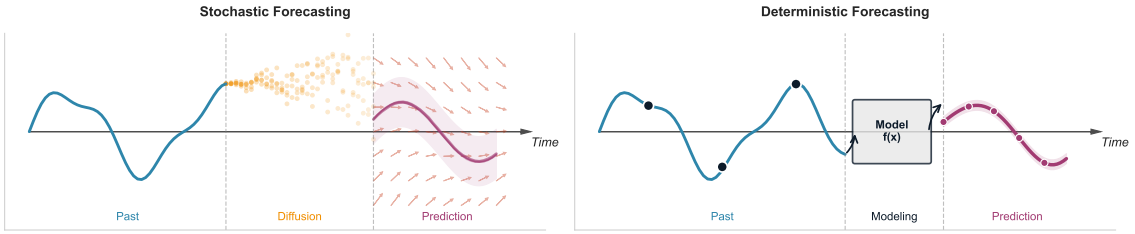
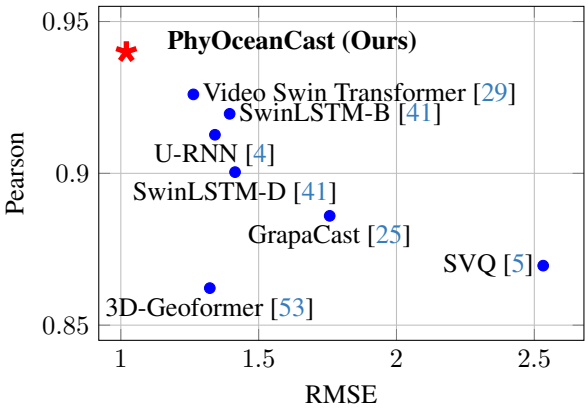
This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

PhyOceanCast: Global Ocean Forecasting with Physics-Informed Diffusion

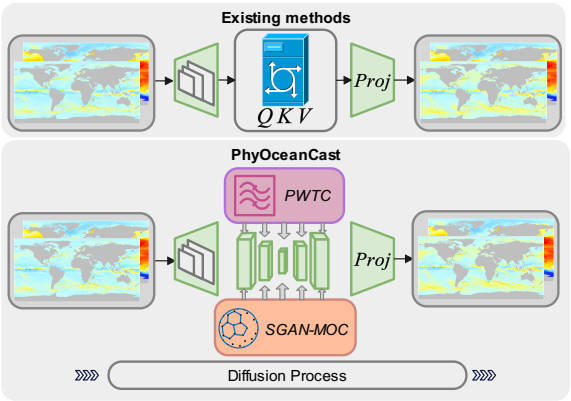
Qixiu Li[1], Xiang Zhu[1,*], Xiaoyong Li[1,*], Xiaolong Xu[2]
[1]National University of Defense Technology, China
[2]School of Software, Nanjing University of Information Science and Technology, China
{zhuxiang, sayingxmu}@nudt.edu.cn



(a) **Stochastic Forecasting vs Deterministic Forecasting**



(b) **RMSE and Pearson for Lead Time 30 day Forecasting**



(c) **PhyOceanCast vs Existing Methods**

Figure 1. (a) Comparison of deterministic and stochastic ocean forecasting paradigms. Probabilistic predictions represent natural complexity beyond finite-grid initial states. (b) The results at 30 day lead time highlight PhyOceanCast's superior forecasting capability. (c) Compared to existing methods, PhyOceanCast is a physics-informed diffusion framework strengthened by the coupling of SGAN-MOC for spherical-aware cross-variable interactions and PWTC for multi-scale temporal dynamics decomposition.

Abstract

Ocean dynamics drive global climate patterns and extreme weather events, making accurate spatiotemporal forecasting essential for climate monitoring and marine operations. Traditional Global Ocean Forecasting Systems (GOFSs) offer high accuracy predictions, yet remain computationally expensive and fail to fully leverage growing historical data. Recent deep learning models have achieved notable success, but still face three fundamental challenges: (1) they homogenize ocean variables despite strong physical coupling via equation-of-state relationships; (2) they neglect spherical geometry, resulting in severe distortions at high latitudes; and (3) they struggle to model multi-scale temporal dynamics. We introduce PhyOceanCast, a physics-informed diffusion model that overcomes these limitations through two key innovations. First, the Spherical Graph Attention Network for Multi-scale Ocean Coupling (SGAN-MOC) preserves spherical topology while enabling cross-variable interactions via heterogeneous encoding and k-hop-constrained attention. Second, the Physics-Informed Wavelet Temporal Coherence (PWTC) module that decomposes ocean dynamics across multiple scales with advection-diffusion constraints. Phy-

*Corresponding author.

OceanCast forecasts 145 ocean variables, including temperature, salinity, and velocity fields, across 36 depth levels plus sea surface height. Extensive experiments demonstrate superior performance over diffusion, transformer, and hybrid baselines, promising a new paradigm for global ocean canonical variable forecasting.

1. Introduction

Oceans regulate Earth's climate by storing over 90% of excess heat [35, 54] and driving weather patterns through complex circulation systems. Accurate ocean forecasting underpins maritime safety [33], disaster warning systems [31], and the blue economy [40], requiring predictions of multiple interconnected variables across spatiotemporal scales, from mesoscale eddies to global circulation.

Ocean variable forecasting constitutes a challenging spatiotemporal prediction task: given historical observations of temperature, salinity, velocity fields, and other variables, predict their future evolution. Deep learning has revolutionized weather prediction, with GraphCast [25] achieving a thousand-fold computational speedup while outperforming traditional numerical methods on 90% of verification targets. Ocean-specific models demonstrate similar promise—WenHai [7] produces eddy-resolving forecasts with 6–10% improved accuracy over GOFSs. Ocean systems exhibit fundamental stochasticity across multiple orders of magnitude [17], with chaotic evolution and irreducible uncertainty that deterministic models cannot capture [24, 43]. Ensemble methods extend predictive skill by quantifying uncertainty, demonstrating the necessity of probabilistic approaches for operational forecasting (Fig. 1 (a)). Yet ocean dynamics present unique challenges [1, 50].

Current spatiotemporal prediction methods fail to meet the core requirements of ocean modeling. Standard architectures assume isotropic dynamics; equirectangular projections introduce area distortion near the poles, and planar convolutions fail to handle longitude wraparound [11, 36, 46]. Many methods ignore the coupling and co-evolution among multiple variables governed by the equation of state, undermining physical consistency across variables [26]. Single-scale temporal modeling is inadequate for dynamics spanning daily internal waves to seasonal–interannual variability [44]. Consequently, existing approaches either oversmooth high-frequency signals, obscuring critical energy transfers, or lack mechanisms for cross-scale information propagation, leading to violations of conservation laws.

Recent advances address individual limitations without achieving comprehensive solutions. GraphCast's [25] icosahedral mesh handles spherical topology, but it applies an isotropic treatment that cannot model the scale separation between horizontal and vertical motions or the vortical motions and stratification induced by latitudinal dynamics [45]. Pangu-Weather [3] customizes a multivariate processing logic yet provides no explicit coupling mechanisms tailored to ocean-specific constraints. Diffusion models generate probabilistic forecasts with realistic variability [6, 28, 37, 48, 51]; however, standard formulations cannot enforce physical consistency. No existing framework simultaneously preserves spherical geometry under anisotropic dynamics, enables physically consistent cross-depth variable coupling, and represents multi-scale temporal evolution while maintaining conservation laws (Fig. 1 (c)).

To address these limitations, our contributions are:

- We propose PhyOceanCast, a physics-informed diffusion model that unifies probabilistic global ocean forecasting with domain-specific constraints.
- We introduce a Spherical Graph Attention Network for Multi-scale Ocean Coupling (SGAN-MOC) that enables cross-variable interactions through heterogeneous node encoding, performs adaptive spherical mapping, and maintains information fusion through k-hop constrained attention mechanisms.
- We develop a Physics-informed Wavelet Temporal Coherence module (PWTC) that performs depth-variable coupling, decomposes dynamics across frequency scales, and applies learnable advection-diffusion operators to enforce conservation constraints. We generate physically plausible ensemble forecasts via iterative refinement guided by these representations.
- Extensive evaluation demonstrates PhyOceanCast's state-of-the-art performance, achieving a 13.7% RMSE improvement for global forecasting of 145 variables at a 30-day lead time (Fig. 1 (b)). Ablation studies validate each module's effectiveness, while case studies confirm that our model delivers highly accurate predictions with strong physical consistency.

2. Related Work

Spatiotemporal prediction & Ocean Variable Coupling. Spatiotemporal prediction has evolved from convolutional-recurrent hybrids to transformer-based architectures, yet fails to adequately model coupled oceanic variables. ConvLSTM [38] and ConvGRU [39] pioneered spatial-temporal integration, enabling direct processing of gridded ocean fields. However, their local receptive fields cannot capture large-scale teleconnections in ocean circulation systems. SimVP [15] enhanced inference speed through pure convolutional architectures, maintaining high accuracy via Inception-based modules, while transformer-based models [2, 6, 10, 14, 20] capture long-range dependencies. SwinLSTM [41] advanced global dependency modeling by integrating shifted window attention with recurrent structures. 3D-Geoformer [53] employs spatiotemporal-separated attention mechanisms to establish multivariate relationships unrestricted by spatial and temporal distance. Beyond deter-

ministic approaches, recent diffusion models [9, 19, 23, 32] have been applied to natural system modeling and prediction, including DiffCast [48] and SEEDS [27]. Despite these architectural advances, existing methods treat temperature, salinity, and velocity as independent variables, fundamentally violating the equation of state governing ocean dynamics. While WenHai [7] partially addresses coupling through bulk formula integration and tendency-based prediction, it lacks explicit cross-variable attention mechanisms necessary for modeling density-driven circulation and thermohaline processes that couple ocean layers across depth [26].

Spherical Geometry Preservation. Planar projections introduce catastrophic distortions for global ocean modeling, with equirectangular grids causing 5-fold area variations between equator and poles [13, 47]. Spherical CNNs [22, 36] maintain $\mathcal{SO}(3)$ equivariance through spectral convolutions in spherical harmonic space, eliminating projection artifacts entirely. GraphCast [25] demonstrates practical scalability using multi-resolution icosahedral meshes: learned edge features capture geodesic relationships, and hierarchical message passing connects local eddies to large-scale circulation. However, most spherical methods assume isotropic dynamics, failing to accommodate the ocean's inherent anisotropy, critically limiting accurate representation of baroclinic instabilities and vertical mixing processes [44, 52].

Multi-scale Dynamics Representation. Existing spatiotemporal prediction models incorporate dynamics decomposition. PhyDNet [16] embeds partial differential equation (PDE) constraints into the recurrent hidden state. Video Swin Transformer [29] hierarchically decomposes temporal patterns through a pyramid architecture, while SVQ [5] introduces vectorization techniques to spatiotemporal prediction. U-RNN [4] approximates complex fluid dynamics through pattern recognition. However, ocean dynamics fundamentally involve nonlinear cross-scale interactions: internal waves exhibit short-term oscillations [8], mesoscale eddies persist for months [18], and thermohaline circulation spans centuries [42]. Current architectures either smooth high-frequency variability or fail to propagate information across temporal scales, precluding accurate representation of the critical energy transfers that shape ocean circulation.

3. Task Definition

We formulate ocean variable forecasting as a probabilistic spatiotemporal prediction problem. Given a sequence of historical ocean states at times $t-2$ and $t-1$, denoted as $\mathbf{X}_{t-2}, \mathbf{X}_{t-1} \in \mathbb{R}^{V\times H\times W}$, the task is to predict the future ocean state $\mathbf{X}_t$, where $V = 145$ represents the total number of ocean variables (detailed in App. 7).

Following the diffusion-based forecasting paradigm, we model the conditional probability distribution $p(\mathbf{X}_t|\mathbf{X}_{t-1}, \mathbf{X}_{t-2})$ through a second order Markov assumption. The model learns to predict the residual representing the state transition as $r_t = \mathbf{X}_t - \mathbf{X}_{t-1}$, where r_t denotes the residual between consecutive time steps.

During training, we corrupt the target residual r_t with noise $\epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ to obtain the noisy residual $\tilde{r}_t = r_t + \sigma\epsilon$, and train a denoiser D_θ to recover the clean residual:

$$\mathcal{L} = \mathbb{E}_{\sigma,\epsilon}\Bigg[\lambda(\sigma)\sum_{v\in V}\sum_{i\in G} w_{v,d(v)}\cdot a_i \cdot \|D_\theta(\tilde{r}_t; \mathbf{X}_{t-2}, \mathbf{X}_{t-1}, \mathcal{F}, \tau_{t-2}, \tau_{t-1}, \sigma) - r_t\|^2\Bigg], \quad (1)$$

where $\mathcal{F}$ denotes external forcing terms, τ_{t-i} encodes the temporal position of state $\mathbf{X}_{t-i}$ within the annual cycle, $w_{v,d(v)}$ represents depth-dependent weights for variable v at depth $d(v)$ following oceanographic principles, a_i denotes the area weight for grid cell i accounting for latitude-dependent grid distortion, and $\lambda(\sigma)$ is the noise-level-dependent loss weight (detailed in App. 8).

4. The Proposed PhyOceanCast

4.1. Overall Framework

The architecture of the proposed PhyOceanCast is illustrated in Fig. 2. PhyOceanCast employs a physics-informed diffusion framework that predicts ocean state evolution through residual learning with a second order Markov rollout strategy. The model integrates two complementary modules: Spherical Graph Attention Network for Multi-Scale Ocean Coupling (SGAN-MOC) that preserves spherical topology with anisotropic characteristics while enabling physics-aware cross-variable interactions (Sec. 4.2), and Physics-Informed Wavelet Temporal Coherence (PWTC) that captures depth-coupled interactions while decomposing ocean dynamics across multiple temporal scales through advection-diffusion constraints (Sec. 4.3).

4.2. SGAN-MOC

1. Heterogeneous Variable Encoding. Ocean variables exhibit distinct physical characteristics and variability across depths and scales. To capture these heterogeneities, SGAN-MOC employs variable-specific encoding strategies that preserve the signatures of each oceanographic quantity.

Each ocean variable is processed through an independent encoder, which is mathematically formulated as:

$$\hbar_\theta^{(v,i)} = \mathcal{E}_v(\mathbf{X}_{t-i}^{(v)}) \quad v \in \{1, ..., V\}, i \in \{1, 2\}, \quad (2)$$

where $\mathcal{E}_v$ denotes the variable-specific encoder for variable v, and $\mathbf{X}_{t-i}^{(v)} \in \mathbb{R}^{H\times W}$ represents the spatial field of variable v at time $t-i$. The heterogeneous encoding $\hbar_\theta^{(v,i)}$ captures variable-specific features through cascaded convolutional layers with progressively increasing receptive fields,

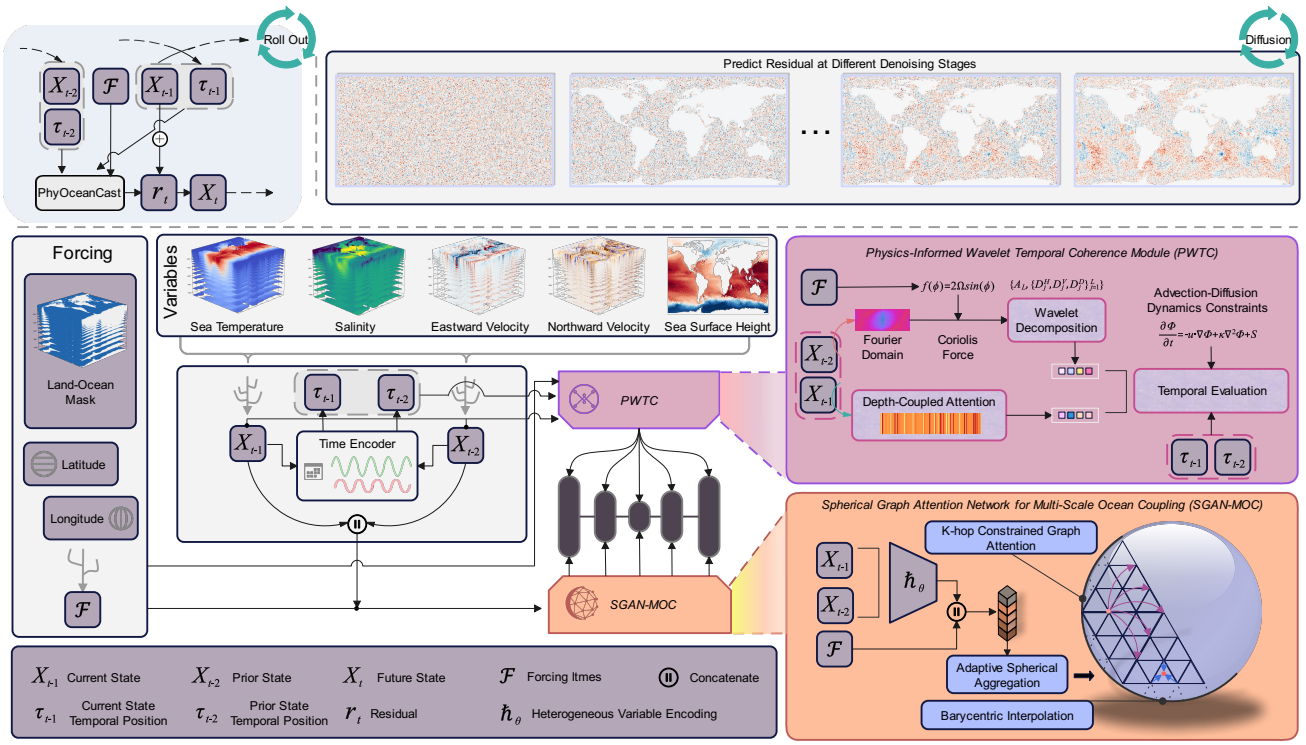


Figure 2. **PhyOceanCast architecture overview.** The model predicts ocean state evolution through residual learning with second order Markov roll-out strategy (top-left) and iterative diffusion refinement (top-right). Given previous states X_{t-1} and X_{t-2} as conditions with Time Encoder providing intra-annual temporal positioning, these inputs are fed into SGAN-MOC and PWTC modules: SGAN-MOC preserves spherical topology while enabling physics-aware cross-variable interactions, and PWTC enables depth-coupled variable interactions while decomposing ocean dynamics across multiple temporal scales with advection-diffusion constraints. The pipeline generates residual r_t through reverse denoising, producing future ocean state X_t.

enabling the extraction of multi-scale oceanic patterns from mesoscale eddies to basin-wide circulation.

To incorporate earth's spherical geometry and account for latitude-dependent dynamics, we augment the encoded features with a sinusoidal geographic positional encoding $\mathbf{PE}_{\text{geo}}^{(p)}$ for $p \in \text{lat}, \text{lon}$. This encoding is computed by concatenating sine and cosine transforms of each coordinate. It provides fine-grained positional cues for resolving mesoscale features and coarse-grained patterns essential for large-scale.

The final heterogeneous variable representation combines the encoded features with positional information:

$$\mathbf{F}_v = \hbar_\theta^{(v,1)} + \hbar_\theta^{(v,2)} + \alpha \cdot (\mathbf{PE}_{\text{geo}}^{\text{lat}} \oplus \mathbf{PE}_{\text{geo}}^{\text{lon}}), \quad (3)$$

where $\oplus$ denotes the addition operation, and α is a learnable scaling factor. This encoding scheme enables the model to learn variable-specific patterns while maintaining awareness of spatial context crucial for accurately representing ocean dynamics that vary significantly with latitude due to Coriolis effects and meridional heat transport [12].

2. Adaptive Spherical Aggregation. To preserve spherical topology while enabling efficient multi-scale processing, SGAN-MOC constructs an icosahedral mesh that approximates the earth's spherical surface without the distortions inherent in planar projections. Following GraphCast [25], we initialize a regular icosahedron and recursively subdivide it through multiple levels of refinement.

The icosahedral mesh refinement process is formulated:

$$\mathcal{M}^{(l+1)} = \{v' \mid v' = \text{proj}_{\mathbb{S}^2}(\frac{v_i + v_j}{2}), \forall (v_i, v_j) \in \mathcal{E}^{(l)}\} \cup \mathcal{M}^{(l)}, \quad (4)$$

where $\mathcal{E}^{(l)}$ denotes the edge set at refinement level l, and $\text{proj}_{\mathbb{S}^2}$ projects vertices (v_i, v_j) onto the unit sphere.

To establish connectivity between the regular grid and icosahedral mesh, we define the bidirectional mapping through geodesic proximity. For each grid point $\mathbf{p}_g$, we identify connected mesh nodes:

$$\Psi(\mathbf{p}_g) = \{\mathbf{m}_i \in \mathcal{M} : d_{geo}(\mathbf{p}_g, \mathbf{m}_i) < r\}, \quad (5)$$

where d_{geo} denotes the great circle distance and r represents the connectivity radius in radians.

The adaptive aggregation from grid features $\mathbf{F}_v$ (Eq. (3)) to mesh nodes employs:

$$\mathbf{H}_m = \frac{\sum_{g:m \in \Psi(\mathbf{p}_g)} \hat{w}_{gm} \cdot \mathbf{F}_v^{(g)}}{\sum_{g:m \in \Psi(\mathbf{p}_g)} \hat{w}_{gm}}, \quad (6)$$

where $\mathbf{F}v^{(g)}$ denotes the heterogeneous variable features at grid point g, and the adaptive weights $\hat{w}_{gm}$ are derived from geodesic distances. This adaptive aggregation scheme preserves the advantages of both representations: the regular grid maintains compatibility with observational data, whereas the icosahedral mesh enables uniform processing across latitudes. Subsequently, spherical-domain features undergo k-hop attention to facilitate feature interaction, followed by equal-area barycentric interpolation to aggregate features on the sphere, after which the results are mapped back to update the grid representation.

4.3. PWTC

1. Multi-scale Wavelet Decomposition. Ocean dynamics span multiple temporal scales from short-term internal waves [8] to seasonal-interannual climate oscillations [42]. PWTC employs a hierarchical decomposition strategy combining spectral analysis in the frequency domain with wavelet transforms to capture these multi-scale interactions while preserving energy cascades across scales.

Temporal scale separation. Given the temporal input sequence, the module performs spectral decomposition to isolate periodic ocean signals:

$$\hat{\mathcal{S}}(\omega_h, \omega_w) = \int \int \mathbf{X}_t(h, w) e^{-2\pi i(\omega_h h + \omega_w w)} dh dw, \quad (7)$$

where $\mathbf{X}_t \in \mathbb{R}^{C \times H \times W}$ represents the temporal ocean features at time t, with (h, w) are spatial coordinates over height H and width W.

Following spectral analysis, multi-scale wavelet decomposition is applied to the inverse-transformed signal:

$$\{\mathbf{A}_L, \{\mathbf{D}_j^H, \mathbf{D}_j^V, \mathbf{D}_j^D\}_{j=1}^L\} = \text{DWT3D}(\mathbb{F}^{-1}\{\widehat{\mathcal{S}}^t\}_{t=1}^T, \psi), \quad (8)$$

we jointly apply DWT3D to the hetero-temporal slices $\{\widehat{\mathcal{S}}^t\}_{t=1}^T$ from Eq. (7), where $\mathbb{F}^{-1}$ denotes the inverse Fourier transform, ψ represents the Daubechies-4 wavelet basis, $\mathbf{A}_L$ is the approximation coefficients at the coarsest scale L, and $\{\mathbf{D}_j^H, \mathbf{D}_j^V, \mathbf{D}_j^D\}$ are the horizontal, vertical, and diagonal detail coefficients at scale j.

Each scale is processed through dedicated networks:

$$\mathbf{H}_j = \mathcal{G}_j([\mathbf{A}_L, \mathbf{D}_j^H, \mathbf{D}_j^V, \mathbf{D}_j^D]), \quad (9)$$

where $\mathcal{G}_j$ learns scale-specific ocean dynamics across decomposition levels.

Coriolis force incorporation. The rotation of Earth fundamentally influences ocean dynamics through latitude-dependent Coriolis effects, which is computed as:

$$f(\phi) = 2\Omega \sin(\phi), \quad (10)$$

where $\Omega = 7.2921 \times 10^{-5}$ rad/s is Earth's rotation rate and ϕ is latitude. The Coriolis encoding is integrated as:

$$\mathbf{H}_{\text{coriolis}} = \mathbf{H}_j + \beta \cdot f(\phi) \odot \mathbf{M}_{\text{velocity}}, \quad (11)$$

where β is a learnable scaling factor and $\mathbf{M}_{\text{velocity}}$ identifies velocity components.

2. Depth-Coupled Attention Mechanism. Ocean vertical structure exhibits strong stratification with distinct dynamical regimes from mixed layer to deep ocean. PWTC incorporates a depth-coupled attention mechanism to model vertical mixing processes and inter-layer energy transfers that are crucial for heat transport and circulation dynamics.

Vertical mixing between adjacent layers is modeled through depthwise convolutions that capture the exchange processes across four depth-varying variables (temperature, salinity, and velocity components) over 36 depth levels:

$$\mathbf{M}_d = \text{Conv1D}_{\text{depth}}(\mathbf{X}_t^{(v,d)}), \quad (12)$$

where $\mathbf{X}_t^{(v,d)}$ represents variable v at depth d.

To capture non-local vertical interactions such as convective plumes and internal wave propagation, the module employs multi-head attention across depth levels. Each depth level is assigned a learnable embedding:

$$\mathbf{E}_d = \mathbf{e}_d + \text{PE}_{\text{depth}}(d) \quad d \in \{1, ..., 36\}, \quad (13)$$

where $\mathbf{e}_d$ is a trainable depth embedding and PE_{depth} encodes the exponential decay of variability with depth.

The depth-coupled attention computes cross-layer interactions. The formal calculation process is as follow:

$$\mathbf{A}_{ij} = \text{softmax}\left(\frac{(\mathbf{M}_i + \mathbf{E}_i)(\mathbf{M}_j + \mathbf{E}_j)^T}{\sqrt{d_k}}\right), \quad (14)$$

where i, j index depth levels and d_k is the key dimension. This attention pattern reveals the learned vertical coupling structure, showing strong interactions within the thermocline and between surface and subsurface layers.

The attended features aggregate information across depths while preserving layer-specific characteristics:

$$\mathbf{H}_{\text{depth}} = \mathbf{M} + \gamma \cdot \sum_{j=1}^{36} \mathbf{A}_{ij} \mathbf{M}_j, \quad (15)$$

where γ is a learnable gating parameter that controls the contribution of non-local vertical interactions.

This depth-coupled attention mechanism enables PWTC to capture critical vertical processes including wind-driven mixing in the surface layer, thermocline ventilation, and the vertical propagation of internal waves.

3. Advection-Diffusion Constraint Temporal Evolution. Ocean transport is fundamentally governed by the advection-diffusion equation, which describes how ocean properties evolve under fluid motion and mixing. PWTC incorporates these physical constraints directly into the temporal evolution module, ensuring predictions respect conservation laws while enabling cross-scale feature fusion.

The module receives multi-scale features from both wavelet decomposition and depth-coupled attention, which are combined through element-wise addition: $\mathbf{H}_{\text{combined}} = \mathbf{H}_{\text{coriolis}} + \mathbf{H}_{\text{depth}}$. The combined features are then processed by a stack of ConvGRU blocks equipped with temporal positional encodings derived from τ_{t-1} and τ_{t-2}, while the ConvGRU hidden states encode temporal evolution.

To enforce advection-diffusion physics, the module models the governing equation, which can be expressed as:

$$\frac{\partial \Phi}{\partial t} = \underbrace{-\mathbf{u} \cdot \nabla \Phi}_{\text{advection}} + \underbrace{\kappa \nabla^2 \Phi}_{\text{diffusion}} + \underbrace{S}_{\text{source/sink}}, \tag{16}$$

where $\mathbf{u}$ is the velocity field, κ is the diffusivity coefficient, and S represents sources and sinks.

Leveraging the ConvGRU hidden states $\mathbf{H}_{\text{ConvGRU}}$, we learn operators corresponding to the advection, diffusion, and source/sink terms of the advection–diffusion equation, yielding physics-informed features $\mathbf{F}_{\text{phys}}$. We implement these operators as learnable convolutional kernels and adaptively weight them to accommodate different ocean regimes. After applying physics constraints, pyramid-scale fusion generates the $\mathbf{F}_{\text{PWTC}}$, and a subsequent linear layer produces the final output, product a comprehensive temporal representation that captures both fast surface dynamics and slow abyssal circulation.

5. Experiments

Datasets. We conduct experiments on the GLORYS12V1 reanalysis dataset [21], which we resample from its native 1/12° resolution to 1°. From this dataset, we select five critical ocean variables: sea surface height (zos) as a surface variable, and four depth-resolved variables including potential temperature (thetao), salinity (so), eastward velocity (uo), and northward velocity (vo) across 36 depth levels from surface to 1062m. We use data from 1993-2018 for training, 2019 for validation, and 2020 for testing. Detailed in App. 7

Implementation Details. We train each model justly on 6 NVIDIA H800 GPUs with a total batch size of 6 for 1K epochs. We employ the AdamW optimizer [30] with an initial learning rate of $1e^{-3}$ and weight decay of $1e^{-2}$. The learning rate uses warmup followed by square root decay. We apply dropout with probability 0.13 and maintain an EMA for stable inference. All experiments are implemented in PyTorch with DDP training, which takes 700–1500 total GPU-hours for the global models. More detailed in App. 10

Evaluation. For comprehensive evaluation, we adopt multiple metrics. Root Mean Square Error (RMSE) measures point-wise accuracy, while latitude-weighted accuracy (ACC_{lat}) accounts for spherical geometry (detailed in App. 9). Pearson correlation captures the linear relationship between predictions and ground truth, and Structural Similarity Index (SSIM) evaluates spatial pattern preservation.

Quantitative Results. We compare our method with several state-of-the-art spatiotemporal prediction and weather forecasting models: ConvLSTM [38], ConvGRU [39], SimVP [15], Video Swin Transformer [29], SwinLSTM [41], 3D-Geoformer [53], GraphCast [25], DiffCast [48], MambaOut [49], SVQ [5] and U-RNN [4]. These baselines cover various architectural paradigms. As shown in Tab. 1, PhyOceanCast achieves superior performance across all forecasting tasks at different lead times. Video Swin Transformer obtains the most second-best results, validating why many current networks [7] adopt it as their backbone model. However, such deterministic models cannot model stochastic natural processes from limited initial conditions, and planar convolutions cannot accommodate International Date Line discontinuities [11, 36, 46]. While U-RNN, GraphCast and 3D-Geoformer are specifically tailored for environmental forecasting tasks, they still exhibit error accumulation at 30-day lead time. In contrast, benefiting from PhyOceanCast's progressive denoising optimization capability, our model maintains robust performance. Notably, while DiffCast employs a diffusion model as its backbone, it shows substantial errors without physics guidance, further highlighting PhyOceanCast's ability to capture physical ocean dynamics. The comparison of skill scores are detailed in App. 12.

As shown in Fig. 3, for the 10-day lead time, RMSE profiles across depth indicate that PhyOceanCast accurately captures thermocline fluctuations in thetao, salinity variability driven by precipitation–evaporation, and Ekman layer changes in horizontal velocity components (uo, vo).

Qualitative Results. Fig. 5 presents the qualitative comparison of thetao predictions at 1.54m depth for October 8, 2020, with a 7-day lead time. Each panel displays the global ocean temperature field with a magnified inset (green box) highlighting the equatorial circulation region. SVQ benefits from vectorization techniques and clearly identifies temperature gradients, but fails to capture the underlying physical patterns in large-scale ocean motions, producing fragmented and physically inconsistent predictions. GraphCast and 3D-Geoformer capture large-scale motions but smooth fine-scale details. In contrast, PhyOceanCast successfully preserves both large-scale circulation patterns and fine-scale oceanic features, maintaining sharp temperature gradients and coherent eddy structures that are consistent with ocean dynamics. SGAN-MOC maintains physical consistency of motions across latitudes, PWTC performs scale decomposition of oceanic phenomena while enforcing diffusion-advection constraints, fitting temporal evolution. This physics-constrained approach, combined with the diffusion model's denoising optimization capability, further enhances PhyOceanCast's ability to capture physical ocean dynamics.

Case Studies. Case studies demonstrate our model's ability to learn physical ocean patterns. As shown in Fig. 4c, the visualization depicts the Eastern Mediterranean during an

Table 1. **Quantitative comparison on global ocean forecasting.** Best results are in **bold**, second best are underlined.

Method	Venue	Lead Time 3 day Avg.					Lead Time 7 day Avg.				
		RMSE↓	MAE↓	ACC_{lat}↑	Pearson↑	SSIM↑	RMSE↓	MAE↓	ACC_{lat}↑	Pearson↑	SSIM↑
ConvLSTM [38]	[NeurIPS'15]	0.6940	1.8010	0.9669	0.9811	0.9808	1.0735	3.0096	0.9318	0.9615	0.9561
ConvGRU [39]	[NeurIPS'17]	0.6534	1.7488	0.9733	0.9846	0.9833	1.0261	2.9615	0.9415	0.9667	0.9600
SimVP [15]	[CVPR'22]	0.5503	1.4909	0.9834	0.9902	0.9893	0.9024	2.6636	0.9578	0.9754	0.9696
SwinLSTM-B [41]	[ICCV'23]	0.7005	1.8150	0.9694	0.9822	0.9818	1.0989	3.0603	0.9354	0.9631	0.9576
SwinLSTM-D [41]	[ICCV'23]	0.6902	1.7908	0.9685	0.9819	0.9814	1.0793	3.0356	0.9309	0.9610	0.9568
MambaOut-T [49]	[CVPR'25]	0.6942	1.8013	0.9668	0.9809	0.9807	1.0739	3.0105	0.9314	0.9613	0.9558
MambaOut-S [49]	[CVPR'25]	0.6940	1.8011	0.9670	0.9812	0.9809	1.0737	3.0104	0.9315	0.9614	0.9560
MambaOut-B [49]	[CVPR'25]	0.6939	1.8010	0.9669	0.9811	0.9808	1.0736	3.0101	0.9317	0.9615	0.9561
Video Swin Transformer [29]	[CVPR'22]	0.5358	1.4399	0.9841	0.9907	0.9901	0.8899	2.5806	0.9604	0.9769	0.9713
3D-Geoformer [53]	[Science Advances'23]	0.5433	1.4687	0.9838	0.9904	0.9898	0.8863	2.6062	0.9606	0.9770	0.9714
SVQ [5]	[SIGKDD'25]	0.6575	2.0151	0.9780	0.9875	0.9868	1.2077	4.0585	0.9418	0.9668	0.9591
U-RNN [4]	[Journal of Hydrology'25]	0.6872	1.8038	0.9677	0.9677	0.9812	1.0658	3.0335	0.9320	0.9618	0.9565
GraphCast [25]	[Science'23]	0.6003	1.6219	0.9778	0.9870	0.9874	1.0509	3.0443	0.9418	0.9658	0.9648
DiffCast (10 members) [48]	[CVPR'25]	0.7325	2.1714	0.9645	0.9794	0.9766	0.9501	2.8989	0.9181	0.9526	0.9447
PhyOceanCast (Ours, 10 members)	[CVPR'26]	**0.4558 (+17.55%)**	**1.3098 (+9.93%)**	**0.9869 (+0.28%)**	**0.9912 (+0.05%)**	**0.9916 (+0.15%)**	**0.7205 (+23.01%)**	**2.1189 (+21.78%)**	**0.9647 (+0.42%)**	**0.9794 (+0.24%)**	**0.9767 (+0.54%)**

Method	Venue	Lead Time 15 day Avg.					Lead Time 30 day Avg.				
		RMSE↓	MAE↓	ACC_{lat}↑	Pearson↑	SSIM↑	RMSE↓	MAE↓	ACC_{lat}↑	Pearson↑	SSIM↑
ConvLSTM [38]	[NeurIPS'15]	1.2263	3.1505	0.8912	0.9394	0.9261	1.3443	2.8909	0.8561	0.9200	0.8976
ConvGRU [39]	[NeurIPS'17]	1.1965	3.1574	0.8956	0.9413	0.9295	1.3566	2.9885	0.8387	0.9079	0.8954
SimVP [15]	[CVPR'22]	1.1098	2.9723	0.9151	0.9506	0.9368	1.2787	2.8497	0.8617	0.9180	0.8979
SwinLSTM-B [41]	[ICCV'23]	1.2701	3.2260	0.8932	0.9415	0.9272	1.3948	2.9747	0.8546	0.9196	0.8978
SwinLSTM-D [41]	[ICCV'23]	1.2578	3.2388	0.8787	0.9322	0.9252	1.4144	3.0486	0.8262	0.9004	0.8946
MambaOut-T [49]	[CVPR'25]	1.2266	3.1519	0.8905	0.9390	0.9260	1.3451	2.8936	0.8541	0.9190	0.8975
MambaOut-S [49]	[CVPR'25]	1.2267	3.1523	0.8902	0.9388	0.9261	1.3453	2.8943	0.8532	0.9183	0.8974
MambaOut-B [49]	[CVPR'25]	1.2265	3.1516	0.8909	0.9392	0.9260	1.3452	2.8939	0.8550	0.9193	0.8973
Video Swin Transformer [29]	[CVPR'22]	1.0931	2.8794	0.9215	0.9549	0.9404	1.2632	2.7769	0.8713	0.9260	0.9034
3D-Geoformer [53]	[Science Advances'23]	1.0971	2.9416	0.9193	0.9533	0.9385	1.3228	2.9766	0.8622	0.9187	0.8929
SVQ [5]	[SIGKDD'25]	1.7927	5.5452	0.8655	0.9213	0.9016	2.5322	6.6601	0.7877	0.8696	0.8401
U-RNN [4]	[Journal of Hydrology'25]	1.2183	3.2009	0.8893	0.9381	0.9260	1.3414	2.9705	0.8463	0.9127	0.8958
GraphCast [25]	[Science'23]	1.3935	3.6417	0.8829	0.9297	0.9289	1.7577	3.8348	0.8177	0.8860	0.8857
DiffCast (10 members) [48]	[CVPR'25]	1.8806	6.5412	0.8339	0.8986	0.8942	4.8514	12.4560	0.6131	0.6922	0.7664
PhyOceanCast (Ours, 10 members)	[CVPR'26]	**0.9677 (+12.95%)**	**2.6639 (+8.08%)**	**0.9289 (+0.79%)**	**0.9564 (+0.15%)**	**0.9479 (+0.79%)**	**1.1109 (+13.70%)**	**2.5139 (+10.46%)**	**0.9023 (+3.34%)**	**0.9431 (+1.81%)**	**0.9293 (+2.78%)**

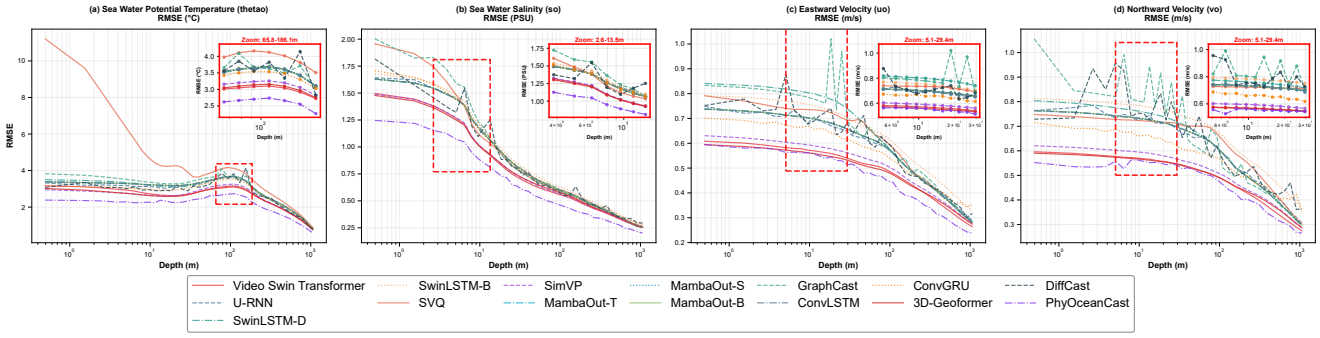
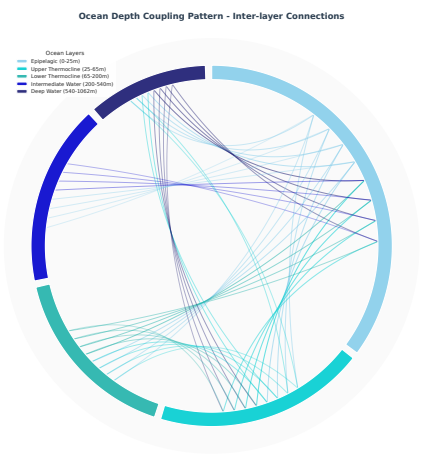


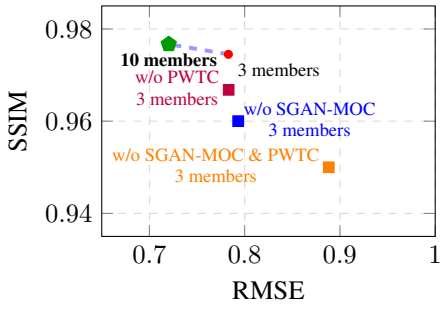
Figure 3. **Depth-resolved RMSE comparison across different ocean variables at lead time 10 day.** we evaluate PhyOceanCast against baseline models for (a) sea water potential temperature (thetao, °C), (b) sea water salinity (so, PSU), (c) eastward velocity (uo, m/s), and (d) northward velocity (vo, m/s) as a function of depth (log scale).

extreme marine heatwave, and PhyOceanCast accurately reproduces the spatial extent and intensity of the warm water mass. Furthermore, we select the Brazil–Malvinas Confluence region (black rectangle), one of the most dynamically complex oceanic areas owing to the collision of the warm Brazil Current with the cold Malvinas Current. Figs. 4d and 4e show forecast-error maps for zos; error magnitudes remain below 0.2 m, demonstrating PhyOceanCast's robustness in complex ocean dynamics. Fig. 4f presents the temporal evolution of the spatial correlation coefficient between PhyOceanCast predictions and ground truth over a 20 day forecast horizon. The correlation stays above 0.6 until day 18, indicating strong forecast skill retention.

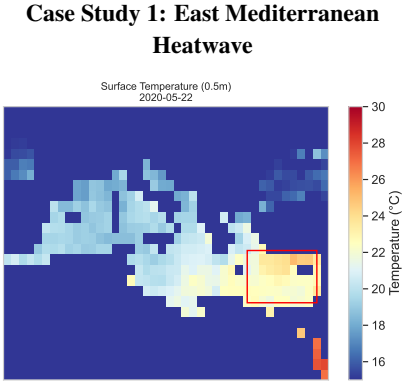
Abliatioin and Hyperparameters Studies. We conduct extensive ablation and hyperparameter studies, demonstrating the model's capacity to capture the stochastic nature of ocean dynamics and validating the effectiveness of each module. As shown in Fig. 4a, the results reveal that our PWTC captures inter-layer coupling patterns, which is crucial for accurately predicting states across different stratified layers, including thermocline and halocline. As shown in Fig. 4b, we evaluate model performance with varying ensemble member sizes and conduct ablation experiments. Multiple ensemble members effectively quantify prediction uncertainty. The ablation studies demonstrate that SGAN-MOC effectively models relationships between heterogeneous variables, while PWTC, through physics-based constraints and scale decomposition, enhances the physical consistency of forecasts over continuous time. More detailed analyses in the App. 11.



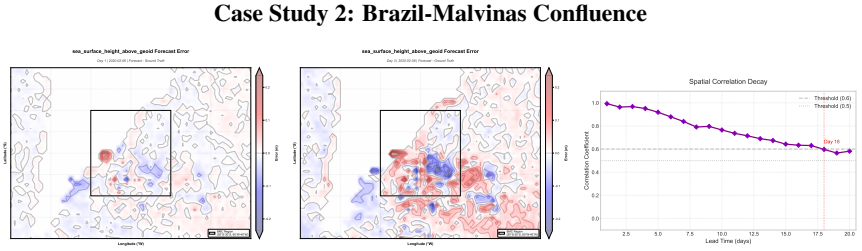
(a) Visualization of depth-coupled attention weights across ocean layers. The chord diagram illustrates inter-layer coupling patterns learned by the Depth-Coupled Attention mechanism in PWTC module. The learned attention patterns reveal strong cross-layer interactions, particularly between thermocline regions and adjacent layers, demonstrating the model's ability to capture vertical mixing processes and stratification dynamics essential for accurate ocean state prediction.



(b) Ablation and hyperparameter studies at lead time 7D.

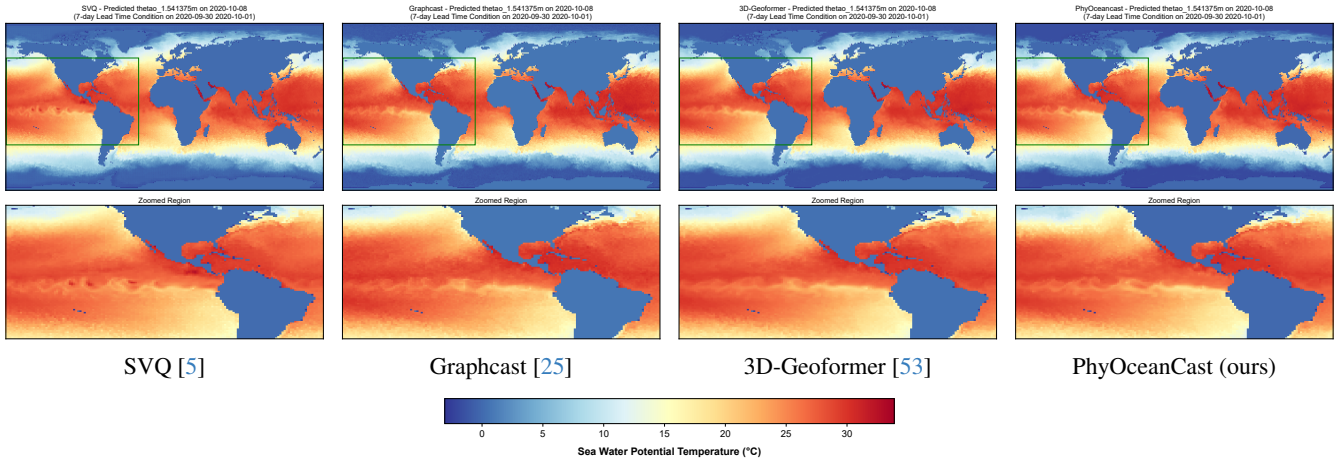


(c) Surface temperature prediction



(d) zos error - Lead Time 1D. (e) zos error - Lead Time 3D. (f) Spatial Correlation Decay

Figure 4. **Comprehensive analysis of PhyOceanCast.** (a) Ocean depth coupling pattern showing inter-layer connections across five depth zones. (b) Ablation and hyperparameter studies at lead time 7 day. Case Study 1: East Mediterranean heatwave event. (c) Surface temperature prediction on 2020-05-22. Case Study 2: Brazil-Malvinas Confluence region. (d-e) zos prediction errors at lead time 1 day and 3 day. (f) Long-range forecast skill retention in the Brazil-Malvinas Confluence region.



SVQ [5] Graphcast [25] 3D-Geoformer [53] PhyOceanCast (ours)

Figure 5. **Qualitative comparison** of thetao on 2020-10-08 at 1.54m with 7-day lead time (conditioned on 2020-09-30 and 2020-10-01).

6. Conclusion and Limitations

We propose PhyOceanCast, a physics-informed diffusion model for global ocean forecasting. PhyOceanCast models ocean dynamics on the sphere using a Spherical Graph Attention Network with heterogeneous encoders, enabling multi-scale ocean coupling. It further incorporates a Physics-Informed Wavelet Temporal Coherence module to analyze multi-scale dynamics and promote temporal coherence in forecasts. Extensive experiments across multiple lead times validate the accuracy of PhyOceanCast, offering a new approach to forecasting ocean state variables.

The presence of landmasses introduces spatial discontinuities in ocean states, reducing the number of effective training samples. In the future, we will focus on accelerating training and extracting high quality features from data.

Acknowledgments

We thank all the anonymous reviewers for their constructive comments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 42275170, 42305170, 62402512), the Science and Technology Innovation Program of Hunan Province (2022RC3070), and the Natural Science Foundation of Hunan Province (Grant No. 2023JJ40678).

References

[1] Matthew Archer, Jinbo Wang, Patrice Klein, Gerald Dibarboure, and Lee-Lueng Fu. Wide-swath satellite altimetry unveils global submesoscale ocean dynamics. *Nature*, 640 (8059):691–696, 2025. 2

[2] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6836–6846, 2021. 2

[3] 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(7970):533–538, 2023. 2

[4] Xiaoyan Cao, Baoying Wang, Yao Yao, Lin Zhang, Yanwen Xing, Junqi Mao, Runqiao Zhang, Guangtao Fu, Alistair GL Borthwick, and Huapeng Qin. U-rnn high-resolution spatiotemporal nowcasting of urban flooding. *Journal of Hydrology*, 659:1–43, 2025. 1, 3, 6, 7

[5] Chao Chen, Tian Zhou, Yanjun Zhao, Hui Liu, Rong Jin, and Liang Sun. Does vector quantization fail in spatio-temporal forecasting? exploring a differentiable sparse soft-vector quantization approach. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 143–154, 2025. 1, 3, 6, 7, 8

[6] Lei Chen, Fei Du, Yuan Hu, Zhibin Wang, and Fan Wang. Swinrdm: integrate swinrnn with diffusion model towards high-resolution and high-quality weather forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 322–330, 2023. 2

[7] Yingzhe Cui, Ruohan Wu, Xiang Zhang, Ziqi Zhu, Bo Liu, Jun Shi, Junshi Chen, Hailong Liu, Shenghui Zhou, Liang Su, et al. Forecasting the eddying ocean with a deep neural network. *Nature Communications*, 16(1):1–11, 2025. 2, 3, 6

[8] Giovanni Dematteis, Arnaud Le Boyer, Friederike Pollmann, Kurt L Polzin, Matthew H Alford, Caitlin B Whalen, and Yuri V Lvov. Interacting internal waves explain global patterns of interior ocean mixing. *Nature Communications*, 15 (1):1–15, 2024. 3, 5

[9] Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped langevin diffusion. In *Proceedings of the International Conference on Learning Representations*, pages 13643–13696, 2022. 3

[10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the International conference on learning representations*, pages 611–631, 2021. 2

[11] 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*, pages 52–68, 2018. 2, 6

[12] Gaël Forget and David Ferreira. Global ocean heat transport dominated by heat export from the tropical pacific. *Nature Geoscience*, 12(5):351–354, 2019. 4

[13] Baylor Fox-Kemper, Alistair Adcroft, Claus W Böning, Eric P Chassignet, Enrique Curchitser, Gokhan Danabasoglu, Carsten Eden, Matthew H England, Rüdiger Gerdes, Richard J Greatbatch, et al. Challenges and prospects in ocean circulation models. *Frontiers in Marine Science*, 6: 1–29, 2019. 3

[14] Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Bernie Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems*, 35:25390–25403, 2022. 2

[15] 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*, pages 3170–3180, 2022. 2, 6, 7

[16] Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11474–11484, 2020. 3

[17] Ian A Hatton, Ryan F Heneghan, Yinon M Bar-On, and Eric D Galbraith. The global ocean size spectrum from bacteria to whales. *Science advances*, 7(46):1–12, 2021. 2

[18] Qingyou He, Weikang Zhan, Ming Feng, Yankun Gong, Shuqun Cai, and Haigang Zhan. Common occurrences of subsurface heatwaves and cold spells in ocean eddies. *Nature*, 634(8036):1111–1117, 2024. 3

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3

[20] Yuan Hu, Lei Chen, Zhibin Wang, and Hao Li. Swinvrnn: A data-driven ensemble forecasting model via learned distribution perturbation. *Journal of Advances in Modeling Earth Systems*, 15(2):1–20, 2023. 2

[21] Lellouche Jean-Michel, Greiner Eric, Bourdallé-Badie Romain, Garric Gilles, Melet Angélique, Drévillon Marie, Bricaud Clément, Hamon Mathieu, Le Galloudec Olivier, Regnier Charly, et al. The copernicus global 1/12 oceanic and sea ice glorys12 reanalysis. *Frontiers in Earth Science*, 9: 1–27, 2021. 6, 1

[22] Chiyu Max Jiang, Jingwei Huang, Karthik Kashinath, Prabhat, Philip Marcus, and Matthias Niessner. Spherical CNNs on unstructured grids. In *Proceedings of the International Conference on Learning Representations*, pages 1204–1219, 2019. 3

[23] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative

models. *Advances in neural information processing systems*, 35:26565–26577, 2022. 3

[24] Robert E Kopp, Michael Oppenheimer, Jessica L O'Reilly, Sybren S Drijfhout, Tamsin L Edwards, Baylor Fox-Kemper, Gregory G Garner, Nicholas R Golledge, Tim HJ Hermans, Helene T Hewitt, et al. Communicating future sea-level rise uncertainty and ambiguity to assessment users. *Nature climate change*, 13(7):648–660, 2023. 2

[25] 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. 1, 2, 3, 4, 6, 7, 8

[26] F. Li, M. S. Lozier, S. Bacon, A. S. Bower, and C. Zhou. Subpolar north atlantic western boundary density anomalies and the meridional overturning circulation. *Nature Communications*, 12(1):1–9, 2021. 2, 3

[27] Lizao Li, Robert Carver, Ignacio Lopez-Gomez, Fei Sha, and John Anderson. Generative emulation of weather forecast ensembles with diffusion models. *Science Advances*, 10(13): 1–11, 2024. 3

[28] Yan Li, Xinjiang Lu, Yaqing Wang, and Dejing Dou. Generative time series forecasting with diffusion, denoise, and disentanglement. *Advances in Neural Information Processing Systems*, 35:23009–23022, 2022. 2

[29] Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. Video swin transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3202–3211, 2022. 1, 3, 6, 7

[30] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proceedings of the International Conference on Learning Representations*, pages 4061–4078, 2019. 6

[31] Nobuhito Mori, Kenji Satake, Daniel Cox, Katsuichiro Goda, Patricio A Catalan, Tung-Cheng Ho, Fumihiko Imamura, Tori Tomiczek, Patrick Lynett, Takuya Miyashita, et al. Giant tsunami monitoring, early warning and hazard assessment. *Nature Reviews Earth & Environment*, 3(9):557–572, 2022. 2

[32] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the International Conference on Machine Learning*, pages 8162–8171, 2021. 3

[33] Fernando S Paolo, David Kroodsma, Jennifer Raynor, Tim Hochberg, Pete Davis, Jesse Cleary, Luca Marsaglia, Sara Orofino, Christian Thomas, and Patrick Halpin. Satellite mapping reveals extensive industrial activity at sea. *Nature*, 625(7993):85–91, 2024. 2

[34] Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, et al. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025. 7

[35] Roland Séférian, Thomas Bossy, Thomas Gasser, Zebedee Nichols, Kalyn Dorheim, Xuanming Su, Junichi Tsutsui, and Yeray Santana-Falcón. Physical inconsistencies in the representation of the ocean heat-carbon nexus in simple climate models. *Communications Earth & Environment*, 5(1):1–10, 2024. 2

[36] Zhengyang Shen, Tiancheng Shen, Zhouchen Lin, and Jinwen Ma. Pdo-es2cnns: Partial differential operator based equivariant spherical cnns. In *Proceedings of the AAAI conference on artificial intelligence*, pages 9585–9593, 2021. 2, 3, 6

[37] Jimeng Shi, Bowen Jin, Jiawei Han, Sundararaman Gopalakrishnan, and Giri Narasimhan. Codicast: Conditional diffusion model for global weather forecasting with uncertainty quantification. In *Proceedings of the International Joint Conferences on Artificial Intelligence Organization*, pages 9853–9861, 2025. 2

[38] 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. *Advances in neural information processing systems*, 28:802–810, 2015. 2, 6, 7

[39] Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in neural information processing systems*, 30:5617–5627, 2017. 2, 6, 7

[40] U Rashid Sumaila, Melissa Walsh, Kelly Hoareau, Anthony Cox, Louise Teh, Patrízia Abdallah, Wisdom Akpalu, Zuzy Anna, Dominique Benzaken, Beatrice Crona, et al. Financing a sustainable ocean economy. *Nature communications*, 12(1): 1–11, 2021. 2

[41] 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*, pages 13470–13479, 2023. 1, 2, 6, 7

[42] Kaustubh Thirumalai, Terrence M Quinn, Yuko Okumura, Julie N Richey, Judson W Partin, Richard Z Poore, and Eduardo Moreno-Chamarro. Pronounced centennial-scale atlantic ocean climate variability correlated with western hemisphere hydroclimate. *Nature Communications*, 9(1):1–11, 2018. 3, 5

[43] Prasad G Thoppil, Sergey Frolov, Clark D Rowley, Carolyn A Reynolds, Gregg A Jacobs, E Joseph Metzger, Patrick J Hogan, Neil Barton, Alan J Wallcraft, Ole Martin Smedstad, et al. Ensemble forecasting greatly expands the prediction horizon for ocean mesoscale variability. *Communications Earth & Environment*, 2(1):1–9, 2021. 2

[44] Hans van Haren. Deep-ocean inertial subrange small bandwidth coherence and ozmidov-frequency separation. *Physics of Fluids*, 31(6), 2019. 2, 3

[45] Tiffany C Vance, Thomas Huang, and Kevin A Butler. Big data in earth science: Emerging practice and promise. *Science*, 383(6688):1–9, 2024. 2

[46] Jonathan A Weyn, Dale R Durran, and Rich Caruana. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9):1–17, 2020. 2, 6

[47] S. Xu, B. Wang, and J. Liu. On the use of schwarz–christoffel conformal mappings to the grid generation for global ocean models. *Geoscientific Model Development*, 8(10):1337–1373, 2015. 3

[48] Demin Yu, Xutao Li, Yunming Ye, Baoquan Zhang, Chuyao Luo, Kuai Dai, Rui Wang, and Xunlai Chen. Diffcast: A

unified framework via residual diffusion for precipitation nowcasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27758–27767, 2024. 2, 3, 6, 7

[49] Weihao Yu and Xinchao Wang. Mambaout: Do we really need mamba for vision? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4484–4496, 2025. 6, 7

[50] Xiaolong Yu, Roy Barkan, and Alberto C Naveira Garabato. Intensification of submesoscale frontogenesis and forward energy cascade driven by upper-ocean convergent flows. *Nature Communications*, 15(1):1–10, 2024. 2

[51] Shiqi Zhang, Pan Mu, Cheng Huang, Jinglin Zhang, and Cong Bai. Tc-diffuser: Bi-condition multi-modal diffusion for tropical cyclone forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1120–1128, 2025. 2

[52] Mengnan Zhao, Rui M Ponte, and Thierry Penduff. Global-scale random bottom pressure fluctuations from oceanic intrinsic variability. *Science Advances*, 9(29):1–5, 2023. 3

[53] Lu Zhou and Rong-Hua Zhang. A self-attention–based neural network for three-dimensional multivariate modeling and its skillful enso predictions. *Science Advances*, 9(10):1–11, 2023. 1, 2, 6, 7, 8

[54] Chenyu Zhu, Saray Sanchez, Zhengyu Liu, Peter U Clark, Chengfei He, Lingfeng Wan, Jiuyou Lu, Chenguang Zhu, Lingwei Li, Shaoqing Zhang, et al. Enhanced ocean heat storage efficiency during the last deglaciation. *Science Advances*, 10(38):1–9, 2024. 2