CORRECTING STATE-DEPENDENT PROPAGATION DEFECTS IN NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural operators provide efficient surrogate models for the time evolution of parameterized partial differential equations (PDEs). Existing spectral neural operators reuse a shared response across input states, while nonlinear dynamics exhibit state-varying local responses. To address this structural discrepancy, we introduce State-Dependent Propagation Defect Correction (SDC), which enables an observable one-step propagation defect between a numerical reference map and a frozen neural propagator. We establish the connection between state-varying local responses and the propagation defect and show that standard prediction-space calibration is equivalent to regression of the observable defect. SDC-based operator learning is designed by a low-rank complex spectral update that adapts global mode-wise channel responses and a bias-free bottleneck that provides nonlinear state-conditioned correction from globally mixed hidden features. Both correction paths are zero-initialized and the pretrained propagator remains frozen during calibration, preserving the original propagation map at initialization. Experiments on time-dependent PDEs and weather forecasting demonstrate improved autoregressive prediction with minimal parameters and computational overhead.

1 INTRODUCTION

Parameterized partial differential equations (PDEs) arise in scientific applications including weather forecasting, fluid dynamics, and geophysical simulation (Karniadakis et al., 2021; Brunton et al., 2020; Dueben & Bauer, 2018). Numerical solvers provide accurate trajectories of these systems, while repeated simulation over long temporal horizons and high-resolution spatial domains can incur substantial computational cost. Neural operators provide a data-driven approach to learning mappings between function spaces (Kovachki et al., 2023; Lu et al., 2021; Li et al., 2021). Message-passing solvers and benchmark suites have extended this operator-learning view to discretized PDE trajectories (Brandstetter et al., 2022; Takamoto et al., 2022). For evolutionary PDEs, these models can be trained as finite-time propagators that map a current physical state to a subsequent state.

Existing neural operators construct nonlocal interactions through different mathematical representations. Fourier Neural Operator (FNO) parameterizes global interactions using learned Fourier multipliers (Li et al., 2021); multiwavelet and Galerkin formulations provide alternative bases and integral transforms (Gupta et al., 2021; Cao, 2021); and learned-deformation Fourier operators modify the spectral parameterization for efficiency or geometry (Li et al., 2023). U-shaped, general transformer, and convolutional operators provide further spatial realizations (Rahman et al., 2023; Hao et al., 2023; Raonic et al., 2023). On spherical domains, equivariant spherical convolutions and spherical-harmonic operators encode the geometry explicitly (Cohen et al., 2018; Esteves et al., 2018; Bonev et al., 2023), while Green-function-based constructions derive integral kernels from solution representations of differential operators (Tang et al., 2026). In these spectral and kernel formulations, the parameters governing the nonlocal response are fixed after training and reused across input states. The complete neural operator remains nonlinear through pointwise transformations, nonlinear activations, and composition of multiple operator blocks.

A structural gap arises from nonlinear governing dynamics: existing spectral neural operators reuse a shared nonlocal response across input states, while nonlinear dynamics exhibit state-varying local responses. Classical dynamical-systems analysis and ensemble forecasting make the same distinction between local linearization and trajectory-specific sensitivity

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.

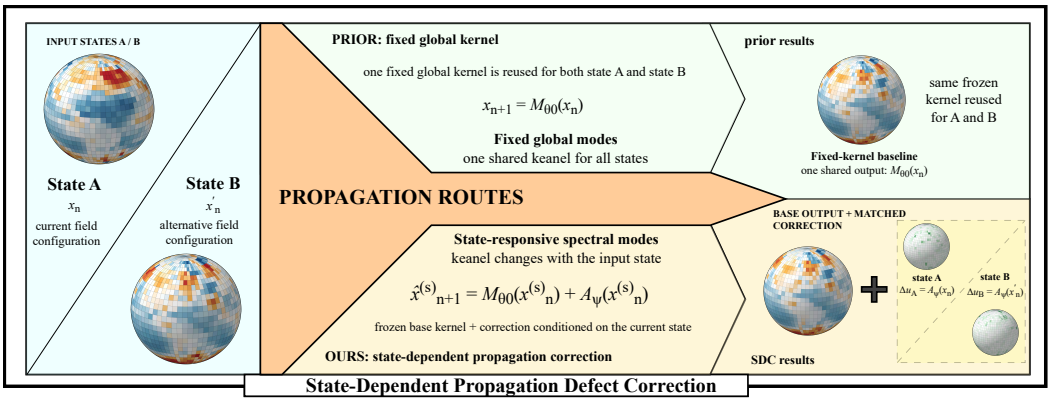

Figure 1: Theory motivation for state-dependent propagation defect correction. A shared nonlocal response reuses the same propagation map for different input states, whereas SDC pairs each state with its own learned defect correction before producing the next state.

(Lorenz, 2017; Leutbecher & Palmer, 2008). Learned forecasters also accumulate mismatch under autoregressive rollout, where local errors are repeatedly fed back into the next state (Rasp et al., 2020; Kochkov et al., 2024). For evolutionary PDEs, state-varying local responses determine the sensitivity of the finite-time propagation map. The discrepancy between a numerical reference map and a trained neural propagator can therefore be characterized at the level of one-step state transitions. We refer to this state-valued discrepancy as the *propagation defect*. Its variation across states provides an observable target for correcting the trained propagator while retaining the original model as a fixed reference.

Figure 1 gives the conceptual picture. A shared nonlocal response reuses one propagation rule for different states, while SDC adds a correction paired with the current state. The difference from the reference update is the observable propagation defect used for calibration.

Based on this formulation, we introduce *State-Dependent Propagation Defect Correction* (SDC). SDC embeds a compact correction into the trained neural operator and calibrates it against the observed propagation defect. A low-rank spectral update adjusts global mode-wise channel responses, while a nonlinear bottleneck applied to globally mixed hidden features provides state-conditioned correction. The original propagator remains frozen during calibration, and both correction paths are zero-initialized so that the adapted model starts from the original propagation map. In this work, we instantiate SDC with a spherical spectral neural operator and evaluate its behavior in autoregressive PDE forecasting.

Our contributions are summarized as follows:

- We characterize the structural gap between shared nonlocal responses in neural operators and state-varying local responses in nonlinear dynamics, and connect this distinction to finite-time PDE propagation.
- We formulate the state-dependent propagation defect between a numerical reference map and a frozen neural propagator as an observable correction target for supervised operator learning.
- We introduce SDC, which parameterizes the propagation defect through complementary low-rank spectral and nonlinear state-conditioned correction paths while preserving the pretrained propagator as the reference model.
- Extensive experiments on time-dependent PDEs and weather forecasting show the improved performance and state-dependent correction behavior.

2 Related Work

Neural operators. Neural operators learn mappings between function spaces through local transformations and nonlocal interactions (Kovachki et al., 2023; Lu et al., 2021). Fourier and multiwavelet

operators parameterize global interactions in transform coordinates (Li et al., 2021; Gupta et al., 2021), while Galerkin and learned-deformation variants change the response parameterization or computational cost (Cao, 2021; Li et al., 2023). U-shaped and convolutional constructions address multiscale spatial structure (Rahman et al., 2023; Raonic et al., 2023); operator transformers extend the same setting to irregular inputs and general token interactions (Hao et al., 2023). Transolver and SineNet are recent baselines for general geometries and time-dependent PDEs, respectively (Wu et al., 2024; Zhang et al., 2024). These constructions use learned nonlocal response parameters that are shared across input states.

Spherical and weather forecasting. Spherical CNNs and equivariant spherical representations provide geometry-aware alternatives to planar convolutions (Cohen et al., 2018; Esteves et al., 2018). SFNO uses spherical harmonics for stable global dynamics, while GSNO derives a spherical operator from Green's functions (Bonev et al., 2023; Tang et al., 2026). WeatherBench established a common data and metric protocol for data-driven forecasting, and ERA5 supplies the reanalysis fields used by many global benchmarks (Rasp et al., 2020; Hersbach et al., 2020). Recent global forecasters include FourCastNet, GraphCast, and NeuralGCM (Kurth et al., 2023; Lam et al., 2023; Kochkov et al., 2024). Our WeatherBench experiment follows this evaluation lineage but tests state-dependent correction around a frozen spherical operator.

Defect and deferred correction. Defect correction improves an approximate solution using its residual or defect. Deferred-correction methods iteratively estimate and correct errors in a provisional numerical solution, and spectral deferred correction extends this principle through high-order quadrature formulations (Dutt et al., 2000). Our use of the term propagation defect follows this predictor–defect–correction lineage, but the learned target is the observable error of a neural one-step map rather than a discretization residual inside a numerical integrator.

Parameter-efficient adaptation. Parameter-efficient adaptation modifies a pretrained model using a small number of trainable parameters while retaining most pretrained parameters. Bottleneck adapters and hypercomplex adapters introduce low-dimensional trainable modules within a frozen network (Houlsby et al., 2019; Karimi Mahabadi et al., 2021), while low-rank adaptation represents weight updates using trainable low-rank factors (Hu et al., 2022). Function-preserving initialization and zero-initialized residual paths provide related ways to add capacity without changing the reference map at initialization (Bachlechner et al., 2021; Zhang et al., 2023). In this work, low-rank and bottleneck adaptations parameterize compact corrections within a frozen neural operator.

3 PRELIMINARIES

3.1 SPECTRAL AND KERNEL REALIZATIONS OF NONLOCAL PROPAGATION

Let Ω be a spatial domain and let v be a vector-valued hidden field. The nonlocal path in a neural operator can be represented in a transform or kernel form. A common transform-domain notation is

$$\mathcal{K}_\vartheta v = \mathcal{T}^{-1}(R_\vartheta \mathcal{T} v)\,, \tag{1}$$

where $\mathcal{T}$ is a domain-adapted transform and R_ϑ is a mode-wise channel response. FNO uses a Fourier transform and truncated Fourier multipliers (Li et al., 2021), while SFNO replaces the Fourier basis by spherical harmonics and applies a symmetry-constrained response to the spherical coefficients (Bonev et al., 2023). A Green-kernel construction uses the integral form (Tang et al., 2026)

$$(\mathcal{K}_\vartheta v)(u) = \int_\Omega G_\vartheta(u,\eta) v(\eta)\, d\eta. \tag{2}$$

For a fixed kernel and an appropriate symmetry-adapted basis, the kernel form admits the same modal-response interpretation as Eq. 1; the complete Green-to-spectral derivation is given in Appendix A.2. The shared-response structure also follows from the underlying linear solution operator. Consider a fixed linear control problem $\mathcal{D}g = f$, with $\mathcal{D} : \mathcal{U} \to \mathcal{Y}$. When the problem is well posed, its inverse response is $\mathcal{G}_0 \coloneqq \mathcal{D}^{-1}$ and $g = \mathcal{G}_0 f$. The same operator $\mathcal{G}_0$ acts on every admissible input and satisfies $\mathcal{G}_0(af_1 + bf_2) = a\mathcal{G}_0(f_1) + b\mathcal{G}_0(f_2)$. When a Green representation exists, $\mathcal{G}_0$ is realized by a fixed kernel under the prescribed operator and boundary conditions; Appendix A.2 gives the corresponding derivation. These formulations share the same structural property used in this work: once trained, the nonlocal response parameters are reused across input states. For fixed

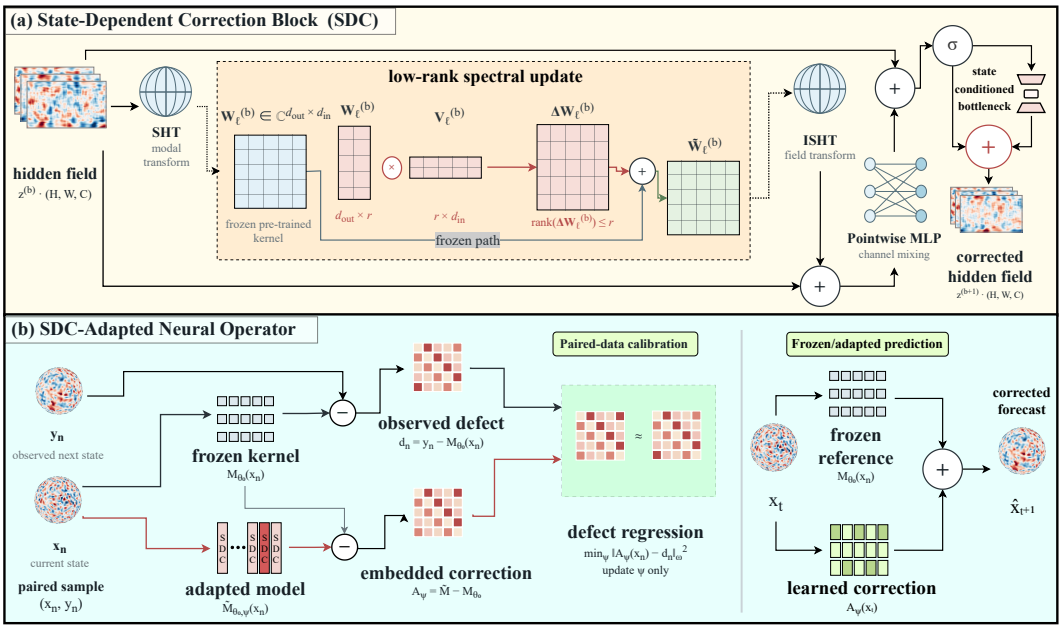

Figure 2: SDC architecture with a frozen GSNO backbone. The left panel shows the spectral and state-conditioned correction block. The right panel shows inference through the frozen GSNO encoder and decoder, together with the paired training path used to learn the correction from one-step targets.

$$\vartheta = \vartheta_0,$$

$$\mathcal{K}_{\vartheta_0} = \mathcal{T}^{-1} \circ R_{\vartheta_0} \circ \mathcal{T}, \qquad \vartheta(x) \equiv \vartheta_0, \qquad \mathcal{K}_{\vartheta_0}(av_1 + bv_2) = a\mathcal{K}_{\vartheta_0}(v_1) + b\mathcal{K}_{\vartheta_0}(v_2). \tag{3}$$

The first relation specifies the response action, the second its cross-state reuse, and the third superposition in the hidden-field argument. In a finite implementation, a projection P_K retains a prescribed set of modes, giving $\mathcal{K}_{\vartheta_0,K} = \mathcal{T}^{-1} P_K R_{\vartheta_0,K} P_K \mathcal{T}$. Mode truncation and quadrature are therefore part of the implemented reference operator.

3.2 Shared response within a nonlinear neural operator

The linearity in Eq. 3 applies to the nonlocal path for fixed parameters. A complete operator block also contains pointwise channel maps and nonlinearities (Bonev et al., 2023; Tang et al., 2026),

$$v^{(b+1)}(u) = \sigma\Big(W^{(b)} v^{(b)}(u) + \mathcal{K}_{\vartheta_b} v^{(b)}(u)\Big), \tag{4}$$

and the resulting state-to-state map is nonlinear through pointwise transformations, normalization, nonlinear activations, and block composition. We therefore use *shared response* specifically for the nonlocal propagation path: the learned response parameters are reused as the current input state changes. Section 4 examines this shared-response structure in relation to the state-varying local response of nonlinear governing dynamics.

4 Methodology

This section develops state-dependent propagation defect correction for neural operators with a shared nonlocal response. We first characterize the local response of a nonlinear governing operator around different solution states. We then transfer this state dependence to finite-time evolution and define an observable one-step propagation defect. Finally, we parameterize the defect inside the trained operator using compact spectral and nonlinear correction paths and calibrate them relative to the frozen reference model.

4.1 FROM SHARED TO STATE-DEPENDENT LOCAL RESPONSES

Let $\mathcal{N} : \mathcal{V} \to \mathcal{Y}$ be a differentiable nonlinear governing operator,

$$\mathcal{N}(g) = f. \tag{5}$$

To characterize its local response, consider a reference solution g^* satisfying $\mathcal{N}(g^*) = f^*$. Perturbing the input from f^* to $f^* + \delta f$ induces a corresponding solution perturbation from g^* to $g^* + \delta g$, such that $\mathcal{N}(g^* + \delta g) = f^* + \delta f$. The Fréchet expansion around g^* gives

$$\mathcal{N}(g^* + \delta g) = \mathcal{N}(g^*) + D\mathcal{N}[g^*](\delta g) + o(\|\delta g\|). \tag{6}$$

Using $\mathcal{N}(g^*) = f^*$, the first-order perturbation satisfies $D\mathcal{N}[g^*](\delta g) = \delta f$. When the derivative is locally invertible, the corresponding local response is

$$\mathcal{G}_{g^*} \coloneqq (D\mathcal{N}[g^*])^{-1}, \qquad \delta g = \mathcal{G}_{g^*}\delta f. \tag{7}$$

The derivative $D\mathcal{N}[g^*]$ depends on the reference state for a general nonlinear operator, yielding a family of local responses $\{\mathcal{G}_{g^*}\}$. If $\mathcal{N}$ is twice Fréchet differentiable and $D\mathcal{N}[g^*]$ remains boundedly invertible along $g^* + \varepsilon h$, its first-order variation is

$$\mathcal{G}_{g^*+\varepsilon h} = \mathcal{G}_{g^*} - \varepsilon\, \mathcal{G}_{g^*} \circ \mathcal{H}_{g^*,h} \circ \mathcal{G}_{g^*} + o(\varepsilon), \qquad \mathcal{H}_{g^*,h}[\delta g] \coloneqq D^2\mathcal{N}[g^*][h, \delta g]. \tag{8}$$

The nonlinear governing equation therefore admits state-indexed local responses, whereas the nonlocal path in Section 3 reuses one learned response parameterization across states. The complete neural operator remains a nonlinear state-to-state map as described in Section 3. This response-level distinction motivates the finite-time formulation developed next.

4.2 FROM LOCAL RESPONSE MISMATCH TO A ONE-STEP PROPAGATION DEFECT

Let $\mathcal{X}$ be the space of complete physical states and let $\mathcal{F}_{\mathrm{dyn}} : \mathcal{X} \to \mathcal{X}$ denote the spatially discretized dynamical vector field,

$$\frac{dx(t)}{dt} = \mathcal{F}_{\mathrm{dyn}}(x(t)). \tag{9}$$

Assuming differentiability along the trajectory, the continuous flow $\phi_{\Delta t}$ generated by Eq. 9 has tangent map

$$D\phi_{\Delta t}[x_n] = \mathcal{P} \exp\left(\int_{t_n}^{t_n+\Delta t} D\mathcal{F}_{\mathrm{dyn}}[x(s)]\, ds \right), \tag{10}$$

where $\mathcal{P}$ preserves the ordering of generally non-commuting Jacobians. The finite-time response accumulates local dynamics along the trajectory and varies with the evolving state.

In supervised operator learning, a numerical solver supplies the discrete reference map $\Phi_{\Delta t} : \mathcal{X} \to \mathcal{X}$, with $x_{n+1} = \Phi_{\Delta t}(x_n)$. Let $M_{\theta_0} : \mathcal{X} \to \mathcal{X}$ denote the frozen single-state neural propagator, including its encoder, nonlocal blocks, pointwise paths, decoder, and residual connections. Their local finite-time response discrepancy at state x is $D\Phi_{\Delta t}[x] - DM_{\theta_0}[x]$. We define the corresponding state-valued propagation defect as

$$A^\star(x) \coloneqq \Phi_{\Delta t}(x) - M_{\theta_0}(x), \qquad \Phi_{\Delta t}(x) = M_{\theta_0}(x) + A^\star(x). \tag{11}$$

For a supervised pair (x_n, x_{n+1}), the target is directly observable as $A^\star(x_n) = x_{n+1} - M_{\theta_0}(x_n)$. The propagation defect measures the complete discrepancy relative to the selected reference map and can include shared-response mismatch, mode truncation, finite model capacity, time discretization, and optimization error. Differentiating Eq. 11 gives the direct link between the observable defect and the local response discrepancy,

$$DA^\star[x] = D\Phi_{\Delta t}[x] - DM_{\theta_0}[x]. \tag{12}$$

The derivative of the state-valued defect therefore represents the difference between the local finite-time responses of the reference and learned propagators. Equation 12 connects the state-indexed response in Section 4.1 to a correction target available directly from supervised trajectories.

4.3 STATE-DEPENDENT DEFECT CORRECTION: PARAMETERIZATION AND CALIBRATION

We approximate $A^\star$ with a compact correction A_ψ embedded in the trained propagator, while keeping θ_0 frozen. The parameterization combines low-rank adaptation for the global spectral response with a nonlinear bottleneck adapter for state-conditioned hidden features. These mechanisms follow established parameter-efficient adaptation strategies (Hu et al., 2022; Houlsby et al., 2019; Karimi Mahabadi et al., 2021) and provide complementary correction coordinates within the original operator graph.

Spectral and state-conditioned correction. For transformed mode k in block b, we augment the frozen response $R_k^{(b)}$ using a low-rank update (Hu et al., 2022),

$$\widetilde{R}_k^{(b)} = R_k^{(b)} + \Delta R_k^{(b)}, \qquad \Delta R_k^{(b)} = \frac{1}{r} U_{R,k}^{(b)} V_{R,k}^{(b)}, \tag{13}$$

where $U_{R,k}^{(b)} \in \mathbb{C}^{d_{\mathrm{out}} \times r}$ and $V_{R,k}^{(b)} \in \mathbb{C}^{r \times d_{\mathrm{in}}}$. The index k may represent Fourier modes, spherical-harmonic degrees, or another modal coordinate. This route provides a compact global modification of the mode-wise channel response. The learned $\Delta R_k^{(b)}$ is fixed after calibration.

State conditioning is introduced through a narrow bottleneck adapter (Houlsby et al., 2019) applied after global mixing. For hidden representation z, we use $H_\psi^{(b)}(z) = U_c^{(b)} \,\mathrm{GELU}(D_c^{(b)} z)$, with $D_c^{(b)} \in \mathbb{R}^{r_c \times d}$ and $U_c^{(b)} \in \mathbb{R}^{d \times r_c}$. Its output depends on the current hidden field and provides the nonlinear state-conditioned component of the correction. For a frozen base block $\mathcal{B}_{\theta_0}^{(b)}$, the complete adapted block is summarized by

$$\widetilde{\mathcal{B}}_\psi^{(b)}(z) = \mathcal{B}_{\theta_0}^{(b)}\left(z; R^{(b)} + \Delta R_\psi^{(b)}\right) + H_\psi^{(b)}\left(\mathcal{B}_{\theta_0}^{(b)}\left(z; R^{(b)} + \Delta R_\psi^{(b)}\right)\right). \tag{14}$$

Across the adapted blocks, the spectral and bottleneck paths jointly induce the function-level correction $A_\psi(x) \coloneqq \widetilde{M}_{\theta_0,\psi}(x) - M_{\theta_0}(x)$. The spectral route modifies global modal response, while the bottleneck route introduces explicit dependence on the current hidden state. The resulting $A_\psi(x)$ provides the state-dependent approximation to the propagation defect in Eq. 11.

Reference-preserving calibration. The pretrained propagator defines the reference in Eq. 11. We initialize the output factors of both correction paths as $U_{R,k}^{(b)} = 0$ and $U_c^{(b)} = 0$ for every adapted block and mode. Zero-initialized residual paths have been used to preserve an existing mapping at initialization (Bachlechner et al., 2021; Zhang et al., 2023). Under this initialization,

$$\widetilde{M}_{\theta_0,\psi_{\mathrm{init}}}(x) = M_{\theta_0}(x), \qquad A_{\psi_{\mathrm{init}}}(x) = 0, \quad \forall x. \tag{15}$$

The frozen θ_0 preserves the original propagator as the reference throughout calibration.

For supervised transitions (x_n, y_n), calibration minimizes the standard prediction objective. Substituting $\widetilde{M}_{\theta_0,\psi} = M_{\theta_0} + A_\psi$ and the observed defect $A^\star(x_n) = y_n - M_{\theta_0}(x_n)$ gives

$$\min_\psi \frac{1}{N} \sum_{n=1}^{N} \left\| \widetilde{M}_{\theta_0,\psi}(x_n) - y_n \right\|_\omega^2 = \min_\psi \frac{1}{N} \sum_{n=1}^{N} \left\| A_\psi(x_n) - A^\star(x_n) \right\|_\omega^2. \tag{16}$$

Prediction-space calibration is therefore equivalent to regression of the observable one-step defect. The correction is evaluated within the same operator graph at inference. Detailed spherical mode sums, block ordering, and the finite-horizon error recursion are provided in Appendix A.6 and Appendix A.9.

5 EXPERIMENTS

This section evaluates the proposed state-dependent propagation defect correction on spherical shallow-water dynamics and global weather forecasting. We report autoregressive accuracy, a state-matching diagnostic, and transfer to WeatherBench.

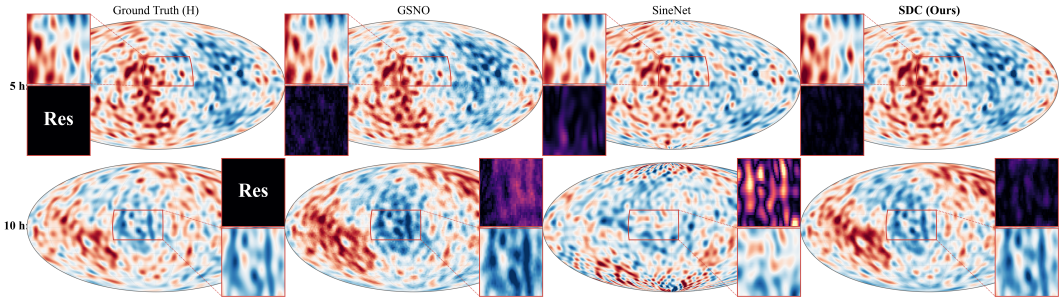

Figure 3: Representative spherical SWE height rollout with residual insets.

5.1 Experimental protocol

We study two autoregressive spherical benchmarks. SWE uses a 128×256 grid and one-hour state updates; WeatherBench follows the ERA5 protocol and latitude-weighted metrics of (Rasp et al., 2020; Hersbach et al., 2020). Implementation and compute details are in Appendices A.6 and A.10.

5.2 Shallow-Water Equations Forecasting

Table 1 compares SDC with GSNO and four operator baselines. SDC uses GSNO as its frozen backbone and trains only the correction in Section 4.3. It gives the lowest MRE in every channel and average at both rollout horizons. The 60-epoch compute-matched comparison is reported in Appendix A.10.

Table 1: Autoregressive SWE MRE at 128×256 under the shared one-hour recursive rollout (lower is better). H, V, and D are the three state channels, and Avg. is their unweighted mean. Every row uses the same held-out trajectories, normalization, sphere weights, and evaluation code. SDC is calibrated from the listed frozen GSNO (50 epochs); bold marks the best value in each column.

	5 h MRE				10 h MRE			
Model	H	V	D	Avg.	H	V	D	Avg.
FNO (2021)	0.006267	0.554734	1.077927	0.546309	0.007020	0.626663	1.110530	0.581404
SFNO (2023)	0.002179	0.268021	0.226895	0.165698	0.002154	0.273537	0.233357	0.169683
SineNet (2024)	0.002524	0.207694	0.849688	0.353302	0.002739	0.225980	0.948915	0.392545
Transolver (2024)	0.004109	0.467028	0.999149	0.490095	0.003975	0.502938	1.000663	0.502525
GSNO (2026)	0.001308	0.210123	0.236194	0.149209	0.001408	0.217549	0.247877	0.155611
SDC (*ours*)	**0.001156**	**0.151654**	**0.170886**	**0.107899**	**0.001222**	**0.157138**	**0.181610**	**0.113323**

Relative to the same 50-epoch GSNO reference, SDC reduces the 5 h and 10 h average MRE by 27.7% and 27.2%, with gains across H, V, and D. Their similar size shows that the correction remains useful under recursive feedback. Figure 3 compares the resulting height fields and local residuals; Appendix A.10 gives the separate 60-epoch compute comparison.

5.3 State-matched correction analysis

We evaluate 500 one-step pairs from the fixed SWE test bank. Following the predictor–defect–correction viewpoint (Dutt et al., 2000), Table 2 tests whether each correction must remain paired with the state that produced it.

Table 2: One-step MRE on 500 fixed-test pairs (lower is better). Fixed channel and fixed spatial apply dataset-level residual templates; Shuffled SDC preserves learned corrections but permutes their state assignments. Matched SDC alone retains the original input–correction pairs.

	Fixed GSNO	Fixed channel	Fixed spatial	Matched SDC	Shuffled SDC
One-step MRE	0.060355	0.060119	0.056219	**0.039420**	0.067596

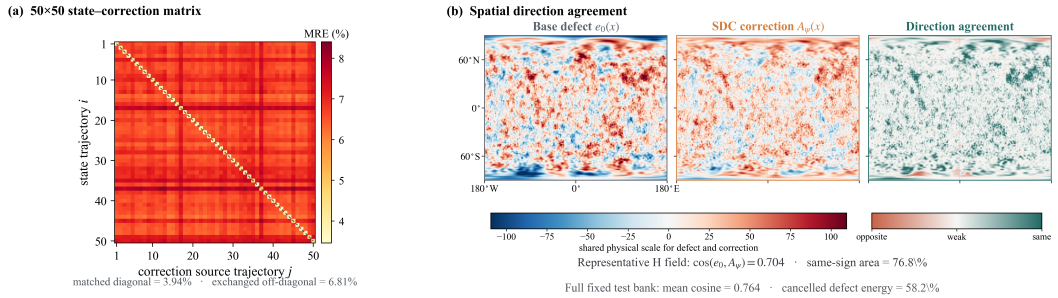

Figure 4: State matching on 500 fixed SWE pairs. Left: correction-source MRE; right: the reference defect, learned correction, and directional agreement.

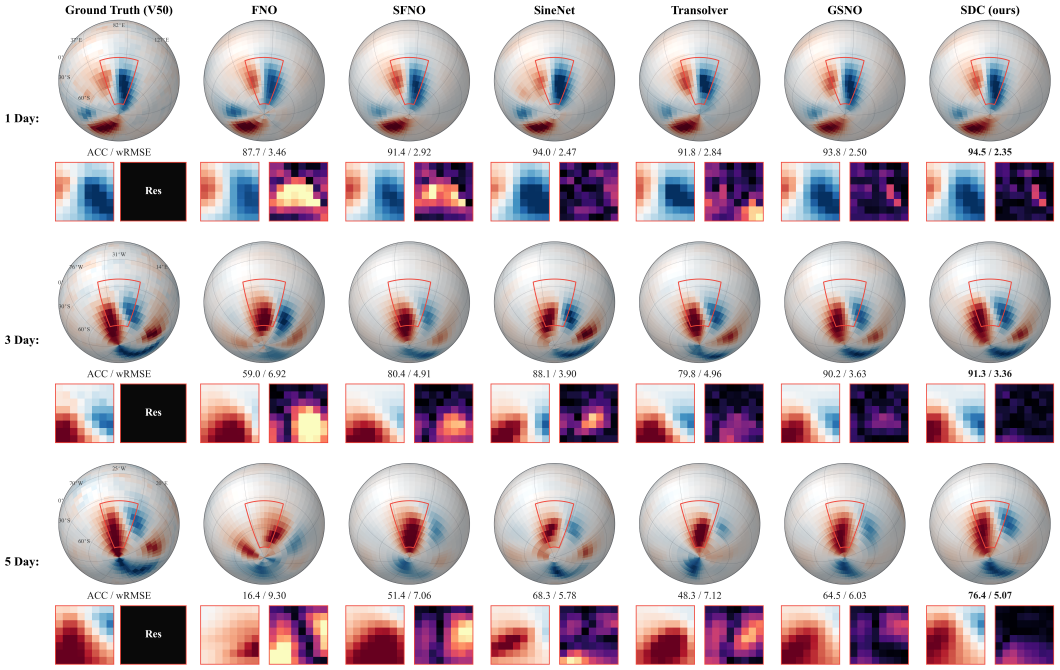

Figure 5: WeatherBench V_{50} predictions for held-out case 243 at 24, 72, and 120 h. Rows follow one initialized trajectory through closed-loop 6 h updates. Values beneath forecasts report ACC/wRMSE; paired insets show the boxed field and its residual.

SDC lowers one-step MRE from 0.060355 to 0.039420 (34.7%), compared with 0.060119 and 0.056219 for the fixed-channel and fixed-spatial controls. Shuffling preserves the learned corrections but reassigns them across states; MRE rises to 0.067596, above the frozen reference. This reversal isolates state pairing rather than correction magnitude. Since the magnitudes are retained, correction energy alone cannot explain the matched gain. Figure 4 shows the same contrast in correction-source and field views. Cosine alignment, defect-energy summaries, and control fitting are given in Appendix A.11. Together, the rollout and one-step studies connect the long-horizon gain to the state-paired defect target rather than to a fixed residual offset.

5.4 Real-World Atmospheric Forecasting: WeatherBench

We evaluate transfer to real atmospheric fields on WeatherBench v1/ERA5 at $5.625°$ (32×64) (Rasp et al., 2020; Hersbach et al., 2020). Models use 1979–2015 for training, 2016 for validation, and 2017–2018 for testing. From 256 fixed starts, each model advances in 6 h steps to 24, 72, and 120 h. Table 3 reports six representative ACC families and latitude-weighted relative L^2 over all 24 dynamical channels.

Table 3: WeatherBench rollout results (mean ± sample SD over three seeds). ACC is higher-is-better and Rel. L^2 lower-is-better; U-Avg, V-Avg, and T-Avg average pressure levels. GSNO and SDC per-family ACC values are reconstructed from rounded per-seed summaries. Each horizon uses the same 256 fixed starts and is generated recursively by one 6 h model, not a separately trained lead-time head. SDC has 4.15M total parameters; approximately 0.12M are calibrated while its 4.03M GSNO backbone stays fixed. Bold values mark the best value in each metric column within a horizon; the final column reports the latitude-weighted full-state error. Here 2T is near-surface temperature and 10U/10V are near-surface wind components. Average pools the six displayed ACC families, whereas Rel. L^2 covers all 24 dynamical channels.

Method (year)	Params	2T	10U	10V	U-Avg	V-Avg	T-Avg	Average	Rel. L^2
24 h forecast									
FNO (2021)	4.74M	86.62 ± 1.63	85.72 ± 0.55	85.50 ± 0.04	89.38 ± 0.29	86.99 ± 0.10	91.25 ± 0.05	87.58 ± 0.12	0.2688 ± 0.0010
SFNO (2023)	3.69M	94.66 ± 0.14	83.94 ± 0.36	83.74 ± 0.21	88.31 ± 0.18	85.69 ± 0.12	94.15 ± 0.01	88.42 ± 0.13	0.2786 ± 0.0016
SineNet (2024)	4.34M	79.12 ± 0.46	$\mathbf{90.60 \pm 0.18}$	$\mathbf{90.49 \pm 0.10}$	$\mathbf{92.75 \pm 0.05}$	$\mathbf{91.01 \pm 0.05}$	88.79 ± 0.52	88.80 ± 0.15	$\mathbf{0.2258 \pm 0.0016}$
Transolver (2024)	4.05M	$\mathbf{96.01 \pm 0.22}$	86.78 ± 0.32	86.13 ± 0.32	89.51 ± 0.20	86.35 ± 0.28	$\mathbf{95.14 \pm 0.07}$	89.99 ± 0.22	0.2698 ± 0.0030
GSNO (2026)	4.03M	95.42 ± 0.03	86.08 ± 0.05	85.55 ± 0.08	89.43 ± 0.10	86.98 ± 0.03	94.75 ± 0.04	89.70 ± 0.03	0.2659 ± 0.0004
SDC (*ours*)	4.15M	95.67 ± 0.02	87.31 ± 0.07	86.85 ± 0.09	90.34 ± 0.04	88.08 ± 0.06	95.14 ± 0.01	$\mathbf{90.57 \pm 0.04}$	0.2549 ± 0.0005
72 h forecast									
FNO (2021)	4.74M	74.97 ± 0.65	57.27 ± 1.65	57.21 ± 0.96	67.02 ± 0.81	61.31 ± 1.10	75.52 ± 0.85	65.55 ± 0.86	0.4706 ± 0.0021
SFNO (2023)	3.69M	88.54 ± 0.42	60.65 ± 1.09	59.22 ± 1.85	71.06 ± 1.03	64.08 ± 1.53	84.97 ± 0.11	71.42 ± 0.95	0.4245 ± 0.0086
SineNet (2024)	4.34M	73.42 ± 2.06	$\mathbf{70.35 \pm 0.13}$	$\mathbf{69.59 \pm 1.06}$	$\mathbf{77.21 \pm 0.40}$	$\mathbf{72.26 \pm 0.90}$	79.23 ± 1.16	73.68 ± 0.73	0.3997 ± 0.0078
Transolver (2024)	4.05M	89.13 ± 1.05	61.09 ± 0.90	59.86 ± 1.46	70.31 ± 0.46	62.97 ± 1.10	84.59 ± 0.16	71.33 ± 0.78	0.4398 ± 0.0045
GSNO (2026)	4.03M	90.58 ± 0.14	64.11 ± 0.62	62.57 ± 0.42	73.26 ± 0.62	66.93 ± 0.34	86.49 ± 0.23	73.99 ± 0.22	0.4075 ± 0.0022
SDC (*ours*)	4.15M	$\mathbf{91.56 \pm 0.13}$	67.12 ± 0.34	65.79 ± 0.29	75.77 ± 0.19	69.79 ± 0.12	$\mathbf{87.84 \pm 0.04}$	$\mathbf{76.31 \pm 0.13}$	$\mathbf{0.3956 \pm 0.0011}$
120 h forecast									
FNO (2021)	4.74M	61.39 ± 2.12	29.59 ± 2.04	28.62 ± 1.13	42.32 ± 1.05	30.97 ± 2.13	57.73 ± 0.83	41.77 ± 1.38	0.6553 ± 0.0032
SFNO (2023)	3.69M	83.10 ± 0.53	40.73 ± 1.68	36.13 ± 2.34	54.31 ± 1.74	40.86 ± 2.41	75.38 ± 0.37	55.09 ± 1.38	0.5303 ± 0.0114
SineNet (2024)	4.34M	67.77 ± 3.16	46.23 ± 0.26	$\mathbf{43.70 \pm 1.69}$	57.75 ± 0.77	46.60 ± 1.62	67.46 ± 1.96	54.92 ± 1.12	0.5514 ± 0.0148
Transolver (2024)	4.05M	82.05 ± 1.78	38.73 ± 1.15	35.25 ± 2.85	52.12 ± 0.73	37.95 ± 2.43	72.81 ± 0.87	53.15 ± 1.53	0.5593 ± 0.0030
GSNO (2026)	4.03M	85.71 ± 0.29	42.76 ± 1.36	39.10 ± 0.38	56.00 ± 1.22	43.16 ± 0.44	76.80 ± 0.44	57.26 ± 0.45	0.5182 ± 0.0029
SDC (*ours*)	4.15M	$\mathbf{87.44 \pm 0.09}$	$\mathbf{46.57 \pm 0.47}$	43.04 ± 0.47	$\mathbf{59.44 \pm 0.28}$	$\mathbf{47.45 \pm 0.36}$	$\mathbf{78.77 \pm 0.06}$	$\mathbf{60.45 \pm 0.23}$	$\mathbf{0.5119 \pm 0.0020}$

Relative to GSNO, SDC improves all six ACC families and lowers mean Rel. L^2 at every horizon. Its Average ACC gains widen from 0.87 points at 24 h to 2.32 at 72 h and 3.19 at 120 h. SineNet retains a slightly higher 10V ACC at 120 h, while SDC leads on the six-family Average and full-state Rel. L^2. The widening gap at later horizons tests repeated use of the same correction under closed-loop feedback. Figure 5 shows how the corresponding V_{50} residuals evolve in one held-out rollout.

Across all the evaluations. The SWE rollout measures accumulated forecast error, the paired diagnostic tests whether correction depends on the current state, and WeatherBench applies the same calibrated design to multichannel atmospheric fields. These results separate predictive accuracy, assignment sensitivity, and cross-system transfer rather than relying on a single aggregate metric. The construction also applies to a planar FNO for a two-dimensional Navier–Stokes equation to further verify its generalization; the supplementary cross-architecture results and rollout visualization are reported in Appendix A.12, Table 5, and Figure 6.

6 Conclusion

We introduced SDC to correct the state-dependent propagation defect of a frozen neural operator. SDC combines a zero-anchored low-rank spectral update with a state-conditioned bottleneck, so calibration changes only the learned defect pathway while retaining the pretrained propagator as reference. This formulation turns the mismatch between a shared nonlocal response and a state-varying local response into an observable one-step target. Prediction- space calibration is therefore supervised regression of that defect with the reference operator fixed. Extensive experiments on spherical PDEs and weather forecasting show that SDC consistently improves autoregressive forecasting while preserving the state pairing required by the defect formulation. The correction adapts a frozen operator without retraining the full propagator, and its complementary spectral and state-conditioned paths keep the added model compact. The same construction can be attached to spectral or kernel-based propagators whenever their one-step defects are available, providing a direct route to efficient adaptation under a fixed rollout and evaluation protocol.

AI USE STATEMENT

We used generative AI tools for language editing, LaTeX consistency checks, and organization of the manuscript text. We did not use them to generate data, run experiments, select reported results, or make scientific decisions. The authors reviewed all AI-assisted text and code suggestions, checked the resulting compilation and numerical claims, and take responsibility for the final manuscript.

ETHICS STATEMENT

This work uses numerical PDE simulations and publicly available ERA5-derived WeatherBench fields. It involves no human or animal subjects, personal data, or interventions. The experiments are computational and do not introduce a new data-collection or deployment risk.

REPRODUCIBILITY STATEMENT

The main paper specifies the training and evaluation protocols. Appendix A provides the mathematical assumptions and derivations, while Appendix A.10 reports parameter and compute accounting and Appendix A.11 defines the state-matching diagnostics.

We includes the code of the proposed model and its implementation in **Supplementary Material**.

REFERENCES

Thomas Bachlechner, Bodhisattwa Prasad Majumder, Henry Mao, Gary Cottrell, and Julian McAuley. Rezero is all you need: Fast convergence at large depth. In *Uncertainty in artificial intelligence*, pp. 1352–1361. PMLR, 2021.

Boris Bonev, Thorsten Kurth, Christian Hundt, Jaideep Pathak, Maximilian Baust, Karthik Kashinath, and Anima Anandkumar. Spherical fourier neural operators: Learning stable dynamics on the sphere. In *International conference on machine learning*, pp. 2806–2823. PMLR, 2023.

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.

Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual review of fluid mechanics*, 52(1):477–508, 2020.

Shuhao Cao. Choose a transformer: Fourier or galerkin. *Advances in neural information processing systems*, 34:24924–24940, 2021.

Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical CNNs. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Hkbd5xZRb.

Peter D Dueben and Peter Bauer. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10):3999–4009, 2018.

Alok Dutt, Leslie Greengard, and Vladimir Rokhlin. Spectral deferred correction methods for ordinary differential equations. *BIT Numerical Mathematics*, 40(2):241–266, 2000.

Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *Proceedings of the european conference on computer vision (ECCV)*, pp. 52–68, 2018.

Gaurav Gupta, Xiongye Xiao, and Paul Bogdan. Multiwavelet-based operator learning for differential equations. *Advances in neural information processing systems*, 34:24048–24062, 2021.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *International conference on machine learning*, pp. 12556–12569. PMLR, 2023.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049, 2020.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pp. 2790–2799. PMLR, 2019.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Rabeeh Karimi Mahabadi, James Henderson, and Sebastian Ruder. Compacter: Efficient low-rank hypercomplex adapter layers. *Advances in neural information processing systems*, 34:1022–1035, 2021.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klöwer, James Lottes, Stephan Rasp, Peter Düben, et al. Neural general circulation models for weather and climate. *Nature*, 632(8027):1060–1066, 2024.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Thorsten Kurth, Shashank Subramanian, Peter Harrington, Jaideep Pathak, Morteza Mardani, David Hall, Andrea Miele, Karthik Kashinath, and Anima Anandkumar. Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In *Proceedings of the platform for advanced scientific computing conference*, pp. 1–11, 2023.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023.

Martin Leutbecher and Tim N Palmer. Ensemble forecasting. *Journal of computational physics*, 227 (7):3515–3539, 2008.

Zongyi Li, Nikola Borislavov 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.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023.

Edward N Lorenz. Deterministic nonperiodic flow. In *Universality in Chaos, 2nd edition*, pp. 367–378. Routledge, 2017.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Md Ashiqur Rahman, Zachary E Ross, and Kamyar Azizzadenesheli. U-NO: U-shaped neural operators. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=j3oQF9coJd.

Bogdan Raonic, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel De Bézenac. Convolutional neural operators for robust and accurate learning of pdes. *Advances in Neural Information Processing Systems*, 36:77187–77200, 2023.

Stephan Rasp, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002203, 2020.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. *Advances in neural information processing systems*, 35:1596–1611, 2022.

Hao Tang, Hao Chen, and Chao Li. Generalized spherical neural operators: Green's function formulation. In *International Conference on Learning Representations*, volume 2026, pp. 11180–11201, 2026.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. *arXiv preprint arXiv:2402.02366*, 2024.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3813–3824. IEEE, 2023.

Xuan Zhang, Jacob Helwig, Yuchao Lin, Yaochen Xie, Cong Fu, Stephan Wojtowytsch, and Shuiwang Ji. Sinenet: Learning temporal dynamics in time-dependent partial differential equations. In *International Conference on Learning Representations*, volume 2024, pp. 11612–11653, 2024.

A Detailed Mathematical Derivation

This appendix gives the complete derivation supporting the Methodology section. The order is important. We first establish what follows from a fixed linear operator, then relax the linearity assumption, connect the resulting response mismatch to a time-propagator defect, and finally prove the identities used by the spherical SDC implementation. The appendix distinguishes mathematical definitions from finite-dimensional neural parameterizations.

A.1 Spaces, maps, and notation

Let $\Omega = \mathbb{S}^2$ be the unit sphere, with surface-area element $d\omega$. The symbols ω, ζ, and η denote points on Ω. A state $x \in \mathcal{X}$ is a vector-valued field on Ω with d channels: $x = (x_1, \dots, x_d)$. The state space $\mathcal{X}$ is assumed to be a Banach space on which the maps considered below are well defined. Norms without a subscript denote an abstract norm on $\mathcal{X}$; $\|\cdot\|_\omega$ denotes the sphere-weighted norm used by the numerical objective.

For a map F between Banach spaces, the Fréchet derivative at z is the bounded linear map $DF[z]$ satisfying

$$F(z+h) = F(z) + DF[z](h) + o(\|h\|) \quad \text{as } \|h\| \to 0. \tag{17}$$

The derivative is linear in the perturbation h after the reference point z has been fixed. The derivative can nevertheless depend on z; this state-dependence is how a nonlinear map differs from one fixed linear map.

A.2 A representative fixed-kernel realization

This appendix section gives one concrete kernel realization of the shared nonlocal-response structure used in the main text. It is a mathematical instance of the operator-level assumptions in Section 4; the method and its defect target do not depend on choosing a Green representation. We begin with the linear equation

$$\mathcal{D}_{\mathrm{pde}} g = f, \tag{18}$$

where g is the unknown field, f is the source term, and $\mathcal{D}_{\mathrm{pde}}$ is a linear differential operator. In this section, fixed means that the coefficients of $\mathcal{D}_{\mathrm{pde}}$, the spatial domain, the boundary conditions, and the function spaces have all been specified and are not changed when f varies.

Assume that the problem is well posed on the subspace of interest. Its inverse operator is

$$\mathcal{G}_0 := \mathcal{D}_{\mathrm{pde}}^{-1}. \tag{19}$$

Applying $\mathcal{G}_0$ to Eq. 18 gives

$$g = \mathcal{G}_0 f. \tag{20}$$

The subscript 0 records that the inverse belongs to one fixed operator. If the inverse has an integral-kernel representation, then

$$(\mathcal{G}_0 f)(\zeta) = \int_\Omega G_0(\zeta, \eta) f(\eta)\, d\eta. \tag{21}$$

Here ζ is the output location, η is the source location, and $G_0(\zeta, \eta)$ is the Green kernel. Fixed means that the same kernel is used for every admissible source for the chosen operator and boundary conditions; it does not mean that the output field is constant in ζ.

The defining relation for the kernel is

$$\mathcal{D}_{\mathrm{pde},\zeta} G_0(\zeta, \eta) = \delta_\Omega(\zeta, \eta), \tag{22}$$

where δ_Ω is the Dirac distribution on Ω and the subscript ζ means that $\mathcal{D}_{\mathrm{pde}}$ acts on the first kernel argument. To verify the representation, let $\bar{g}$ be the right-hand side of Eq. 21. Under the regularity

conditions that allow the derivative to pass through the integral,

$$\begin{aligned}\mathcal{D}_{\mathrm{pde}}\bar{g}(\zeta) &= \mathcal{D}_{\mathrm{pde},\zeta}\int_\Omega G_0(\zeta,\eta)f(\eta)\,d\eta \\ &= \int_\Omega \mathcal{D}_{\mathrm{pde},\zeta}G_0(\zeta,\eta)f(\eta)\,d\eta \\ &= \int_\Omega \delta_\Omega(\zeta,\eta)f(\eta)\,d\eta \\ &= f(\zeta). \end{aligned} \tag{23}$$

The last equality is the sampling property of the Dirac distribution. Thus a fixed linear operator produces one fixed inverse response.

For the sphere-specific form used in the main text, let $n \in \mathbb{S}^2$ be the north pole and represent a source location as $\eta = Rn$ for a rotation $R \in SO(3)$. If the fixed kernel is rotationally equivariant, it depends on the relative position only:

$$G_0(u, Rn) = \overline{G}_0(R^{-1}u). \tag{24}$$

With $d\mu(R)$ denoting the rotation-invariant measure induced on the sphere, the Green representation can then be written as

$$g(u) = \int_{SO(3)} \overline{G}_0(R^{-1}u)f(Rn)\,d\mu(R) \tag{25}$$

$$= \big(f * \overline{G}_0\big)(u). \tag{26}$$

Under the adopted spherical-transform and measure normalization, the spherical convolution theorem gives

$$\widehat{g}_{\ell m} = c_\ell\,\widehat{\overline{G}}_{0,\ell 0}\,\widehat{f}_{\ell m}, \qquad c_\ell = 2\pi\sqrt{\frac{4\pi}{2\ell+1}}. \tag{27}$$

This is the exact point where a fixed Green response becomes a shared spectral response: for every input f, the same coefficient $c_\ell\widehat{\overline{G}}_{0,\ell 0}$ multiplies the degree-ℓ input coefficient. The spherical spectral operator learns a channel-valued analogue of these degree-wise coefficients, while retaining the same transform–contraction–inverse-transform ordering.

The same conclusion follows from superposition. Let a, b be scalars and f_1, f_2 be two source terms. Linearity of $\mathcal{D}_{\mathrm{pde}}$ gives

$$\begin{aligned}\mathcal{D}_{\mathrm{pde}}\left(a\mathcal{G}_0 f_1 + b\mathcal{G}_0 f_2\right) &= a\mathcal{D}_{\mathrm{pde}}\mathcal{G}_0 f_1 + b\mathcal{D}_{\mathrm{pde}}\mathcal{G}_0 f_2 \\ &= af_1 + bf_2. \end{aligned} \tag{28}$$

Uniqueness of the solution therefore implies

$$\mathcal{G}_0(af_1 + bf_2) = a\mathcal{G}_0 f_1 + b\mathcal{G}_0 f_2. \tag{29}$$

The first theoretical conclusion is therefore

$$\mathcal{D}_{\mathrm{pde}} \text{ fixed and linear} \quad\Longrightarrow\quad \mathcal{G}_0 \text{ and } G_0 \text{ fixed.} \tag{30}$$

A.3 Relaxing linearity and obtaining local responses

We next replace the fixed linear equation by

$$\mathcal{N}(g) = f, \tag{31}$$

where $\mathcal{N}$ is nonlinear. Choose a reference state g^* and a small perturbation δg. Applying the definition of the Fréchet derivative gives

$$\mathcal{N}(g^* + \delta g) = \mathcal{N}(g^*) + D\mathcal{N}[g^*](\delta g) + o(\|\delta g\|). \tag{32}$$

For a fixed reference state g^*, $D\mathcal{N}[g^*]$ provides the linear response to the perturbation δg. The governing map remains nonlinear through the dependence of this derivative on the reference state: changing g^* changes the local linear response.

Suppose the source is perturbed from f to $f + \delta f$. Since $\mathcal{N}(g^*) = f$, Eq. 32 becomes

$$D\mathcal{N}[g^*](\delta g) = \delta f \quad \text{to first order.} \tag{33}$$

If the derivative is locally invertible, the corresponding perturbation response is

$$\delta g = \mathcal{G}_{g^*}(\delta f), \qquad \mathcal{G}_{g^*} \coloneqq (D\mathcal{N}[g^*])^{-1}. \tag{34}$$

The response carries the reference state in its subscript. At another reference state $\widetilde{g}^*$, the corresponding response is

$$\mathcal{G}_{\widetilde{g}^*} = (D\mathcal{N}[\widetilde{g}^*])^{-1}, \tag{35}$$

and, in general,

$$\mathcal{G}_{g^*} \neq \mathcal{G}_{\widetilde{g}^*}. \tag{36}$$

If $\mathcal{N}$ is twice Fréchet differentiable and its derivative remains boundedly invertible along the perturbation, define the directional variation of its derivative by

$$\mathcal{H}_{g^*,h}[\delta g] \coloneqq D^2\mathcal{N}[g^*][h, \delta g]. \tag{37}$$

Differentiating in operator norm $D\mathcal{N}[g] \circ \mathcal{G}_g = I$ along $g = g^* + \varepsilon h$ gives

$$\mathcal{G}_{g^*+\varepsilon h} = \mathcal{G}_{g^*} - \varepsilon\, \mathcal{G}_{g^*} \circ \mathcal{H}_{g^*,h} \circ \mathcal{G}_{g^*} + o(\varepsilon). \tag{38}$$

This identity makes the state dependence quantitative: the local inverse changes through the curvature of the governing operator, while each fixed inverse remains linear in the source perturbation. We call $\mathcal{G}_{g^*}$ a local inverse response. It describes the instantaneous response to a source perturbation at the selected reference state. A nonlinear governing operator can therefore require a family of such responses indexed by the current state, whereas the fixed linear construction supplies one shared response.

A.4 Time propagation, tangent response, and the reference operator

The preceding argument concerns a spatial response. The learning task uses a time-step map. Let $\mathcal{X}$ be the space of complete current spherical states and let $\mathcal{F}_{\mathrm{dyn}} : \mathcal{X} \to \mathcal{X}$ denote the state-dependent part of the spatially discretized evolutionary PDE:

$$\frac{dx(t)}{dt} = \mathcal{F}_{\mathrm{dyn}}(x(t)). \tag{39}$$

Assume that $\mathcal{F}_{\mathrm{dyn}}$ is differentiable along the trajectory. For a perturbation δx_n of the state at t_n, linearization along the trajectory gives

$$\frac{d}{dt}\delta x(t) = D\mathcal{F}_{\mathrm{dyn}}[x(t)]\, \delta x(t), \qquad \delta x(t_n) = \delta x_n. \tag{40}$$

Let $\mathcal{T}(t, t_n; x_n)$ be the transition operator of this variational equation:

$$\frac{d}{dt}\mathcal{T}(t, t_n; x_n) = D\mathcal{F}_{\mathrm{dyn}}[x(t)]\mathcal{T}(t, t_n; x_n), \qquad \mathcal{T}(t_n, t_n; x_n) = I. \tag{41}$$

If $\phi_{\Delta t}$ denotes the continuous flow, its local sensitivity is

$$D\phi_{\Delta t}[x_n] = \mathcal{T}(t_n + \Delta t, t_n; x_n). \tag{42}$$

Equivalently, using the chronological ordering operator $\mathcal{P}$,

$$D\phi_{\Delta t}[x_n] = \mathcal{P} \exp\left(\int_{t_n}^{t_n+\Delta t} D\mathcal{F}_{\mathrm{dyn}}[x(s)]\, ds \right). \tag{43}$$

The ordering is required because Jacobians evaluated at different times need not commute. Thus the tangent response depends on the entire reference path, not only on its endpoint. Thus $\mathcal{G}_{g^*}$ and $\mathcal{T}$ are related but distinct: the former maps a source perturbation through an instantaneous linearized equation, while the latter accumulates the local linearized response along a time trajectory.

Let Δt denote the interval represented by one neural prediction. The numerical solver used to generate the data supplies a reference discrete flow

$$x_{n+1} = \Phi_{\Delta t}(x_n), \qquad x_n \in \mathcal{X}. \tag{44}$$

This map approximates the continuous flow $\phi_{\Delta t}$, and the supervised target y_n is its numerical output. When the numerical update is differentiable, its derivative $D\Phi_{\Delta t}[x_n]$ is the discrete counterpart of the tangent map in Eq. 42; it contains the response accumulated by the solver over one time step.

The trained single-state operator is the reference operator:

$$M_{\theta_0} : \mathcal{X} \to \mathcal{X}. \tag{45}$$

The phrase "reference operator" means one fixed trained parameter instance, including its encoder, spherical blocks, decoder, and outer residual. Its standard time-stepping form is

$$M_{\theta_0}(x) = x + R_{\theta_0}(x), \tag{46}$$

where the identity term preserves the current state and $R_{\theta_0}(x)$ is the learned one-step increment.

The reference operator contains nonlinear normalization and MLP operations. Its connection to the preceding Green argument is at the level of the global spherical propagation path: the trained spectral parameters are shared across input states, while the complete one-step map has a state-dependent local derivative.

A.5 Reference propagation defect and its state dependence

Both $\Phi_{\Delta t}(x)$ and $M_{\theta_0}(x)$ belong to $\mathcal{X}$. Their difference is therefore a well-defined state-valued function:

$$A^\star(x) \coloneqq \Phi_{\Delta t}(x) - M_{\theta_0}(x). \tag{47}$$

Adding $M_{\theta_0}(x)$ to both sides gives

$$\Phi_{\Delta t}(x) = M_{\theta_0}(x) + A^\star(x). \tag{48}$$

This is an algebraic identity relative to the chosen reference map. It is unique for that choice: if another function B satisfies $\Phi_{\Delta t}(x) = M_{\theta_0}(x) + B(x)$ for every x, subtracting $M_{\theta_0}(x)$ gives

$$B(x) = \Phi_{\Delta t}(x) - M_{\theta_0}(x) = A^\star(x). \tag{49}$$

To see why this defect can depend on the current state, choose x^* and a small perturbation δx. Assuming both maps are Fréchet differentiable at x^*, write

$$\Phi_{\Delta t}(x^* + \delta x) = \Phi_{\Delta t}(x^*) + D\Phi_{\Delta t}[x^*](\delta x) + o(\|\delta x\|), \tag{50}$$

$$M_{\theta_0}(x^* + \delta x) = M_{\theta_0}(x^*) + DM_{\theta_0}[x^*](\delta x) + o(\|\delta x\|). \tag{51}$$

Subtract the second equation from the first. The constant terms combine into $A^\star(x^*)$ by Eq. 47, giving

$$A^\star(x^* + \delta x) = A^\star(x^*) + \left(D\Phi_{\Delta t}[x^*] - DM_{\theta_0}[x^*]\right)(\delta x) + o(\|\delta x\|). \tag{52}$$

Equivalently,

$$DA^\star[x^*](\delta x) = \left(D\Phi_{\Delta t}[x^*] - DM_{\theta_0}[x^*]\right)(\delta x). \tag{53}$$

The right-hand side is the local response mismatch between the reference flow and the reference operator. It may collect several effects, including shared spectral response, spectral truncation, finite capacity, time discretization, and optimization error. The fixed linear-operator construction supplies the motivation for one component of this mismatch, while the observable defect definition accommodates the complete discrepancy.

The residual representation gives the same defect in increment coordinates. Define the exact reference increment

$$\Delta^\star(x) \coloneqq \Phi_{\Delta t}(x) - x. \tag{54}$$

Substituting Eq. 46 into Eq. 47 gives, step by step,

$$A^\star(x) = \Phi_{\Delta t}(x) - (x + R_{\theta_0}(x)) \tag{55}$$

$$= (\Phi_{\Delta t}(x) - x) - R_{\theta_0}(x) \tag{56}$$

$$= \Delta^\star(x) - R_{\theta_0}(x). \tag{57}$$

Thus the reference operator learns the main increment, while the propagation defect is the remaining difference relative to that increment. This explains the role of the residual connection without making residual learning a separate contribution.

A.6 SPHERICAL SPECTRAL PARAMETERIZATION OF THE CORRECTION

We now describe how a finite neural function approximates $A^\star$. Let $z^{(b)}$ be the hidden spherical field entering operator block b. For each input channel i, degree ℓ, and order m, the spherical-harmonic coefficient is

$$\widehat{z}^{(b)}_{i\ell m} = \int_\Omega z^{(b)}_i(\omega) Y^{m*}_\ell(\omega)\, d\omega, \tag{58}$$

and the channel-wise spherical mean is

$$q^{(b)}_i = \frac{1}{4\pi}\int_\Omega z^{(b)}_i(\omega)\, d\omega. \tag{59}$$

The learned global integral injection is

$$\widetilde{z}^{(b)}_{i\ell m} = \widehat{z}^{(b)}_{i\ell m} + q^{(b)}_i P^{(b)}_{0,i\ell m}. \tag{60}$$

The Driscoll–Healy degree-wise channel contraction is

$$s^{(b)}_{o\ell m} = \sum_i W^{(b)}_{0,oi\ell}\widetilde{z}^{(b)}_{i\ell m}, \tag{61}$$

followed by inverse spherical-harmonic reconstruction:

$$\mathcal{S}^{(b)}_{W_0,P_0}[z^{(b)}]_o(\omega) = \sum_{\ell=0}^{L-1}\sum_{m=-\ell}^{\ell} s^{(b)}_{o\ell m} Y^m_\ell(\omega). \tag{62}$$

The original operator block also contains its normalization, MLP and skip paths. Equations 58–62 isolate the global spherical propagation component.

Low-rank spectral update. For each block and degree, SDC introduces a low-rank complex update:

$$\widetilde{W}^{(b)}_{0,oi\ell} = W^{(b)}_{0,oi\ell} + \Delta W^{(b)}_{oi\ell}, \tag{63}$$

$$\Delta W^{(b)}_{oi\ell} = \frac{1}{r}\sum_{\rho=1}^{r} U^{(b)}_{o\rho\ell} V^{(b)}_{\rho i\ell}. \tag{64}$$

In matrix form, for fixed b and ℓ, $\Delta W^{(b)}_\ell = r^{-1} U^{(b)}_\ell V^{(b)}_\ell$, with $U^{(b)}_\ell \in \mathbb{C}^{d_{\text{out}}\times r}$ and $V^{(b)}_\ell \in \mathbb{C}^{r\times d_{\text{in}}}$. Hence the update has rank at most r. The spherical transform, global integral injection, contraction and inverse transform are not replaced; only the channel response uses $\widetilde{W}_0$.

Bias-free hidden bottleneck. The second route operates on the hidden field after the frozen block computation with its adapted spectral response:

$$H^{(b)}_\psi(z)(\omega) = U^{(b)}_c\,\mathrm{GELU}\Big(D^{(b)}_c z(\omega)\Big), \qquad D^{(b)}_c \in \mathbb{R}^{r_c\times d}, \quad U^{(b)}_c \in \mathbb{R}^{d\times r_c}. \tag{65}$$

The matrix $D^{(b)}_c$ compresses the d channels to r_c directions, GELU introduces a nonlinear transformation, and $U^{(b)}_c$ maps the result back to d channels. There is no additive bias. The operation is pointwise in implementation, but its input has already been globally mixed by Eq. 62; the response is conditioned on the current full-field state.

Fused block and function-level correction. Let $r^{(b)}$ denote the feature used by the original skip path. The adapted block preserves the original ordering:

$$a^{(b)} = \mathrm{Norm}^{(b)}_0\left(\mathcal{S}^{(b)}_{\widetilde{W},P_0}[z^{(b)}]\right) + R^{(b)}_{\text{in}}(r^{(b)}), \tag{66}$$

$$\bar{z}^{(b)} = \mathrm{DropPath}^{(b)}\left(\mathrm{Norm}^{(b)}_1\left(\mathrm{MLP}^{(b)}(a^{(b)})\right)\right) + R^{(b)}_{\text{out}}(r^{(b)}), \tag{67}$$

$$\check{z}^{(b)} = \rho^{(b)}\left(\bar{z}^{(b)}\right), \tag{68}$$

$$z^{(b+1)} = \check{z}^{(b)} + H^{(b)}_\psi(\check{z}^{(b)}). \tag{69}$$

Here $\rho^{(b)}$ is the activation already present in the frozen reference block and is the identity for its configured last layer. Thus the bottleneck receives exactly the block output used by the implementation. The encoder, decoder and original block parameters remain those of the reference operator. Running the adapted blocks defines $\widetilde{M}_{\theta_0,\psi}$. The engineering correction is the exact function difference

$$A_\psi(x) \coloneqq \widetilde{M}_{\theta_0,\psi}(x) - M_{\theta_0}(x), \tag{70}$$

so that

$$\widehat{x}_{n+1} = \widetilde{M}_{\theta_0,\psi}(x_n) = M_{\theta_0}(x_n) + A_\psi(x_n). \tag{71}$$

The ideal $A^\star$ is defined by the reference flow, whereas A_ψ is its restricted neural approximation.

A.7 Zero anchoring and calibration

The SDC correction is initialized at the reference operator. Specifically, for every block and degree, set the output spectral factor and upper bottleneck projection to zero:

$$U_\ell^{(b)} = 0, \qquad U_c^{(b)} = 0. \tag{72}$$

From the low-rank construction,

$$U_\ell^{(b)} = 0 \Longrightarrow \Delta W_\ell^{(b)} = 0 \Longrightarrow \widetilde{W}_{0,\ell}^{(b)} = W_{0,\ell}^{(b)}. \tag{73}$$

From the bottleneck construction,

$$U_c^{(b)} = 0 \Longrightarrow H_\psi^{(b)}(z) = 0 \quad \text{for every } z. \tag{74}$$

Therefore, if the input to an adapted block equals the input to the corresponding reference-operator block, its spectral path is identical and its added bottleneck is zero. The complete-network statement follows by induction. At the encoder output, $z_{\text{adapted}}^{(0)} = z_{\text{base}}^{(0)}$. If $z_{\text{adapted}}^{(b)} = z_{\text{base}}^{(b)}$, then Eqs. 73 and 74 imply $z_{\text{adapted}}^{(b+1)} = z_{\text{base}}^{(b+1)}$. The decoder and outer residual are shared, so

$$\psi = \psi_{\text{init}} \Longrightarrow \widetilde{M}_{\theta_0,\psi}(x) = M_{\theta_0}(x) \quad \text{and} \quad A_\psi(x) = 0 \quad \forall x. \tag{75}$$

Here ψ_{init} denotes the initialization in Eq. 72; the remaining factors need not be zero. This is a structural identity rather than a loss penalty.

During calibration, θ_0 is held fixed and only ψ is optimized. Thus the reference operator remains the same function throughout training, and the learned change remains identifiable relative to that function.

A.8 Loss equivalence: prediction training as defect regression

For each training pair (x_n, y_n), define

$$\widehat{y}_n = \widetilde{M}_{\theta_0,\psi}(x_n) = M_{\theta_0}(x_n) + A_\psi(x_n). \tag{76}$$

The sphere-weighted squared prediction loss is

$$\mathcal{L}(\psi) = \frac{1}{N}\sum_{n=1}^{N} \|\widehat{y}_n - y_n\|_\omega^2 . \tag{77}$$

Substitute Eq. 76 into the error:

$$\widehat{y}_n - y_n = M_{\theta_0}(x_n) + A_\psi(x_n) - y_n \tag{78}$$

$$= A_\psi(x_n) - [y_n - M_{\theta_0}(x_n)] . \tag{79}$$

Therefore the same loss is

$$\mathcal{L}(\psi) = \frac{1}{N}\sum_{n=1}^{N} \|A_\psi(x_n) - [y_n - M_{\theta_0}(x_n)]\|_\omega^2 . \tag{80}$$

The quantity in square brackets is fixed once the reference operator and the training data are fixed. Thus the calibration can be interpreted as defect regression, but no additional defect label and no additional loss term have been introduced.

A.9 FINITE-HORIZON ROLLOUT RECURSION

Let the reference and adapted trajectories start from the same state:

$$x_{k+1} = \Phi_{\Delta t}(x_k), \qquad \widehat{x}_{k+1} = \widetilde{M}_{\theta_0,\psi}(\widehat{x}_k), \qquad x_0 = \widehat{x}_0. \tag{81}$$

Assume that the reference flow is L-Lipschitz on a domain containing the states visited by both trajectories:

$$\|\Phi_{\Delta t}(a) - \Phi_{\Delta t}(b)\| \le L\|a - b\|. \tag{82}$$

Assume also that the adapted one-step error on this same domain is uniformly bounded:

$$\left\|\widetilde{M}_{\theta_0,\psi}(z) - \Phi_{\Delta t}(z)\right\| \le \varepsilon_\psi. \tag{83}$$

Set $e_k = \|x_k - \widehat{x}_k\|$. Add and subtract $\Phi_{\Delta t}(\widehat{x}_k)$:

$$e_{k+1} = \left\|\Phi_{\Delta t}(x_k) - \widetilde{M}_{\theta_0,\psi}(\widehat{x}_k)\right\| \tag{84}$$

$$\le \|\Phi_{\Delta t}(x_k) - \Phi_{\Delta t}(\widehat{x}_k)\| + \left\|\Phi_{\Delta t}(\widehat{x}_k) - \widetilde{M}_{\theta_0,\psi}(\widehat{x}_k)\right\| \tag{85}$$

$$\le Le_k + \varepsilon_\psi. \tag{86}$$

Since $e_0 = 0$,

$$e_1 \le \varepsilon_\psi, \tag{87}$$

$$e_2 \le L\varepsilon_\psi + \varepsilon_\psi, \tag{88}$$

$$e_3 \le L^2\varepsilon_\psi + L\varepsilon_\psi + \varepsilon_\psi. \tag{89}$$

By induction,

$$e_k \le \varepsilon_\psi \sum_{i=0}^{k-1} L^i. \tag{90}$$

For $L \neq 1$, this is

$$e_k \le \varepsilon_\psi \frac{L^k - 1}{L - 1}, \tag{91}$$

and for $L = 1$ it becomes $e_k \le k\varepsilon_\psi$. The bound holds on a shared state domain where the reference flow is Lipschitz and the adapted one-step error is uniformly bounded. Under these conditions, reducing the one-step propagation defect reduces the corresponding finite-horizon error upper bound.

A.10 SWE COMPUTE AND PARAMETER ACCOUNTING

Table 4 reports the total wall time for the complete from-scratch protocol used by each model. Both runs use the same SWE resolution and evaluation protocol as the main experiment. The SDC protocol generates the 50-epoch cache, trains GSNO for 50 epochs, and then calibrates the adapter for 50 epochs.

Table 4: SWE accuracy and compute accounting. Total end-to-end time includes fixed-cache trajectory generation and the corresponding training schedule: 60 GSNO epochs, or 50 GSNO epochs followed by 50 adapter epochs for SDC. Rollout evaluation is excluded. Lower MRE is better.

						Total wall time (s)
Model	Optimization schedule	Trainable params	Total params	5 h Avg. MRE	10 h Avg. MRE	End-to-end time
GSNO (60 epochs)	60 backbone epochs	797,537	797,537	0.124450	0.132355	843.587
SDC	GSNO-50 + 50 adapter epochs	58,176	855,713	**0.107899**	**0.113323**	**826.391**

Relative to GSNO trained for 60 epochs, SDC reduces the 5 h and 10 h average MRE by 13.3% and 14.4%. The complete SDC protocol uses 58,176 trainable parameters and takes 826.391 s, compared with 797,537 trainable parameters and 843.587 s for GSNO. The two totals differ by about 2.0% under the same from-scratch accounting.

A.11 ADDITIONAL STATE-DEFECT DIAGNOSTICS

The mechanism analysis reported with the experiments uses 500 one-step pairs drawn from the fixed SWE test trajectory bank. The reference-operator checkpoint and the SDC checkpoint are evaluated without changing their parameters. The state-matched correction is compared with two state-independent controls: a constant channel residual and a constant spatial residual. Each fixed control is fitted on the opposite trajectory fold and evaluated frozen on held-out states. This comparison isolates what state-independent residuals capture without access to the evaluation states.

The matched and shuffled one-step MRE values are

Predictor	MRE	Description
Fixed reference operator	0.060355	fixed operator
Matched SDC	0.039420	state paired with its correction
Fixed channel residual	0.060119	state independent
Fixed spatial residual	0.056219	state independent
Shuffled SDC	0.067596	correction reassigned by state

For each pair, let $e_n = y_n - M_{\theta_0}(x_n)$ be the observed defect of the reference operator and let $a_n = A_\psi(x_n)$ be the learned correction. The mean defect energy is 130.2944, the mean correction energy is 102.9058, and the mean residual defect energy after correction is 84.3296. These energies are squared L^2 norms of error fields, not physical energy conservation quantities. For each pair we compute cosine alignment and the net fraction of defect energy removed by the correction as

$$\rho_n = \frac{\langle e_n, a_n \rangle}{\|e_n\|_2 \, \|a_n\|_2}, \qquad \kappa_n = 1 - \frac{\|e_n - a_n\|_2^2}{\|e_n\|_2^2} = \frac{2\langle e_n, a_n \rangle - \|a_n\|_2^2}{\|e_n\|_2^2}. \tag{92}$$

Thus positive ρ_n means that the correction points toward the observed defect, while κ_n measures the resulting reduction in squared defect norm. The reported mean cosine alignment is 0.7636 and the mean per-pair cancellation fraction is 0.5821; the latter is averaged before aggregation, so it is not the ratio of the three mean energies above. For the displayed H-channel field, we also report the same-sign area

$$s_H = \frac{\mathrm{area}\{p \in \Omega : e_H(p) a_H(p) > 0\}}{\mathrm{area}(\Omega)}, \tag{93}$$

which counts locations where the correction and the defect have the same sign, irrespective of magnitude. The representative panel has cosine 0.704 and $s_H = 76.8\%$, alongside the full-bank means above.

The two fixed controls isolate what can be explained without state dependence. For a training fold $\mathcal{S}_{\mathrm{tr}}$, write $a_n(c, p)$ for the SDC correction at channel c and spherical location p. The fixed channel control first averages over both states and locations,

$$\bar{a}_{\mathrm{ch}}(c) = \frac{1}{|\mathcal{S}_{\mathrm{tr}}||\Omega|} \sum_{n \in \mathcal{S}_{\mathrm{tr}}} \sum_{p \in \Omega} a_n(c, p), \qquad \tilde{a}_{\mathrm{ch}}(c, p) = \bar{a}_{\mathrm{ch}}(c),$$

and broadcasts this three-component vector to every state and location. It captures a global channel-wise bias. The fixed spatial control instead averages over states,

$$\bar{a}_{\mathrm{sp}}(c, p) = \frac{1}{|\mathcal{S}_{\mathrm{tr}}|} \sum_{n \in \mathcal{S}_{\mathrm{tr}}} a_n(c, p), \qquad \tilde{a}_{\mathrm{sp}}(c, p) = \bar{a}_{\mathrm{sp}}(c, p),$$

so it captures a stationary spatial pattern. Neither control varies with the input state.

The fixed channel residual gives an MRE of 0.060119, close to the frozen reference at 0.060355; the fixed spatial residual lowers it to 0.056219. Matched SDC reaches 0.039420. Its channel-wise MREs are 0.000243, 0.049226, and 0.068792 for H, V, and D. After state shuffling, these values rise to 0.000420, 0.082849, and 0.119520, respectively.

The compatibility matrix used in the main text contains one entry for every pair of the 50 target states and 50 correction-source states. Its diagonal evaluates a_i on the state that produced it, while an off-diagonal entry evaluates a_j on state i. The matrix is therefore a direct state-conditionality

diagnostic: a lower diagonal than off diagonal means that the correction carries information about the state that generated it.

The trajectory-level percentile bootstrap uses complete trajectories as the resampling unit, retaining the correlation among the ten forecast leads from one initial condition. The 95% intervals are

reference-operator MRE:	$[0.059147, 0.061679]$,	matched SDC MRE:	$[0.038572, 0.040340]$,
relative improvement:	$[0.341314, 0.349233]$,	cosine alignment:	$[0.758675, 0.768636]$,
cancellation fraction:	$[0.574227, 0.589937]$.		

The state-shuffled control was repeated with 32 independent shuffle seeds. Its mean MRE was approximately 0.06746, with across-shuffle standard deviation 0.00058 and a cluster-bootstrap 95% interval of $[0.06654, 0.06846]$. The interval summarizes trajectory-level uncertainty for the evaluated checkpoints.

A.12 Supplementary cross-architecture study: frozen-backbone SDC on a planar FNO

This experiment evaluates frozen-backbone SDC with a planar FNO on a pseudo-spectral 2-D Navier–Stokes dataset generated for this study. It contains 1,200 trajectories, with 1,000 for training and 200 for testing. The solver uses viscosity 10^{-3} and step size 0.001 at 128×128, after which states are regularly downsampled to 64×64. Each trajectory contains 20 recorded frames: the first 10 form the input container and the next 10 are rollout targets.

The backbone is a residual current-only FNO with 12 Fourier modes and width 20, trained for 100 epochs with a ten-step autoregressive rollout loss. Its update and the calibrated update are

$$\widehat{x}_{n+1}^{\mathrm{FNO}} = x_n + F_\theta(x_n), \tag{94}$$

$$\widehat{x}_{n+1}^{\mathrm{FNO+SDC}} = x_n + F_{\theta_0}(x_n) + A_\psi(x_n), \tag{95}$$

where the complete θ_0 is frozen and only ψ is calibrated for 25 epochs. Both stages use batch size 20 and Adam with learning rate 10^{-3} and weight decay 10^{-4}. The FNO has 467,681 parameters; SDC adds 184,960 trainable parameters, giving 652,641 parameters in the adapted model. Although ten historical frames are retained in the data container, both the residual FNO and SDC forward paths read only the current state.

Table 5 reports relative L^2 computed per test trajectory, summarized first over rollout steps and then over the 200 fixed test trajectories. Across three independent training seeds, 25 SDC epochs reduce the ten-step mean error by $2.62\% \pm 0.43\%$ and the final-step error by $2.51\% \pm 0.30\%$. Figure 6 pairs the seed-9301 field predictions with rollout and final-step improvement summaries over the full test bank.

Table 5: Planar FNO cross-architecture study over three training seeds. Mean $\pm$ sample SD is computed across seeds. Lower relative L^2 is better.

	Ten-step mean relative L^2			Final-step relative L^2		
Seed	FNO	FNO+SDC	Improvement	FNO	FNO+SDC	Improvement
9301	0.002069	0.002006	3.04%	0.003216	0.003125	2.82%
9302	0.002204	0.002146	2.65%	0.003526	0.003438	2.51%
9303	0.002229	0.002181	2.17%	0.003580	0.003501	2.21%
Mean $\pm$ SD	0.002167 ± 0.000086	0.002111 ± 0.000093	$\mathbf{2.62\% \pm 0.43\%}$	0.003441 ± 0.000197	0.003355 ± 0.000201	$\mathbf{2.51\% \pm 0.30\%}$

Implementation checks. Before calibration, zero-initialized SDC reproduces the backbone exactly: the maximum output difference is 0.0 for every seed. The frozen FNO parameters also remain unchanged during calibration (maximum parameter change 0.0). Changing the nine older input frames while holding the current state fixed leaves the output unchanged (maximum difference 0.0). These checks confirm zero anchoring, backbone freezing, and the current-state correction path.

Training and calibration budgets. For all three seeds, the FNO reaches its lowest observed training loss at epoch 100; the mean loss over its final five epochs is lower than over the preceding

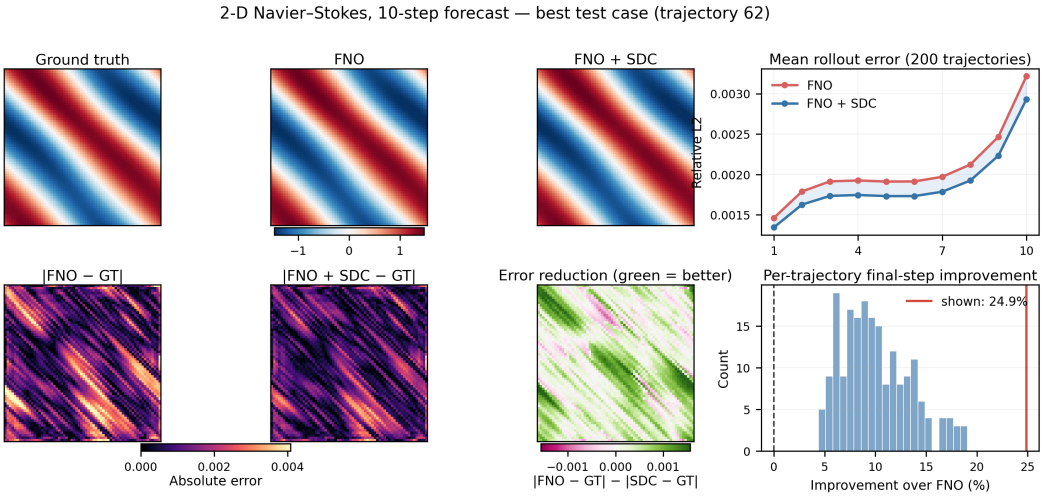

Figure 6: Planar Navier–Stokes 10-step forecast for seed 9301. Field panels show the selected high-improvement test trajectory (case 62), including ground truth, FNO and FNO+SDC predictions, absolute errors, and their error difference. The upper-right plot reports mean relative L^2 by rollout step over all 200 test trajectories. The lower-right histogram shows their final-step improvements, with case 62 marked at 24.9%.

five by 1.98%, 1.80%, and 1.79%, respectively. On these 100-epoch checkpoints, 25 SDC calibration epochs reduce the ten-step mean and final-step errors across all three seeds, as summarized in Table 5.

Extending SDC calibration to 50 epochs for seed 9301 reduces the mean-step error from 0.002069 to 0.001878 (9.20%) and the final-step error from 0.003216 to 0.002929 (8.90%). Paired bootstrap intervals over the 200 fixed test trajectories are $[8.85\%, 9.56\%]$ and $[8.34\%, 9.51\%]$, respectively. These intervals summarize trajectory-level uncertainty for this seed and checkpoint.