OA-BRIDGE: ONE-STEP COUPLED OCEAN-ATMOSPHERE FORECASTING WITH A VISION–LANGUAGE MODEL

Anonymous authors
Paper under double-blind review

ABSTRACT

Atmosphere–ocean coupling is a fundamental physical process of the Earth system, yet most AI-based forecasting models treat the atmosphere and ocean as independent components. Learning a unified forecasting model is challenging because atmospheric and oceanic variables evolve on substantially different timescales, while straightforward joint training can cause negative transfer and degrade the predictive skill of otherwise strong component models. We propose OA-Bridge, the first VLM-based framework for explicitly coupled global atmosphere–ocean forecasting at daily resolution. Given the current atmospheric and oceanic states and a specified lead time, OA-Bridge directly predicts their joint future states without recursively applying a fixed-step transition model. A shared frozen Qwen2.5-VL backbone jointly processes the two representations conditioned on the requested lead time. Component-specific encoders, output adapters, and decoders preserve specialized representations, while a coupling adapter uses bidirectional cross-attention to exchange information between the two systems. Experiments from 1 to 60 days against a jointly trained FuXi baseline show that OA-Bridge substantially improves long-range ocean RMSE and ACC across depths, extreme-event skill, and SSH-derived geostrophic currents, while also improving atmospheric skill at extended lead times.

1 INTRODUCTION

The atmosphere and ocean form an intrinsically coupled system, continuously exchanging momentum, heat, and moisture across the air–sea interface. Their interactions shape weather across a broad range of timescales, making accurate representation of their joint evolution essential for reliable prediction and weather-risk mitigation (He et al., 2025; Huang et al., 2025a). Physics-based coupled numerical models have long served as the primary tool for simulating these interactions and forecasting the atmosphere–ocean system. Their computational cost, however, increases rapidly with spatial resolution and ensemble size (Hewitt et al., 2017; Luo et al., 2026).

AI-based models have recently emerged as computationally efficient alternatives, achieving competitive predictive skill for atmospheric and oceanic forecasting at substantially lower inference costs (Pathak et al., 2022; Bi et al., 2023; Chen et al., 2023b; Lam et al., 2023; Huang et al., 2025a; Niu et al., 2026). Progress in atmospheric and oceanic forecasting has nevertheless occurred largely in isolation: most AI models forecast either the atmosphere or the ocean without explicitly representing their bidirectional interactions. As a result, they cannot fully capture the joint evolution of the coupled system.

This limitation has motivated recent efforts to develop explicitly coupled AI models for medium-range and subseasonal-to-seasonal (S2S) forecasting. Ola (Wang et al., 2024a) couples independently trained atmospheric and oceanic models through predefined exchanges of surface variables and adopts component-specific update intervals, whereas FengShun-CSM (Zhou et al., 2025) introduces learned latent-feature exchange between separate component branches at each daily autoregressive step. Although these approaches retain specialized representations for the two components, they model coupled evolution through autoregressive rollouts with prescribed update intervals. AIFS (Hahner et al., 2026) instead uses a shared backbone to jointly advance a unified atmospheric

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.

and surface-ocean state at six-hour intervals. However, it remains tied to fixed-step recursive forecasting, and preserving atmospheric skill requires careful balancing of the marine and atmospheric training objectives.

Together, these approaches highlight a central challenge in unified atmosphere–ocean forecasting: how to represent the coupled evolution of two systems that operate on substantially different timescales. The atmosphere varies rapidly, whereas many oceanic processes evolve more slowly and exhibit stronger persistence. A unified model must therefore exchange coupling-relevant information without erasing the specialized representations required by each component, while accommodating their distinct evolution across forecast horizons.

To meet these requirements, we propose OA-Bridge, a VLM-based framework for explicitly coupled global atmosphere–ocean forecasting at daily resolution. OA-Bridge combines component-specific representations and explicit feature exchange with direct multi-horizon prediction, mapping current atmospheric and oceanic states to their joint future states at a requested lead time in one forward pass, without fixed-step recursion. Component-specific encoders extract separate representations, which the coupling adapter exchanges through residual bidirectional cross-attention and projects into the image-token embedding space of a shared frozen Qwen2.5-VL backbone (Bai et al., 2025). The backbone jointly processes these features with a prompt specifying the target lead time. Component-specific output adapters and decoders reconstruct the atmospheric and oceanic forecast fields.

Our contributions are summarized as follows:

- We introduce OA-Bridge, to our knowledge the first VLM-based framework for explicitly coupled global atmosphere–ocean forecasting at daily resolution. It combines a shared frozen Qwen2.5-VL backbone with component-specific encoders, adapters and decoders, using bidirectional cross-attention to exchange information between atmospheric and oceanic representations.

- We formulate coupled forecasting as lead-time-conditioned direct prediction, using a single model across 1–60-day horizons without fixed-step recursion. Latency per target horizon remains nearly constant, with a $15.83\times$ speedup over the autoregressive baseline at day 60.

- Experiments from 1 to 60 days show that OA-Bridge improves ocean RMSE, ACC, extreme-event skill, and SSH-derived geostrophic currents over a jointly trained FuXi baseline. Ocean gains extend across depths, alongside improvements in selected atmospheric variables at extended lead times.

2 Related Work

Separate atmosphere and ocean forecasting Data-driven global weather forecasting has advanced rapidly. Early work by Pathak et al. (2022) demonstrated efficient global forecasting through neural operators. Subsequent systems adopted Transformer- and graph-based architectures, including Pangu-Weather (Bi et al., 2023), GraphCast (Lam et al., 2023), FengWu (Chen et al., 2023a), and FuXi (Chen et al., 2023b). Building on the success of AI-based atmospheric forecasting, ocean forecasting has also begun to transition from traditional physics-based approaches, including ocean general circulation models such as the Hybrid Coordinate Ocean Model (HYCOM) (Chassignet et al., 2007) and the Nucleus for European Modelling of the Ocean (NEMO) (Gurvan et al., 2017), as well as data-assimilative operational systems such as PSY4 (Lellouche et al., 2018; Tonani et al., 2015), toward data-driven models. Recent data-driven models, many of which are trained on high-resolution ocean reanalyses such as GLORYS12 (Lellouche et al., 2021), substantially accelerate ocean forecasting but differ in how they incorporate atmospheric information. GLONET (El Aouni et al., 2025) and FuXi-Ocean (Niu et al., 2026) operate without explicit forecast-time atmospheric forcing, whereas XiHe (Wang et al., 2024c) and WenHai (Cui et al., 2025) incorporate atmospheric information into ocean prediction. Overall, these models predict one component, omitting the other or using it as external forcing, without jointly forecasting atmospheric and oceanic states.

Numerical and AI-based coupled forecasting Beyond atmosphere-conditioned ocean forecasting, recent studies have begun to jointly evolve atmospheric and oceanic states. Physics-based coupled forecasting includes regional systems (Warner et al., 2010; Zhang, 2015; Hu et al., 2019; Kim et al., 2024; Yu et al., 2025; He et al., 2025) and global operational systems such as the Climate Forecast System Version 2 (CFSv2) (Saha et al., 2014) and Seasonal Forecast System 5 (SEAS5) (Johnson et al., 2019). AI-based coupled models such as DLESyM (Cresswell-Clay et al.,

2025) and SamudrACE (Duncan et al., 2026) link separate atmospheric and oceanic components through predefined physical interfaces for century-scale climate simulations. At monthly resolution, Transformer-based models (Zhou & Zhang, 2023; Wang et al., 2024b; Zhang et al., 2024) and hybrid neural–physical systems (Du & Zhang, 2024; Du et al., 2025) target SST and ENSO forecasts up to two years ahead. Across medium-range to seasonal forecasting, Ola (Wang et al., 2024a) links separately trained neural operators through predefined bidirectional state exchange. FengShun-CSM (Zhou et al., 2025) introduces learned latent coupling between separate component branches, while AIFS (Hahner et al., 2026) further adopts a unified backbone for joint atmosphere–surface-ocean prediction. In contrast, OA-Bridge combines a shared frozen VLM backbone with component-specific adapters and decoders, using bidirectional cross-attention between encoded atmospheric and oceanic representations.

LLM-based and lead-time-aware forecasting Beyond coupling architecture, recent studies have explored alternatives to fixed-step forecasting for representing different forecast horizons. ARROW (Tian et al., 2026) adaptively selects forecast intervals within an autoregressive atmospheric forecasting framework. LangYa (Yang et al., 2025) uses a large language model (LLM) to encode lead-time information for direct ocean forecasting at lead times from 1 to 7 days. While LangYa predicts oceanic states using atmospheric forcing, OA-Bridge jointly forecasts atmospheric and oceanic states with a single shared lead-time-conditioned model over horizons from 1 to 60 days.

3 METHODOLOGY

This section presents OA-Bridge for coupled atmosphere–ocean forecasting. We first describe the data and forecasting task (Section 3.1), then introduce component-specific pretraining (Section 3.2) and coupled training with a frozen VLM conditioned on lead time (Section 3.3), and finally define the shared training objective (Section 3.4). The training pseudocode is provided in Appendix A.

3.1 DATA AND FORECASTING TASK

We use ERA5 atmospheric reanalysis (Hersbach et al., 2020) and GLORYS12 ocean reanalysis (Lellouche et al., 2021). Following the spatial resolution used in AviaSafe (Zhu et al., 2026) for ERA5 and FuXi-ONS (Huang et al., 2026) for GLORYS12, we resample both datasets to a common 1° global grid with $H = 180$ latitude points and $W = 360$ longitude points. We use daily states and standardize each channel using training-set statistics.

The atmospheric state contains $C_a = 72$ channels. Geopotential (`z`), temperature (`t`), zonal and meridional wind (`u`, `v`), and specific humidity (`q`) are represented at 13 pressure levels: 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa. The remaining seven channels are mean sea level pressure (`msl`), 2-m temperature and dewpoint temperature (`t2m`, `d2m`), 10-m wind components (`u10m`, `v10m`), surface solar radiation downwards (`ssrd`), and total precipitation (`tp`). Precipitation remains a prediction target, but its input channel is set to zero.

The oceanic state contains $C_o = 93$ channels: salinity (`ss`), temperature (`st`), and zonal and meridional currents (`su`, `sv`) at 23 depths (0, 2, 5, 7, 11, 15, 21, 29, 40, 55, 77, 92, 109, 130, 155, 186, 222, 266, 318, 380, 453, 541, and 643 m), together with sea surface height (`ssh`).

At each daily time step t_k, the state of component $s \in \{a, o\}$ is denoted by $X^s_{t_k} \in \mathbb{R}^{B\times 1\times C_s\times H\times W}$, with a and o indexing the atmosphere and ocean, respectively. Here, B is the batch size and the second axis contains a single input frame. Given a lead time of n days, we learn

$$\big(\widehat{X}^a_{t_{k+n}}, \widehat{X}^o_{t_{k+n}}\big) = \mathcal{F}_\theta\big(X^a_{t_k}, X^o_{t_k}, n\big), \tag{1}$$

where each output retains the channel and spatial dimensions of its input. As shown in Figure 1, training first learns one-day forecasts for each component independently, then couples the two branches to predict at the requested lead time.

3.2 ONE-STEP PRETRAINING

We first pretrain the atmospheric and oceanic encoders and decoders independently for one-day forecasting without the VLM backbone, providing component-specific representations and initializing

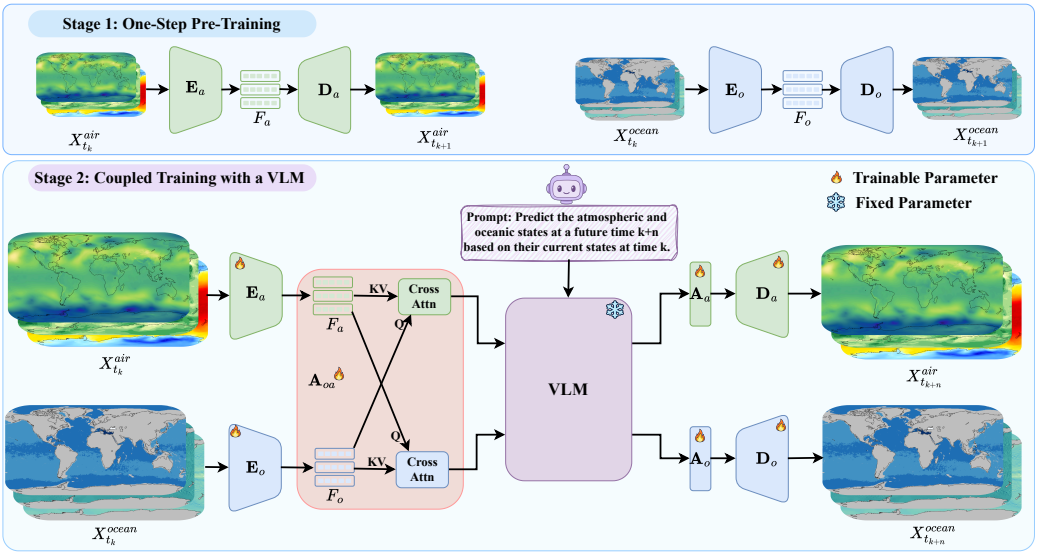

Figure 1: Architecture and training of OA-Bridge. Top: One-Step Pre-Training. Component-specific encoders E_a, E_o and decoders D_a, D_o are pretrained for one-day forecasting of atmospheric and oceanic states, respectively. Bottom: Coupled Training with a Vision-Language Model. The coupling adapter A_{oa} exchanges atmospheric and oceanic features through residual bidirectional cross-attention and projects them into the image-token embedding space of a frozen Qwen2.5-VL-7B backbone. The prompt template shown in the figure specifies the target lead time. Output adapters A_a, A_o and decoders recover the atmospheric and oceanic forecast fields. Q denotes queries, and KV denotes keys and values. Flame and snowflake icons indicate trainable and frozen parameters.

the coupled model. Following the convolutional embedding and Swin-based design of FuXi (Chen et al., 2023b), each encoder E_s uses a 2×2 convolution with stride 2 to embed non-overlapping spatial patches, followed by 12 Swin Transformer blocks (Liu et al., 2021). Window attention models spatial dependencies on the $H/2 \times W/2$ feature grid, while shifted windows connect neighboring regions. Each decoder D_s reconstructs physical fields using 12 Swin Transformer blocks followed by a 2×2 transposed convolution with stride 2. Both branches use the same architecture, with separate parameters and component-specific input/output channels.

For each training pair, the encoder produces features F_s and the decoder predicts the next daily state:

$$F_s = E_s(X^s_{t_k}) \in \mathbb{R}^{B \times N \times d}, \qquad \widehat{X}^s_{t_{k+1}} = D_s(F_s), \tag{2}$$

where $N = HW/4 = 16{,}200$ is the number of spatial tokens and $d = 1{,}536$ their feature dimension. The predicted fields are supervised with the loss in Section 3.4. Both encoders and decoders then initialize the coupled model.

3.3 Coupled Training with a Vision-Language Model

Frozen Qwen2.5-VL-7B $\mathcal{Q}$ (Bai et al., 2025) couples the pretrained branches using the Figure 1 prompt, instantiated with integer lead time n and tokenized with its native tokenizer.

The coupling adapter A_{oa} exchanges information between the encoded features through residual cross-attention:

$$\widetilde{F}_a = F_a + \mathrm{CA}_a(F_o, F_a, F_a), \qquad \widetilde{F}_o = F_o + \mathrm{CA}_o(F_a, F_o, F_o), \tag{3}$$

where the three arguments of CA_s supply queries, keys, and values through learned projections. Each block uses 32-head non-causal cross-attention within unshifted 18×18 windows. Both outputs retain shape $[B, N, d]$. Projections P_s within the adapter map them to the backbone's embedding dimension, yielding $Z_s = P_s(\widetilde{F}_s) \in \mathbb{R}^{B \times N \times d_q}$, where $d_q = 3{,}584$. All N projected tokens per component occupy the Qwen2.5-VL's image-token positions, totaling $2N = 32{,}400$ spatial tokens alongside prompt p_n specifying an n-day lead.

The backbone $\mathcal{Q}$ returns hidden states $G_s \in \mathbb{R}^{B\times N\times d_q}$ at the image-token positions for component s. The forecasts are decoded as

$$(G_a, G_o) = \mathcal{Q}(p_n, Z_a, Z_o), \qquad \widehat{X}^s_{t_{k+n}} = D_s(A_s(G_s)), \tag{4}$$

where output adapters A_a, A_o map hidden states to the d-dimensional decoder inputs. The encoders, coupling adapter, output adapters, and decoders are trained jointly, with the backbone parameters held fixed. At inference, the same model predicts the requested lead time in a single forward pass.

3.4 Training objective

Both stages use the same latitude-weighted Charbonnier loss, combining the latitude weighting of FuXi (Chen et al., 2023b) with the smooth absolute-error penalty used in FuXi-Ocean (Huang et al., 2025a). For component s, let $e_s(n) = \widehat{X}^s_{t_{k+n}} - X^s_{t_{k+n}}$ denote the prediction error in standardized units, with the singleton time axis omitted. We define

$$\mathcal{L}^{(n)}_s = \frac{1}{BC_sHW} \sum_{b,c,i,j} w_i M^s_{c,i,j} \sqrt{e_{s,b,c,i,j}(n)^2 + \epsilon^2}, \tag{5}$$

where b, c, i, j index samples, channels, latitudes, and longitudes, and $\epsilon > 0$ smooths the penalty near zero. The latitude weight is $w_i = H\cos\phi_i / \sum_{r=1}^{H}\cos\phi_r$, with ϕ_i the latitude of row i. For the atmosphere, $M^a_{c,i,j} = 1$, and for the ocean, $M^o_{c,i,j} = 1$ at valid ocean grid points and 0 otherwise.

The two component losses receive equal weights:

$$\mathcal{L}^{(n)} = \mathcal{L}^{(n)}_a + \mathcal{L}^{(n)}_o. \tag{6}$$

Pretraining optimizes each branch independently with $n = 1$, while coupled training jointly minimizes this objective at the requested lead time n.

4 Experiments

4.1 Implementation details

Data and training. We train on ERA5 and GLORYS12 from 1993 to 2021, validate on 2022, and test on 2023. Normalization statistics, ACC climatology, and extreme-event thresholds are estimated from the training period. We train OA-Bridge on 48 NVIDIA H200 GPUs with batch size 1 per GPU, using AdamW with $(\beta_1, \beta_2) = (0.9, 0.95)$, initial learning rate 2.5×10^{-4}, and weight decay 0.1. Training comprises 100,000 iterations of one-step component pretraining and 60,000 iterations of coupled training (Section 3). Coupled training samples lead times uniformly from $\{1, \ldots, L_{\max}\}$ days, with $L_{\max}$ increasing from 1 to 60 by one every 1,000 iterations.

Baselines and ablation. Following FuXi (Chen et al., 2023b), the baseline extends its input and output to 165 channels, concatenating 72 atmospheric and 93 oceanic variables on the same 1° grid. It jointly forecasts both components through daily autoregressive rollout, whereas OA-Bridge directly predicts each requested horizon. For a fair comparison, we increase the number of Swin blocks in the baseline to match the 2 billion (2B) parameters of OA-Bridge excluding the frozen VLM. The FuXi baseline also undergoes one-step pretraining followed by multi-step autoregressive rollout training, with the same number of training iterations per stage as OA-Bridge. Persistence holds the initial state fixed as a reference for slow ocean evolution. For the ablation, we replace the frozen Qwen2.5-VL-7B backbone with a trainable multilayer perceptron (MLP), retaining the atmospheric and oceanic branches and coupling adapters.

Forecast accuracy. Following FuXi (Chen et al., 2023b) and FuXi-ONS (Huang et al., 2026), we evaluate RMSE and ACC after converting forecasts back to physical units. For component s and channel c, define normalized spatial weights $\omega^s_{c,i,j} = w_i M^s_{c,i,j} / \sum_{i,j} w_i M^s_{c,i,j}$, using the latitude weights and masks in Section 3.4. Ocean metrics include only valid ocean grid points at the corresponding depth; land values contribute to neither the score nor its normalization. For K forecast initializations, the root mean square error is

$$\mathrm{RMSE}^s_c(n) = \frac{1}{K}\sum_{k=1}^{K}\left[\sum_{i,j}\omega^s_{c,i,j}\left(\widehat{X}^s_{t_{k+n},c,i,j} - X^s_{t_{k+n},c,i,j}\right)^2\right]^{1/2}. \tag{7}$$

Let $X^{s\prime} = X^s - \overline{X}^s$ and $\widehat{X}^{s\prime} = \widehat{X}^s - \overline{X}^s$ denote anomalies relative to the training-period climatology. The anomaly correlation coefficient is

$$\mathrm{ACC}_c^s(n) = \frac{1}{K}\sum_{k=1}^{K} \frac{\sum_{i,j} \omega_{c,i,j}^s \widehat{X}_{t_{k+n},c,i,j}^{s\prime} X_{t_{k+n},c,i,j}^{s\prime}}{\sqrt{\sum_{i,j} \omega_{c,i,j}^s (\widehat{X}_{t_{k+n},c,i,j}^{s\prime})^2}\sqrt{\sum_{i,j} \omega_{c,i,j}^s (X_{t_{k+n},c,i,j}^{s\prime})^2}}. \tag{8}$$

Lower RMSE indicates smaller errors; higher ACC indicates better agreement in anomaly patterns. Following FuXi-RTM (Huang et al., 2025b), we express the RMSE difference of OA-Bridge relative to the baseline as

$$\mathrm{NRMSE}_c^s(n) = \frac{\mathrm{RMSE}_c^s(n) - \mathrm{RMSE}_{c,\mathrm{base}}^s(n)}{\mathrm{RMSE}_{c,\mathrm{base}}^s(n)} \times 100\%. \tag{9}$$

Thus, negative NRMSE indicates improvement over the baseline. For the ablation, the MLP variant serves as the reference in Equation 9; the relative ACC change is $\mathrm{NACC} = 100(\mathrm{ACC} \,/\, \mathrm{ACC}_{\mathrm{MLP}} - 1)\%$, where ACC denotes the score of OA-Bridge.

Extreme events. We use the symmetric extremal dependence index (SEDI) (Ferro & Stephenson, 2011) to evaluate exceedances of the local 90th and 95th percentiles estimated from the training data. The hit and false alarm rates are $h = N_{\mathrm{hit}}/(N_{\mathrm{hit}} + N_{\mathrm{miss}})$ and $r = N_{\mathrm{fa}}/(N_{\mathrm{fa}} + N_{\mathrm{cn}})$, where the four counts denote the numbers of correctly forecast exceedances, missed exceedances, falsely forecast exceedances, and correctly forecast non-exceedances, respectively. The score is

$$\mathrm{SEDI} = \frac{\log r - \log h - \log(1-r) + \log(1-h)}{\log r + \log h + \log(1-r) + \log(1-h)}. \tag{10}$$

Larger values indicate better discrimination of threshold exceedances. Ocean events are evaluated only at valid ocean grid points. The same training-derived thresholds are applied to predictions and verifying states.

Geostrophic currents. Following FuXi-Ocean (Niu et al., 2026), we derive surface geostrophic velocities from predicted and reference SSH:

$$u_g(\eta) = -\frac{g}{f(\phi)}\frac{\partial \eta}{\partial y}, \qquad v_g(\eta) = \frac{g}{f(\phi)}\frac{\partial \eta}{\partial x}, \qquad f(\phi) = 2\Omega \sin\phi, \tag{11}$$

where η is SSH, x and y are eastward and northward distances, g is gravitational acceleration, and Ω is Earth's angular velocity. We apply Equation 7 separately to $u_g(\widehat{\eta})$ versus $u_g(\eta)$ and $v_g(\widehat{\eta})$ versus $v_g(\eta)$. The mask includes ocean points with valid SSH gradients and excludes $5°$S to $5°$N, following Niu et al. (2026). This diagnostic evaluates SSH-derived pressure gradients.

4.2 Forecast skill analysis

Skill across forecast horizons. Figure 2 shows that forecasts obtained by holding the initial state fixed lose accuracy more rapidly for most atmospheric variables than for ocean variables. The different lead times over which these fixed-initial-state forecasts retain skill reflect differences in the temporal evolution of the ocean and atmosphere. Despite these differences, OA-Bridge achieves lower RMSE and higher ACC than the baseline at extended lead times in both components, with atmospheric gains emerging after an initial short-range tradeoff.

For ocean salinity and temperature, baseline RMSE increases sharply beyond approximately day 30, accompanied by a pronounced decline in ACC, while OA-Bridge maintains more gradual error growth. At day 60, OA-Bridge's salinity and temperature RMSEs are approximately 0.17 PSU and $0.74\,°$C, compared with 1.71 PSU and $4.94\,°$C for the baseline. For these two variables, OA-Bridge maintains ACCs of approximately 0.67 and 0.65, respectively, while the baseline ACCs are both below 0.2. SSH exhibits a similar late-stage deterioration in the baseline, particularly after approximately day 45. At longer leads, baseline forecasts for salinity, temperature, and SSH become worse than persistence in both metrics, whereas its forecasts for the two current components generally outperform or remain comparable to persistence. OA-Bridge outperforms persistence across all five ocean variables at extended lead times, indicating predictive skill beyond retaining the initial state.

Atmospheric improvements depend more strongly on lead time. The baseline often performs better during the first one to two weeks, whereas OA-Bridge achieves lower RMSE across the displayed atmospheric variables at longer leads. Following rapid initial growth, OA-Bridge's errors

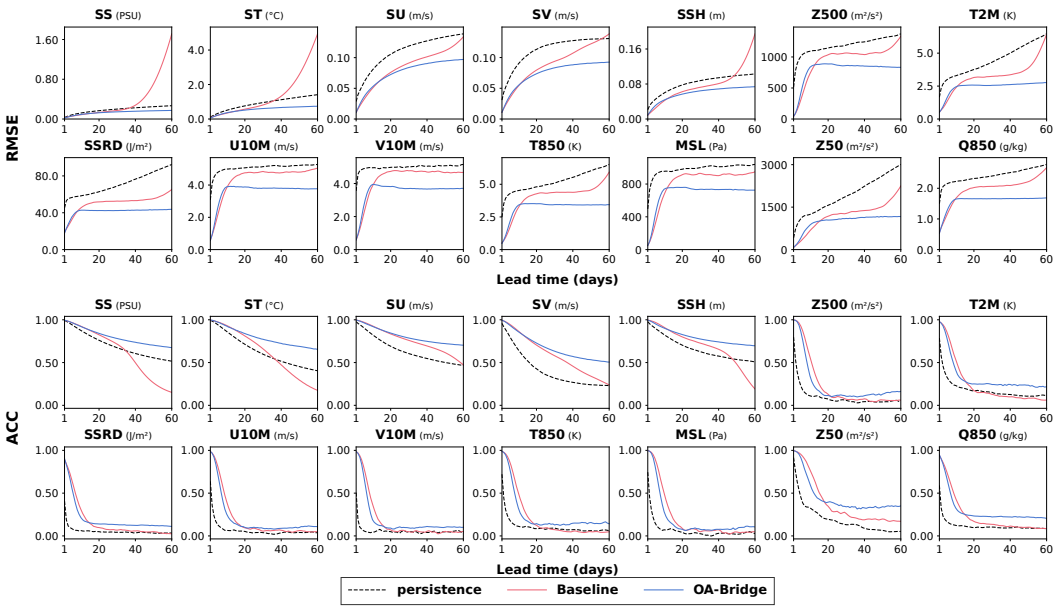

Figure 2: RMSE and ACC over 1–60 days for OA-Bridge (blue), joint FuXi (red), and persistence (black dashed). Ocean masks apply; SS, ST, SU, and SV scores are averaged over 23 depths; SSH is evaluated at the surface. Pressure suffixes denote hPa. Units apply only to RMSE.

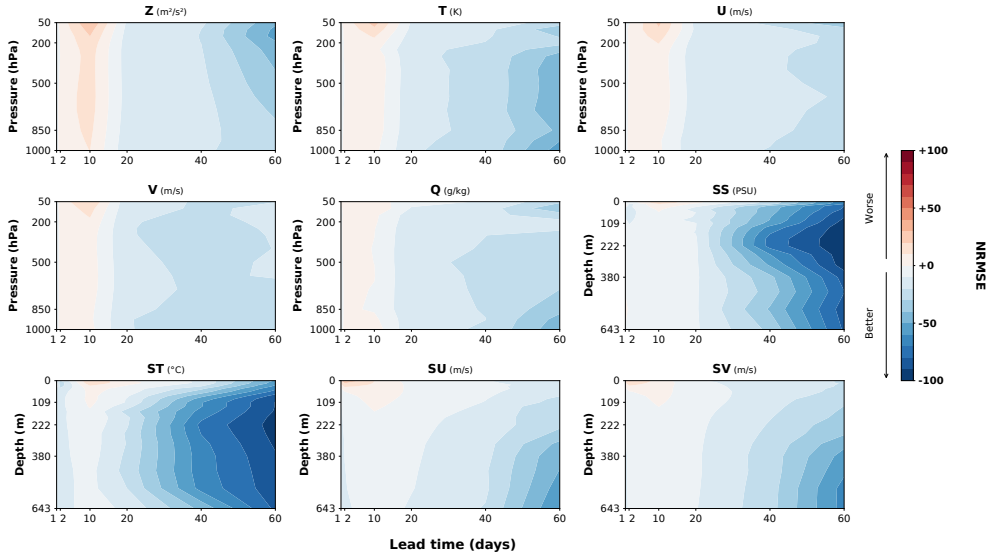

Figure 3: NRMSE (%) across lead times and vertical levels, as defined in Equation 9. Blue favors OA-Bridge and red favors the baseline. Atmospheric fields span 13 pressure levels, and ocean fields span 23 depths. Ocean scores are computed over valid ocean grid points at each depth.

generally level off earlier. Several baseline curves, including T2M and Z50, exhibit renewed error growth around days 45–50, while the corresponding OA-Bridge curves remain comparatively flat. Extended-range ACC also improves, particularly for T2M, Z50, and Q850, although absolute correlations remain low for several other fields. Overall, OA-Bridge outperforms the baseline at extended lead times in both oceanic and atmospheric forecasting, although short-range atmospheric performance is sometimes weaker and long-range atmospheric ACC remains low for several variables.

Vertical distribution of improvements. Figure 3 shows that atmospheric improvements depend primarily on lead time and extend across most pressure levels. Short-range disadvantages generally give way to lower errors after approximately 15–20 days. Relative reductions become more pronounced at longer leads for geopotential, temperature, and humidity, while wind improvements remain more modest. The broadly consistent transition across pressure levels indicates that the extended-range advantage is not confined to a particular atmospheric layer.

Ocean improvements strengthen more markedly with lead time and extend well below the surface. Salinity and temperature exhibit large relative error reductions across much of the sampled water column at longer leads. Improvements in both current components are more modest near the surface and become larger at depth late in the forecast period. These profiles show that the gains observed in the aggregated metrics extend across atmospheric levels and ocean depths.

Inference Efficiency. For a specified target lead time, OA-Bridge's inference time stays near 2.68 s, while baseline latency grows approximately linearly, making OA-Bridge faster from day 4 onward. At 30 and 60 days, baseline latency reaches 21.28 and 42.60 s, yielding speedups of $7.91\times$ and $15.83\times$, respectively.

4.3 Extreme events and physical consistency

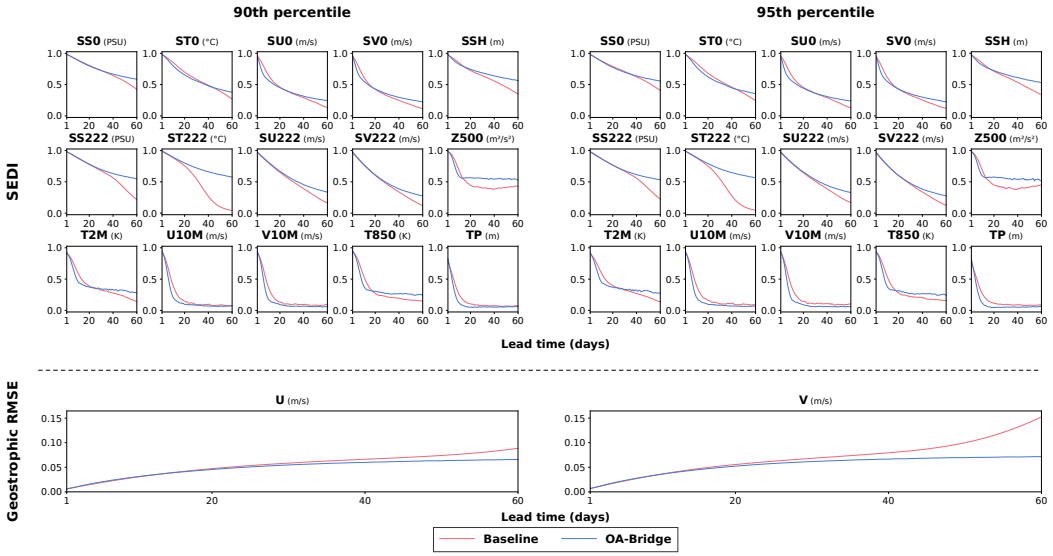

Figure 4: Extreme-event and circulation diagnostics for OA-Bridge (blue) and the baseline (red). Top: dimensionless SEDI at the 90th (left) and 95th (right) percentiles; ocean suffixes indicate depth (m). Bottom: RMSE of geostrophic velocities derived from predicted versus reference SSH. Ocean masks apply throughout, with $5°$S–$5°$N additionally excluded for geostrophic scores.

Extreme-event skill. The SEDI results in Figure 4 show consistent patterns at the 90th- and 95th-percentile thresholds, with the largest extended-range advantages over the baseline in subsurface salinity and temperature. The gaps over the baseline are substantially larger at 222 m than at the surface, particularly for temperature. At day 60, OA-Bridge retains a temperature SEDI near 0.6 at 222 m, whereas the baseline falls below 0.1. For salinity, the larger subsurface advantage mainly reflects stronger baseline deterioration. Both current components also retain somewhat higher long-range scores at depth, although the differences are smaller. SSH maintains a sustained advantage, with SEDI near 0.55 versus approximately 0.35 for the baseline at day 60.

Atmospheric improvements are more variable-dependent. Z500 retains an extended-range SEDI near 0.55, compared with approximately 0.4 for the baseline. T2M and T850 show modest advantages beyond approximately day 20 following a short-range disadvantage, but their absolute scores remain lower. In contrast, 10-m winds and precipitation exhibit rapidly declining SEDI in both

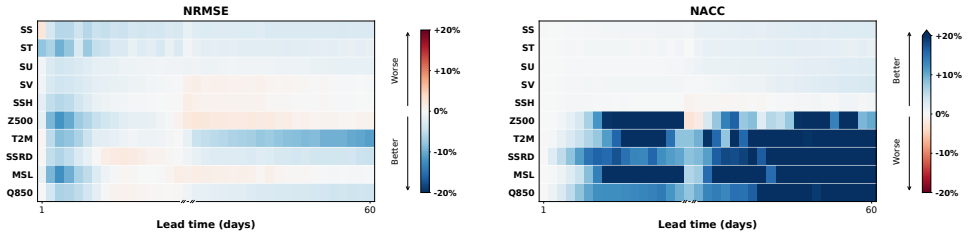

Figure 5: Relative RMSE (left) and ACC (right) differences between OA-Bridge and an ablation replacing the VLM with an MLP using one-hot lead-time encoding. Blue favors OA-Bridge. SS, ST, SU, and SV scores are averaged over 23 depths.

models, with little separation at longer leads. The benefits therefore extend to selected atmospheric events, while long-range wind and precipitation prediction remains limited.

Circulation inferred from SSH. The geostrophic RMSEs in Figure 4 are similar at short leads but increasingly favor OA-Bridge after about day 20. The difference is particularly pronounced for the meridional component, whose baseline error accelerates after roughly day 40 while that of OA-Bridge grows more slowly. At day 60, zonal RMSE is approximately 0.066 $\mathrm{m\,s^{-1}}$ for OA-Bridge versus 0.089 $\mathrm{m\,s^{-1}}$ for the baseline, while meridional RMSE is approximately 0.072 versus 0.152 $\mathrm{m\,s^{-1}}$. These results indicate that the SSH improvements extend to the spatial gradients used to diagnose surface geostrophic circulation. Together with the atmospheric and ocean-profile results, these findings support the overall forecasting advantage of OA-Bridge's complete framework, which combines feature coupling with direct lead-time conditioning, over the autoregressive baseline.

4.4 Ablation Study

To assess the contribution of the VLM backbone, we compare OA-Bridge with an ablated variant that replaces it with an MLP conditioned on a one-hot encoding of the target lead time, while retaining the other architectural components and the same training strategy. Figure 5 shows that the clearest advantages of OA-Bridge occur in atmospheric ACC at extended lead times. The relative gains for Z500, T2M, SSRD, MSL, and Q850 are larger than those for ocean variables. RMSE improvements are more variable-dependent. T2M exhibits increasing error reductions toward day 60, whereas Z500 and MSL show improvements at some leads and degradations at others. These results indicate that the VLM-based configuration improves atmospheric anomaly prediction relative to the MLP alternative, without uniformly reducing RMSE across variables and lead times.

Additional results. Additional analyses of cross-component sensitivity to input perturbations and spatial ocean forecast errors are provided in Appendices B and C, respectively.

5 Conclusion

We presented OA-Bridge, a VLM-based framework for coupled global atmosphere–ocean forecasting at daily resolution. Bidirectional feature exchange and a shared frozen VLM conditioned on lead time enable direct prediction of joint future states. Across 1–60-day forecasts, it improves extended-range ocean skill and selected atmospheric predictions over the jointly trained autoregressive baseline. Oceanic improvements extend across depths, to extreme-event prediction, and to SSH-derived geostrophic circulation. Direct prediction also maintains nearly constant inference time across target lead times, achieving a $15.83\times$ speedup at day 60. These results demonstrate the framework's advantages in forecast accuracy and inference efficiency at extended horizons. Two directions remain for future research. First, improving short-range atmospheric accuracy and the limited long-range skill of some atmospheric variables requires further investigation. Second, the FuXi-derived encoder–decoder components offer a starting point for probabilistic extensions using VAE-based latent perturbations. Repeated sampling could generate coupled ensembles, with direct prediction potentially reducing the cost of reaching individual subseasonal target days.

AI USE STATEMENT

We used generative AI tools for language polishing, LaTeX assistance, table formatting, and literature discovery for related work. We did not use generative AI tools to generate or alter data, run experiments, replace author decisions about the method, or make unsupported scientific claims. All AI-assisted text and code were reviewed by the authors. The authors designed the method and experiments, executed and tested the code, checked all numerical results against the recorded outputs, and manually verified the relevant literature. We take responsibility for the final content of this work, including any text, claims, code, or other artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses publicly available ERA5 and GLORYS12 reanalysis data and a frozen pretrained VLM, without collecting human-subject or personal data. Forecast skill varies across regions, variables, and lead times. We report the data sources, evaluation protocols, and observed limitations, and recommend further validation before operational use, particularly for extreme-event forecasting.

REPRODUCIBILITY STATEMENT

The paper and appendices describe the data sources and splits, preprocessing, model architecture, training objectives, optimization settings, and evaluation protocols, with training pseudocode in Appendix A. Normalization statistics, climatology, and extreme-event thresholds are computed exclusively from the training data. Upon acceptance, we will release the source code, configuration files, and preprocessing, training, and evaluation scripts to facilitate reproduction of the experiments.

REFERENCES

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*, 2025. URL https://arxiv.org/abs/2502.13923.

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.

Eric P Chassignet, Harley E Hurlburt, Ole Martin Smedstad, George R Halliwell, Patrick J Hogan, Alan J Wallcraft, Remy Baraille, and Rainer Bleck. The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *Journal of Marine Systems*, 65:60–83, 2007.

Kang Chen, Tao Han, Junchao Gong, Lei Bai, Fenghua Ling, Jing-Jia Luo, Xi Chen, Leiming Ma, Tianning Zhang, Rui Su, et al. Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv preprint arXiv:2304.02948*, 2023a.

Lei Chen, Xiaohui Zhong, Feng Zhang, Yuan Cheng, Yinghui Xu, Yuan Qi, and Hao Li. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, 6:190, 2023b. doi: 10.1038/s41612-023-00512-1.

Nathaniel Cresswell-Clay, Bowen Liu, Dale R. Durran, Zihui Liu, Zachary I. Espinosa, Raul A. Moreno, and Matthias Karlbauer. A Deep Learning Earth System Model for Efficient Simulation of the Observed Climate. *AGU Advances*, 6(4):e2025AV001706, 2025. doi: 10.1029/2025AV001706.

Yingzhe Cui, Ruohan Wu, Xiang Zhang, Ziqi Zhu, Bo Liu, Jun Shi, Junshi Chen, Hailong Liu, Shenghui Zhou, Liang Su, Zhao Jing, Hong An, and Lixin Wu. Forecasting the Eddying Ocean with a Deep Neural Network. *Nature Communications*, 16:2268, 2025.

Shuangying Du and Rong-Hua Zhang. U-Net models for representing wind stress anomalies over the tropical pacific and their integrations with an intermediate coupled model for ENSO studies. *Advances in Atmospheric Sciences*, 41(7):1403–1416, 2024. doi: 10.1007/s00376-023-3179-2.

Shuangying Du, Rong-Hua Zhang, and Chuan Gao. Retrospective ENSO predictions using an intermediate ocean–atmosphere coupled model by integrating deep-learning sea surface wind stress. *Journal of Oceanology and Limnology*, 2025. doi: 10.1007/s00343-025-5166-1.

James P. C. Duncan, Elynn Wu, Surya Dheeshjith, Adam Subel, Troy Arcomano, Spencer K. Clark, Brian Henn, Anna Kwa, Jeremy McGibbon, W. Andre Perkins, William Gregory, Carlos Fernandez-Granda, Julius Busecke, Oliver Watt-Meyer, William J. Hurlin, Alistair Adcroft, Laure Zanna, and Christopher Bretherton. SamudrACE: Fast and Accurate Coupled Climate Modeling With 3D Ocean and Atmosphere Emulators. *Geophysical Research Letters*, 53:e2025GL119340, 2026.

Anass El Aouni, Quentin Gaudel, Charly Regnier, Simon Van Gennip, Olivier Le Galloudec, Marie Drevillon, Yann Drillet, and Jean-Michel Lellouche. GLONET: Mercator's End-to-End Neural Global Ocean Forecasting System. *Journal of Geophysical Research: Machine Learning and Computation*, 2:e2025JH000686, 2025.

Christopher A. T. Ferro and David B. Stephenson. Extremal dependence indices: Improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, 26(5): 699–713, 2011. doi: 10.1175/WAF-D-10-05030.1.

Madec Gurvan, Romain Bourdallé-Badie, Clément Bricaud, Diego Bruciaferri, Daley Calvert, Jérôme Chanut, Emanuela Clementi, Andrew Coward, Damiano Delrosso, Christian Ethé, et al. Nemo ocean engine. 2017.

Sara Hahner, Lorenzo Zampieri, Jean-Raymond Bidlot, Philip Browne, Matthew Chantry, Mariana C. A. Clare, Harrison Cook, Peter Dueben, Rachel Furner, Sarah Keeley, Josh Kousal, Simon Lang, Christian Lessig, Gert Mertes, Kristian Mogensen, Gabriel Moldovan, Charles Pelletier, Florian Pinault, Ana Prieto Nemesio, Baudouin Raoult, Irina Sandu, Mario Santa Cruz, Jakob Schloer, Steffen Tietsche, and Hao Zuo. Representing the Surface Ocean in ECMWF's data-driven forecasting system AIFS, 2026.

Juanxiong He, Jiangbo Jin, Xiao Dong, Dongling Zhang, Xiaodong Zeng, and Guangqing Zhou. The Development of a High-Resolution Regional Ocean–Atmosphere Forecast System for East Asia and the Western North Pacific. *Frontiers in Marine Science*, 12:1715903, 2025.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: 10.1002/qj.3803.

Helene T. Hewitt, Michael J. Bell, Eric P. Chassignet, Arnaud Czaja, David Ferreira, Stephen M. Griffies, Pat Hyder, Julie L. McClean, Adrian L. New, and Malcolm J. Roberts. Will High-resolution Global Ocean Models Benefit Coupled Predictions on Short-range to Climate Timescales? *Ocean Modelling*, 120:120–136, 2017.

Junya Hu, Rong-Hua Zhang, and Chuan Gao. A Hybrid Coupled Ocean-Atmosphere Model and Its Simulation of ENSO and Atmospheric Responses. *Advances in Atmospheric Sciences*, 36: 643–657, 2019.

Qiusheng Huang, Yuan Niu, Xiaohui Zhong, Anboyu Guo, Lei Chen, Dianjun Zhang, Xuefeng Zhang, and Hao Li. FuXi-Ocean: A global ocean forecasting system with sub-daily resolution. In *Advances in Neural Information Processing Systems*, volume 38, pp. 32656–32678, 2025a. doi: 10.52202/085713-1095.

Qiusheng Huang, Xiaohui Zhong, Xu Fan, and Hao Li. FuXi-RTM: A physics-guided prediction framework with radiative transfer modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8852–8862, 2025b.

Qiusheng Huang, Xiaohui Zhong, Anboyu Guo, Ziyi Peng, Lei Chen, and Hao Li. Data-driven ensemble prediction of the global ocean. *arXiv preprint arXiv:2603.19591*, 2026. URL `https://arxiv.org/abs/2603.19591`.

S. J. Johnson, T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo, S. P. E. Keeley, K. Mogensen, H. Zuo, and B. M. Monge-Sanz. SEAS5: The New ECMWF Seasonal Forecast System. *Geoscientific Model Development*, 12:1087–1117, 2019.

Hyun-Sook Kim, Bin Liu, Biju Thomas, Daniel Rosen, Weiguo Wang, Andrew Hazelton, Zhan Zhang, Xuejin Zhang, and Avichal Mehra. Ocean component of the first operational version of Hurricane Analysis and Forecast System: Evaluation of HYbrid Coordinate Ocean Model and hurricane feedback forecasts. *Frontiers in Earth Science*, 12:1399409, 2024.

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.

Jean-Michel Lellouche, Eric Greiner, Olivier Le Galloudec, Charly Regnier, Mounir Benkiran, Charles-Emmanuel Testut, Romain Bourdalle-Badie, Marie Drevillon, Gilles Garric, and Yann Drillet. The mercator ocean global high-resolution monitoring and forecasting system. *New Frontiers in Operational Oceanography*, 1:563–592, 2018.

Jean-Michel Lellouche, Eric Greiner, Romain Bourdallé-Badie, Gilles Garric, Angélique Melet, Marie Drévillon, Clément Bricaud, Mathieu Hamon, Olivier Le Galloudec, Charly Regnier, Tony Candela, Charles-Emmanuel Testut, Florent Gasparin, Giovanni Ruggiero, Mounir Benkiran, Yann Drillet, and Pierre-Yves Le Traon. The Copernicus global $1/12^{\circ}$ oceanic and sea ice GLORYS12 reanalysis. *Frontiers in Earth Science*, 9:698876, 2021. doi: 10.3389/feart.2021.698876.

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*, pp. 10012–10022, 2021.

Jing-Jia Luo, Jiangjiang Xia, Baoxiang Pan, Yoo-Geun Ham, Xiaofeng Li, Wei Shangguan, Wei Xue, Yaqiang Wang, Bin Mu, Youngjoon Hong, Hao Li, Xiaohui Zhong, Kan Dai, Lei Bai, Fenghua Ling, Niklas Boers, Christopher Bretherton, Bin Chen, Dongjin Cho, Pierre Gentine, Zijie Guo, Xiaomeng Huang, Daehyun Kang, Hyunwoo J Kim, Jeong-Hwan Kim, Lili Lei, Fan Meng, Seol-Hee Oh, Bo Qin, Zixiong Shen, Qiming Sun, Yuheng Tang, Xuan Tong, Bingcheng Wan, Lina Wang, Ya Wang, Yiming Wang, Jiye Wu, Yi Xiao, Lina Yao, Song Yang, Chaoxia Yuan, Shijin Yuan, Tingzhao Yu, and Mengchu Zhao. AI for atmosphere–ocean sciences: advancements, challenges and ways forward. *National Science Review*, 13:nwag063, 2026.

Yuan Niu, Qiusheng Huang, Xiaohui Zhong, Anboyu Guo, Lei Chen, Dianjun Zhang, Zhen-Qiang Zhou, Xiaoyan Jia, Libo Wu, Renhe Zhang, Jiawei Qi, Xuefeng Zhang, and Hao Li. A deep learning global ocean forecasting model with sub-daily and eddy-resolving resolution. *npj Climate and Atmospheric Science*, 2026. doi: 10.1038/s41612-026-01444-2.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators, 2022.

Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui ya Chuang, Mark Iredell, Michael Ek, Jesse Meng, Rongqian Yang, Malaquías Peña Mendez, Huug van den Dool, Qin Zhang, Wanqiu Wang, Mingyue Chen, and Emily Becker. The NCEP Climate Forecast System Version 2. *Journal of Climate*, 27:2185–2208, 2014.

Jindong Tian, Yifei Ding, Ronghui Xu, Hao Miao, Chenjuan Guo, and Bin Yang. ARROW: An Adaptive Rollout and Routing Method for Global Weather Forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 54823–54847, 2026.

Marina Tonani, Magdalena Balmaseda, Laurent Bertino, Ed Blockley, Gary Brassington, Fraser Davidson, Yann Drillet, Pat Hogan, Tsurane Kuragano, Tong Lee, et al. Status and future of global and regional ocean prediction systems. *Journal of Operational Oceanography*, 8:s201–s220, 2015.

Chenggong Wang, Michael S. Pritchard, Noah Brenowitz, Yair Cohen, Boris Bonev, Thorsten Kurth, Dale Durran, and Jaideep Pathak. Coupled Ocean-Atmosphere Dynamics in a Machine Learning Earth System Model, 2024a.

Longhao Wang, Xuanze Zhang, L. Ruby Leung, Francis H. S. Chiew, Amir AghaKouchak, Kairan Ying, and Yongqiang Zhang. CAS-Canglong: A skillful 3D transformer model for sub-seasonal to seasonal global sea surface temperature prediction. *arXiv preprint arXiv:2409.05369*, 2024b. doi: 10.48550/arXiv.2409.05369. URL https://arxiv.org/abs/2409.05369.

Xiang Wang, Renzhi Wang, Ningzi Hu, Pinqiang Wang, Peng Huo, Guihua Wang, Huizan Wang, Senzhang Wang, Junxing Zhu, Jianbo Xu, Jun Yin, Senliang Bao, Ciqiang Luo, Ziqing Zu, Yi Han, Weimin Zhang, Kaijun Ren, Kefeng Deng, and Junqiang Song. XiHe: A Data-Driven Model for Global Ocean Eddy-Resolving Forecasting, 2024c.

John C. Warner, Brandy Armstrong, Ruoying He, and Joseph B. Zambon. Development of a Coupled Ocean–Atmosphere–Wave–Sediment Transport (COAWST) Modeling System. *Ocean Modelling*, 35:230–244, 2010.

Nan Yang, Chong Wang, Meihua Zhao, Zimeng Zhao, Huiling Zheng, Bin Zhang, Jianing Wang, and Xiaofeng Li. LangYa: Revolutionizing Cross-Spatiotemporal Ocean Forecasting, 2025. arXiv:2412.18097v3, revised 31 Mar 2025.

Yang Yu, Yin-Nan Li, Rong-Hua Zhang, Shu-Hua Chen, Yu-Heng Tseng, Wenzhe Zhang, and Hongna Wang. A flexible regional ocean modeling system-based hybrid coupled model for el niño–southern oscillation studies – model formulation and performance evaluation. *Geoscientific Model Development*, 18:5527–5547, 2025.

Rong-Hua Zhang. A Hybrid Coupled Model for the Pacific Ocean–Atmosphere System. Part I: Description and Basic Performance. *Advances in Atmospheric Sciences*, 32:301–318, 2015.

Rong-Hua Zhang, Lu Zhou, Chuan Gao, and Lingjiang Tao. A transformer-based coupled ocean–atmosphere model for ENSO studies. *Science Bulletin*, 69(15):2323–2327, 2024. doi: 10.1016/j.scib.2024.04.048.

Chenguang Zhou, Lei Chen, Xiaohui Zhong, Bo Lu, Hao Li, Libo Wu, Jie Wu, Jiahui Hu, Zesheng Dou, Pang-Chi Hsu, and Xiaoye Zhang. A Machine Learning Model for Skillful Climate System Prediction, 2025.

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):eadf2827, 2023. doi: 10.1126/sciadv.adf2827.

Zijian Zhu, Qiusheng Huang, Anboyu Guo, Xiaohui Zhong, and Hao Li. AviaSafe: A physics-informed data-driven model for aviation safety-critical cloud forecasts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 33143–33152, 2026.

A TRAINING PIPELINE

Algorithm 1 summarizes the two training stages in Section 3. Inputs use the training-set normalization and channel definitions in Section 3.1; the precipitation input is set to zero while precipitation remains a prediction target. Component pretraining learns one-day forecasts independently for the atmosphere and ocean. Coupled training then initializes both branches from these weights and introduces the coupling adapter A_{oa}, the frozen backbone $\mathcal{Q}$, and output adapters A_a, A_o. The same latitude-weighted Charbonnier objective is used in both stages, with ocean masks applied as in Equation 5.

Algorithm 1 Two-stage training of OA-Bridge

Require: Training states X^a, X^o, ocean masks, pretrained backbone $\mathcal{Q}$
Ensure: Coupled predictor $\mathcal{F}_\theta$

Stage 1: component pretraining
for $s \in \{a, o\}$ **do**
  for $j = 1, \ldots, 100{,}000$ **do**
    Sample valid initial states $X^s_{t_k}$ and one-day targets $X^s_{t_{k+1}}$
    $\widehat{X}^s_{t_{k+1}} \leftarrow D_s(E_s(X^s_{t_k}))$
    Compute $\mathcal{L}^{(1)}_s$ using Equation 5
    Update the parameters of E_s, D_s with $\mathcal{L}^{(1)}_s$
  end for
end for
Stage 2: coupled training
Initialize the two branches from Stage 1 and freeze the weights of $\mathcal{Q}$
for $j = 0, \ldots, 59{,}999$ **do**
  $n_{\max} \leftarrow \min\{60,\ 1 + \lfloor j/1000 \rfloor\}$
  Sample n uniformly from $\{1, \ldots, n_{\max}\}$
  Sample paired inputs $(X^a_{t_k}, X^o_{t_k})$ and targets $(X^a_{t_{k+n}}, X^o_{t_{k+n}})$
  $F_s \leftarrow E_s(X^s_{t_k})$ for $s \in \{a, o\}$
  $(Z_a, Z_o) \leftarrow A_{oa}(F_a, F_o)$ ▷ Cross-attention and projection
  $(G_a, G_o) \leftarrow \mathcal{Q}(p_n, Z_a, Z_o)$
  $\widehat{X}^s_{t_{k+n}} \leftarrow D_s(A_s(G_s))$ for $s \in \{a, o\}$
  $\mathcal{L} \leftarrow \mathcal{L}^{(n)}_a + \mathcal{L}^{(n)}_o$
  Update $E_a, E_o, A_{oa}, A_a, A_o, D_a, D_o$ with $\mathcal{L}$; keep $\mathcal{Q}$ fixed
end for
return $\mathcal{F}_\theta$

The maximum sampled lead time increases by one day every 1,000 coupled-training iterations. Thus, the first interval uses only one-day targets, the second samples uniformly between one and two days, and the final interval permits all daily horizons up to 60 days. We use 48 NVIDIA H200 GPUs with a batch size of 1 per GPU. Following FuXi-RTM (Huang et al., 2025b), AdamW uses $(\beta_1, \beta_2) = (0.9, 0.95)$, an initial learning rate of 2.5×10^{-4}, and weight decay of 0.1.

Freezing $\mathcal{Q}$ prevents updates to its weights while allowing gradients to pass through it to the encoders and coupling adapter. Inference uses the requested lead-time prompt p_n and a single forward pass through the coupled model, as in Equation 4.

B INPUT PERTURBATION ANALYSIS

Experimental protocol. To examine how information passes between the two components, we add noise to one input variable of OA-Bridge at inference while leaving the remaining inputs and model parameters fixed. Noise values are sampled independently at valid grid points from $\mathcal{N}(0, 0.1^2)$ and added to the input after training-set standardization. The perturbation standard deviation is therefore 10% of the variable's training standard deviation, approximately $0.527\ \mathrm{m\,s^{-1}}$ for U10M. Invalid ocean grid points receive no noise. We compare perturbed and unperturbed forecasts from the same initial states at the same lead time. Using the RMSE definition in Equation 7, the relative increase for output channel c in component s when input variable p is perturbed is

$$\Delta^{s}_{p\to c}(n) = \frac{\mathrm{RMSE}^{s,(p)}_{c}(n) - \mathrm{RMSE}^{s,(0)}_{c}(n)}{\mathrm{RMSE}^{s,(0)}_{c}(n)} \times 100\%, \tag{12}$$

where superscripts (p) and (0) denote perturbed and unperturbed inputs, respectively, and n is the lead time in days. Ocean output errors are evaluated only over valid ocean grid points. For each input variable and lead time, we rank output channels in descending order of $\Delta^{s}_{p\to c}(n)$. Table 1 lists the ten highest-ranking outputs for each setting. We report day 1 for V10M and T850, day 4 for Z500, U10M, and SS0, day 12 for SU0, day 20 for ST0 and SV0, and day 40 for SSH.

Table 1: Output sensitivity to input perturbations in OA-Bridge. Each row specifies the perturbed input; columns list output variables in descending order of relative RMSE increase. Rank 1 indicates the largest increase. Light blue and light green identify oceanic and atmospheric variables, respectively.

Input	Rank									
variable	1	2	3	4	5	6	7	8	9	10
Z500	Z500	T50	T100	SS29	SS0	SS2	SSH	SS5	SS7	T400
U10M	V50	T100	U50	Q100	Z50	SS29	U100	SS0	SS15	SSH
V10M	V10M	SS15	SS318	SS7	D2M	ST55	SS21	SS155	SS266	Q200
T850	T850	Q850	Z150	Q925	Q1000	Z100	Z200	Z250	TP	Z300
SS0	SS0	Z50	SS11	SS7	SS5	SS15	SS2	T50	U50	V50
ST0	ST0	T200	V700	V600	T150	V500	V850	V925	V400	V10M
SU0	SS15	SS11	U500	U400	SS0	U600	U300	U700	SS2	V1000
SV0	T200	T150	V700	V850	V600	V500	V925	V400	SS0	V1000
SSH	V100	V50	V150	V600	V700	V500	Z300	Z250	Z600	V400

Atmospheric pressure-level suffixes are in hPa; ocean suffixes are depths in meters. SS, ST, SU, and SV correspond to `s`, `st`, `su`, and `sv`; SSH corresponds to `sel`. Surface atmospheric labels follow Section 3.1.

Sensitivity across components. The rankings reveal responses in both directions. For Z500 perturbations at day 4, six of the ten leading outputs are oceanic, including SSH and salinity at several depths. Perturbing V10M places seven oceanic channels among the ten leading outputs at day 1, whereas the T850 ranking at the same lead time contains only atmospheric outputs. Thus, atmospheric perturbations can affect both surface and subsurface ocean predictions, with responses depending on the input variable.

Ocean perturbations also affect atmospheric predictions. For ST0 and SV0 perturbations at day 20, nine of the ten leading outputs are atmospheric, with wind components appearing at multiple pressure levels. Perturbing SSH at day 40 yields an entirely atmospheric top ten, comprising winds and geopotential. These responses provide complementary evidence that OA-Bridge uses information from both components during inference. Since the rows report relative rankings at different lead times, they describe which outputs are most sensitive within each setting rather than the absolute strength of influence across settings.

C SPATIAL DISTRIBUTIONS OF OCEAN FORECAST ERRORS

C.1 SPATIAL DISTRIBUTION OF SALINITY ERRORS

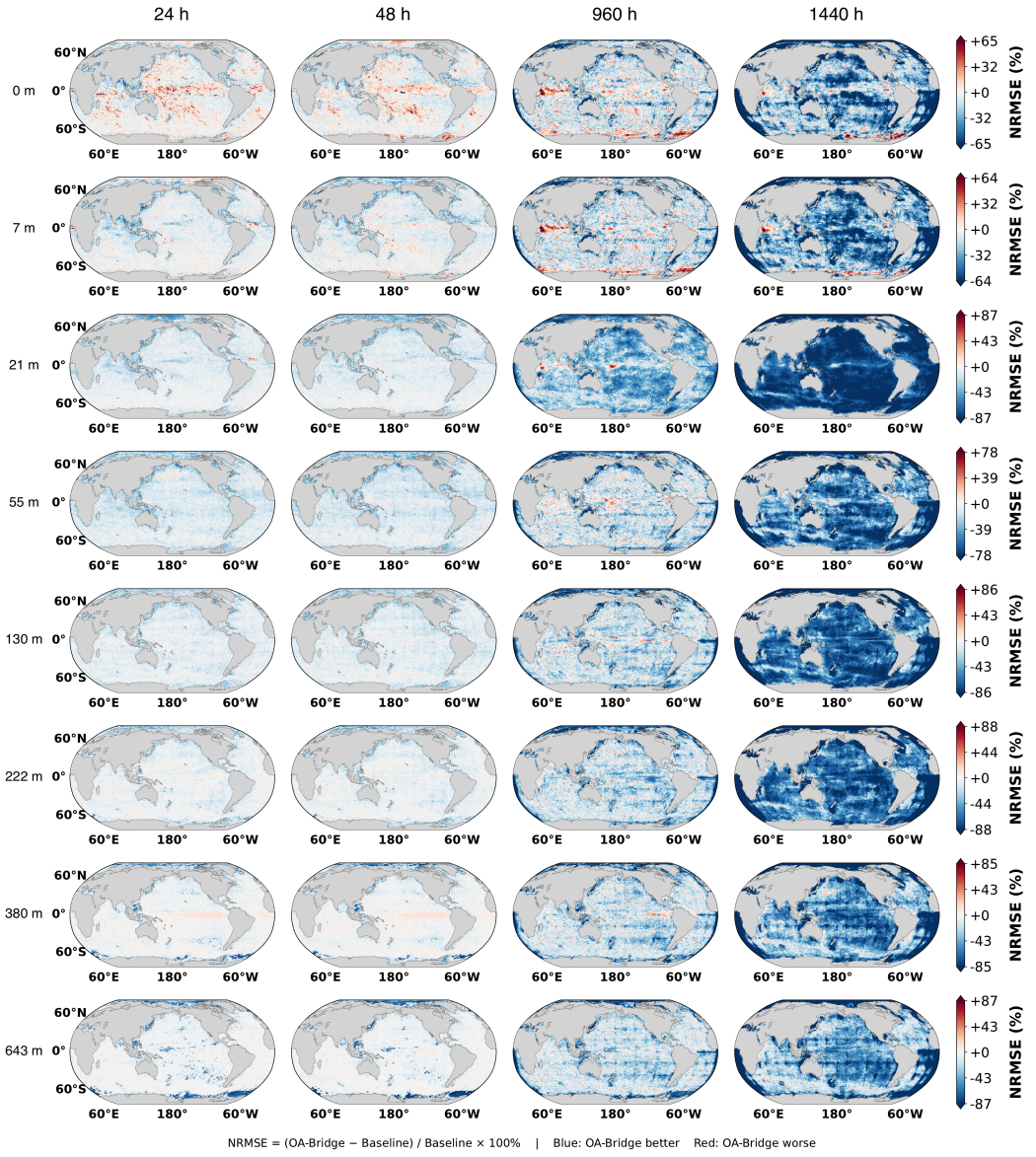

Figure 6: Spatial NRMSE (%) for ocean salinity, computed as $100(\mathrm{RMSE}_{\text{OA-Bridge}} - \mathrm{RMSE}_{\text{Baseline}})/\mathrm{RMSE}_{\text{Baseline}}$ at valid ocean grid points. Columns show 24, 48, 960, and 1440 h (1, 2, 40, and 60 days); rows show depths of 0, 7, 21, 55, 130, 222, 380, and 643 m. Blue favors OA-Bridge; red favors the baseline. Each depth has its own color scale. Gray regions are masked.

Figure 6 shows mixed differences at one and two days, with positive values in parts of the tropical surface ocean and mostly small changes at depth. By day 40, negative NRMSE is more widespread, particularly at 21 m. At day 60, improvements extend across large areas of the Pacific, Atlantic, and Indian oceans and through the sampled water column. Local positive regions remain in the uppermost layers, including parts of the equatorial Pacific and Southern Ocean. The long-range salinity gains therefore extend beyond a single region while retaining spatial variation that is not visible in the global averages.

C.2 SPATIAL DISTRIBUTION OF TEMPERATURE ERRORS

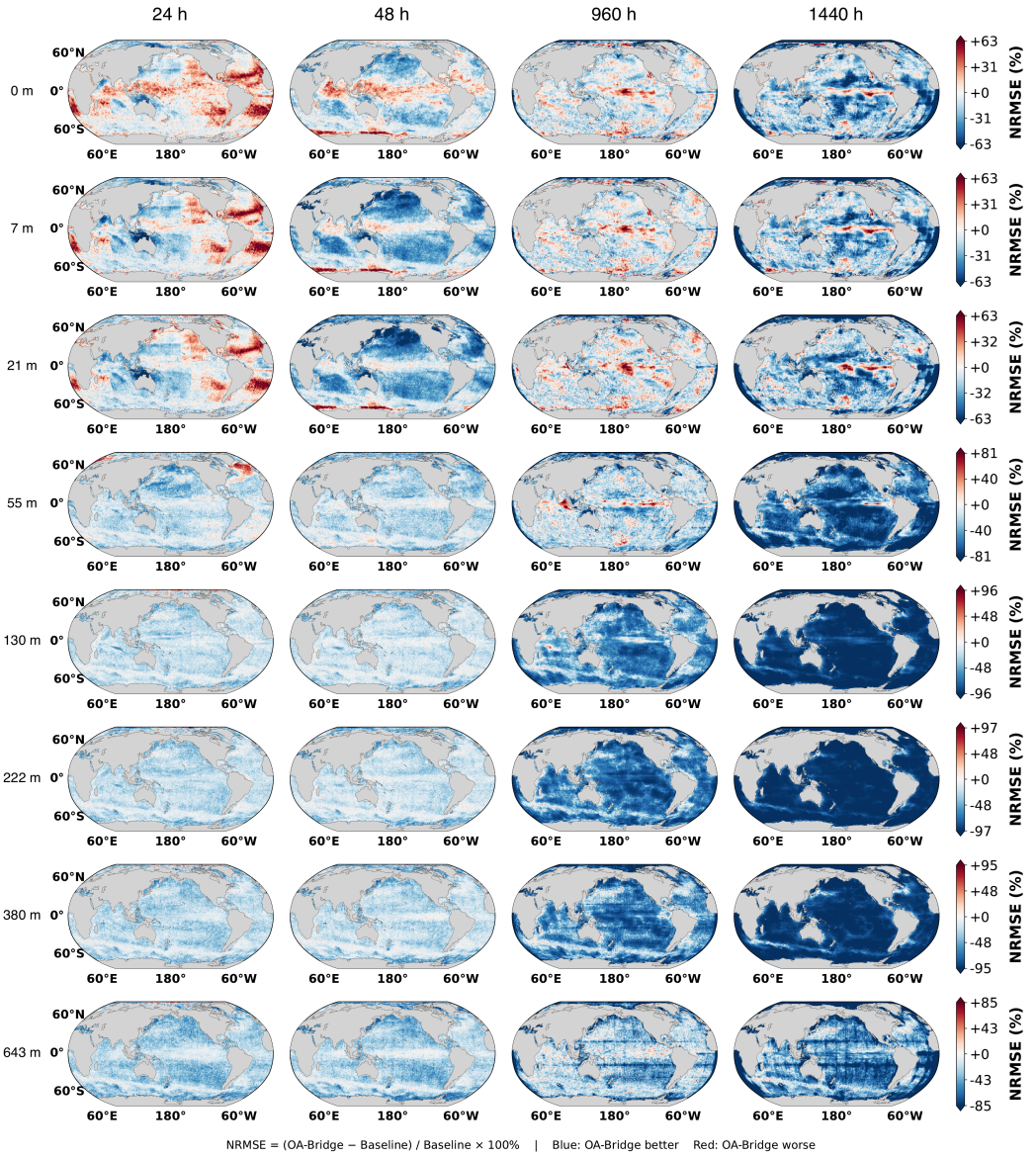

Figure 7: Spatial NRMSE (%) for ocean temperature relative to the baseline. Columns show 24, 48, 960, and 1440 h (1, 2, 40, and 60 days); rows show depths of 0, 7, 21, 55, 130, 222, 380, and 643 m. The definition and sign convention are the same as in Figure 6. Color limits vary by depth; blue indicates lower RMSE for OA-Bridge, and red indicates higher RMSE. Gray regions are masked.

Temperature skill varies more strongly with depth (Figure 7). At day 1, broad positive regions appear in the Atlantic at 0, 7, and 21 m, while much of the deeper ocean already has negative NRMSE. At days 40 and 60, improvements are extensive at 130, 222, and 380 m. Shallow fields remain less uniform, with positive bands persisting in parts of the equatorial Pacific even at day 60. These maps support the vertical error profiles in Figure 3: the coupled framework's oceanic gains are particularly widespread below the surface, while upper-ocean forecasts retain regional tradeoffs. The comparisons assess the full forecasting framework; the perturbation experiment in Appendix B separately examines its dependence on inputs from both components.