	−7%	0.826 ± 0.535	−17%	
+MP(PA)+RC	0.027	15%	0.056	−17%	0.137	−33%	$\mathbf{0.692 \pm 0.361}$	**2%**	8^V
+MP+SC	**0.023**	**27%**	**0.044**	**8%**	**0.093**	**10%**	1.150 ± 0.616	−63%	
+MP+SC+RC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
+MP(PA)+SC	0.025	20%	0.048	1%	0.102	1%	0.821 ± 0.518	−16%	
+MP(PA)+SC+RC	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 9: Test VRMSE on active matter. Ranking and reporting follow Table 7.

Rollout method Components	Step 1	Improv.	6–12	Improv.	13–30	Improv.	Full ± SE	Improv.	RC starts
Model FNO									
One-step / Baseline	0.186		**1.143**		2.200		1.529±0.435		
+MP	**0.176**	**5%**	1.255	−10%	2.139	3%	1.503±0.417	2%	
+RC	0.186	0%	**1.143**	**0%**	1.937	12%	1.377±0.367	10%	16
+MP+RC	**0.176**	**5%**	1.255	−10%	**1.889**	**14%**	**1.371±0.370**	**10%**	14
Unrolling / Baseline	0.661		0.954		1.672		1.426±0.306		
+MP	**0.650**	**2%**	0.873	8%	1.599	4%	1.402±0.289	2%	
+RC	0.661	0%	0.940	2%	1.618	3%	1.400±0.298	2%	8^V
+MP+RC	**0.650**	**2%**	**0.862**	**10%**	**1.541**	**8%**	**1.369±0.275**	**4%**	8^V
Pushforward / Baseline	0.539		0.511		1.121		1.162±0.133		
+MP	**0.517**	**4%**	0.517	−1%	1.201	−7%	1.145±0.135	1%	
+RC	0.539	0%	**0.501**	**2%**	1.070	5%	1.121±0.118	4%	8^V
+MP+RC	**0.517**	**4%**	0.506	1%	1.149	−3%	1.119±0.122	4%	8^V
+MP(PA)	0.534	1%	0.514	−1%	1.061	5%	1.133±0.111	3%	
+MP(PA)+RC	0.534	1%	0.502	2%	**1.003**	**10%**	**1.082±0.094**	**7%**	8^V
Model CNextU-Net									
One-step / Baseline	**0.090**		**0.568**		1.649		1.370±0.262		
+MP	0.091	−1%	0.665	−17%	1.758	−7%	1.464±0.298	−7%	
+RC	**0.090**	**0%**	**0.568**	**0%**	**1.499**	**9%**	1.292±0.235	6%	15
+MP+RC	0.091	−1%	0.665	−17%	1.578	4%	**1.204±0.240**	**12%**	15
Unrolling / Baseline	0.617		0.824		1.727		1.391±0.278		
+MP	**0.540**	**12%**	**0.772**	**6%**	1.833	−6%	1.435±0.300	−3%	
+RC	0.617	0%	0.815	1%	**1.565**	**9%**	1.279±0.244	8%	11
+MP+RC	**0.540**	**12%**	**0.772**	**6%**	1.719	0%	**1.269±0.261**	**9%**	14
Pushforward / Baseline	0.373		**0.719**		1.722		1.431±0.294		
+MP	0.366	2%	0.743	−3%	1.660	4%	1.388±0.262	3%	
+RC	0.373	0%	**0.719**	**0%**	1.588	8%	1.279±0.239	11%	13
+MP+RC	0.366	2%	0.743	−3%	1.628	5%	1.287±0.237	10%	20
+MP(PA)	**0.351**	**6%**	0.720	0%	1.727	0%	1.383±0.286	3%	
+MP(PA)+RC	**0.351**	**6%**	0.719	0%	**1.588**	**8%**	**1.266±0.247**	**12%**	12

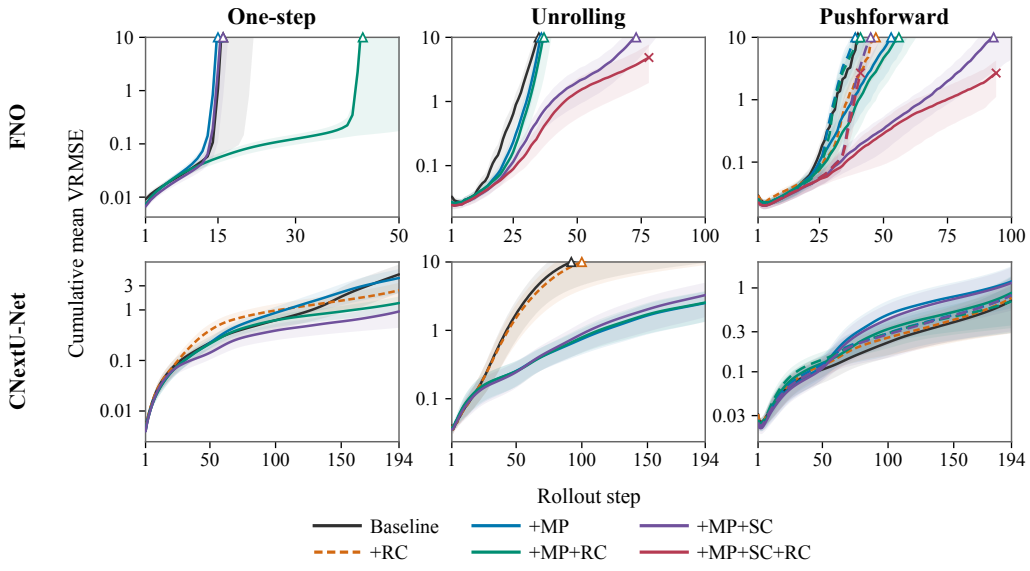


Figure 7: Cumulative mean VRMSE on shear flow (13 test trajectories). FNO panels show steps 1–50 for one-step training and 1–100 for unrolling and pushforward. Beyond these steps, the cumulative means exceed 10 or are undefined. CNextU-Net panels show all 194 steps.

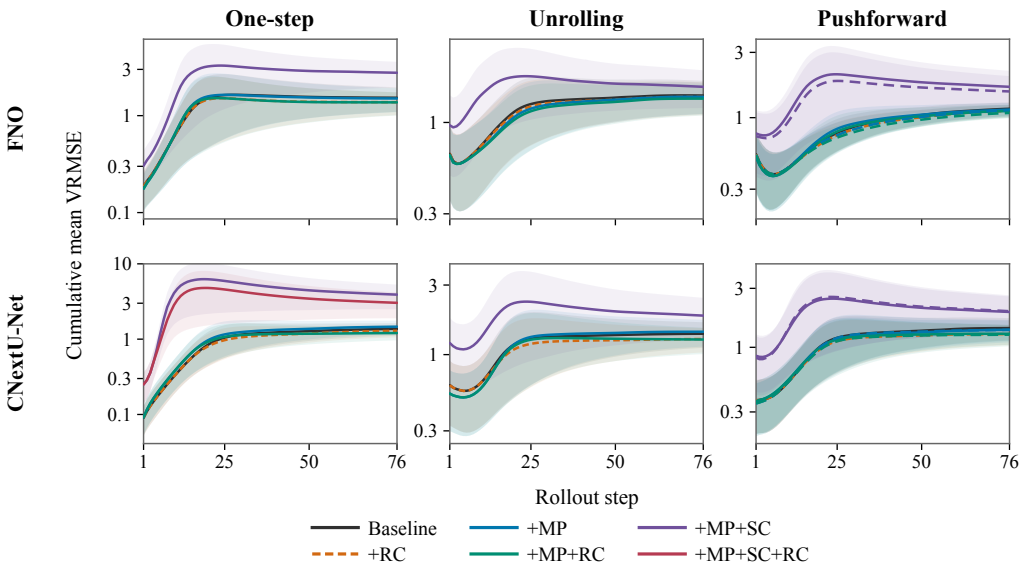


Figure 8: Cumulative mean VRMSE on active matter over 76 forecast steps (26 test trajectories).

E COMPONENT AND MATCHED-TRAINING CONTROLS

E.1 PROJECTION AND TRAINING OBJECTIVE

Training through the projection. A separate one-step CNextU-Net control on Rayleigh–Bénard compares the raw model, projection applied only during rollout, and training through MP. Training through the projected transition gives the lowest error. Over steps 13–30, VRMSE is 8.749 for the raw model, 7.279 for rollout-only projection and 4.731 for MP. MP has lower error than rollout-only projection on 28 of 35 trajectories. Section 4.2 reports the FNO comparisons with MP and PA.

Discarded-velocity penalty. A one-step FNO+MP control on Rayleigh–Bénard applies the discarded-velocity penalty outside the pushforward recurrence. Its full-rollout VRMSE is 2.505, compared with 1.535 for +MP without the penalty. PA is evaluated as the combination of raw-proposal feedback and the penalty.

E.2 SCALAR TRANSPORT ACROSS PHYSICAL SYSTEMS

Table 10 compares SC with MP and the same MP model trained one more hour. Continued training and SC use update budgets of 60 and 50 minutes, respectively. On shear flow, SC lowers steps 13–30 error relative to both controls in seven of eight models, averaging 38% against MP and 37% against continued training over bounded pairs. Full-rollout gains are narrower, occurring in three models against MP and one against continued training, with four pairs above 10. Table 11 reports all long-horizon means.

The coupled systems reveal a different dependence on horizon. On Rayleigh–Bénard, SC lowers step-1 and full-rollout errors in all eight MP models, including a 20% mean full-rollout reduction against one-hour continued training. However, it raises steps 6–12 error in all eight and improves steps 13–30 in only two. On active matter, the extension transports concentration and orientation with a local orientation law and learned closure. It raises mean error in every window for all eight models against both controls. These results delimit the benefit of the evaluated SC extension when the carried fields also drive the flow.

Table 10: SC on all 24 MP models. Counts give the number of models with lower mean error than MP or MP trained one more hour (MP + 1 h), out of eight models per dataset. Parentheses count pairs that are equal or both exceed 10.

Dataset	Compared with	Step 1	6–12	13–30	Full
Shear flow	MP	8/8	7/8	7/8 (1)	3/8 (4)
	MP + 1 h	8/8	6/8	7/8 (1)	1/8 (4)
Rayleigh–Bénard	MP	8/8	0/8	2/8 (1)	8/8
	MP + 1 h	8/8	0/8	2/8 (1)	8/8
Active matter	MP	0/8	0/8	0/8	0/8
	MP + 1 h	0/8	0/8	0/8	0/8

Table 11: SC on all 24 MP models over long horizons. Mean VRMSE of the 12-hour MP model (MP), the same model trained one more hour (+1 h), and SC after joint adaptation (+SC). Bold marks the best result within each row and window.

		Steps 13–30			Full rollout		
Model	Rollout method	MP	+1 h	+SC	MP	+1 h	+SC
Shear flow (passive tracer, periodic domain, main configuration)							
FNO	One-step	> 10	> 10	> 10	> 10	> 10	> 10
FNO	Unrolling	1.561	1.503	**0.388**	> 10	> 10	> 10
FNO	Pushforward	0.361	0.337	**0.112**	> 10	> 10	> 10
FNO	Pushforward (PA)	1.233	1.401	**0.120**	> 10	> 10	> 10
CNextU-Net	One-step	0.144	0.145	**0.124**	4.283	4.816	**0.933**
CNextU-Net	Unrolling	0.231	0.227	**0.214**	2.513	**2.255**	3.279
CNextU-Net	Pushforward	0.098	0.095	**0.093**	1.196	**1.137**	1.150
CNextU-Net	Pushforward (PA)	0.110	0.108	**0.102**	0.826	**0.821**	0.821
Rayleigh–Bénard (buoyancy-driven, heated from below)							
FNO	One-step	2.939	**2.841**	3.097	1.535	1.492	**1.277**
FNO	Unrolling	1.419	1.398	**0.659**	3.220	3.238	**0.981**
FNO	Pushforward	> 10	> 10	> 10	5.524	5.523	**4.405**
FNO	Pushforward (PA)	3.818	**3.788**	5.462	2.008	1.998	**1.641**
CNextU-Net	One-step	4.731	4.327	**2.625**	3.036	2.986	**2.236**
CNextU-Net	Unrolling	**3.099**	3.153	3.910	2.756	2.789	**2.747**
CNextU-Net	Pushforward	2.582	**2.466**	3.571	2.583	2.553	**2.479**
CNextU-Net	Pushforward (PA)	2.571	**2.559**	3.256	2.796	2.791	**2.588**
Active matter (orientation-source extension)							
FNO	One-step	**2.139**	2.227	3.952	**1.503**	1.565	2.769
FNO	Unrolling	1.599	**1.598**	1.976	1.402	**1.393**	1.599
FNO	Pushforward	**1.201**	1.215	2.563	**1.145**	1.155	1.681
FNO	Pushforward (PA)	**1.061**	1.087	2.334	**1.133**	1.148	1.555
CNextU-Net	One-step	1.758	**1.735**	5.770	1.464	**1.428**	3.901
CNextU-Net	Unrolling	**1.833**	1.880	2.718	**1.435**	1.446	1.851
CNextU-Net	Pushforward	**1.660**	1.663	2.915	1.388	**1.387**	1.933
CNextU-Net	Pushforward (PA)	1.727	**1.704**	2.979	1.383	**1.380**	1.960

E.3 Correction direction

The half-innovation control uses the same response target and correction bounds without the input Jacobian. It is also eligible for deployment under the selection procedure in Appendix C.3. Table 12 compares the selected rule with the alternative at the same starting step on the same frozen model. Alternative-rule scores are available for 42 deployments, including 12 SC models. The selected rule lowers steps 13–30 error in 32 cases by 13% on average over bounded pairs, and full-rollout error in 28.

Table 12: Selected RC update rule versus the alternative on the same frozen model and starting step. Scores for both rules are available for 42 deployments selected on held-out training trajectories, including 12 SC models. Lower/higher/same counts use a 0.5% tolerance and treat pairs above 10 as the same.

Dataset	Deployments	13–30 (lower/higher/same)	Improv.	Full (lower/higher/same)
Rayleigh–Bénard	22	15/4/3	15%	18/2/2
Shear flow	10	9/1/0	18%	1/1/8
Active matter	10	8/1/1	4%	9/1/0
All	42	32/6/4	13%	28/4/10

Table 13 compares RC with direct state correction and a random admissible direction. At the same step bound, RC has lower steps 13–30 error than direct correction on all four tested models and lower full-rollout error in all three bounded comparisons. The random control raises error relative to the model alone in every bounded comparison. The shear-flow full-rollout means all exceed 10. RC nevertheless improves on the random control for ten of thirteen shear-flow trajectories, with the other three tied.

The direct-correction companion uses the same initialization, data and one-hour training budget to predict the ground-truth next state, with output $\widehat{\mathbf{y}}$. Its proposal is $\widehat{\mathbf{x}} + \beta P_c(\widehat{\mathbf{y}} - \widehat{\mathbf{x}})$, subject to RC's 5% bound. We report the best tested $\beta \in \{0.25, 0.5, 1\}$ for each model. On Rayleigh–Bénard, the bound is active throughout, making all three fractions equivalent. The random control projects a random direction onto the same admissible set and matches the norm of RC's bounded proposal

before acceptance. Both controls begin at the same rollout step as RC and retain finite state updates, while RC additionally requires a finite response and decreased response misfit. These comparisons therefore assess direction and acceptance together.

Table 13: RC versus direct state correction at the same step bound and a random admissible direction with the norm of RC's bounded proposal. Means and trajectory counts use paired test evaluations with the same frozen model and starting step. Bold marks the best bounded result per row.

Model (+MP)	Window	Model alone	RC	Direct correction same bound	Random direction RC step norm
Rayleigh–Bénard, FNO, unrolling	13–30	1.419	**0.418**	0.621	> 10
	Full	3.220	**0.901**	1.989	7.480
Rayleigh–Bénard, CNextU-Net, one-step	13–30	4.731	**0.739**	1.657	4.734
	Full	3.036	**0.871**	1.295	3.048
Shear flow, FNO, one-step	13–30	> 10	**0.179**	0.206	> 10
	Full	> 10	> 10	> 10	> 10
Active matter, CNextU-Net, one-step	13–30	1.758	**1.578**	1.642	3.296
	Full	1.464	**1.204**	1.398	3.138

Trajectories with lower RC error (13–30, full)	Against direct correction	Against random direction
Rayleigh–Bénard, FNO	34/35, 33/35	35/35, 35/35
Rayleigh–Bénard, CNextU-Net	35/35, 33/35	35/35, 35/35
Shear flow, FNO	9/13, —	13/13, 10/13
Active matter, CNextU-Net	20/26, 26/26	26/26, 26/26

E.4 Matched continued-training controls

All 42 Baseline and MP models have a one-hour continued-training control, and all 24 MP models also have a second consecutive hour. Each control starts from the model's last checkpoint and keeps its training strategy, batch size and, for MP models, the projection. AdamW uses 0.05 times the original peak learning rate, a 5% warm-up and cosine decay. Pushforward resumes the curriculum from the saved checkpoint. Hour two starts from hour one's last checkpoint.

Table 14 compares additional training with RC and SC+RC. Over 40 RC deployments, one extra training hour lowers steps 13–30 error by 2% on average and changes full-rollout error by less than 0.5%, relative to the same 12-hour model. RC reduces these errors by 20% and 26% and outperforms the one-hour control in 32 and 31 cases. SC+RC outperforms two additional training hours over steps 13–30 in nine of twelve deployments, averaging 29% over bounded pairs. The three shear-flow deployments improve by 82% on average in this window, but all full-rollout pairs exceed 10. The active-matter deployment is worse than continued training in both windows. For convection FNO with unrolling, SC+RC improves on the two-hour control on 35/35 trajectories over steps 13–30 and 34/35 over the full rollout.

Table 14: Matched continued-training controls on the test split. RC and SC+RC are compared with the same 12-hour model trained for one and two additional hours, respectively. (a) Counts give lower/higher/same, with same indicating a difference within 0.5% or both errors above 10. (b) Mean ± SE for Rayleigh–Bénard FNO with unrolling. RC starts at step 3 for +MP+SC+RC. Bold marks the best in each window.

(a) Summary of matched controls

Comparison	Group	Models	13–30	Improv.	Full	Improv.
RC vs +1 h	Rayleigh–Bénard	14	12/0/2	51%	14/0/0	46%
	Shear flow	12	6/6/0	5%	4/2/6	24%
	Active matter	14	14/0/0	7%	13/0/1	8%
	Baseline models	16	12/3/1	16%	12/1/3	28%
	MP models	24	20/3/1	20%	19/1/4	25%
	All	40	32/6/2	19%	31/2/7	26%
SC+RC vs +2 h	Rayleigh–Bénard	8	6/0/2	34%	8/0/0	45%
	Shear flow	3	3/0/0	82%	0/0/3	—
	Active matter	1	0/1/0	−165%	0/1/0	−110%
	All	12	9/1/2	29%	8/1/3	27%

(b) Rayleigh–Bénard FNO, unrolling

Configuration	Step 1	6–12	13–30	Full rollout
+MP, two more training hours	0.836±0.091	**0.535±0.044**	1.419±0.120	3.233±0.667
+MP+SC	**0.820±0.088**	0.730±0.090	0.659±0.063	0.981±0.127
+MP+SC+RC	**0.820±0.088**	0.722±0.089	**0.623±0.060**	**0.798±0.045**

F TRANSOLVER++

MP and SC also apply to the attention-based Transolver++ model (Luo et al., 2025). Our implementation uses the released Transolver++ architecture, including its Gumbel-softmax slice assignment. Transolver++ uses the training budget, hardware and checkpoint selection described in Appendix A. The model has eight layers, width 128, eight attention heads, 64 slices, multilayer perceptron (MLP) width ratio two and 0.86 million parameters. For autoregressive prediction, inputs contain four frames of every field, a constant node-type channel and both grid coordinates. The learning rate is 10^{-3} and the batch size is four, except for the unrolled MP and MP+SC models, which use three. Evaluation covers Rayleigh–Bénard under all three strategies and both pushforward variants, with SC adapted within one additional hour on each MP model.

Table 15 gives all evaluated configurations. MP lowers error wherever at least one compared mean is at most 10, including step-1 reductions of 17–42%. SC further lowers each MP model's full-rollout error, by 42% under standard pushforward and 51% under PA, and brings steps 13–30 below 10 for both pushforward variants. One-step and unrolled models still exceed 10 in that window. RC is not applied because this implementation resamples its slice assignment on each forward pass, while RC requires a fixed response map.

Table 15: Transolver++ with MP and SC on Rayleigh–Bénard. Bounded windows report mean ± SE. SC is adapted within one additional hour and follows the continuation policy in Appendix B.3. Improvements are relative to the matching Baseline, shared by both pushforward variants. Bold/underline mark the best/second-best across configurations in each window.

Components	Step 1	Improv.	6–12	Improv.	13–30	Improv.	Full	Improv.
Rayleigh–Bénard One-step								
Baseline	0.36 ± 0.03		> 10		> 10		> 10	
+MP	$\underline{0.21 \pm 0.02}$	42%	> 10	—	> 10	—	7.59 ± 0.21	bnd.
+MP+SC	$\mathbf{0.20 \pm 0.02}$	43%	> 10	—	> 10	—	7.50 ± 0.22	bnd.
Rayleigh–Bénard Unrolling								
Baseline	1.48 ± 0.10		8.83 ± 0.31		> 10		> 10	
+MP	1.02 ± 0.07	31%	6.87 ± 0.18	22%	> 10	—	> 10	—
+MP+SC	0.96 ± 0.05	35%	4.42 ± 0.13	50%	> 10	—	6.02 ± 0.61	bnd.
Rayleigh–Bénard Pushforward								
Baseline	1.34 ± 0.11		6.65 ± 0.19		> 10		9.61 ± 1.02	
+MP	1.09 ± 0.07	19%	$\underline{2.47 \pm 0.04}$	63%	> 10	—	5.04 ± 0.50	48%
+MP+SC	1.05 ± 0.06	21%	$\mathbf{1.98 \pm 0.08}$	70%	$\mathbf{4.96 \pm 0.56}$	bnd.	$\underline{2.91 \pm 0.44}$	70%
Rayleigh–Bénard Projection-aware pushforward								
+MP	1.11 ± 0.10	17%	2.99 ± 0.12	55%	> 10	—	5.18 ± 0.57	46%
+MP+SC	1.02 ± 0.08	24%	2.84 ± 0.11	57%	$\underline{7.12 \pm 0.57}$	bnd.	$\mathbf{2.52 \pm 0.17}$	74%