HELMGEONET: PHYSICS-EMBEDDED FORECASTING OF GLOBAL GEOSTROPHIC CURRENTS VIA HELMHOLTZ DECOMPOSITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate and stable long-horizon forecasting of global ocean currents remains difficult. Data-driven models typically learn the velocity field as a whole and forecast it autoregressively. Without explicit guidance on the evolution of its physically distinct components, they capture the overall trend but learn each component inaccurately, so forecasts of currents and kinetic energy lose accuracy and stability at long horizons. To address this, we propose HELMGEONET for more energy-faithful forecasting of geostrophic currents. Specifically, we embed the Helmholtz decomposition into the network architecture, so that the model explicitly learns the evolution of each physically distinct component. We also introduce a physics-guided relaxation mechanism that filters out redundant information from the input field, mitigating the error accumulation. Experiments on daily geostrophic currents confirm that HELMGEONET outperforms eight spatiotemporal backbones in currents forecasting. At the 20-day lead, its eddy kinetic energy RMSE is 0.0029 m^2/s^2 lower than that of the best baseline, and its EKE remains stable across the forecast horizon.

1 INTRODUCTION

Ocean currents store roughly 80 percent of the ocean's kinetic energy in the mesoscale eddy field, where the flow is approximately in geostrophic balance (Ferrari & Wunsch, 2009; Yu et al., 2021; Garabato et al., 2022). They transport mass and energy across the ocean and shape the global climate, so their accurate long-horizon forecast is valuable. Yet forecasting the evolution of these currents and their kinetic energy is extremely difficult: the chaotic currents dynamics make energy conservation over long horizons especially hard.

Numerical ocean models follow the governing physics and conserve energy by construction, but their computational cost limits routine high-resolution global forecasting. Data-driven forecasters are far cheaper, and trained to minimize a direct velocity-component error, they can produce satisfactory velocity forecasts with their powerful learning capability. Yet they typically learn the velocity field as a whole field, so one network must model several physically distincet components with very different dynamics at once (Bhatia et al., 2013; Xing et al., 2024): it captures the bulk trend but learns the individual physically distinct components poorly. The kinetic energy, however, is far harder to forecast as errors in each component of the currents progressively grow with the forecast horizon. As a squared quantity of the velocity, it is far more sensitive to velocity errors, so the error accumulated over autoregressive rollout strikes the energy first and hardest. This failure of pure data-driven rollouts has been repeatedly observed. On the spectral side, neural networks are biased toward low frequencies and under-estimate the high-wavenumber energy (Rahaman et al., 2019), and learned turbulence predictors amplify high-frequency errors until the energy spectrum is visibly distorted (Stachenfeld et al., 2022). On the rollout side, one-step errors compound over long horizons so that learned forecasters drift, diverge, or dissipate energy (Brandstetter et al., 2022; Lippe et al., 2023; Musale et al., 2026; Mikuš et al., 2026), and in weather and climate applications the forecast energy is systematically under-predicted (Ravuri et al., 2021; Scher & Messori, 2019; Watson, 2022). The consequence is that, over long leads, the velocity metrics of a learned forecaster can remain satisfactory while its kinetic-energy forecast has already drifted and become unstable, sometimes falling behind a trivial persistence baseline that simply copies the observed energy.

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.

To alleviate the inaccurate and unstable long-horizon forecasts of currents and kinetic energy caused by mixing their distinct physical components, we design HELMGEONET, a network structured around the Helmholtz decomposition. HELMGEONET decomposes the flow according to Helmholtz's theorem into a divergence-free, a curl-free, and a uniform component, each handled by a dedicated branch. A physics-guided relaxation mechanism additionally keeps the informative part of the input field and discards the rest, so the subsequent network modules learn from a cleaner, higher-quality signal. This mitigates the error accumulation and energy drift that plague autoregressive learning, without any physics penalty term.

Our contributions are threefold. **First, a physics-as-architecture forecaster:** we introduce HELMGEONET, which embeds the Helmholtz decomposition of the flow into the forward pass and forecasts a divergence-free, a curl-free, and a uniform component through dedicated branches, so that each physically distinct component explicitly learns its own unique evolution law, thereby capturing the dynamics more precisely. **Second, a physics-guided relaxation mechanism:** a physics-guided relaxation mechanism retains the effective and high-quality information in the input field, which makes the subsequent network modules learn more efficiently. **Third, improved accuracy and stability of the velocity and energy forecasts:** HELMGEONET improves both the velocity and the eddy-kinetic-energy forecast over eight spatiotemporal backbones and keeps the forecast energy stable across the forecast horizon.

2 RELATED WORK

Spatiotemporal predictive learning. A long line of work learns sequence-to-sequence maps for gridded dynamics with convolutional, recurrent, or attention backbones, including ConvLSTM (Shi et al., 2015), PredRNN (Wang et al., 2017), SimVP (Gao et al., 2022a), TAU (Tan et al., 2023), SwinLSTM (Tang et al., 2023; Liu et al., 2021), PredFormer (Tang et al., 2026), Earthformer (Gao et al., 2022b), and VideoMamba (Li et al., 2024). These methods achieve satisfactory long-horizon forecasting, but when forecasting physical quantities they can violate physical laws because they lack physical constraints.

Physics-informed and operator learning. PINNs (Raissi et al., 2019) and physics-informed machine learning (Karniadakis et al., 2021; Wang et al., 2021) regularize networks with equation residuals, while FNOs (Li et al., 2021) parameterize solution operators in the spectral domain, and a parallel line embeds differentiable solvers into the learning loop (Sanchez-Gonzalez et al., 2020; de Avila Belbute-Peres et al., 2021; Ummenhofer et al., 2020; Holl et al., 2020; Raissi et al., 2020). However, the PINN constraint is soft, whereas ClimODE (Verma et al., 2024) and HybridOM (Shu et al., 2026) embed hard physical structure into the architecture, offering a paradigm of physics-assisted learning that keeps forecasts from violating physical laws.

Learned Earth-system forecasting. FourCastNet (Kurth et al., 2023), GraphCast (Lam et al., 2023), and Pangu-Weather (Bi et al., 2023) demonstrate the strong potential of deep learning for Earth-system forecasting, providing a promising direction for global ocean currents.

Helmholtz decomposition in learning. The Helmholtz-Hodge decomposition splits a vector field into divergence-free and curl-free components (Bhatia et al., 2013), and has been used to post-process flows (Schoder et al., 2023), to structure flow estimation with a differentiable Helmholtz layer, and to build interpretable fluid predictors that evolve the two components separately (Xing et al., 2024). Gaussian-process models place independent priors on the two components to reconstruct ocean currents (Berlinghieri et al., 2023), and atmospheric downscaling benchmarks use the same decomposition for subgrid fluxes (Shinde et al., 2026). These works target reconstruction or idealized flows rather than real global ocean forecasting.

3 METHOD

3.1 HELMHOLTZ DECOMPOSITION

Let $\mathbf{u} = (u, v)$ be the surface geostrophic velocity on a longitude-latitude grid. Under geostrophic balance, the horizontal pressure gradient is balanced by the Coriolis force. The derived sea-surface geostrophic velocity, however, is only *approximately* non-divergent: the divergent motion

and the spatial-mean flow, though energetically small, are not negligible at daily forecast horizons. Helmholtz's theorem decomposes any velocity field on the ocean mask into three mutually orthogonal parts,

$$\mathbf{u} = \mathbf{u}_{\mathrm{sol}} + \mathbf{u}_{\mathrm{irr}} + \bar{\mathbf{u}}, \qquad \mathbf{u}_{\mathrm{sol}} = \nabla^{\perp}\psi, \quad \mathbf{u}_{\mathrm{irr}} = \nabla\chi, \quad \bar{\mathbf{u}} = \langle \mathbf{u} \rangle, \tag{1}$$

where ψ is the streamfunction and $\mathbf{u}_{\mathrm{sol}}$ the solenoidal (divergence-free) component, χ the velocity potential and $\mathbf{u}_{\mathrm{irr}}$ the irrotational (curl-free) component, and $\bar{\mathbf{u}}$ the spatially uniform flow, with $\langle \cdot \rangle$ the area-weighted ocean mean. The three components evolve on very different time scales and carry very different shares of the kinetic energy, which we exploit by forecasting each through a dedicated branch; these component-wise statistics are established in §4.2.

3.2 Solenoidal component and Biot-Savart inversion

The solenoidal component is written in terms of a streamfunction as

$$u = -\frac{\partial \psi}{\partial y}, \qquad v = \frac{\partial \psi}{\partial x}, \tag{2}$$

so that $\nabla \cdot \mathbf{u}_{\mathrm{sol}} = 0$ holds by construction, and its relative vorticity is the single scalar

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi, \tag{3}$$

which determines the entire solenoidal component. The vorticity, unlike the divergent motion, is governed by the vorticity equation and is conserved to leading order; at daily timescales it is strongly persistent, so the zero-order forecast $\zeta(t + \Delta t) \approx \zeta(t)$ is already accurate, and only the residual evolution remains to be learned. Forecasting the vorticity and inverting it to velocity therefore recovers the solenoidal evolution from a quantity that does not spuriously decay.

Given the vorticity ζ, the velocity is recovered by inverting the Poisson equation 3, which is the classical Biot-Savart inversion (Vallis, 2017; Cottet & Koumoutsakos, 2000; Janny et al., 2023). We evaluate it spectrally: writing $\hat{\zeta}(\mathbf{k})$ for the two-dimensional Fourier transform of ζ,

$$\hat{\psi}(\mathbf{k}) = -\frac{\hat{\zeta}(\mathbf{k})}{|\mathbf{k}|^2}, \quad \hat{u} = -ik_y\hat{\psi}, \quad \hat{v} = ik_x\hat{\psi}, \tag{4}$$

with the zero-frequency mean component set to zero. This step is exact for a doubly-periodic domain and carries only numerical error; its output is divergence-free to machine precision, so the non-divergence constraint of geostrophy is enforced exactly. The equivalence with the classical Biot-Savart integral is derived in Appendix B.

3.3 Irrotational and uniform components

The irrotational component is the gradient of a scalar potential, $\mathbf{u}_{\mathrm{irr}} = \nabla\chi$ (equation 1), so that $\nabla \times \mathbf{u}_{\mathrm{irr}} = \nabla \times \nabla\chi \equiv 0$: it is curl-free by construction. Unlike the solenoidal component, the irrotational component has no conservative restoring mechanism and relaxes toward zero under the dominant vortical flow; its one-day autocorrelation is $\rho \approx 0.97$, so the linear forecast $\mathbf{u}_{\mathrm{irr}}(t+\Delta t) = \rho\, \mathbf{u}_{\mathrm{irr}}(t)$ captures its geometric relaxation (§4.2). The uniform flow is quasi-static and is simply persisted. Together with the observation that the solenoidal component carries the overwhelming majority of the evolution energy (Table 1, Fig. 2), these statistics dictate the branch structure of HelmGeoNet below: a large corrector for the solenoidal component, a small damped-persistence corrector for the irrotational component, and persistence for the uniform flow.

3.4 The HelmGeoNet architecture

The forward pass of HelmGeoNet decomposes the input velocity by Helmholtz's theorem (equation 1) and forecasts the three components through dedicated branches (Fig. 1). **Solenoidal branch.** The relative vorticity ζ_t is computed from the current velocity by equation 3; the large VorCorr (Vorticity Corrector) learns the residual evolution of the vorticity, the nonlinear advection, ageostrophic coupling, and external forcing omitted by the baseline, and the corrected vorticity is inverted by the Biot-Savart solve to a divergence-free velocity. **Divergent branch.** The smaller DivCorr (Divergence Corrector) emits a scalar potential whose gradient is the curl-free velocity increment, added

on top of a damped persistence of the input divergent component. **Uniform branch.** The uniform flow is persisted from the input. All branches are zero-initialized on top of physically consistent baselines, so the model starts from the pure persistent solver.

Divergent and uniform branches. The uniform flow is simply persisted, $\bar{\mathbf{u}}_{t+\Delta t} = \bar{\mathbf{u}}_t$. The smaller DivCorr emits a scalar potential χ and forms an irrotational velocity increment

$$\delta\mathbf{u} = \nabla\chi, \qquad \widehat{\delta u} = ik_x\hat{\chi}, \quad \widehat{\delta v} = ik_y\hat{\chi}, \tag{5}$$

which is curl-free to numerical precision by construction, since $\nabla \times \nabla\chi \equiv 0$. The divergent forecast is then

$$\mathbf{u}_{\mathrm{irr},\,t+\Delta t} = \rho\,\mathbf{u}_{\mathrm{irr},\,t} + \delta\mathbf{u}, \tag{6}$$

where $\rho \approx 0.97$ is the measured one-day relaxation (Fig. 2) and $\delta\mathbf{u}$ is zero-initialized so that the model begins from damped persistence. The full forecast is the sum $\mathbf{u}_{t+\Delta t} = \mathbf{u}_{\mathrm{sol}} + \bar{\mathbf{u}} + \mathbf{u}_{\mathrm{irr}}$, and each component is divergence-free, uniform, or curl-free respectively by construction.

Encoder-translator-decoder. Both VorCorr and DivCorr share a common backbone: a three-level convolutional encoder-decoder whose bottleneck stacks L temporal-attention blocks in the style of TAU (Tan et al., 2023), chosen for its strong accuracy and efficiency in spatiotemporal forecasting. Each block applies, in parallel, a frame-static spatial attention and a channel attention,

$$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{x} \odot \Big[\sigma\big(\mathrm{DW}_k(\mathbf{x})\big) \odot \sigma\big(\mathrm{MLP}(\mathrm{pool}(\mathbf{x}))\big)\Big], \tag{7}$$

where DW_k is a large-kernel depthwise convolution that is longitudinally periodic, pool is global average pooling, MLP a two-layer channel bottleneck, and σ the sigmoid. The seasonal channels $\mathbf{t} = [\sin(2\pi\tau/T), \cos(2\pi\tau/T)]$ are concatenated to the input. The solenoidal branch uses a larger width and more translator blocks, reflecting its faster evolution and larger residual correction, while the divergent branch uses a reduced width and depth. The final projection is zero-initialized, so both correctors vanish at initialization and the model begins from the pure persistent solver. Normalization and denormalization are set to keep the inputs well scaled.

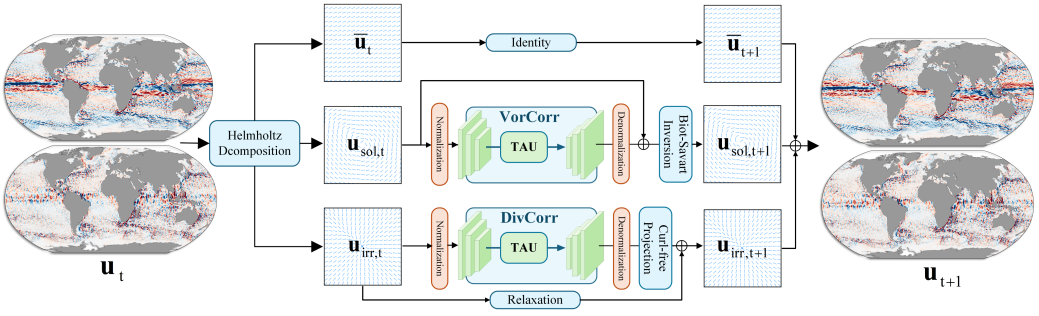


Figure 1: Overview of HELMGEONET. The velocity $\mathbf{u}_t$ is decomposed into a uniform, a divergence-free (solenoidal), and a curl-free (irrotational) component, each forecast through a dedicated branch and summed to $\mathbf{u}_{t+\Delta t}$.

4 EXPERIMENTS

4.1 DATA

We use the AVISO/DUACS delayed-time product (Pujol et al., 2016; Taburet et al., 2019), which provides daily global maps of geostrophic currents derived from multi-mission satellite altimetry, mapped onto a regular grid by optimal interpolation (Le Traon et al., 1998). The data span 25 years, from 2000 to 2024, at $0.25°$ resolution on a 680×1440 longitude-latitude grid, with the zonal and meridional geostrophic velocity components (u, v) stored per day. We split the record chronologically into a training set spanning 2000 to 2020, a validation set spanning 2021 to 2022, and a test set spanning 2023 to 2024. Land points are masked throughout. Each model consumes $T = 1$ day of history, the most recent velocity field, and forecasts the next day; longer horizons are produced by autoregressive rollout.

Table 1: Component-wise persistence errors of the Helmholtz decomposition, measured on a decade of daily AVISO currents, from 2000 to 2009. E_{raw} is the total persistence error, E_{sol}, E_{irr}, $E_{\bar{\mathbf{u}}}$ the solenoidal, irrotational, and uniform errors, and $\rho(\Delta)$ the optimal linear forecast coefficient of the divergent component.

Δ (day)	E_{raw}	E_{sol}	E_{irr}	$E_{\bar{\mathbf{u}}}$	$\rho(\Delta)$
1	6.99×10^{-4}	6.33×10^{-4}	6.30×10^{-5}	~ 0	0.9717
5	6.67×10^{-3}	6.24×10^{-3}	3.69×10^{-4}	~ 0	0.8359
10	1.53×10^{-2}	1.45×10^{-2}	7.15×10^{-4}	~ 0	0.6839
20	2.57×10^{-2}	2.45×10^{-2}	1.03×10^{-3}	~ 0	0.5463

4.2 HELMHOLTZ DECOMPOSITION STATISTICS

The branch structure of HELMGEONET rests on the statistical evolution of the three Helmholtz components, which we quantify on a decade of daily AVISO currents, from 2000 to 2009, using the latitude-weighted inner product $\langle \mathbf{a}, \mathbf{b} \rangle = \frac{\sum_{\text{ocean}} w\, \mathbf{a}\cdot\mathbf{b}}{\sum_{\text{ocean}} w}$ with $w = \cos\phi$. For a lag Δ, the evolution of a component is measured by the mean squared error $E(\Delta) = \langle \|\mathbf{u}(t+\Delta) - \mathbf{u}(t)\|^2 \rangle$, which splits into solenoidal, irrotational, uniform, and cross terms; the share of each term reveals which component carries the forecastable evolution.

Figure 2 analyzes these statistics over a decade of daily currents. Panel *(a)* shows the uniform flow: its zonal and meridional ocean means vary by only ~ 0.003 and ~ 0.001 m s^{-1} over the year, so it is quasi-static and is persisted. Panels *(b)* and *(c)* decompose the evolution energy (Table 1): the solenoidal component accounts for about 90% of the total evolution, the uniform flow contributes essentially nothing, and the irrotational component carries the remaining 10%, while the cross term lies one to two orders of magnitude below the irrotational part; this small cross term shows that, in a statistical sense, the two components evolve almost independently, so they can be forecast in separate branches. Panel *(d)* compares the measured optimal linear forecast coefficient with a geometric decay: the irrotational component matches it closely, while the solenoidal component does not.

For the divergent component, this independence justifies a simple relaxational forecast. The optimal linear forecast of $\mathbf{u}_{\text{irr}}(t+\Delta)$ from $\mathbf{u}_{\text{irr}}(t)$ is the scalar multiple

$$\hat{\mathbf{u}}_{\text{irr}}(t+\Delta) = \rho(\Delta)\, \mathbf{u}_{\text{irr}}(t), \qquad \rho(\Delta) = \frac{\langle \mathbf{u}_{\text{irr}}(t),\, \mathbf{u}_{\text{irr}}(t+\Delta) \rangle}{\langle \mathbf{u}_{\text{irr}}(t),\, \mathbf{u}_{\text{irr}}(t) \rangle}, \tag{8}$$

which minimizes the mean squared error among all linear forecasts, since the residual $\varepsilon = \mathbf{u}_{\text{irr}}(t+\Delta) - \rho \mathbf{u}_{\text{irr}}(t)$ is orthogonal to $\mathbf{u}_{\text{irr}}(t)$. At one day the measured coefficient is $\rho(1) \approx 0.97$, and it decays nearly geometrically as $\rho(1)^{\Delta}$ (Fig. 2, Table 1), so the divergent component is well described by the damped persistence $\mathbf{u}_{\text{irr}}(t+\Delta t) = \rho\, \mathbf{u}_{\text{irr}}(t)$ with $\rho = 0.97$. The solenoidal component, by contrast, has a one-day autocorrelation of $\rho_{\text{sol}} \approx 0.99$ (Fig. 2), so persisting it exactly and learning only the residual is the natural baseline rather than a scalar damping. The damping coefficient $\rho = 0.97$ is further confirmed by a pure-data ablation over the physics-only baseline, which minimizes the eddy-kinetic-energy error at $\rho = 0.97$ (Appendix Tables S1 and S2). The relaxation is thus physics-guided rather than learned: the coefficient is read off from the measured autocorrelation of the historical real flow, so the divergent branch inherits the observed decay of the ocean as a fixed inductive prior.

4.3 BASELINES

We compare against eight recent spatiotemporal backbones: ConvLSTM (Shi et al., 2015), SwinLSTM (Tang et al., 2023), TAU (Tan et al., 2023), SimVP (Gao et al., 2022a), PredFormer (Tang et al., 2026), VideoMamba (Li et al., 2024), FourCastNet (Kurth et al., 2023), and Earthformer (Gao et al., 2022b). A persistence forecast serves as a non-learned lower bound. Each baseline retains its core architecture but is scaled down to a small parameter count of roughly 4M. Our model is smaller still, requiring fewer than 2M parameters.

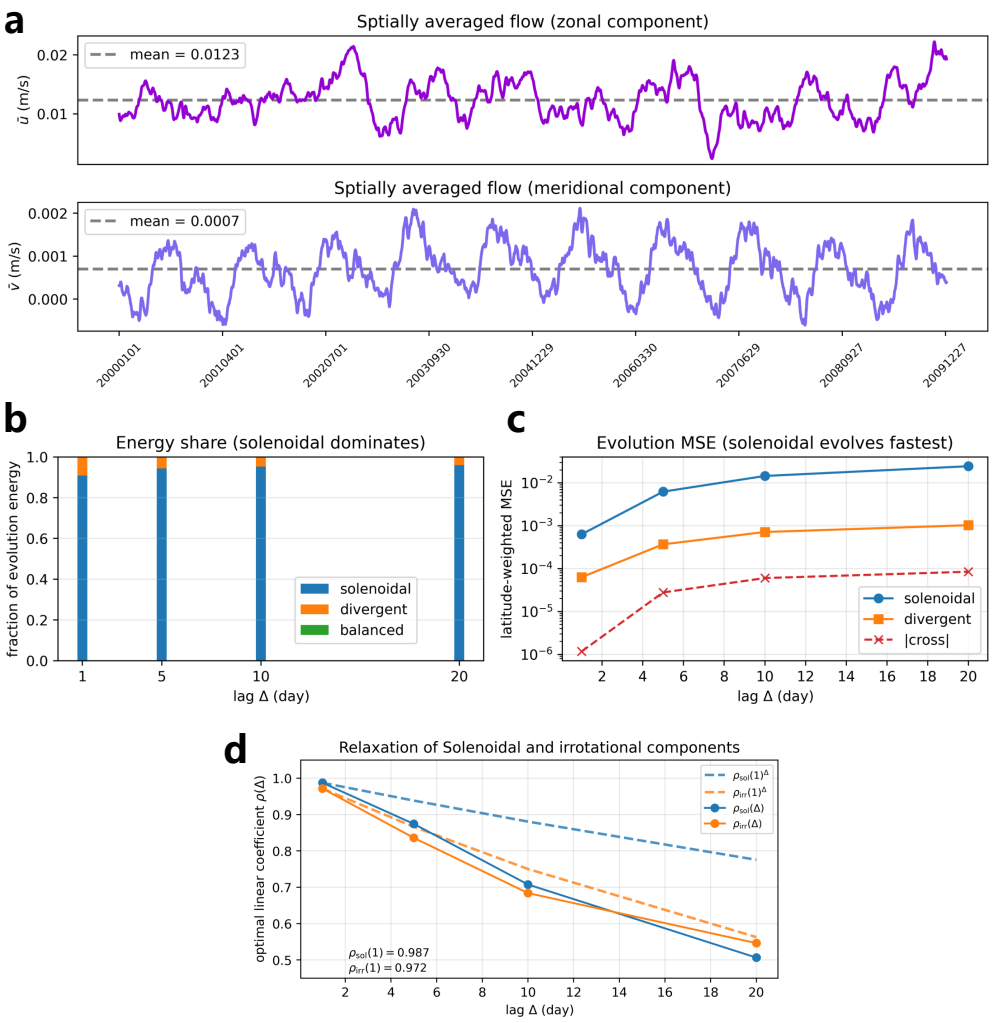


Figure 2: Helmholtz decomposition statistics on a decade of daily AVISO currents. *(a)* Uniform ocean-mean flow. *(b)* Share of each component in the evolution energy. *(c)* Component-wise mean squared error versus lag. *(d)* Optimal linear forecast coefficient $\rho(\Delta)$.

4.4 METRICS

We measure the forecast accuracy of the velocity components u and v and of the eddy kinetic energy EKE by the latitude-weighted root-mean-squared error RMSE and the anomaly correlation coefficient ACC against the climatology, with $\mathrm{EKE} = \frac{1}{2}\langle|\mathbf{u} - \mathbf{u}^{\mathrm{clim}}|^2\rangle$. To measure the stability of the EKE forecast across the $T = 20$ lead days, we form the per-lead normalized energy error $e(t) = \left[\bar{K}^{\mathrm{pred}}(t) - \bar{K}^{\mathrm{obs}}(t)\right]/\bar{K}^{\mathrm{obs}}(t)$ and report its standard deviation

$$\sigma = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\big(e(t) - \bar{e}\big)^2}, \qquad \bar{e} = \frac{1}{T}\sum_{t=1}^{T} e(t), \tag{9}$$

and the standard deviation of its first differences

$$\sigma_\Delta = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T-1}\big(\Delta e(t) - \overline{\Delta e}\big)^2}, \qquad \Delta e(t) = e(t{+}1) - e(t), \tag{10}$$

4.5 Implementation details

All losses and metrics are restricted to ocean points and weighted by $\cos\phi$, and the training loss is the latitude-weighted MSE over ocean points. Within 5° of the equator the AVISO data does not satisfy geostrophic balance and carries large errors, so this band is excluded from evaluation. All models are trained with AdamW and a fixed learning rate of 10^{-4}, with a batch size of 16, on 0.25° data with early stopping on validation MSE. All experiments are run in PyTorch 2.5 on a single Hygon BW1000 DCU with 64 GB of memory.

5 Results

5.1 Sensitivity of the divergence damping coefficient

Fig. 2 fixes the relaxation coefficient at $\rho = 0.97$, the measured optimal linear forecast coefficient of the divergent component. To verify that this choice is necessary rather than incidental, we retrain the full model with $\rho \in \{0.8, 1.1\}$ and report the skill in Table 2.

Table 2: Sensitivity of HelmGeoNet to the divergence damping coefficient ρ, reporting the eddy-kinetic-energy anomaly correlation and latitude-weighted RMSE. Best value per column is bolded and second-best is underlined.

	lead 5d		lead 10d		lead 20d	
	EKE ACC	EKE RMSE	EKE ACC	EKE RMSE	EKE ACC	EKE RMSE
$\rho = 0.8$	0.8664	0.0188	0.4445	0.0710	0.0375	0.5077
$\rho = 0.97$	<u>0.8700</u>	**0.0185**	**0.6872**	**0.0283**	**0.4194**	**0.0382**
$\rho = 1.1$	**0.8705**	<u>0.0187</u>	<u>0.6775</u>	<u>0.0297</u>	<u>0.2527</u>	<u>0.0683</u>

The single-step forecast masks this sensitivity: the divergent network adapts to whatever damping it is trained with, so the short-lead error stays comparable across ρ, and the sensitivity appears only in the autoregressive rollout; at the 5-day lead, $\rho = 1.1$ even slightly surpasses $\rho = 0.97$ in EKE ACC (0.8705 versus 0.8700), because over such short rollouts DivCorr can learn to compensate for the relaxation it is trained with, leaving the coefficients indistinguishable. Setting $\rho = 0.8$ or $\rho = 1.1$ produces markedly worse long-lead skill: the EKE RMSE at 20 days rises from 0.038 to 0.068 for $\rho = 1.1$ and to 0.508 for $\rho = 0.8$, even though the coefficient is perturbed by only 0.1 to 0.2. A damping that is too strong removes too much divergent energy, while one that is too weak retains too much, and either physically inconsistent signal misleads the network over the long rollout.

5.2 Multi-day forecast accuracy

We assess forecast skill over a 20-day horizon by autoregressive rollout: each model, trained for single-day prediction, feeds its own output back as the input for the next step. Two numerical precautions are applied. First, a minority of grid points can diverge to unphysical magnitudes during rollout and would otherwise dominate the global error statistics; we therefore clamp every predicted velocity component to $|u|, |v| \le 5\,\mathrm{m\,s^{-1}}$, a bound far above the geostrophic-current range and chosen solely to exclude numerical blow-up. Second, because geostrophic balance degenerates near the equator, all metrics are evaluated over the daily valid non-NaN ocean grid excluding the band $|\phi| < 5^\circ$; the equatorial points are nevertheless retained as autoregressive inputs so that the input field remains spatially continuous.

Table 3 and Figure 3 report the forecast accuracy of the velocity components u and v and of the eddy kinetic energy. At day 20, HelmGeoNet attains the highest EKE ACC, 0.4194, and the lowest EKE RMSE, 0.0382 m^2/s^2, outperforming every baseline and demonstrating a stronger ability to forecast the kinetic energy. Across the horizon, although most baselines keep a velocity RMSE comparable to, or better than, persistence, their EKE RMSE already exceeds that of persistence after roughly ten days, because the kinetic energy is far more sensitive to velocity errors; in contrast, HelmGeoNet outperforms all baselines on every velocity metric and remains ahead of persistence on the energy metrics across the entire horizon.

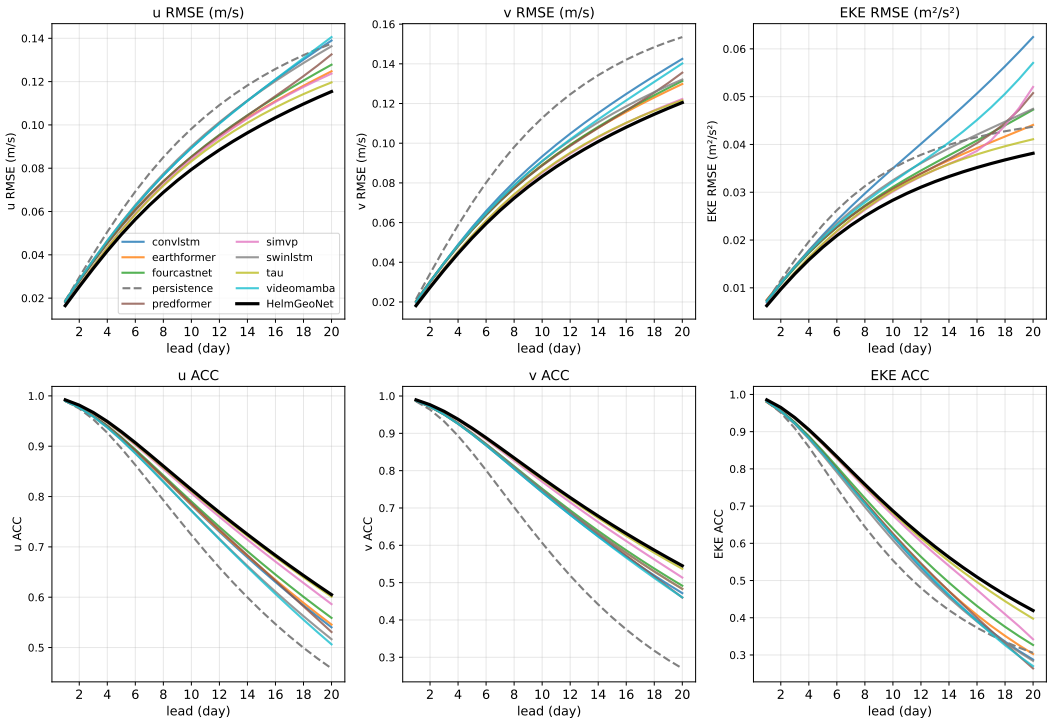


Figure 3: Multi-day forecast skill versus lead time on the 2023-2024 test set, restricted to $|\phi| > 5°$ with latitude weighting. Top row: RMSE of u, v, and EKE; bottom row: anomaly correlation.

Table 3: Day-20 EKE forecast skill on the test set spanning 2023 to 2024. EKE ACC and EKE RMSE are latitude-weighted over valid ocean points excluding $|\phi| < 5°$; the best value per column is bolded and the second-best is underlined.

Model	EKE ACC ↑	EKE RMSE (m^2/s^2)
Persistence	0.3057	0.0437
ConvLSTM (Shi et al., 2015)	0.2877	0.0625
SwinLSTM (Tang et al., 2023)	0.2842	0.0475
TAU (Tan et al., 2023)	<u>0.3975</u>	<u>0.0411</u>
SimVP (Gao et al., 2022a)	0.3422	0.0520
PredFormer (Tang et al., 2026)	0.2638	0.0508
VideoMamba (Li et al., 2024)	0.2701	0.0571
FourCastNet (Kurth et al., 2023)	0.3268	0.0473
Earthformer (Gao et al., 2022b)	0.3026	0.0441
HELMGEONET (ours)	**0.4194**	**0.0382**

Beyond the pointwise skill curves, we diagnose the stability of the energy forecast with the metrics of §4.4. Figure 4(a) shows, for each model, the two-year mean of the day-20 eddy kinetic energy field relative to the corresponding label mean, with the equatorial band $|\phi| < 5°$ masked; this two-year mean exposes the overall error trend of the energy forecast rather than a single-day accuracy. Several baselines, most visibly ConvLSTM, accumulate a large spurious energy excess concentrated in the western boundary currents, whereas HELMGEONET, TAU, and SimVP stay close to zero. Figure 4(b) reports the standard deviation σ and the first-difference standard deviation σ_Δ of equation 9 and equation 10 for every model: HELMGEONET attains the smallest of both, so its energy forecast stays closest to the label energy level and evolves most smoothly with lead, neither diverging nor dissipating as the horizon grows. This result demonstrates that HELMGEONET retains the forecasting stability of eddy kinetic energy.

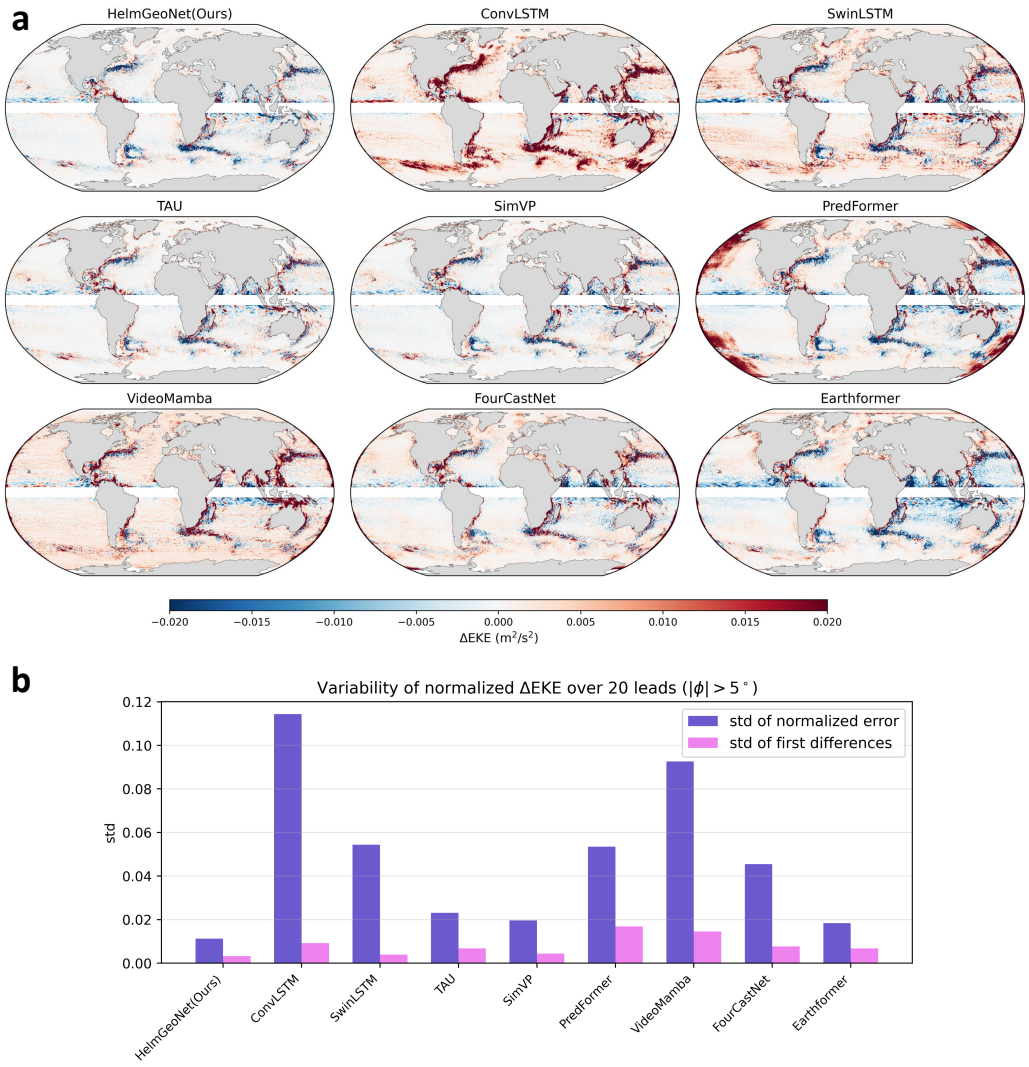


Figure 4: Energy stability of the 20-day autoregressive forecast on the 2023-2024 test set. *(a)* Two-year mean of the day-20 eddy kinetic energy, shown as the difference from the label mean, with $|\phi| < 5°$ masked. *(b)* Standard deviation σ (purple) and first-difference standard deviation σ_Δ (pink) of the per-lead normalized energy error $e(t)$; HELMGEONET attains the smallest of both.

6 CONCLUSION

In this work, we introduced HELMGEONET, a physics-embedded forecaster that mitigates forecast errors primarily by separately learning the physically distinct components of the flow velocity that follow different evolution laws, each through a dedicated branch. In our verifications on daily geostrophic currents, HELMGEONET improves both the velocity and the eddy-kinetic-energy forecast over eight spatiotemporal backbones, and keeps the forecast energy stable across the horizon. Three directions still remain for future work. First, part of the non-divergence of geostrophic currents stems from numerical error, and although the relaxation coefficient proves effective, whether it acts more on numerical error or on physical divergence is unclear; studying sea-surface absolute velocity, whose divergence is dominated by real physics, would clarify this. Second, the repeated Fourier transforms in HELMGEONET limit its computational efficiency and inference speed, so improving efficiency while preserving accuracy is worth pursuing. Third, the current $0.25°$ resolution is coarse, and validating the model at higher resolution is a natural next step.

AI USE STATEMENT

We used a generative AI assistant to help implement the baseline models, evaluation code, and to polish portions of this manuscript. We did not use generative AI tools to produce experimental results, figures, tables, or scientific claims, which are the authors' own. All AI-assisted code and text were reviewed, verified, and corrected by the authors, who take full responsibility for the final content.

ETHICS STATEMENT

This work uses publicly available satellite-derived ocean data and involves no human subjects, personally identifiable information, or harmful applications; we foresee no ethical concerns.

REPRODUCIBILITY STATEMENT

We describe the data source in §4.1 and the preprocessing and training protocol in §4.5. The model, baselines, and evaluation scripts are implemented in PyTorch and will be released upon acceptance, together with the exact data splits and hyperparameters. The AVISO/DUACS product is publicly available from the Copernicus Marine Service.

REFERENCES

Renato Berlinghieri, Brian L. Trippe, David R. Burt, Ryan Giordano, Kaushik Srinivasan, Tamay Özgökmen, Junfei Xia, and Tamara Broderick. Gaussian processes at the helm(holtz): A more fluid model for ocean currents. In *International Conference on Machine Learning (ICML)*, 2023.

Harsh Bhatia, Gregory Norgard, Valerio Pascucci, and Peer-Timo Bremer. The helmholtz-hodge decomposition—a survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(8): 1386–1404, 2013.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619:533–538, 2023.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations (ICLR)*, 2022.

Georges-Henri Cottet and Petros D. Koumoutsakos. *Vortex Methods: Theory and Practice*. Cambridge University Press, 2000.

Filipe de Avila Belbute-Peres, Thomas D. Economon, and J. Zico Kolter. Combining differentiable PDE solvers and graph neural networks for fluid flow prediction. In *International Conference on Learning Representations (ICLR)*, 2021.

Raffaele Ferrari and Carl Wunsch. Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, 41:253–282, 2009.

Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z. Li. SimVP: Simpler yet better video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022a.

Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Exploring space-time transformers for earth system forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022b.

Alberto C. Naveira Garabato, Xiaolong Yu, Jörn Callies, Roy Barkan, Kurt L. Polzin, Eleanor E. Frajka-Williams, Christian E. Buckingham, and Stephen M. Griffies. Kinetic energy transfers between mesoscale and submesoscale motions in the open ocean's upper layers. *Journal of Physical Oceanography*, 52(1):75–97, 2022. doi: 10.1175/JPO-D-21-0099.1.

Philipp Holl, Vladlen Koltun, and Nils Thuerey. Learning to control PDEs with differentiable physics. In *International Conference on Learning Representations (ICLR)*, 2020.

Steeven Janny, Madiha Nadri, Julie Digne, and Christian Wolf. Learning vortex dynamics for fluid inference and prediction. In *International Conference on Learning Representations (ICLR)*, 2023.

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

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*, PASC '23, pp. 1–11, New York, NY, USA, 2023. Association for Computing Machinery. doi: 10.1145/3592979.3593412.

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.

Pierre-Yves Le Traon, F. Nadal, and N. Ducet. An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology*, 15(2):522–534, 1998.

Kunchang Li, Xinhao Li, Yi Wang, Yinan He, Yali Wang, Limin Wang, and Yu Qiao. VideoMamba: State space model for efficient video understanding. In *European Conference on Computer Vision (ECCV)*, 2024.

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 (ICLR)*, 2021.

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 (NeurIPS)*, 2023.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Iva Mikuš, Boris Muha, and Domagoj Vlah. Deep convolutional-transformer models for stable autoregressive evolution of parametric partial differential equations. In *World Congress on Computational Mechanics (WCCM-ECCOMAS)*, 2026.

Ashish Musale, Ajay Waghumbare, T. Vinduja, and Upasna Singh. Nigo: Lyapunov-stable continuous-time neural operators for long-horizon pde dynamics. In *International Conference on Machine Learning (ICML)*, 2026.

Marie-Isabelle Pujol, Yannice Faugère, Guillaume Taburet, Stéphanie Dupuy, Camille Pelloquin, Michaël Ablain, and Nicolas Picot. DUACS DT2014: the new multi-mission altimeter data set reprocessed over 20 years. *Ocean Science*, 12(5):1067–1090, 2016.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning (ICML)*, 2019.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, et al. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597:672–677, 2021.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning (ICML)*, 2020.

Sebastian Scher and Gabriele Messori. Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geoscientific Model Development*, 12(7):2797–2809, 2019.

Stefan Schoder, Florian Kraxberger, Eniz Mušeljić, and Andreas Wurzinger. Post-processing subsonic flows using physics-informed neural networks. In *AIAA AVIATION 2023 Forum*, 2023.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

Rajat C. Shinde, Kumar Ankur, Christopher Phillips, Aman Gupta, Simon Pfreundschuh, Sujit Roy, Sheyenne Kirkland, Vishal Gaur, Venkatesh Kolluru, Amy Lin, Prajun Trital, Aditi Sheshadri, Udaysankar Nair, Manil Maskey, and Rahul Ramachandran. Wxc-bench: A novel dataset for weather and climate downstream tasks. *Scientific Data*, 13, 2026.

Ruiqi Shu, Xiaohui Zhong, Qiusheng Huang, Ruijian Gou, Tianrun Gao, Hao Li, and Xiaomeng Huang. Hybridom: Hybrid physics-based and data-driven global ocean modeling with efficient spatial downscaling. In *International Conference on Machine Learning (ICML)*, 2026.

Kimberly Stachenfeld, Drummond B. Fielding, Dmitrii Kochanski, 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 (ICLR)*, 2022.

Guillaume Taburet, Antonio Sanchez-Roman, Maxime Ballarotta, Marie-Isabelle Pujol, Jean-François Legeais, Florent Fournier, Yannice Faugère, and Gérald Dibarboure. DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Science*, 15(5):1207–1224, 2019.

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

Song Tang, Chuang Li, Pu Zhang, and RongNian Tang. SwinLSTM: Improving spatiotemporal prediction accuracy using swin transformer and LSTM. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Yujin Tang, Lu Qi, Fei Xie, Xiangtai Li, Chao Ma, and Ming-Hsuan Yang. Video prediction transformers without recurrence or convolution. *Transactions on Machine Learning Research*, 2026. arXiv:2410.04733.

Benjamin Ummenhofer, Lukas Prantl, Nils Thuerey, and Vladlen Koltun. Lagrangian fluid simulation with continuous convolutions. In *International Conference on Learning Representations (ICLR)*, 2020.

Geoffrey K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*. Cambridge University Press, 2nd edition, 2017.

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

Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2021.

Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S. Yu. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Peter A. G. Watson. Applying machine learning to improve simulations of a chaotic dynamical system using empirical error correction. *Journal of Advances in Modeling Earth Systems*, 14(5): e2021MS002910, 2022.

Lanxiang Xing, Haixu Wu, Yuezhou Ma, Jianmin Wang, and Mingsheng Long. Helmfluid: Learning helmholtz dynamics for interpretable fluid prediction. In *International Conference on Machine Learning (ICML)*, 2024.

Xiaolong Yu, Aurélien Ponte, Noé Lahaye, Zoé Caspar-Cohen, and Dimitris Menemenlis. Geostrophy assessment and momentum balance of the global oceans in a tide- and eddy-resolving model. *Journal of Geophysical Research: Oceans*, 2021.

APPENDIX

A HELMHOLTZ DECOMPOSITION STATISTICS

We describe the numerical procedure behind Fig. 2 and the statistics quoted in §3.1. All quantities are computed on the ocean mask with the same spectral, filling, and masking semantics as the forward pass, in double precision. Given a velocity field $\mathbf{u}$, its uniform component is the area-weighted ocean mean $\bar{\mathbf{u}} = \langle \mathbf{u} \rangle$, and its solenoidal component is the Biot-Savart inversion of its vorticity, $\mathbf{u}_{\mathrm{sol}} = \mathrm{BS}^{-1}(\zeta)$, with $\zeta = \partial_x v - \partial_y u$ computed spectrally as $\hat{\zeta} = ik_x\hat{v} - ik_y\hat{u}$, and the $k = 0$ mode set to zero. The divergent component is the residual $\mathbf{u}_{\mathrm{irr}} = \mathbf{u} - \mathbf{u}_{\mathrm{sol}} - \bar{\mathbf{u}}$, which is orthogonal to the other two under the same latitude weighting as the training loss.

The latitude-weighted inner product and mean squared error used throughout are

$$\langle \mathbf{a}, \mathbf{b} \rangle = \frac{\sum_{\mathrm{ocean}} w\, \mathbf{a} \cdot \mathbf{b}}{\sum_{\mathrm{ocean}} w}, \qquad E(\mathbf{a}, \mathbf{b}) = \langle \mathbf{a} - \mathbf{b}, \mathbf{a} - \mathbf{b} \rangle, \tag{11}$$

where $w = \cos\phi$ is the area weight and the sum runs over valid non-NaN ocean grid points only. For a lag Δ, with $\Delta\mathbf{x} = \mathbf{x}(t + \Delta) - \mathbf{x}(t)$, the evolution of each component is measured by $E_{\mathrm{sol}}(\Delta) = \langle \|\Delta\mathbf{u}_{\mathrm{sol}}\|^2 \rangle$, and likewise for E_{irr} and $E_{\bar{\mathbf{u}}}$. Writing $\mathbf{x}_0 = \mathbf{x}(t)$ and $\mathbf{x}_\Delta = \mathbf{x}(t + \Delta)$, the Helmholtz decomposition $\mathbf{x} = \mathbf{x}_{\mathrm{sol}} + \mathbf{x}_{\mathrm{irr}} + \bar{\mathbf{x}}$ yields the exact account identity

$$E_{\mathrm{raw}}(\Delta) = \langle \|\mathbf{x}_\Delta - \mathbf{x}_0\|^2 \rangle = E_{\mathrm{sol}}(\Delta) + E_{\mathrm{irr}}(\Delta) + E_{\bar{\mathbf{u}}}(\Delta) + 2 \sum_{a<b} \langle \Delta\mathbf{x}_a, \Delta\mathbf{x}_b \rangle, \tag{12}$$

which closes to 10^{-6} in double precision. The optimal linear forecast coefficient of the divergent component is

$$\rho(\Delta) = \frac{\langle \mathbf{u}_{\mathrm{irr}}(t), \mathbf{u}_{\mathrm{irr}}(t + \Delta) \rangle}{\langle \mathbf{u}_{\mathrm{irr}}(t), \mathbf{u}_{\mathrm{irr}}(t) \rangle}, \tag{13}$$

whose one-day value $\rho(1) \approx 0.97$ is fixed in the model as the damping factor ρ in equation 6. The residual floor after this optimal linear forecast is

$$F(\Delta) = E_{\mathrm{irr}}(\Delta)\,\big(1 - r^2(\Delta)\big), \qquad r(\Delta) = \frac{\langle \mathbf{x}_{\mathrm{irr},0}, \mathbf{x}_{\mathrm{irr},\Delta} \rangle}{\sqrt{\langle \|\mathbf{x}_{\mathrm{irr},0}\|^2 \rangle \langle \|\mathbf{x}_{\mathrm{irr},\Delta}\|^2 \rangle}}, \tag{14}$$

where $r(\Delta)$ is the correlation coefficient. The one-day coefficient $\rho(1) = 0.9717$ and its slow decay with Δ (Table 1) justify the damped-persistence forecast of equation 6.

B FROM THE BIOT-SAVART INTEGRAL TO THE SPECTRAL SOLUTION

We make explicit the equivalence between the Biot-Savart integral and the FFT-based spectral Poisson solve used in §3.2. The velocity induced by a vorticity field $\omega(\mathbf{x})$, $\mathbf{x} = (x, y)$, is the Biot-Savart integral

$$\mathbf{u}(\mathbf{x}) = \int K(\mathbf{x} - \mathbf{x}')\, \omega(\mathbf{x}')\, \mathrm{d}\mathbf{x}', \qquad K(\mathbf{x}) = \frac{1}{2\pi}\, \frac{(-y,\ x)}{x^2 + y^2}, \tag{15}$$

where $\mathbf{u} = (u, v)$ and K denotes the singular Biot-Savart kernel.

Step 1: the kernel as a perpendicular gradient. Let $G(\mathbf{x}) = -\frac{1}{2\pi} \ln|\mathbf{x}|$ be the fundamental solution of the two-dimensional Laplacian,

$$\nabla^2 G = -\delta(\mathbf{x}), \qquad \text{since} \qquad \nabla^2 \ln|\mathbf{x}| = 2\pi\, \delta(\mathbf{x}). \tag{16}$$

The Biot-Savart kernel is precisely the perpendicular gradient $\nabla^\perp \equiv (\partial_y,\ -\partial_x)$ of G:

$$K(\mathbf{x}) = \nabla^\perp G(\mathbf{x}) = -\frac{1}{2\pi} \nabla^\perp \ln|\mathbf{x}| = \frac{1}{2\pi}\, \frac{(-y,\ x)}{|\mathbf{x}|^2}. \tag{17}$$

Step 2: recovering the stream function and the Poisson equation. Substituting this into equation 15 and using the associativity of convolution,

$$\mathbf{u} = (\nabla^\perp G) * \omega = \nabla^\perp (G * \omega) \equiv \nabla^\perp \psi, \qquad \psi \equiv G * \omega. \tag{18}$$

Taking the Laplacian and using $\nabla^2 G = -\delta$,

$$\nabla^2 \psi = (\nabla^2 G) * \omega = -\delta * \omega = -\omega, \tag{19}$$

i.e., the Biot-Savart integral is exactly the Green's-function solution of the Poisson equation $\nabla^2 \psi = -\omega$, followed by $\mathbf{u} = \nabla^\perp \psi$.

Step 3: Fourier transform of the kernel. Since derivatives map to multiplications, $\mathcal{F}[\partial_j f] = \mathrm{i} k_j \hat{f}$, and transforming $\nabla^2 \ln|\mathbf{x}| = 2\pi\delta$ yields $-|\mathbf{k}|^2 \mathcal{F}[\ln|\mathbf{x}|] = 2\pi$, i.e., in the distributional sense with the zero mode excluded,

$$\mathcal{F}[\ln|\mathbf{x}|] = -\frac{2\pi}{|\mathbf{k}|^2}. \tag{20}$$

The Fourier transform of the Biot-Savart kernel therefore reads

$$\widehat{K}(\mathbf{k}) = \mathcal{F}\big[\nabla^\perp G\big] = -\frac{1}{2\pi}\,(\mathrm{i}k_y,\, -\mathrm{i}k_x)\,\mathcal{F}[\ln|\mathbf{x}|] = \frac{\mathrm{i}\,(k_y,\, -k_x)}{k_x^2 + k_y^2}. \tag{21}$$

Step 4: the integral becomes a product. By the convolution theorem, the global integral equation 15 is equivalent, in spectral space, to the pointwise product

$$\hat{u}(\mathbf{k}) = \frac{\mathrm{i}k_y}{|\mathbf{k}|^2}\,\hat{\omega}(\mathbf{k}), \qquad \hat{v}(\mathbf{k}) = \frac{-\mathrm{i}k_x}{|\mathbf{k}|^2}\,\hat{\omega}(\mathbf{k}). \tag{22}$$

Defining $\hat{\psi} = \hat{\omega}/|\mathbf{k}|^2$, equation 22 is the spectral image of $\mathbf{u} = \nabla^\perp \psi$ together with $-|\mathbf{k}|^2 \hat{\psi} = -\hat{\omega}$, i.e., the transform of the Poisson equation $\nabla^2 \psi = -\omega$.

Step 5: discrete algorithm. On a periodic grid, equation 22 is evaluated with the FFT: (i) $\hat{\omega} = \mathrm{FFT}(\omega)$; (ii) $\hat{\psi} = \hat{\omega}/|\mathbf{k}|^2$, with the zero mode set to zero, as it corresponds to an arbitrary constant of ψ; (iii) $(\hat{u}, \hat{v}) = \mathrm{i}(k_y, -k_x)\,\hat{\psi}$; (iv) $(u, v) = \mathrm{Re}\{\mathrm{IFFT}(\hat{u}, \hat{v})\}$. This evaluates the *untruncated* Biot-Savart integral in $O(N \log N)$ operations instead of $O(N^2)$, and preserves incompressibility exactly, since

$$k_x \hat{u} + k_y \hat{v} = \mathrm{i}\,(k_x k_y - k_y k_x)\,\hat{\psi} = 0. \tag{23}$$

A Gaussian-regularized vortex-blob kernel $K_\delta(\mathbf{x}) = \frac{1}{2\pi|\mathbf{x}|^2}\big(1 - e^{-|\mathbf{x}|^2/\delta^2}\big)(-y,\, x)$ corresponds, in spectral space, to multiplying equation 21 by a Gaussian factor,

$$\widehat{K}_\delta(\mathbf{k}) = \frac{\mathrm{i}\,(k_y,\, -k_x)}{|\mathbf{k}|^2}\, e^{-\delta^2|\mathbf{k}|^2/4}, \tag{24}$$

i.e., mollifying the point vortex in physical space is equivalent to damping high wavenumbers in spectral space.

C SENSITIVITY OF THE DIVERGENCE DAMPING COEFFICIENT

Table S1: Anomaly correlation of the physics-only baseline as a function of the divergence damping coefficient ρ. Each block reports ACC u, ACC v, and EKE ACC at lead 5, 10, and 20 days on the 2023-2024 test set; the best value per column is bolded.

	lead 5d			lead 10d			lead 20d		
	ACC u	ACC v	EKE ACC	ACC u	ACC v	EKE ACC	ACC u	ACC v	EKE ACC
$\rho = 0.8$	0.8904	0.8441	0.7947	0.7171	0.6004	0.5412	0.4512	0.2661	0.2961
$\rho = 0.97$	**0.8962**	**0.8487**	**0.8043**	**0.7247**	**0.6062**	**0.5528**	**0.4608**	**0.2697**	**0.3032**
$\rho = 1.1$	0.8893	0.8430	0.7958	0.6924	0.5798	0.5176	0.3435	0.1623	0.1820

We fix the divergence damping coefficient ρ through a pure-data ablation rather than by tuning the full model. On the physics-only baseline, in which every neural module is disabled, the model reduces to vorticity persistence, mean-flow persistence, and a damped persistence $\rho\,\mathbf{u}_{\mathrm{irr}}$ of the divergent component, with no learnable parameters. Varying ρ over $\{0.8, 0.97, 1.1\}$ and evaluating the 20-day autoregressive forecast on the 2023-2024 test set yields Tables S1 and S2. Two observations follow. First, the eddy-kinetic-energy RMSE is minimized at $\rho = 0.97$ at every lead, matching the measured optimal linear forecast coefficient $\rho(1) = 0.9717$ of the divergent component. Second, $\rho = 1.1$ is clearly inferior: a damping larger than unity amplifies the divergent component at every autoregressive step, so it grows without bound, contradicting the physical tendency of divergence to decay, and the long-lead error accumulates sharply. Fixing $\rho = 0.97$ therefore anchors the numerical tendency of the divergent branch to the physical optimum and shortens the path that the residual network must optimize.

Table S2: Latitude-weighted RMSE of the physics-only baseline as a function of ρ. Each block reports RMSE u, RMSE v, and EKE RMSE at lead 5, 10, and 20 days; the best value per column is bolded. The EKE RMSE is minimized at $\rho = 0.97$ at every lead, whereas $\rho = 1.1$ amplifies the divergence and the error accumulates sharply.

	lead 5d			lead 10d			lead 20d		
	RMSE u	RMSE v	EKE RMSE	RMSE u	RMSE v	EKE RMSE	RMSE u	RMSE v	EKE RMSE
$\rho = 0.8$	0.0609	0.0723	0.0236	0.0973	0.1165	0.0353	0.1360	0.1584	0.0439
$\rho = 0.97$	**0.0599**	**0.0701**	**0.0231**	**0.0973**	**0.1134**	**0.0349**	**0.1354**	**0.1552**	**0.0435**
$\rho = 1.1$	0.0639	0.0705	0.0243	0.1123	0.1148	0.0404	0.2178	0.1836	0.1051