A data-driven global ocean forecasting model with sub-daily and eddy-resolving resolution

Yuan Niu[1,4†], Qiusheng Huang[2,3,4†], Xiaohui Zhong[2*], Anboyu Guo[5], Lei Chen[2,4], Xiaoyan Jia[1], Jiawei Qi[1], Dianjun Zhang[1*], Hao Li[2,3,4*], Xuefeng Zhang[1*]

[1]School of Marine Science and Technology, Tianjin University,Tianjin, 300072, China.
[2]Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai, 200433, China.
[3]Shanghai Innovation Institute, Shanghai, 200232, China.
[4]Shanghai Academy of Artificial Intelligence for Science, Shanghai, Shanghai, 200433, China.
[5]Department of Atmospheric Sciences, Fudan University, Shanghai, 200433, China.

*Corresponding author(s). E-mail(s): x7zhong@gmail.com; zhangdianjun@tju.edu.cn; lihao_lh@fudan.edu.cn; xuefeng.zhang@tju.edu.cn;
†These authors contributed equally to this work.

Abstract

High-fidelity ocean forecasting at high spatial and temporal resolution is essential for capturing fine-scale dynamical features, with profound implications for hazard prediction, maritime navigation, and sustainable ocean management. While conventional numerical models can generate sub-daily, eddy-resolving forecasts, they demand substantial computational resources and often struggle to maintain predictive skill at such fine scales. Data-driven models offer a promising alternative with significantly higher computational efficiency; however, most are constrained to daily outputs and show a rapid decay in accuracy when extended to sub-daily timescales.

Here, we introduce TianHai, the first-of-its-kind global data-driven 6-hour forecasting model, which delivers predictions at 1/12° eddy-resolving resolution with a vertical extent down to 1,500 m. A key feature of TianHai is the integration of atmospheric forcings through FuXi-Atmosphere, a data-driven atmospheric

forecasting system, which enables the explicit representation of air-sea coupling effects. Unlike conventional approaches, TianHai does not rely on numerical atmospheric models or external meteorological forecasts, making it a fully data-driven framework for coupled prediction.
Benchmark experiments demonstrate that TianHai delivers state-of-the-art performance in forecasting temperature and salinity profiles, zonal and meridional currents, sea surface temperature, and sea level anomalies for lead times ranging from 1 to 10 days. Relative to the latest deep learning based GOFS, it achieves accuracy improvements of 3.4%–19.3%. By advancing from daily to 6-hour coupled global forecasts with atmospheric forcing, TianHai represents a significant advance in AI-driven ocean prediction. Beyond improvements in statistical accuracy, it exhibits enhanced capability in resolving extreme events such as typhoons, as well as sub-daily oceanic phenomena including fronts, thermoclines, and mesoscale eddies, thereby paving the way for real-time, fine-scale Earth system monitoring and decision support.

1 Introduction

The ocean is a complex, nonlinear dynamical system exhibiting chaotic behavior, which fundamentally constrains its predictability and presents substantial challenges for accurate forecasting [1, 2]. Reliable ocean dynamics forecasting is essential for advancing our understanding of oceanic processes and their role in regulating the climate system [3]. One of the primary trends in the evolution of ocean forecasting technology is the pursuit of high spatiotemporal resolution, which enables more faithful simulation of key dynamical features such as ocean fronts, mesoscale and submesoscale eddies, and a variety of small- to medium-scale processes that strongly influence ocean circulation and biogeochemical cycles [4, 5].

Numerical global ocean forecasting systems (GOFS) predict ocean states by solving the governing equations of ocean dynamics [6]. Although they have achieved significant progress in capturing large-scale ocean dynamics [7–9], simultaneously achieving high temporal resolution while maintaining comprehensive global coverage remains a formidable challenge [10, 11]. This challenge primarily arises from two factors: first, resolving fine-scale variability across the global ocean requires immense computational resources, which limits the operational temporal and spatial resolution [12]; second, effectively assimilating heterogeneous, multimodal observational data into numerical models is highly complex [13–15]. Parameterizations of sub-grid-scale processes introduce uncertainties that can accumulate and degrade forecast skill. Although modern data assimilation techniques integrate diverse observational datasets, effectively incorporating heterogeneous data into models remains nontrivial. Together, these factors constrain the ability of traditional GOFS to accurately represent the present ocean state and reliably predict future changes, highlighting the need for new methodologies and enhanced computational frameworks.

At present, deep learning ocean forecasting technology is advancing rapidly, enabling the prediction of a wide range of ocean variables including satellite-derived

sea surface temperature(SST), sea surface height(SSH), wave height, temperature, salinity, 15-m zonal and meridional current [16–21]. Moreover, these methods have demonstrated capability in forecasting seasonal to interannual oceanic phenomena such as the Indian Ocean Dipole (IOD), the El Niño–Southern Oscillation (ENSO) [22–25] and climate-scale three-dimensional global ocean structures [26]. Deep learning approaches have emerged as promising alternatives to traditional ocean general circulation models (OGCMs), and have been successfully applied to global medium-range ocean forecasts with increasing accuracy [27–33]. Studies such as [34] demonstrate that these models can achieve comparable or even superior predictive skill at daily temporal resolution compared to operational numerical forecast systems, while improving computational efficiency by orders of magnitude. Despite these advances, existing data-driven ocean forecasting methodologies face inherent limitations that constrain their practical utility. Chief among these is the restriction of temporal resolution predominantly to daily intervals. A major reason is that most current approaches are trained on GLORYS reanalysis products, which provide only daily-mean fields and thus lack sub-daily variability. As a result, these models are inherently unable to resolve ocean processes evolving on sub-daily timescales. This limitation is particularly critical given that ocean parameters evolve over a wide range of timescales, with surface temperature exhibiting strong diurnal variability, whereas subsurface currents evolve more gradually over longer periods. Without mechanisms to both overcome the daily-mean constraint of training data and adaptively learn multiscale temporal dependencies, current models remain fundamentally limited in their ability to faithfully represent the full spectrum of ocean dynamics.

In this study, we introduce TianHai, the first deep learning global ocean forecasting system capable of producing six-hourly predictions at 1/12° eddy-resolving resolution, enabled by the utilization of HYCOM reanalysis and analysis data with six-hourly temporal resolution. The system addresses the challenges of sub-daily ocean forecasting through a variable-aware autoregressive architecture and the incorporation of near-surface atmospheric forcing, and, unlike WenHai and other GOFS, it can be deployed without relying on outputs from physical atmospheric models. TianHai delivers superior forecast skill for temperature and salinity profiles, zonal and meridional currents, sea surface temperature, and sea level anomalies, consistently outperforming both state-of-the-art numerical GOFS and the deep learning GOFS WenHai over 1–10 day lead times. Beyond improvements in traditional statistical metrics, TianHai's six-hourly predictions enable the resolution of sub-daily ocean phenomena, such as diurnal variability in surface temperature, fronts, thermoclines, and mesoscale eddies, as well as extreme events like typhoons. Moreover, it achieves a substantial increase in computational efficiency compared to traditional numerical GOFS, enabling faster and more scalable operational deployment.

2 Results

This study explores the root mean square error (RMSE) of TianHai and HYCOM analysis data between horizontal and vertical directions. By analyzing the fourth type of IV-TT framework data, the RMSE between TianHai and numerical forecasting

models, as well as the deep learning model WenHai, was compared. In addition, the predictive performance of the model was evaluated under extreme weather conditions.

2.1 Comparison with HYCOM GOFS3.1 analysis data

We utilize the HYCOM 2022 analysis dataset as the reference to conduct a comprehensive vertical evaluation of TianHai's forecast performance. Vertical profiles of oceanic variables are critical for thoroughly assessing the accuracy and physical realism of ocean prediction systems [35, 36]. In Figure 1, we present the depth-dependent variation of RMSE for both salinity and temperature.

Our analysis reveals several key insights. First, TianHai demonstrates robust and consistent predictive skill throughout the entire water column, with RMSE values increasing gradually with depth. This indicates the model's ability to capture vertical ocean structure across a wide range of depths. Second, the thermocline region, roughly between 100 and 300 meters, exhibits relatively higher errors compared to the surface and twilight zone layers, consistent with known challenges in predicting this dynamically complex interface [37, 38]. Nevertheless, despite this intrinsic difficulty, TianHai maintains RMSE values within acceptable and operationally useful bounds, underscoring its effective vertical generalization.

Furthermore, the evolution of forecast errors with increasing lead time exhibits a pronounced depth-dependent pattern. Leveraging the MoT module, which adaptively learns and weights temporal dependencies specific to each variable and depth, TianHai successfully mitigates error amplification over time. Surface and subsurface layers (0–300m), which are highly sensitive to rapidly varying conditions such as diurnal heating and atmospheric forcing [39], do not show disproportionately greater error growth compared to the deeper, more stable ocean layers beyond 500 m [40]. This results in a relatively uniform error growth profile across depths, highlighting the MoT module's effectiveness in balancing multi-timescale dynamics and enhancing forecast stability.

In addition to thermodynamic variables, we assess TianHai's performance in predicting ocean currents, specifically 0-1500 m zonal current and meridional current, with RMSE results displayed in Figure 1. These evaluations are conducted using the HYCOM 2022 analysis data as the reference. The model consistently achieves high accuracy and demonstrates strong stability in error accumulation for ocean current forecasts. As the forecast lead time increases, errors grow gradually and evenly without abrupt degradation, and there are no substantial differences in RMSE between intra-day (e.g., 6-hour) and daily forecast intervals. Particularly noteworthy is the 6-hour forecast horizon, where the vertical stratification of ocean current variability near the surface becomes apparent. This reflects the pronounced influence of recent and rapidly evolving forcings such as sea surface winds [41], which modulate near-surface current changes on short timescales and are effectively captured by TianHai's coupled framework [42].

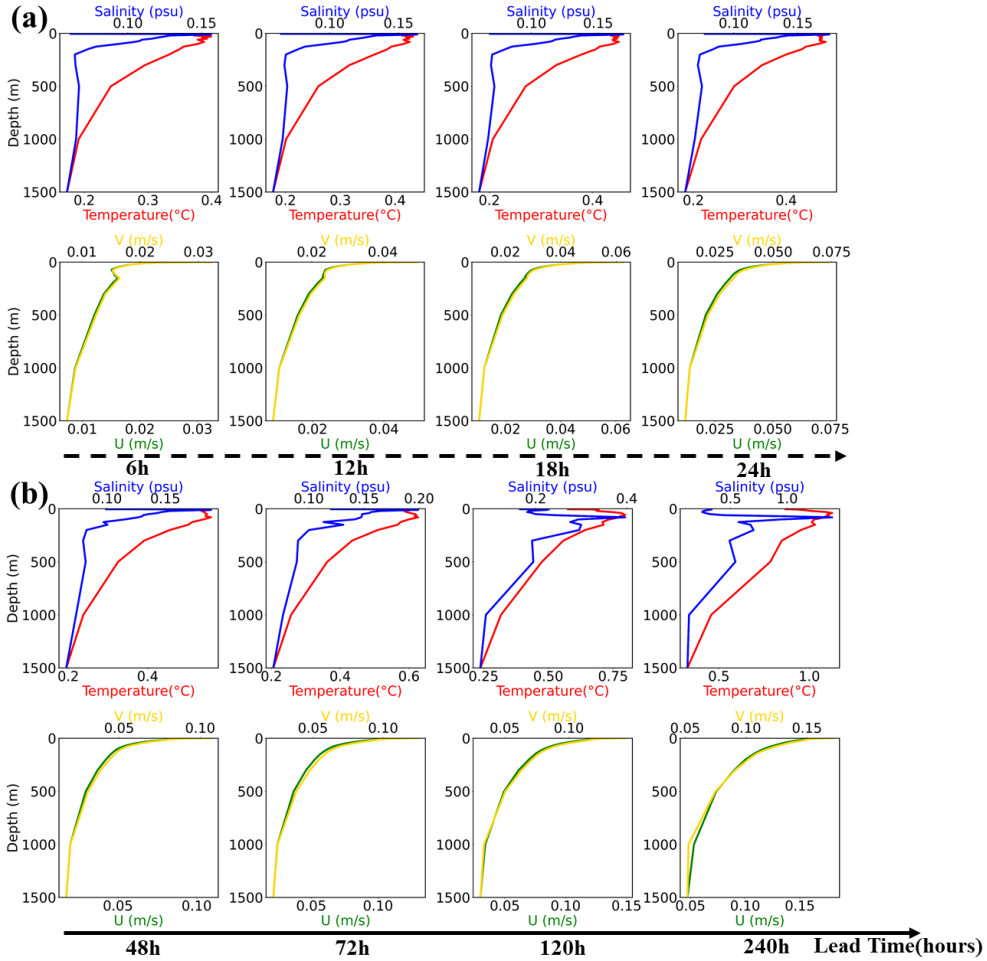

Fig. 1: **Depth-dependent RMSE distributions of salinity, temperature, 15-m zonal current and 15-m meridional current) between TianHai forecasts and HYCOM analysis data for the year 2022.** RMSE profiles (0–1500m) for each lead time within the 24-hour forecast window are shown in panel (a), while profiles for lead times ranging from 48 to 240 hours are shown in panel (b). In both panels, lower RMSE values indicate better agreement with the reference analysis. Salinity is shown in blue, temperature in red, the 15-m zonal current in green, and the 15-m meridional current in gold.

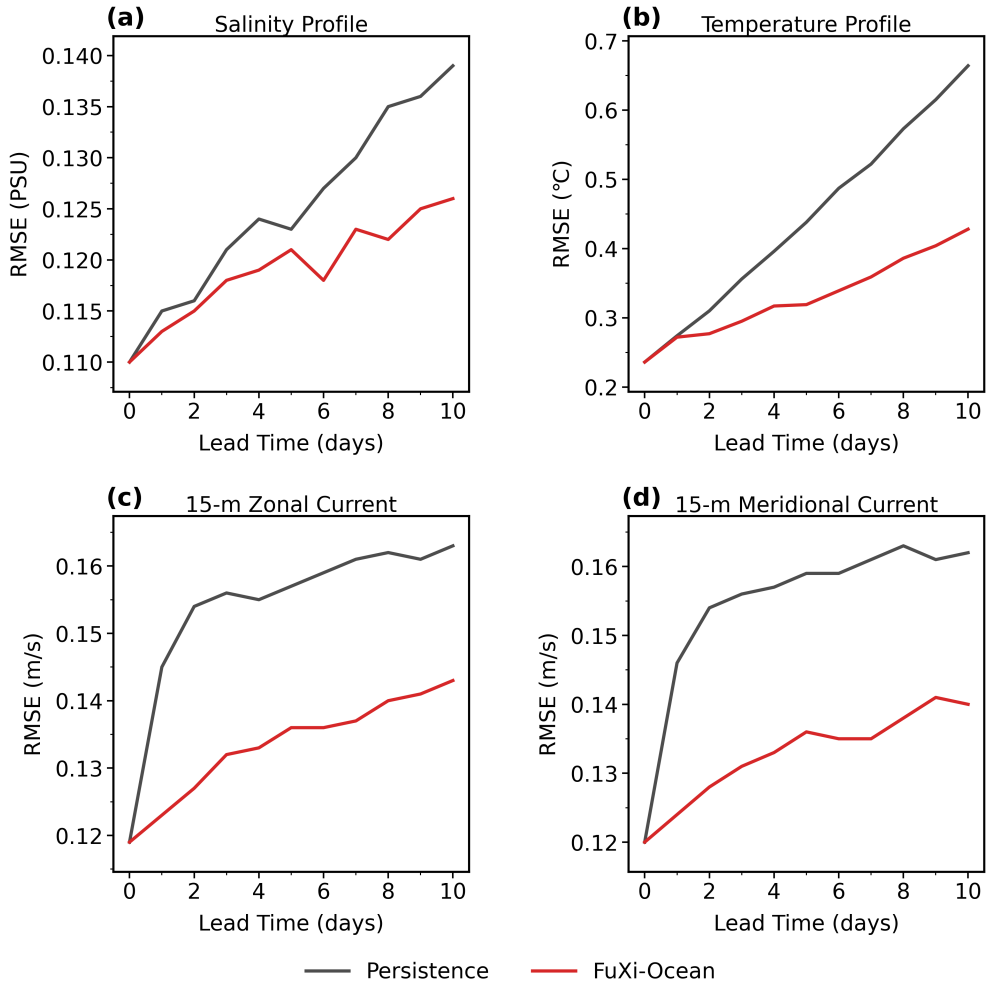

Fig. 2: **Comparison of root mean square error (RMSE) between TianHai and persistent forecast as a function of forecast lead time.** Globally averaged RMSE (the lower, the better) of the forecast temperature profile (a), salinity profile (b), 15-m zonal current (c), and 15-m meridional current (d) as a function of forecast lead time. The zero-lead time represents the initial conditions. The red and grey lines correspond to TianHai and persistent forecast, respectively. For the temperature and salinity profile forecast, the RMSE is vertically averaged over the upper 1500 meters.

Figure.B2 and B3 illustrates the global spatial distribution of RMSE for key temperature and salinity at multiple forecast lead times (6, 24, 120, and 240 hours). For SST, TianHai consistently maintains low error magnitudes in the tropical and subtropical ocean basins throughout all forecast horizons. At the 6-hour lead time, RMSE values generally remain below 0.3°C, demonstrating the model's strong skill in these dynamically stable regions. However, error intensities notably increase in regions dominated by western boundary currents, such as the Gulf Stream and Kuroshio Current, as well as the Antarctic Circumpolar Current (ACC). These areas are characterized by vigorous mesoscale eddy activity and complex dynamical interactions, which present intrinsic predictability limits for ocean forecasting systems [43–45]. The chaotic nature of these eddy-rich regions complicates both model predictions and observational consistency, as satellite retrievals and in situ measurements often exhibit increased uncertainty there [46]. Such inherent variability poses a formidable challenge for any forecasting framework aiming to capture fine-scale ocean dynamics. Salinity forecast errors exhibit spatial patterns broadly similar to temperature but with pronounced relative increases in areas subject to substantial freshwater forcing. These include major river discharge zones such as the Amazon and Congo river plumes, and regions experiencing intense precipitation or meltwater input [47–49]. Despite these challenges, the model performs robustly in the equatorial Pacific region, sustaining low salinity RMSE values even for extended forecast lead times up to 10 days. This region's strong stratification and relatively stable salinity regimes facilitate forecast skill [50]. Moreover, TianHai maintains operationally acceptable accuracy levels in the more challenging riverine and high-precipitation areas, underscoring its resilience and practical utility for marine applications requiring freshwater flux monitoring.

While the initial evaluation compared TianHai forecasts directly against HYCOM analysis products, this approach alone does not provide a clear baseline for judging relative forecast skill. Analysis fields, although used as verification datasets, are themselves the product of model-data assimilation and contain inherent uncertainties that can vary spatially and temporally [51]. Consequently, a direct comparison between the forecast and analysis primarily reflects differences between two model estimates, which may not fully capture the absolute skill of the forecast system. To address this, we introduce a persistence forecast as a simple and widely used benchmark in forecast verification [52, 53]. The persistence assumption posits that the ocean state at the initial time will remain unchanged throughout the forecast period, effectively representing a "no change" scenario [54]. Despite its simplicity, persistence often provides surprisingly competitive skill for variables with high temporal autocorrelation, especially at short lead times [55]. As such, it serves as a physically interpretable reference against which the added value of a forecasting system can be quantified. Models that fail to outperform persistence are generally considered to have limited practical utility for operational or scientific applications. Including this benchmark thus enables a more robust and interpretable assessment of TianHai's predictive advantage across variables, spatial domains, and forecast lead times.

The RMSE for temperature profile, salinity profile, 15-m zonal current, and 15-m meridional current as a function of forecast lead time is illustrated in Figure 2, which clearly shows the TianHai's significant improvement over persistence forecasts.

This improvement aligns with recent studies where machine learning models showed superior predictive skill relative to naive baselines, especially at short lead times. Globally averaged RMSE for upper ocean temperature and salinity profiles, vertically integrated over the upper 1500 meters, reveal that data-driven approaches can capture thermohaline variability more accurately than persistence. Although surface currents generally present higher variability and forecasting challenges, the TianHai model reduces RMSE compared to persistence, consistent with results demonstrating improved ocean current forecasts using neural networks [56]. RMSE increases with forecast lead time for all variables, reflecting the intrinsic growth of forecast uncertainty, but the FuXi model maintains a clear advantage. The zero-lead time RMSE corresponds to initial condition errors, indicating that assimilation quality critically impacts forecast accuracy [57].

2.2 Outperformance of TianHai against numerical and deep learning GOFS

To rigorously assess the predictive skill of TianHai, we conducted a comprehensive comparison against three internationally recognized numerical GOFS: the Hybrid Coordinate Ocean Model (HYCOM) [58, 59], the BLUElink Ocean Model Analysis and Prediction System (OceanMAPS) from the Australian Bureau of Meteorology [60], and the Forecast Ocean Assimilation Model (FOAM) operated by the UK Met Office [61].

We evaluated TianHai against both a conventional numerical model and the large-scale WenHai model using the IV-TT Category IV benchmark dataset. The assessment covers six key variables: sea surface temperature (SST), temperature and salinity profiles, 15-m zonal and meridional current, and sea level anomalies (SLA). The evaluation used independent in situ observations from Argo profiling floats during 2022 [62], providing vertically resolved measurements of temperature and salinity profiles as ground truth. Within the IV-TT framework, we extracted daily gridded forecasts of temperature and salinity from each system. For temporal consistency, TianHai's native 6-hour outputs were averaged to produce daily mean fields, matching the daily forecast cycles of the benchmark GOFS. The primary quantitative metric was the RMSE, computed against the daily Argo-derived fields. To separate forecast skill from initial condition accuracy, the RMSE was calculated both at the initial forecast time (day 0) and over subsequent lead times.

As shown in Figure 3, TianHai consistently achieved lower RMSE values than HYCOM, OceanMAPS, and FOAM for both temperature and salinity profiles. At Day 0, TianHai's RMSE was significantly smaller, indicating a more accurate initial state estimation, this may be related to the initial field data we selected for the wind field. More importantly, RMSE growth over time was slower, implying better representation of large-scale and mesoscale dynamics, and reduced amplification of uncertainties during forecast integration [63, 64]. The slower error growth is a crucial factor in extending the useful forecast horizon for operational ocean applications.

In addition to the comparison with numerical GOFS, we evaluated TianHai against the latest deep learning global ocean forecasting system, WenHai [20]. Unlike

WenHai, which uses multi-decade historical training data (1993–2020) at 24-hour resolution, TianHai was trained on a shorter but higher-frequency dataset (2006–2019) at 6-hour resolution. Moreover, TianHai explicitly incorporates atmospheric forcing variables—10-meter zonal and meridional wind components (U_{10}, V_{10})—in addition to oceanic state variables (T, S, 15-m zonal current, 15-m meridional current, SSH). The inclusion of accurate wind initial fields is known to reduce spin-up time and initial forecast errors, while improving the simulation of wind-driven ocean currents, particularly under rapidly evolving weather conditions such as tropical cyclones and strong convective systems [65]. For the 2022 testing data, TianHai demonstrated substantial improvements over WenHai across multiple variables. TianHai outperformed WenHai in both temperature and salinity profile predictions, achieving consistently lower RMSE. For currents, TianHai reduced RMSE by 19%, indicating better simulation of momentum balances, subgrid-scale processes, and wind stress coupling [5, 66].

At initialization, TianHai consistently outperformed numerical and deep learning GOFS, achieving lower RMSE across all variables. While its advantage narrowed over time—and was occasionally surpassed by WenHai at specific lead times—TianHai maintained superior overall accuracy throughout the 10-day forecast window. To quantify this, we computed the time-averaged RMSE for each model and variable. TianHai demonstrated better mean performance than both the numerical model and WenHai across most variables. Notably, WenHai achieved the lowest average RMSE for SST and meridional flow velocity.

Over the 10-day horizon, TianHai reduced RMSE relative to WenHai by 14.4% for SST, 6.7% for SLA, 8.1% for temperature profile, 3.4% for salinity profiles, 19.3% for 15-m zonal velocity, and 19.2% for 15-m meridional velocity. These results, detailed in Supplementary Fig.B4, highlight the model's strong and stable performance across both surface and subsurface ocean parameters.

Overall, TianHai outperformed both numerical GOFS and the latest deep learning GOFS, achieving lower errors in core oceanographic variables. This consistent improvement across temperature profile, salinity profile, 15-m zonal velocity, 15-m meridional velocity, SST, and SLA underscores TianHai's potential as a reliable operational forecasting tool and a valuable resource for climate research and marine applications.

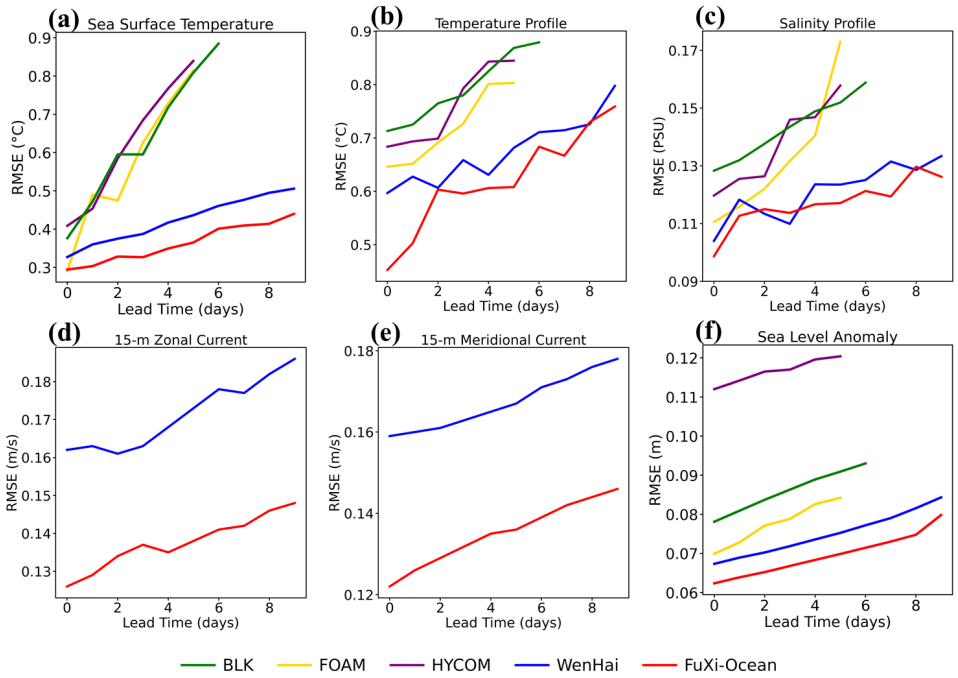

Fig. 3: **Comparison of RMSE for multiple variables between TianHai, the numerical GOFS, and the deep learning GOFS over a 10-day forecast horizon.** Panels show global mean RMSE as a function of forecast lead time (lower values indicate better performance) for sea surface temperature (a), temperature profile (b), salinity profiles (c), 15-m zonal current (d), 15-m meridional current (e), and sea level anomaly (f). The zero lead time corresponds to the initial condition. TianHai is inherently a six-hourly forecasting system; for a fair comparison with daily-output models, its forecasts were averaged to daily resolution prior to evaluation. Due to the absence of 15-m zonal and meridional current data in the 2022 IV-TT dataset, the velocity comparison is conducted only between TianHai and WenHai. This figure highlights how forecast skill degrades with increasing lead time and underscores TianHai's relative advantages across multiple ocean state variables.

2.3 Accuracy testing of forecasts under extreme weather conditions

Extreme weather events such as typhoons pose substantial challenges for ocean forecasting systems due to their rapid evolution, strong wind forcing, and intense air-sea interactions. These events often lead to sharp sea surface temperature (SST) changes, vertical mixing, and dynamic circulation responses that occur on sub-daily timescales. Accurate and timely SST forecasts under such conditions are crucial for disaster preparedness, marine operations, and understanding upper ocean thermodynamics. Therefore, evaluating a forecast system's performance during typhoon conditions is

essential to assess its ability to capture high-frequency ocean responses and to support operational readiness. To this end, we examine the SST forecasting performance of TianHai during the passage of Typhoon Aere in 2022, with a focus on comparing its high-frequency (6-hour) forecasts against daily-resolution models and observational analyses. During the passage of Typhoon Aere, the 6-hour stepwise SST forecasts provided by TianHai demonstrated significant advantages in capturing rapid ocean responses to extreme weather. Compared with the daily SST forecasts produced by the WenHai model, TianHai was able to more accurately characterize the sharp changes in sea surface temperature induced by typhoon activity, including localized cooling effects caused by upwelling and intense vertical mixing.

Figure B10 shows the track of Typhoon Aere in 2022. The typhoon originated as a tropical disturbance east of the Philippines on June 30, moved across the Ryukyu Islands into the East China Sea by July 2, and made landfall in Kyushu, Japan, on the morning of July 5 before dissipating. The storm passed through regions where it significantly altered the thermal structure of the upper ocean. Figure 4 presents the SST forecast results from TianHai and WenHai during the typhoon's impact. The black line marks the typhoon's path. From left to right, the panels show: TianHai's SST forecasts at different lead times (6h, 12h, 18h, 24h), WenHai's daily SST forecast, HYCOM-analyzed SST, and the corresponding RMSE distributions. The HYCOM SST data used for comparison were both raw and daily-averaged (to match WenHai's temporal resolution), and were interpolated to WenHai's sea surface depth (0.49 m).

The TianHai model's higher temporal resolution enables it to resolve the sub-daily evolution of SST under rapidly changing atmospheric conditions. The RMSE distributions clearly show that TianHai outperforms WenHai in the vicinity of the typhoon path and its impact region, with lower errors and better-defined spatial gradients. In contrast, WenHai's daily-averaged forecasts exhibit larger errors and fail to capture transient cooling events, mainly due to the smoothing effect of temporal averaging.

This result underscores the critical necessity of high-frequency (e.g., 6-hour) ocean forecasting in the context of extreme weather events like typhoons. Such events evolve on sub-daily timescales and interact with the ocean in complex and rapidly changing ways. Only models with high temporal resolution can adequately capture the short-term dynamics of ocean-atmosphere coupling, such as sudden upwelling, rapid SST drops, and intensified surface mixing.

Moreover, the 6-hour forecast interval is better aligned with the requirements of modern operational oceanography, which increasingly demands timely, accurate, and high-resolution forecasts for applications such as disaster early warning, maritime safety, and air-sea flux estimation. Daily-average models like WenHai, while useful for climatological assessment, lack the responsiveness needed in dynamic, real-time ocean forecasting scenarios. The performance of FuXi Ocean during Typhoon Aere exemplifies its ability to dynamically track and respond to high-impact atmospheric disturbances with enhanced temporal fidelity and physical realism. These strengths make the model more suitable for operational deployment in marine forecasting systems, particularly for events requiring rapid situational awareness and proactive response.

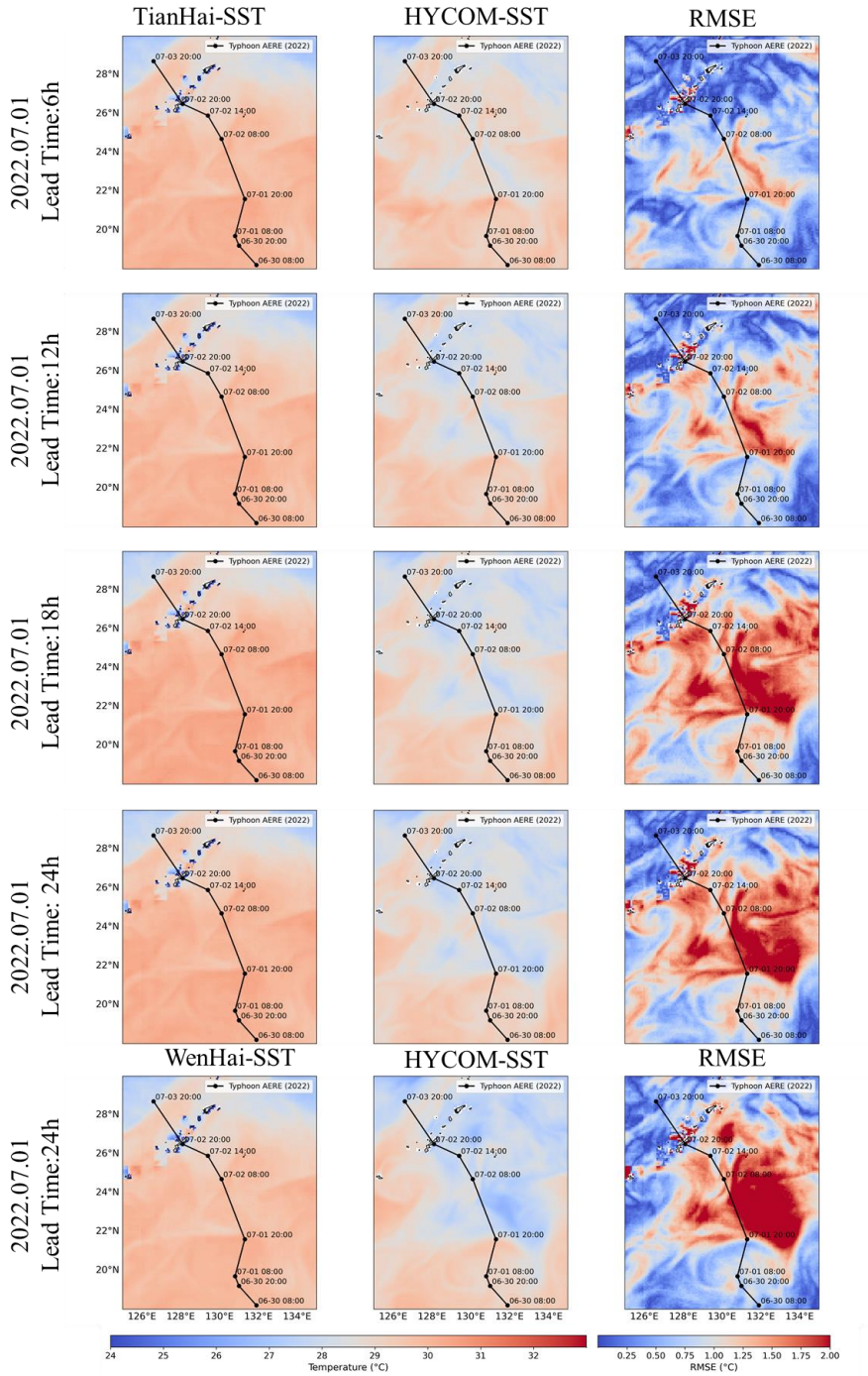

Fig. 4: **SST forecast from TianHai under the influence of Typhoon Aere.** The black line indicates the typhoon's track. From left to right, the panels show sea surface temperatures (SST) forecasted by TianHai and WenHai, followed by SST and RMSE calculated from HYCOM data. From top to bottom, the SST forecasts from TianHai are shown at lead times of 6, 12, 18, and 24 hours, while WenHai provides daily forecasts at a 1-day lead time.

3 Discussion

We introduce TianHai, the first data-driven global ocean forecasting system capable of producing six-hourly predictions at a 1/12° spatial resolution, spanning from the surface down to 1500 meters depth. This high-resolution, sub-daily forecasting system represents a significant advancement over existing models in both temporal granularity and vertical coverage, essential for accurately resolving mesoscale and upper-ocean dynamics [4]. A key innovation in TianHai is its hybrid MoT module, which adaptively integrates multi-timescale dependencies of ocean variables, overcoming the major limitation of previous AI models constrained by fixed temporal resolutions. Through comprehensive benchmarking against state-of-the-art numerical GOFS such as HYCOM [59] and the latest deep learning based GOFS where available, TianHai demonstrates superior predictive skill across critical variables including SST, temperature and salinity profiles, 15-m zonal and meridional current velocities, and sla over forecast horizons extending up to ten days. Notably, TianHai achieves this while requiring substantially fewer computational resources, largely attributable to its data-driven framework and efficient temporal modeling. The system further incorporates oceanic variables as input to constrain the initial wind field, effectively reducing initial forecast errors and improving short-term forecast accuracy—a crucial factor often overlooked in prior data-driven ocean models. Remarkably, the six-hourly predictions by TianHai surpass the accuracy of daily-averaged forecasts generated by advanced numerical models, validating the efficacy of high-frequency deep learning ocean forecasting methods [35].

Despite these promising achievements, several limitations merit further investigation. First, TianHai's training relies predominantly on reanalysis data, such as HYCOM's historical fields, which inherently carry biases from the underlying numerical models. The scarcity of publicly available, high-quality sub-daily ocean reanalysis datasets further constrains the robustness of model training. To mitigate this, future efforts will focus on integrating HYCOM analysis with in-situ observations from Argo floats and other autonomous platforms to build a more comprehensive and bias-reduced training dataset [36]. Additionally, while TianHai maintains robust forecast skill through ten-day lead times, its performance on longer-term forecasts spanning seasonal to interannual scales remains unexplored, presenting a critical avenue for future research.

The success of deep learning in sub-daily ocean forecasting opens several exciting future directions. Extending vertical coverage beyond 1500 m into abyssal depths would allow more complete ocean system representation, crucial for climate and carbon cycle studies. Incorporating physical constraints, such as conservation laws of mass, momentum, and energy, within the model architecture could enhance prediction stability and physical consistency. Moreover, increasing forecast frequency toward hourly or even higher temporal resolution would benefit applications requiring fine-scale temporal fidelity, including tidal forecasting and coastal hazard warning systems.

Our MoT module demonstrates strong adaptive learning of variable-specific temporal reliabilities. Experimental results demonstrate state-of-the-art performance for relatively stable variables, such as temperature, salinity, and sea surface height. However, MoT is less effective for fast-changing, strongly physics-driven variables subject

to spatial errors, notably 15-m zonal and meridional current. This results in relatively muted gains in current velocity predictions compared to thermohaline variables. Nonetheless, this limitation does not diminish the overall impact of our approach. We anticipate that augmenting MoT with physical constraints—such as enforcing momentum conservation and boundary conditions—will better capture the complex dynamics of ocean currents and further enhance model accuracy [5, 66]. Exploring these physics-informed hybrid modeling approaches is a natural and promising extension of this work.

4 Methods

4.1 HYCOM dataset

The HYCOM is a state-of-the-art, high-performance ocean circulation model jointly developed by multiple research institutions in the United States, including the Naval Research Laboratory (NRL), the University of Miami, and the Los Alamos National Laboratory [59]. HYCOM employs a hybrid vertical coordinate system, seamlessly transitioning between isopycnal (density-following), z-level (depth-following), and sigma (terrain-following) coordinates. This design allows the model to efficiently represent both open-ocean stratification and complex coastal bathymetry, minimizing spurious numerical mixing and improving thermohaline structure representation.

The model supports high horizontal resolutions, reaching 1/12° (9 km) or finer in operational configurations, with up to 40 vertical layers. HYCOM integrates a wide range of observational data—such as satellite-derived sea surface temperature (SST), sea surface height anomalies (SSHA), and ocean color, along with in-situ measurements from Argo profiling floats, moored buoys, and ship surveys—through advanced data assimilation systems (e.g., the Navy Coupled Ocean Data Assimilation, NCODA). This assimilation framework significantly enhances the accuracy of model fields by continuously constraining them to match real-world observations.

HYCOM produces comprehensive global and regional ocean analyses and forecasts, delivering key physical parameters including three-dimensional velocity fields, mixed layer depth, thermocline structure, temperature and salinity distributions, and surface currents. Its temporal resolution can range from hourly forecasts to monthly climatological means, enabling multi-scale research and applications.

Because of its robust design and broad data coverage, HYCOM has been extensively applied in ocean circulation studies, climate system simulations, and marine hazard prediction—including tropical cyclone impact assessment, storm surge modeling, and search-and-rescue drift predictions. It also plays a critical role in naval operations, supporting submarine routing, acoustic propagation studies, and maritime situational awareness. In recent years, HYCOM has served as the backbone for numerous GOFS, including the U.S. Navy's operational ocean prediction capability.

4.2 TianHai Model

TianHai employs an autoregressive architecture designed to capture the multiscale temporal dynamics of oceanic variables. The model comprises three primary components: a feature extraction module that encodes multi-temporal inputs, a prediction module that captures temporal evolution patterns, and a feature remapping module that reconstructs the target variables. TianHai addresses the challenging task of forecasting future oceanic states by leveraging sequential historical observations with both high spatial and temporal fidelity. Formally, given a sequence of historical observations spanning N consecutive time points $\{\mathrm{X}^{t-N+1}, \mathrm{X}^{t-N+2}, \ldots, \mathrm{X}^{t}\} \in \mathbb{R}^{N\times C\times H\times W}$, we aim to predict the state at the subsequent time step $\mathrm{X}_{t+1} \in \mathbb{R}^{C\times H\times W}$. Here, $H = 2160$ and $W = 4320$ represent the spatial dimensions of our global grid at $1/12^{\circ}$ resolution, providing eddy-resolving capability essential for capturing mesoscale ocean dynamics.

The model focuses on five fundamental oceanic variables that comprehensively characterize the physical state of the global ocean: temperature profile(T), salinity profile (S), 15-m zonal and meridional current, and sea surface height (SSH). Except for SSH, which is defined at the ocean surface, the remaining four variables exhibit significant vertical heterogeneity, necessitating predictions across multiple depth levels to fully capture the vertical structure of ocean dynamics. The vertical domain is discretized into 26 strategically chosen depth layers ranging from the surface (0 m) down to 1500 m, specifically at: 0, 2, 4, 6, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 200, 300, 400, 500, 600, 800, 1000, 1250, and 1500 meters. This vertical resolution is particularly fine near the surface to resolve upper-ocean processes that strongly influence air-sea interactions and biogeochemical exchanges.

In total, the TianHai model ingests 107 variables, including one sea surface variable (SSH), four ocean variables each spanning 26 vertical layers (T, S, 15-m zonal and meridional current), and two atmospheric surface variables corresponding to the zonal and meridional components of sea surface wind. The oceanic inputs are derived from the HYCOM analysis and reanalysis datasets, while the atmospheric wind fields come from the initial condition fields of FuXi's atmospheric component, providing critical information to capture ocean-atmosphere coupling effects and forcing mechanisms.

The temporal resolution is fixed at 6-hour intervals, consistent with operational ocean forecasting standards and striking a balance between temporal detail and computational practicality. Through empirical analysis, we find that a historical window of $N = 4$ consecutive time steps, equivalent to 24 hours of past data, sufficiently captures key temporal patterns such as diurnal cycles and inertial oscillations, while maintaining manageable computational complexity. To ensure physical consistency and avoid spurious predictions over land, a land-sea mask is applied uniformly across all input and output variables, restricting computations solely to valid ocean grid points.

TianHai's architecture is built around an autoregressive framework tailored to model the multiscale temporal dependencies inherent in ocean dynamics. The system comprises three main components: (1) a feature extraction module that encodes spatiotemporal information from the historical input sequence, (2) a predictive module designed to learn temporal evolution patterns and inter-variable interactions, and (3) a feature remapping module responsible for reconstructing the predicted ocean state variables at the target forecast time. An overview of the architecture is illustrated

in Figure 5, highlighting the model's end-to-end pipeline and its capacity for high-fidelity, multi-variable ocean forecasting.

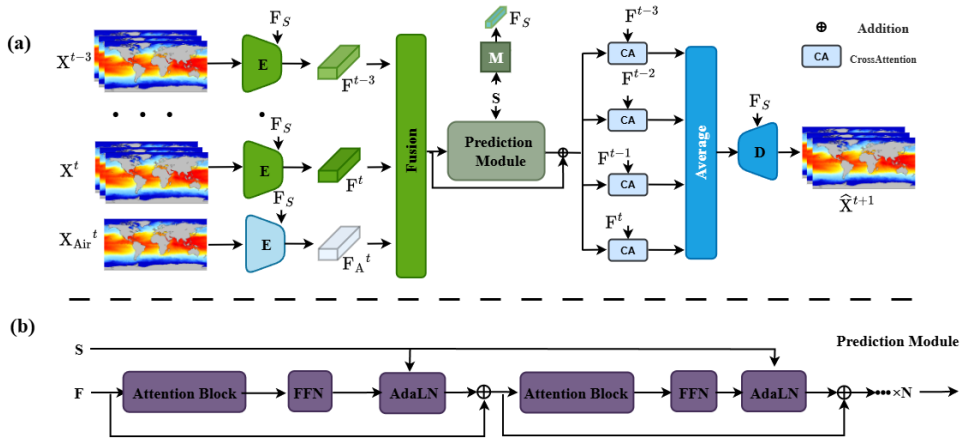

Fig. 5: **Architecture of our ocean forecasting framework.** Our model processes sequential ocean states X^{t-3} through X^{t} to predict $\widehat{\mathrm{X}}^{t+1}$. (a) The main pipeline consists of a shared encoder **E** that transforms input states into latent representations, modulated by spatiotemporal features F_S from network **M**. The fused representations feed into the prediction module, whose outputs are processed by decoders **D** with skip connections. (b) The prediction module employs stacked attention blocks with adaptive layer normalization (AdaLN) and feed-forward networks (FFN), capturing complex temporal dynamics.

Feature extraction module. Ocean forecasting requires robust feature representations that capture both spatial dependencies and temporal correlations. We design a feature extraction pipeline that combines context-aware encoding with efficient feature fusion, shown in Fig. 5(a).

Our design employs a shared encoder **E** for patch embedding. This encoder uses convolutional layers with matched kernel size and stride for spatial downsampling, followed by layer normalization. This approach balances computational efficiency with representation power when processing global ocean data. To leverage spatiotemporal context, we implement a prior information network **M** that processes:

$$\mathrm{F}_S = \mathbf{M}(\mathrm{S}) \tag{1}$$

where S contains temporal information (diurnal, seasonal patterns) and spatial data (coordinates, bathymetry). The network **M** uses sinusoidal positional encodings for temporal components and learnable embeddings for spatial coordinates. We then modulate the encoder weights using these contextual features through:

$$\mathrm{F}^t = \mathbf{E}(\mathrm{X}^t, \mathrm{F}_S) = \mathbf{Norm}(\mathbf{Conv}(\mathrm{X}^t, \mathrm{W} \odot \mathrm{F}_S)) \tag{2}$$

where W represents the learnable convolution parameters and $\odot$ denotes element-wise multiplication. This modulation enhances the encoder's sensitivity to region-specific patterns, such as boundary currents or seasonal mixed layer dynamics. The encoder processes each input time step $t-i$ ($i \in \{0,1,2,3\}$) to produce feature tensors $\mathrm{F}^{t-i} \in \mathbb{R}^{C' \times H' \times W'}$. Here, C' is the feature dimension while H' and W' represent the downsampled spatial dimensions.

Our feature fusion module concatenates these tensors along the channel dimension, preserving their temporal characteristics. We then apply 1×1 convolutions with normalization layers to integrate information while maintaining spatial coherence. This fusion approach enables the model to capture complex spatiotemporal patterns essential for accurate forecasting, without compromising computational efficiency.

Prediction Module. Building upon the extracted features, our prediction process involves a a shared decoder for feature reconstruction, and a novel MoT module that adaptively integrates multi-temporal predictions across multiple temporal scales, as illustrated in Figure 5(b).

The prediction module processes the fused representations to model the non-linear evolution of oceanic states, which consists of stacked attention blocks[67] paired with feed-forward networks. Each attention block captures dependencies across spatiotemporal features, while adaptive layer normalization (AdaLN)[68] incorporates contextual information F_S to modulate normalization parameters. This design allows the model to focus selectively on relevant oceanic patterns while maintaining computational efficiency.

For feature reconstruction, we employ a shared decoder $\mathbf{D}$ that transforms latent representations back into the physical variable space. The decoder consists of transposed convolutional layers with normalization operations. We establish skip connections between encoder features and corresponding decoder layers, preserving fine-grained spatial details often lost during prediction—a critical factor when modeling mesoscale ocean dynamics.

Training Strategy. Training deep learning models for ocean forecasting presents unique challenges due to Earth's spherical geometry. To address the varying grid cell areas across latitudes, we implement a latitude-weighted Charbonnier loss[69]:

$$\mathcal{L}_{\text{pred}} = \frac{1}{C \times H \times W} \sum_{c=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} \alpha_i \sqrt{(\widehat{\mathrm{X}}_{c,i,j}^{t+1} - \mathrm{X}_{c,i,j}^{t+1})^2 + \epsilon^2} \tag{3}$$

where $\alpha_i = H \times \frac{\cos \Phi_i}{\sum_{i=1}^{H} \cos \Phi_i}$ is the latitude-dependent weighting factor at latitude Φ_i, and ϵ is a small constant ensuring differentiability. This approach prevents bias toward high-latitude regions in error calculations, which would otherwise be overrepresented in the rectangular grid projection.

Besides, we employ a two-stage training approach to enhance model stability and performance. In the first stage, we pre-train the model using a single-step prediction objective. In the second stage, we fine-tune the model with a multi-step loss

that penalizes predictions across multiple consecutive time steps, mitigating error accumulation in autoregressive forecasting. This approach follows recent advances in autoregressive modeling [20, 70], which demonstrate the effectiveness of multi-step training for improving forecast stability.

Training and validation datasets. We train and evaluate TianHai using the HYCOM analysis and Reanalysis datasets [71], which is currently the only publicly available global ocean dataset offering 6-hour temporal resolution. HYCOM provides a $1/12^{\circ}$ horizontal resolution with up to 40 vertical layers, spanning from the sea surface to a depth of 5,000 m. For this study, we utilize approximately 14 years of data (January 2006 to June 2019) for model training, one year (January 2020 to December 2020) for validation, and one year (January 2021 to December 2021) for testing. To balance computational efficiency with the need for capturing critical subsurface processes, we focus on 26 strategically selected vertical layers extending from the surface to 1,500 m depth, a range that effectively resolves key ocean dynamics such as mixed-layer variability, thermocline structure, and upper-ocean circulation patterns.

Land-sea mask. To ensure physically consistent predictions and properly handle missing values over non-ocean regions, TianHai employs a land–sea masking strategy throughout the entire forecasting process. At each time step, all grid points corresponding to land are forcefully reset to zero for all ocean variables, effectively removing spurious values that could otherwise arise from numerical artifacts. This approach allows the model to focus computational and representational capacity exclusively on oceanic regions, rather than wasting resources on irrelevant land points.

Unlike several existing deep learning GOFS that apply a uniform sea-surface mask for all depth layers, TianHai incorporates depth-dependent masks tailored to each vertical level. This design more accurately reflects the true three-dimensional bathymetric structure of the ocean floor, ensuring that predictions at deeper layers respect the underlying seabed topography and coastline geometry. As a result, the model is better able to reproduce realistic subsurface features, such as continental shelf breaks, seamounts, and deep-ocean basins, which play a critical role in shaping local and regional circulation patterns.

This layer-specific masking mechanism not only improves the physical plausibility of subsurface predictions but also helps prevent the propagation of erroneous values into deeper layers during autoregressive forecasting, thereby enhancing both the stability and accuracy of long-lead forecasts.

4.3 Observational datasets

The observational data from Argo is not provided on a regular grid; instead, it consists of irregularly spaced profiles collected by autonomous profiling floats. Globally, the Argo network maintains a relatively uniform and comprehensive distribution of stations, particularly in key oceanic regions, ensuring robust coverage for large-scale climate and ocean monitoring.

For temperature and salinity profiles, we utilize quality-controlled (QC) near–real-time Argo observations from the E.U. Copernicus Marine Service Information (CMEMS). These profiles undergo standardized real-time QC procedures to remove spurious values and ensure reliability for operational validation.

For sea level anomaly (SLA), we use an along-track Level-3 near–real-time product from AVISO, with data provided by the Ssalto/Duacs altimetry system and distributed through CMEMS (http://www.marine.copernicus.eu). This product combines multi-mission satellite altimeter measurements, including Sentinel-6A-HR, Jason-3 interleaved, Sentinel-3A, Sentinel-3B, SARAL-DP/AltiKa, CryoSat-2, HaiYang-2B, and SWOT-nadir. We adopt the filtered version of the dataset to minimize high-frequency noise and instrumental errors, improving signal quality for mesoscale and large-scale ocean dynamics.

For SST and 15-m zonal and meridional current, we use near–real-time drifter observations from the Global Drifter Program (GDP). Similar to the Argo and SLA datasets, only records that pass all real-time QC tests are retained for analysis.

To evaluate model performance against real-world observations, we adopt the GODAE Ocean View IV-TT Class 4 framework [72]. This widely used operational ocean forecast validation framework provides in-situ and satellite observational datasets for 2022, together with interpolated outputs from participating global ocean forecasting systems. Within the IV-TT Class 4 framework, each forecasting system interpolates its predictions—specifically temperature profile, salinity profiles, and SLA—to the exact locations and times of the observations. This allows direct point-to-point comparison between model predictions and independent observations, facilitating consistent cross-system performance evaluation. In our case, temperature and salinity profile observations come from Argo floats, while SLA observations are derived from multi-satellite altimetry.

4.4 Verification metrics

Point-to-point verification metrics. The GOFSs forecasts are categorized based on their initial time c and forecast lead time τ, and are then compared to observations at time $c + \tau$. Since the observations are not regularly spaced and do not coincide with the grid points of the GOFS forecasts, the forecast values are interpolated to the observation locations. This allows for point-to-point comparisons between the forecasts and observations, enabling the calculation of forecast errors, as expressed by:

$$\epsilon_i^{c,\tau} = \hat{V}_i^{c+\tau} - V_i^c \tag{4}$$

where V_i^c represents the observation at time c, and $\hat{V}_i^{c+\tau}$ is the forecast value at time $c + \tau$ interpolated to the i-th observation site. To ensure a fair comparison between the different GOFS forecasts, only those match-ups where all GOFSs have corresponding forecast values are considered.

RMSE Calculation. To rigorously assess the predictive accuracy of TianHai, we calculate the RMSE between model forecasts and observations across the global ocean grid. Given the spherical geometry of the Earth, the physical area represented by each

grid cell varies with latitude, becoming progressively smaller toward the poles. To prevent biased skill assessments due to uneven spatial sampling, we apply a latitude weighting scheme when aggregating errors, ensuring that each grid cell's contribution is proportional to its surface area.

Let $\mathcal{D}$ denote the set of forecast initialization times within the testing period. For each forecast initialized at time $t_0 \in \mathcal{D}$ and forecast lead time τ, the latitude-weighted RMSE for ocean variable c is computed as:

$$\mathrm{RMSE}(c,\tau) = \frac{1}{|\mathcal{D}|} \sum_{t_0 \in \mathcal{D}} \sqrt{\frac{\sum_{i=1}^{H} \sum_{j=1}^{W} w_i \left(\hat{X}_{c,i,j}^{t_0+\tau} - X_{c,i,j}^{t_0+\tau} \right)^2}{\sum_{i=1}^{H} \sum_{j=1}^{W} w_i}} \tag{4}$$

where $\hat{X}_{c,i,j}^{t_0+\tau}$ is the predicted value of variable c at grid cell (i,j) at forecast lead time τ after initialization time t_0; $X_{c,i,j}^{t_0+\tau}$ is the corresponding observed (or reanalysis) value used as the verification reference;H and W denote the total number of latitudinal and longitudinal grid points, respectively;$w_i = \cos(\phi_i)$ is the latitude weight assigned to the i-th latitude band, where ϕ_i represents the central latitude of the grid cell row i; $|\mathcal{D}|$ is the total number of forecast initializations in the test dataset.

The cosine of latitude weighting corrects for the non-uniformity of surface area represented by each grid cell, allowing a spatially unbiased global RMSE estimate. The RMSE is computed individually for each ocean state variable (e.g., temperature, salinity, currents, sea surface height) and at each forecast lead time. Subsequently, the errors are averaged over all forecast initializations to obtain a robust, statistically representative measure of forecast skill.

Persistence Forecast as a Baseline. To provide a baseline for model performance assessment, we calculate the persistence forecast error. Persistence assumes the ocean state remains static at its initial condition value throughout the forecast period, representing a simple yet informative naive prediction method commonly used in operational oceanography as a lower bound of forecast skill.

The Performance Skill Score (PSS) is defined as:

$$\mathrm{PSS}_{c,\tau} = 1 - \frac{\mathrm{RMSE}(c,\tau)}{\mathrm{RMSE}_{(p,c,\tau)}} \tag{6}$$

where $\mathrm{RMSE}_{p,c,\tau}$ is the RMSE of a persistent forecast.

Once $\mathrm{RMSE}_{c,\tau}$ and $\mathrm{PSS}_{c,\tau}$ are obtained, their median values among different initial times c are computed and expressed as a function of forecast lead time τ.

4.5 Computation of ocean front variabilities

Using gradient method to identify and extract oceanic fronts. The principle of gradient method is to calculate pixel gradients and select pixels above a specific threshold to extract ocean fronts. The principle of gradient method is simple and the calculation speed is fast, so it is widely used in the detection of early ocean fronts. Ocean fronts are often accompanied by drastic temperature changes, and temperature gradients can describe the intensity of ocean front movement. Therefore, it is necessary to calculate

the gradient of sea surface temperature data in order to convert the temperature data in the corresponding latitude and longitude coordinates into gradient data, thereby better characterizing the oceanic front. Gradient calculation is performed within a 3x3 neighborhood of each data point. The expression for gradient calculation is

$$G = \sqrt{D_x^2 + D^2} \quad [\mathrm{m}^2/\mathrm{s}^2] \tag{5}$$

$$D_x = \frac{T(i, j+1) - T(i, j-1)}{2\,\Delta x} \tag{6}$$

$$D_y = \frac{T(i+1, j) - T(i-1, j)}{2\,\Delta y} \tag{7}$$

Where G is the sea surface temperature gradient; ΔX, ΔY is the pixel size in the X and Y directions, respectively; D_x and D_y are the gradient size in the horizontal and vertical directions, respectively; and i and j are the pixel positions in the image, respectively.

4.6 Vertical Gradient Method for Thermocline Detection

To detect the thermocline, we calculate the vertical gradient of the sea surface temperature. The vertical temperature gradient D_z is computed using the central difference method:

$$D_z = \frac{T(z+1) - T(z-1)}{2\Delta z} \tag{8}$$

where $T(z+1)$ and $T(z-1)$ represent the temperature at the depths $z+1$ and $z-1$, respectively, and Δz is the vertical resolution (the depth spacing between consecutive data points).

To identify the thermocline, a threshold value for the gradient is set. If the vertical gradient exceeds a specified threshold $D_z^{\text{threshold}}$, then the corresponding depth is considered to be part of the thermocline. This can be expressed as:

$$\text{If} \quad D_z > D_z^{\text{threshold}}, \quad \text{then this depth is part of the thermocline.} \tag{9}$$

Data availability

HYCOM Global 1/12° Reanalysis data are obtained https://www.hycom.org/dataserver/gofs-3pt1/reanalysis/. HYCOM Global 1/12° analysis data are obtained https://www.hycom.org/dataserver/gofs-3pt1/analysis/. Argo observations are obtained https://data.marine.copernicus.eu/product/INSITU_GLO_PHYBGCWAV_DISCRETE_MYNRT_013_030/ . The HYCOM Global 1/12 ° forecast data are obtained https://www.hycom.org/dataserver/espc-d-v02/. The drifter observations are obtained from https://data.marine.copernicus.eu/product/INSITU_GLO_PHYBGCWAV_DISCRETE_MYNRT_013_048//description. The SLA observations are obtained from https://data.marine.copernicus.eu/product/SEALEVEL_GLO_PHY_L3_NRT_008_044/description. The ERA5 reanalysis datasets are obtained from https://doi.org/10.24381/cds.adbb2d47.

Code availability

The details of TianHai, including network architectures, modules are available in the paper. The source code employed for training and running TianHai models in this research is accessible within a specific Google Drive folder(https://doi.org/*)

References

[1] Schindling, E., Wilson, D.: Deterministic nonperiodic flow (1963). Gotenborg, Sweden (2006)

[2] Maskey, M.: Rethinking ai for science: An evolution from data-driven to data-centric framework. Perspectives of Earth and Space Scientists **4**(1), 2023–000222 (2023) https://doi.org/10.1029/2023CN000222

[3] Tonani, M., Balmaseda, M., Bertino, L., Blockley, E., Brassington, G., Davidson, F., Drillet, Y., Hogan, P., Kuragano, T., Lee, T., *et al.*: Status and future of global and regional ocean prediction systems. Journal of Operational Oceanography **8**(sup2), 201–220 (2015) https://doi.org/10.1080/1755876X.2015.1049892

[4] Chelton, D.B., Schlax, M.G., Samelson, R.M.: Global observations of nonlinear mesoscale eddies. Progress in Oceanography **91**(2), 167–216 (2011) https://doi.org/10.1016/j.pocean.2011.01.002

[5] McWilliams, J.C.: Submesoscale currents in the ocean. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **472**(2189), 20160117 (2016) https://doi.org/10.1098/rspa.2016.0117

[6] Griffies, S.M.: Fundamentals of Ocean Climate Models. Princeton University Press, ??? (2018). https://www.torrossa.com/en/resources/an/5558718

[7] Adcroft, A., Hallberg, R., Harrison, M.: A finite volume discretization of the pressure gradient force using analytic integration. Ocean Modelling **22**(3), 106–113 (2008) https://doi.org/10.1016/j.ocemod.2008.02.001

[8] Chassignet, E.P., *et al.*: The hycom (hybrid coordinate ocean model) data assimilative system. Journal of Marine Systems **65**, 60–83 (2007) https://doi.org/10.1016/j.jmarsys.2005.09.016

[9] Griffies, S.M., Harrison, M.J., Pacanowski, R.C., Rosati, A., *et al.*: A technical guide to mom4. GFDL Ocean Group Tech. Rep **5**(5), 371 (2004)

[10] Griffies, S., Adcroft, A., Banks, H., Böning, C., Chassignet, E., Danabasoglu, G., Danilov, S., Deleersnijder, E., Drange, H., England, M., *et al.*: Problems and prospects in large-scale ocean circulation models. Proceedings of OceanObs **9**, 410–431 (2009)

[11] Fox-Kemper, B., et al.: Challenges and prospects in ocean circulation models. Frontiers in Marine Science **Volume 6 - 2019** (2019) https://doi.org/10.3389/fmars.2019.00065

[12] Smith, N.: The global ocean data assimilation experiment. Advances in Space Research **25**(5), 1089–1098 (2000) https://doi.org/10.1016/S0273-1177(99)00868-6

[13] BEll, M.J., Lefebvre, M., Le Traon, P.-Y., Smith, N., Wilmer-Becker, K.: Godae: the global ocean data assimilation experiment. Oceanography **22**(3), 14–21 (2009) https://doi.org/https://www.jstor.org/stable/24860986

[14] Smith, R., Gent, P.: Reference manual for the parallel ocean program (pop). Los Alamos unclassified report LA-UR-02-2484 (2002)

[15] Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., Dukowicz, J., Eden, C., Fox-Kemper, B., Gent, P., *et al.*: The parallel ocean program (pop) reference manual: Ocean component of the community climate system model (ccsm). Rep. LAUR-01853 **141**, 1–141 (2010)

[16] Zheng, G., Li, X., Zhang, R.-H., Liu, B.: Purely satellite data–driven deep learning forecast of complicated tropical instability waves. Science advances **6**(29), 1482 (2020)

[17] Quach, B., Glaser, Y., Stopa, J.E., Mouche, A.A., Sadowski, P.: Deep learning for predicting significant wave height from synthetic aperture radar. IEEE Transactions on Geoscience and Remote Sensing **59**(3), 1859–1867 (2020)

[18] Wang, X., Wang, R., Hu, N., Wang, P., Huo, P., Wang, G., Wang, H., Wang, S., Zhu, J., Xu, J., et al.: Xihe: A data-driven model for global ocean eddy-resolving forecasting. arXiv preprint arXiv:2402.02995 (2024)

[19] Yang, N., Wang, C., Zhao, M., Zhao, Z., Zheng, H., Zhang, B., Wang, J., Li, X.: Langya: Revolutionizing cross-spatiotemporal ocean forecasting. arXiv preprint arXiv:2412.18097 (2024)

[20] Cui, Y., Wu, R., Zhang, X., Zhu, Z., Liu, B., Shi, J., Chen, J., Liu, H., Zhou, S., Su, L., *et al.*: Forecasting the eddying ocean with a deep neural network. Nature Communications **16**(1), 2268 (2025) https://doi.org/10.1038/s41467-025-57389-2

[21] Xiong, W., Xiang, Y., Wu, H., Zhou, S., Sun, Y., Ma, M., Huang, X.: Ai-goms: Large ai-driven global ocean modeling system. arxiv. arXiv preprint arXiv:2308.0315 (2023)

[22] Ham, Y.-G., Kim, J.-H., Luo, J.-J.: Deep learning for multi-year enso forecasts. Nature **573**(7775), 568–572 (2019)

[23] Zhou, L., Zhang, R.-H.: A hybrid neural network model for enso prediction in combination with principal oscillation pattern analyses. Advances in Atmospheric Sciences **39**(6), 889–902 (2022) https://doi.org/10.1007/s00376-021-1368-4

[24] Zhou, L., Zhang, R.-H.: A self-attention–based neural network for three-dimensional multivariate modeling and its skillful enso predictions. Science Advances **9**(10), 2827 (2023) https://doi.org/10.1126/sciadv.adf2827

[25] Santoki, M., George, S., Sharma, R., Joshipura, K., Basu, S.: Assimilation of satellite-derived ocean surface current in an indian ocean circulation model. Remote Sensing Letters **4**(5), 475–484 (2013) https://doi.org/10.1080/2150704X.2012.750036

[26] Guo, Z., Lyu, P., Ling, F., Bai, L., Luo, J.J., Boers, N., Yamagata, T., Izumo, T., Cravatte, S., Capotondi, A., Ouyang, W.: Data-driven global ocean modeling for seasonal to decadal prediction. Science Advances **11**(33), 2488 (2025) https://doi.org/10.1126/sciadv.adu2488 . Epub 2025 Aug 13

[27] Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., Miele, A., Kashinath, K., Anandkumar, A.: Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In: Proceedings of the Platform for Advanced Scientific Computing Conference, pp. 1–11 (2023)

[28] Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., Chen, X., Ma, L., Zhang, T., Su, R., et al.: Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. arXiv preprint arXiv:2304.02948 (2023)

[29] Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., *et al.*: Learning skillful medium-range global weather forecasting. Science **382**(6677), 1416–1421 (2023)

[30] Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q.: Accurate medium-range global weather forecasting with 3d neural networks. Nature **619**(7970), 533–538 (2023)

[31] Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J.K., Grover, A.: Climax: A foundation model for weather and climate. arXiv preprint arXiv:2301.10343 (2023)

[32] Bodnar, C., Bruinsma, W.P., Lucic, A., Stanley, M., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J., Dong, H., Vaughan, A., et al.: Aurora: A foundation model of the atmosphere. arXiv preprint arXiv:2405.13063 (2024)

[33] Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T.R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., *et al.*: Probabilistic weather forecasting with machine learning. Nature **637**(8044), 84–90 (2025)

[34] Aouni, A.E., Gaudel, Q., Regnier, C., Gennip, S.V., Drevillon, M., Drillet, Y., Lellouche, J.-M.: GLONET: Mercator's End-to-End Neural Forecasting System. Preprint at https://arxiv.org/abs/2412.05454 (2024)

[35] Griffies, S.: Fundamentals of Ocean Climate Models. Princeton university press, ??? (2018). https://www.torrossa.com/en/resources/an/5558718

[36] Roemmich, D., *et al.*: On the future of argo: A global, full-depth, multi-disciplinary array. Frontiers in Marine Science **6**, 439 (2019) https://doi.org/10.3389/fmars.2019.00439

[37] Wu, A., Hsieh, W.W., Tang, B.: Neural network forecasts of the tropical pacific sea surface temperatures. Neural networks **19**(2), 145–154 (2006) https://doi.org/10.1016/j.neunet.2006.01.004

[38] Gill, A., Turner, J.: A comparison of seasonal thermocline models with observation. In: Deep Sea Research and Oceanographic Abstracts, vol. 23, pp. 391–401 (1976). https://doi.org/10.1016/0011-7471(76)90836-6 . Elsevier

[39] Bernie, D., Guilyardi, E., Madec, G., Slingo, J.M., Woolnough, S.J., Cole, J.: Impact of resolving the diurnal cycle in an ocean–atmosphere gcm. part 2: A diurnally coupled cgcm. Climate dynamics **31**(7), 909–925 (2008) https://doi.org/10.1007/s00382-008-0429-z

[40] Levitus, S., Antonov, J.: Observational evidence of interannual to decadal-scale variability of the subsurface temperature-salinity structure of the world ocean. Climatic Change **31**(2), 495–514 (1995) https://doi.org/10.1007/BF01095159

[41] Chelton, D.B., *et al.*: Global observations of large oceanic eddies. Science **303**, 974–978 (2004) https://doi.org/10.1029/2007GL030812

[42] Brown, R.A.: Surface fluxes and remote sensing of air-sea interactions. In: Surface Waves and Fluxes: Volume I—Current Theory, pp. 7–27. Springer, ??? (1990). https://doi.org/10.1007/978-94-009-2069-9_2

[43] Morrow, R., Fu, L.-L., Farrar, J.T., Seo, H., Le Traon, P.-Y.: Ocean eddies and mesoscale variability. In: Satellite Altimetry over Oceans and Land Surfaces, pp. 315–342. CRC press, ??? (2017). https://www.taylorfrancis.com/chapters/edit/10.1201/9781315151779-10/ocean-eddies-mesoscale-variability-rosemary-morrow-lee-lueng-fu-thomas-farrar-hyodae-seo-pierre-yves-le-traon

[44] Kirtman, B.P., Perlin, N., Siqueira, L.: Ocean eddies and climate predictability. Chaos: An Interdisciplinary Journal of Nonlinear Science **27**(12) (2017) https://doi.org/10.1063/1.4990034

[45] Dong, C., You, Z., Dong, J., Ji, J., Sun, W., Xu, G., Lu, X., Xie, H., Teng, F.,

Liu, Y., *et al.*: Oceanic mesoscale eddies. Ocean-Land-Atmosphere Research **4**, 0081 (2025) https://doi.org/10.34133/olar.0081

[46] Morrow, R., Le Traon, P.-Y.: Recent advances in observing mesoscale ocean dynamics with satellite altimetry. Advances in Space Research **50**(8), 1062–1076 (2012) https://doi.org/10.1016/j.asr.2011.09.033

[47] Lagerloef, G., Schmitt, R., Schanze, J., Kao, H.-Y.: The ocean and the global water cycle. Oceanography **23**(4), 82–93 (2010)

[48] Kara, A.B., Wallcraft, A.J., Metzger, E.J., Gunduz, M.: Impacts of freshwater on the seasonal variations of surface salinity and circulation in the caspian sea. Continental Shelf Research **30**(10-11), 1211–1225 (2010) https://doi.org/10.1016/j.csr.2010.03.011

[49] Durack, P.J., Wijffels, S.E.: Fifty-year trends in global ocean salinities and their relationship to broad-scale warming. Journal of Climate **23**(16), 4342–4362 (2010) https://doi.org/10.1175/2010JCLI3377.1

[50] Yu, X., McPhaden, M.J.: Seasonal variability in the equatorial pacific. Journal of Physical Oceanography **29**(5), 925–947 (1999) https://doi.org/10.1175/1520-0485(1999)029

[51] Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., *et al.*: The era-interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society **137**(656), 553–597 (2011) https://doi.org/10.1002/qj.828

[52] Wilks, D.S.: Statistical Methods in the Atmospheric Sciences, 3rd edn. Academic Press, ??? (2011)

[53] Murphy, A.H.: What is a good forecast? an essay on the nature of goodness in weather forecasting. Weather and Forecasting **8**(2), 281–293 (1993) https://doi.org/10.1175/1520-0434(1993)008

[54] Jolliffe, I.T., Stephenson, D.B.: Forecast Verification: A Practitioner's Guide in Atmospheric Science. John Wiley & Sons, ??? (2012)

[55] Hamill, T.M.: Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review **129**(3), 550–560 (2001) https://doi.org/10.1175/1520-0493(2001)129⟨0550:IORHFV⟩2.0.CO;2

[56] Deshmukh, A.N., Deo, M., Bhaskaran, P.K., Nair, T.B., Sandhya, K.: Neural-network-based data assimilation to improve numerical ocean wave forecast. IEEE Journal of Oceanic Engineering **41**(4), 944–953 (2016) https://doi.org/10.1109/JOE.2016.2521222

[57] Thomas, A., Dietrich, J., Loveland, M., Samii, A., Dawson, C.: Improving coastal flooding predictions by switching meshes during a simulation. Ocean Modelling **164**, 101820 (2021) https://doi.org/10.1016/j.ocemod.2021.101820

[58] Chassignet, E.P., *et al.*: The hycom (hybrid coordinate ocean model) data assimilative system. Journal of Marine Systems **65**(1–4), 60–83 (2007) https://doi.org/10.1016/j.jmarsys.2005.09.015

[59] Bleck, R.: An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. Ocean Modelling **4**(1), 55–88 (2002) https://doi.org/10.1016/S1463-5003(01)00012-9

[60] Schiller, A., Brassington, G.B., Oke, P., Cahill, M., Divakaran, P., Entel, M., Freeman, J., Griffin, D., Herzfeld, M., Hoeke, R., *et al.*: Bluelink ocean forecasting australia: 15 years of operational ocean service delivery with societal, economic and environmental benefits. Journal of Operational Oceanography **13**(1), 1–18 (2020) https://doi.org/10.1080/1755876X.2019.1685834

[61] Blockley, E., Martin, M., McLaren, A., Ryan, A., Waters, J., Lea, D., Mirouze, I., Peterson, K., Sellar, A., Storkey, D.: Recent development of the met office operational ocean forecasting system: an overview and assessment of the new global foam forecasts. Geoscientific Model Development **7**(6), 2613–2638 (2014)

[62] Roemmich, D., *et al.*: The argo program: Observing the global ocean with profiling floats. Oceanography **22**(2), 34–43 (2009)

[63] Martin, M.J., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., Fujii, Y., Lea, D., Lellouche, J.-M., Mogensen, K., *et al.*: Status and future of data assimilation in operational oceanography. Journal of Operational Oceanography **8**(sup1), 28–48 (2015) https://doi.org/10.1080/1755876X.2015.1022055

[64] Griffies, S.M.: Developments in ocean climate modelling. Ocean Modelling **2**(3-4), 123–192 (2000) https://doi.org/10.1016/S1463-5003(00)00014-7

[65] Eymard, L., Taconet, O.: The methods for inferring surface fluxes from satellite data, and their use for atmosphere model validation. International Journal of Remote Sensing **16**(11), 1907–1930 (1995)

[66] Fox-Kemper, B., Ferrari, R., Hallberg, R.: Parameterization of mixed layer eddies. part i: Theory and diagnosis. Journal of Physical Oceanography **38**(6), 1145–1165 (2008) https://doi.org/10.1175/2007JPO3792.1

[67] Liu, Z., *et al.*: Swin transformer v2: Scaling up capacity and resolution. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11999–12009 (2022)

[68] Guo, Y., Wang, C., Yu, S.X., McKenna, F., Law, K.H.: Adaln: a vision transformer for multidomain learning and predisaster building information extraction from images. Journal of Computing in Civil Engineering **36**(5), 04022024 (2022)

[69] Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proceedings of 1st International Conference on Image Processing, vol. 2, pp. 168–1722 (1994). https://doi.org/10.1109/ICIP.1994.413553

[70] Chen, L., et al.: Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast. npj Climate and Atmospheric Science, 1–11 (2023)

[71] Chassignet, E.P., Hurlburt, H.E., Smedstad, O.M., Halliwell, G.R., Hogan, P.J., Wallcraft, A.J., Baraille, R., Bleck, R.: The hycom (hybrid coordinate ocean model) data assimilative system. Journal of Marine Systems **65**(1-4), 60–83 (2007) https://doi.org/10.1016/j.jmarsys.2005.09.016

[72] Ryan, A., Régnier, C., Divakaran, P., Spindler, T., Mehra, A., Smith, G.C., Davidson, F., Hernandez, F., Maksymczuk, J., Liu, Y.: Godae oceanview class 4 forecast verification framework: global ocean inter-comparison. Journal of Operational Oceanography **8**, 111–98 (2015)

Acknowledgements

This work was supported in part by the Key Research and Development Program, sponsored by the Ministry of Science and Technology (MOST), under Grant 2023YFC3107701 and Grant 2023YFC3107901; in part by the National Natural Science Foundation of China under Grant 42375143. We thank the HYCOM consortium for providing the HYCOM reanalysis, HYCOM analysis data (https://www.hycom.org/dataserver/gofs-3pt1/analysis) . We thank the Mercator Océan and E.U. Copernicus Marine Service Information for providing the observational data (https:// doi.org/10.48670/moi-00036, https://doi.org/10.48670/moi-00041, https://doi.org/10.48670/moi-00147). We thank WenHai team for making the model weights and inference code of WenHai available (https://github.com/Cuiyingzhe/WenHai). The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it (https://argo.ucsd.edu, https://www.ocean-ops.org). The Argo Program is part of the Global Ocean Observing System.The altimeter products were produced by Ssalto/Duacs and distributed by AVISO+, with support from CNES (https://www.aviso.altim etry.fr). The surface drifter data are produced by the Global Drifter Program (https://www.aoml.noaa.gov/phod/gdp/data.php).

Author contributions

Y.N., **Q.H.**, **A.G.**, and **X.Z.** conceived the study. **Y.N.**, **Q.H.** and **X.Z.** developed the methodology. **Y.N.**, **J.Q.**, and **X.J.** conducted the investigation. Visualization was carried out by **Y.N.**, **Q.H.**, **A.G.**, and **X.Z.**. **H.L.**, **L.C.**, and **X.Z.** supervised the project. **Y.N.**, **Q.H.**, **A.G.**, and **X.Z.** wrote the original draft, and **X.Z.**. **H.L.**, **D.Z.** , **X.Z.** reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.