Towards An Early Warning System for Ocean Heat Extremes through AI-Ocean Dynamics Synergy

Zheng Jiang[1], Wei Wang[1,*], Gaowei Zhang[1], Yifei Bao[1], Zengzhou Hao[2], Lingyu Xu[3], Suixiang Shi[4], Lei Wang[5], Yi Wang[1,6]
[1]Beijing University of Posts and Telecommunications, Beijing, China
[2]State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou, China
[3]Shanghai University, Shanghai, China
[4]National Marine Data and Information Service, Tianjin, China
[5]East China Sea Forecasting and Disaster Reduction Center, Shanghai, China
[6]Shandong Key Laboratory of Advanced Computing, Inspur Computer Technology Co., Ltd
{jiangzheng, weiwang, zhanggaowei, byffly}@bupt.edu.cn, hzyx80@sio.org.cn, xly@shu.edu.cn, shisuixiang@hotmail.com, wangleidett727@163.com, wang@cocolabs.org

Abstract

Ocean heat extremes, including marine heatwaves and the El Niño–Southern Oscillation (ENSO), exert profound impacts on marine ecosystems and socio-economic stability. Establishing robust early warning systems is critical for proactive risk management; however, conventional predictive models often fail to generalize to the intensifying, non-stationary extremes driven by rapid global warming. This project introduces a novel AI-Ocean Dynamics synergy designed to provide an integrated early warning system. By synthesizing multi-source observations with physics-informed neural networks, it ensures predictions remain constrained by fundamental physical laws. The system forecasts event onset, intensity, duration, and spatial extent while simultaneously attributing the underlying mechanisms, such as ocean advection and air–sea heat exchange. To validate performance, we establish a specialized ocean heat extremes benchmark to assess predictive skill and attribution reliability. Furthermore, the system incorporates an incremental learning mechanism, enabling continuous adaptation to long-term climatic and environmental evolutions. This project advances the development of reliable, interpretable, and adaptive early warning systems, providing a vital tool for informed policy and maritime decision-making.

1 Introduction

In recent decades, episodes of ocean heat extremes have been associated with more intense and frequent impacts on marine organisms, ecosystems and reliant human industries around the world [Capotondi *et al.*, 2024; Wernberg *et al.*, 2025]. These extremes occur ranging from short-term marine heatwaves (MHWs) lasting days to weeks [Wernberg *et al.*, 2025], to prolonged anomalies for months, such as the El Niño-Southern Oscillation (ENSO) [Chen *et al.*, 2024]. Under ongoing global warming, the frequency and intensity of these events have reached record highs [Dong *et al.*, 2025]. Therefore, developing reliable ocean heat warning systems is essential to predict and understand these extremes in advance, enabling us to take preventive actions to reduce ecological and socio-economic impacts.

Despite growing importance, many countries have yet to implement ocean heat warning systems; even such systems running in a few countries are often poorly defined and impact-based [Nairn and Mason, 2025]. The core functions of such systems rely on the accurate temperature forecasting for further detecting anomalies. Most systems depend on numerical ocean models to simulate extreme events under varying scenarios, such as the Hybrid Coordinate Ocean Model (HYCOM) [Chassignet *et al.*, 2007]. While physically consistent, they are computationally expensive and often accumulate biases that cause their simulations to diverge from real-world ocean evolution. Recent AI advances in data-driven forecasting models, such as the emerging ocean foundation models [Wang *et al.*, 2024; Cui *et al.*, 2025; Huang *et al.*, 2025], have shown promise in efficiently learning large-scale ocean temperature variability from reanalysis data [Camps-Valls *et al.*, 2025]. However, integrating these models into practical ocean heat warning systems remains a significant challenge. Ocean heat extremes differ fundamentally from normal ocean temperature variability. They are rare, data-limited, and are often driven by large-scale air-sea interactions [Sen Gupta *et al.*, 2020; Gregory *et al.*, 2024], resulting in purely data-driven models that perform well under typical conditions tend to overfit mean variability and under-represent the severity of extremes [Prochaska *et al.*, 2021; Xu *et al.*, 2025]. Moreover, their black-box nature limits physical interpretability and hinders understanding of the key drivers underlying ocean heat extremes. Their long-term deployment is further challenged by the non-stationary nature of the climatic and environmental evolution, in which ocean

*Corresponding author.

temperature baselines and extreme characteristics evolve. Given the high computational cost of retraining, commonly used static models struggle to adapt to rapid recent warming trends, becoming unreliable for predicting the emerging, record-breaking extremes [Ryu *et al.*, 2025; Singh *et al.*, 2021]. Thus, it is imperative to develop a specialized AI-powered ocean heat warning system that can reliably forecast ocean heat extremes while maintaining physical interpretability and robustness to climatic and environmental evolutions.

Recent advances in physics-informed AI offer a promising pathway for capturing ocean thermodynamics and heat extremes by coupling data-driven efficiency with intrinsic physical mechanisms [Kashinath *et al.*, 2021; Jiang *et al.*, 2026; Shu *et al.*, 2025]. Motivated by this potential, we propose a more reliable, interpretable, and adaptive early warning system for ocean heat extremes through AI-Ocean Dynamics synergy. To support this system, we first curate a comprehensive multi-source data repository that integrates heat state variables ranging from the atmosphere to the deep ocean. The core of the system is to develop physics-informed AI models that learn spatiotemporal representations with governing ocean thermodynamic laws. Building upon these models, the system targets two core tasks: ocean heat extremes prediction and physical driver attribution, identifying how extremes are driven by atmospheric fluxes or oceanic transport. To rigorously evaluate this system, we design a specialized ocean heat extremes benchmark covering MHWs and ENSO extremes. Finally, the system incorporates an incremental learning mechanism to continuously adapt to climatic and environmental evolutions without retraining from scratch.

The goals of this project have been defined through a cross-disciplinary collaboration between AI researchers, geophysicists, and physical oceanography experts from *East China Sea Forecasting and Disaster Reduction Center* and the *National Marine Data and Information Service (NMDIS)* in China. Through joint discussions, our multidisciplinary team identified a persistent "trust gap" stemming from the lack of physical consistency and interpretability in black-box AI models. We aim to bridge the gap between advanced deep learning and physical oceanography by synthesizing ocean data archives and developing physics-informed AI approaches. This collaborative foundation ensures that the proposed system is not only algorithmically innovative but also practically feasible for real-world deployment.

2 Project Goal

The overall goal is to develop a reliable, interpretable and adaptive early warning system specifically tailored to ocean heat extremes. This project not only targets improving general ocean temperature forecasts as common warning systems do, but also the challenges of monitoring, predicting, and understanding those rare, high-impact ocean heat extremes. By integrating physics-informed AI with incremental learning, the project seeks to enable timely warnings and a deeper understanding of ocean heat extremes, supporting both scientific insight and practical decision-making.

Data Foundation Objectives. Ocean temperature variability is not limited to surface anomalies but is strongly influenced by the thermal process in the deep ocean and air-sea interactions [Sun *et al.*, 2023; Wu *et al.*, 2024b]. Therefore, comprehensively studying ocean heat extremes across air-surface-subsurface requires integrating multiple complementary data sources, including satellite remote sensing data, in-situ measurements, and reanalysis products. The project aims to assemble these heterogeneous, multi-source data to establish a data repository. Note that the geographical distribution of available in-situ observations is highly uneven. Publicly available datasets are often concentrated in regions such as Europe and North America [Zhu *et al.*, 2026], which may introduce regional biases if used directly as the data foundation. Thus, our data strategy needs to deliberately balance global coverage with targeted sampling in key regions of interest, enabling the model to capture both large-scale and localized patterns in ocean heat extremes.

Methodological Objectives. Recent interdisciplinary research has demonstrated the potential of physics-informed AI, utilizing architectures such as Neural Ordinary Differential Equations (Neural ODEs) and Physics-Informed Neural Networks (PINNs) to model complex fluid dynamics [Jiang *et al.*, 2026; Hu *et al.*, 2025]. Building upon these advancements, the methodological objective is to extend and explore physics-informed models for ocean heat extremes, mitigating the insensitivity of traditional deep learning models to rare extremes. We will develop targeted modeling strategies that differentiate between the rapid onset of short-term marine heatwaves and the slowly evolving teleconnections of ENSO.

Evaluation Objectives. The evaluation objectives focus on comprehensively assessing the system's early warning capability for both short-term marine heatwaves and long-term, basin-scale ENSO extremes. Most existing ocean benchmarks primarily evaluate deterministic prediction accuracy [Aouni *et al.*, 2025; Nathaniel *et al.*, 2024]. While several recent benchmarks target extreme events in the atmospheric domain [Zhao *et al.*, 2025], standardized benchmarks for ocean heat extremes remain largely absent. To address this gap, this project proposes the construction of a dedicated benchmark for ocean heat extremes to rigorously evaluate system reliability. In addition to predictive accuracy, the evaluation will also assess physical consistency, the validity of driver attribution, and risk assessment.

Application and Deployment Objectives. The global ocean currents are an essential driver of Earth's climate system, linking distant ocean basins and coupling local temperature anomalies with large-scale teleconnections [Bian *et al.*, 2024]. Our project needs to build upon a global ocean modeling architecture capable of capturing both basin-scale and cross-regional interactions. As a demonstration of practical deployment, the proposed system will be applied in several representative regions, including the Northwest Pacific, the South China Sea, and the East China Sea, as well as the Equatorial Pacific ENSO zones. These regional deployments serve as case studies to evaluate real-world practical value and provide transferable experience for the development of ocean heat warning systems in other regions and countries.

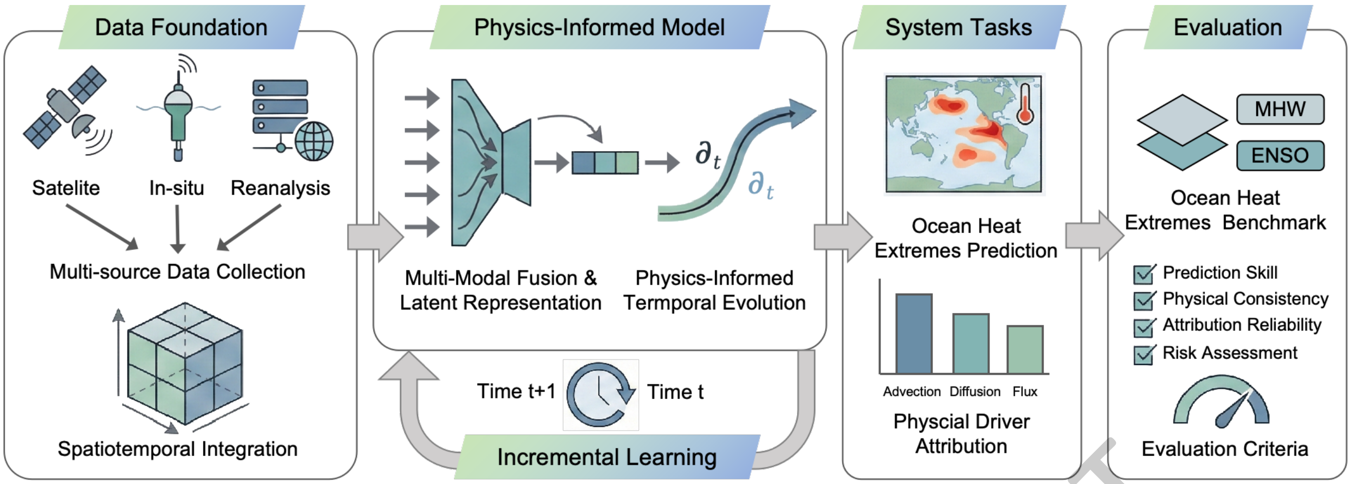

Figure 1: Overview of the early warning system for ocean heat extremes through AI-Ocean Dynamics synergy. The system integrates multi-source data, constrained representation learning, extreme event prediction and attribution, systematic evaluation, and incremental learning for adapting to climatic and environmental evolutions.

3 Strategy

Figure 1 presents a schematic for the construction of an early warning system for ocean heat extremes that integrates five key components: data foundation, physics-informed model, system tasks, systematic evaluation, and an incremental learning mechanism. This system begins by collecting heterogeneous, multi-source ocean data and integrating them into a unified spatiotemporal data representation. These data are then processed by a physics-informed deep model constrained by governing thermodynamic laws. The resulting representations support two downstream tasks, including ocean heat extremes prediction and physical driver attribution. These model outputs and analytical insights are evaluated using both technical benchmarks and their broader societal impact. Finally, the system further incorporates an incremental learning mechanism to enable long-term applicability under a non-stationary climate.

3.1 Data Foundation

To construct a holistic view of the ocean thermal state, we curate comprehensive multi-source datasets that capture not only the temperature field but also the thermodynamic drivers governing its evolution. Table 1 summarizes the multi-source datasets used to characterize the ocean thermal state, spanning surface and subsurface temperatures and their physical drivers. These heterogeneous data sources are subsequently integrated into a unified spatiotemporal representation to support consistent downstream learning and analysis.

Multi-source Data Collection

Surface Temperature Field. Sea surface temperature (SST) can be directly observed at the global scale through satellite remote sensing. We integrate long-term satellite SST products from the NOAA Optimum Interpolation Sea Surface Temperature (OISST) products [Reynolds *et al.*, 2007], including both daily and monthly SST and SST anomaly fields spanning from 1982 to the present. These data provide a high-resolution (0.25°) horizontal view, enabling a precise tracking of the spatial development and transport of surface MHWs at global scales.

Category	Variable	Source	Data Type
Surface	SST & Anomalies	NOAA OISST	Satellite
Subsurface	Temp. Profiles	Argo Floats	In-situ
	3D Temp. Fields	GLORYS12V1	Reanalysis
Drivers	Heat Fluxes	ERA5	Reanalysis
	Ocean Currents	GLORYS12V1	Reanalysis

Table 1: Summary of multi-source datasets used in the ocean heat extreme warning system

Subsurface Temperature Field. Ocean heat extremes are frequently driven by subsurface heat transport and vertical redistribution processes [Sun *et al.*, 2023]. However, satellite observations are restricted to the surface layer. To reconstruct the 3D ocean thermal structure, we integrate the subsurface temperature field from both in-situ observations and reanalysis products. Specifically, we utilize temperature profiles from the Argo float array [Roemmich *et al.*, 2009], which provide high-quality vertical measurements from the surface to approximately 2000m depth, along with additional in-situ data from NMDIS. Since in-site data are spatially sparse and irregularly distributed, we supplement them with the 3D ocean reanalysis products such as GLORYS12V1 reanalysis product [Jean-Michel *et al.*, 2021]. These reanalysis products assimilate satellite and in-situ data into physical models, yielding a dense, volumetrically complete estimate of subsurface temperature.

Heat Fluxes and Currents. Air–sea interactions play a crucial role in the initiation, amplification, and decay of ocean heat extremes by regulating surface energy exchange [Wu *et al.*, 2024b]. To characterize these external forcing processes, we incorporate surface heat flux variables from the ERA5 reanalysis dataset [Hersbach *et al.*, 2020], including net heat flux and its constituent components. Recent physics-informed learning studies indicate that explicitly modeling heat fluxes can improve the representation of ocean thermal evolution, motivating their inclusion as key physical drivers in our dataset [Jiang *et al.*, 2026; Shu *et al.*, 2025]. In addition, ocean circulation redistributes heat horizontally through advection [Gao *et al.*, 2026]. We also acquire zonal and meridional current velocity data (u, v) from GLORYS12V1 reanalysis product [Jean-Michel *et al.*, 2021], which can be utilized to explicitly model the thermal advection process.

Spatiotemporal Integration

The collected multi-source observations are inherently heterogeneous in both structure and sampling. Satellite and reanalysis data are dense, regular 2D/3D grids, whereas in-situ measurements are sparse, irregularly distributed point profiles. To make these datasets jointly usable as model inputs, we should manipulate all datasets within a shared spatiotemporal coordinates system.

To bridge these differences, we establish a standardized metadata schema and a unified 4D coordinate standard. Following the FAIR (Findable, Accessible, Interoperable, Reusable) data principles [Zhu *et al.*, 2026], we align all input data onto this shared spatiotemporal scale. For dense satellite and reanalysis datasets, we employ bilinear interpolation to regrid all variables onto daily and monthly grids with a spatial resolution of $0.25°$. All gridded datasets are processed into spatially congruent, pixel-for-pixel. For sparse in-situ observations, we place each profile at the nearest spatiotemporal grid point and use a binary mask to indicate whether a grid cell contains ground truth or missing data. This mask allows the downstream model to distinguish real measurements from unobserved regions during training. The resulting output is a unified spatiotemporal datacube ($\mathcal{X} \in \mathbb{R}^{T\times C\times D\times H\times W}$), where T, C, D, H, and W denote the dimensions for time, variable channels, depth, height (latitude), and width (longitude), respectively. This datacube provides a consistent, machine-readable representation of heterogeneous ocean observations and serves as the foundational input for subsequent physics-informed models and analyses.

3.2 Physical-Informed Model Learning

The physical-informed model is the computational core of ocean heat warning system, connecting raw heterogeneous observations with the prediction and analysis of extreme events. As illustrated in Figure 2, the architecture follows an encoder-decoder paradigm with a physical-informed temporal evolution backbone: (1) a multi-modal encoding module that fuses heterogeneous observations into a unified latent state; (2) a physics-informed temporal evolution backbone network that simulates the dynamics changes of ocean heat under physical constraints of ocean dynamics; and (3) a decoder module that maps the evolved latent states back to temperature variables for downstream prediction and attribution tasks. While the overall architecture is shared, the temporal evolution backbone needs to be instantiated differently depending on the target extreme phenomenon. Next, we first describe the multi-modal fusion and latent representation learning shared in all settings, and then present two representative case studies of the temporal evolution module: one is a Neural ODE-based formula, which applies to short-term, rapidly evolving marine heat waves; the other is a constraint-based formula, which applies to long-term ENSO patterns.

Multi-Modal Fusion and Latent Representation

The model takes as input the unified spatiotemporal grids in the data integration stage. While the preceding integration step ensures that all variables are mapped onto a common spatiotemporal coordinate system, it does not resolve the differences of physical meaning and sampling density. To accommodate data heterogeneity, we utilize distinct encoders where separate encoders are tailored to the unique modality of each input source. For example, gridded surface and subsurface temperature fields can be handled by convolutional-based [Qiao *et al.*, 2025] or transformer-based architectures encoders [Zhou and Zhang, 2023] to capture spatial patterns, while sparse in-situ observations are encoded using GNN-based networks that can naturally accommodate irregular sampling [Yang *et al.*, 2023]. In addition, we incorporate a meta-encoding module to embed metadata as auxiliary information such as time, geographic location, and seasonality. We can also employ spherical coordinate encoding to better reflect the Earth's geometry [Lam *et al.*, 2023]. The outputs of all encoders are combined into a latent state vector $\mathbf{z}(t)$ at time t. This training is an end-to-end process encompassing atmospheric, surface, and subsurface variables, with in-situ observations and reanalysis data providing supervision for future time steps.

Case Study I: Short-Term Marine Heatwave

Short-term MHWs are characterized by the rapid onset and intensification of thermal anomalies. The key challenge in modeling such events lies in accurately capturing these abrupt temporal dynamics. The commonly used approaches treat time as a sequence of discrete snapshots [Qiao *et al.*, 2025; Prochaska *et al.*, 2021], which may smooth the rapid heat transport processes and underestimate peak intensities during extreme heat events. To overcome this, we model the ocean temperature evolution by leveraging the physics-informed neural ODE–based temporal evolution module developed in our previous work [Jiang *et al.*, 2026]. Rather than directly predicting discrete future states, this model learns the instantaneous rate of change of the latent state $(\partial\mathbf{z}/\partial t)$ and integrates it continuously forward in time. The learned dynamics are constrained by well-known physical principles in Eq. 1: the evolution of the 3D velocity field $\mathbf{v}$ is governed by the Navier-Stokes equations, while ocean temperature evolution follows the heat budget equations.

$$
\begin{aligned}
&\frac{\partial T}{\partial t} + \underbrace{(\mathbf{v}\cdot\nabla)T}_{\text{Advection}} = \underbrace{\kappa\nabla^2 T}_{\text{Diffusion}} + \underbrace{Q_{\text{net}}}_{\text{Flux}},\\
&\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\frac{1}{\rho}\nabla p + \nu\nabla^2\mathbf{v} + \mathbf{f}.
\end{aligned}
\tag{1}
$$

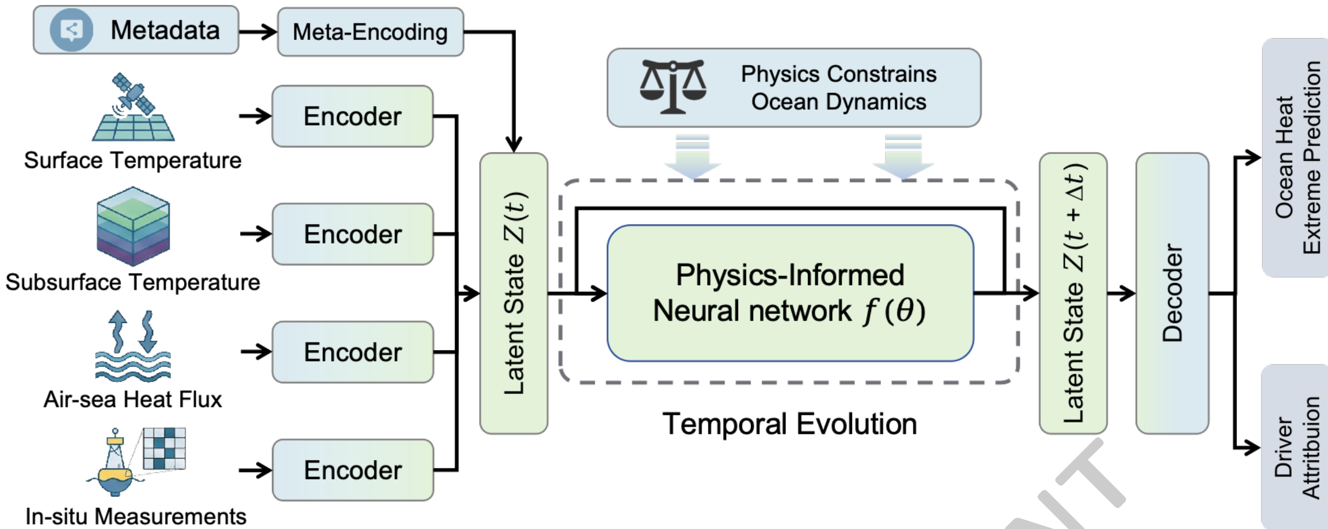

Figure 2: Architecture of the physics-Informed ocean heat prediction model. The model encodes multi-source data and metadata into a latent representation. This state evolves via a physics-informed temporal evolution module and is decoded into downstream states tailored to ocean heat extremes prediction and physical driver attribution.

In practice, due to the complexity of the N-S equations and the unavailability of 3D pressure fields, the evolution of the velocity field is modeled using a dynamic neural network, while the heat budget equation is embedded within the neural ODE formulation that governs temperature evolution. This formulation offers two key advantages. First, the adaptive ODE solver can accurately track continuous abrupt temporal gradients and the peak intensity of rapidly evolving heatwaves. Second, by explicitly representing 3D heat transport, diffusion and air–sea energy exchange processes, the system naturally supports attribution of the underlying drivers. By examining the relative contributions of these components across different directions, we can identify the dominant physical drivers underlying every marine heatwave event.

Case Study II: Long-Term Climate Patterns (ENSO)

In contrast to the rapid and short-lived nature of marine heatwaves, the global climate phenomenon ENSO is governed by long-term dependence and basin-scale air-sea interactions [Chen *et al.*, 2024]. Accurately modeling ENSO therefore requires preserving physical consistency over longer timescales. To address this challenge, we explore methods to improve ENSO prediction capabilities by explicitly incorporating ocean heat budget dynamics and using the neural network operator to simulate long-term ocean processes. By integrating these current velocity data, the system can input key components of the mixed-layer heat budget equation for learning the temporal evolution of ocean heat states, including horizontal heat transport (advection) and vertical heat transport (entrainment) and air-sea heat flux. For the backbone of temporal evolution, we can employ neural operators (e.g., Fourier neural operators [Li *et al.*, 2020] or Transolver [Wu *et al.*, 2024a]) that learn the underlying ocean dynamics equations governing fluid flow. To explicitly enforce sensitivity to extreme ENSO phases, we introduce the magnitude-weighted Niño Index loss function as an additional optimized objective:

$$\mathcal{L}_{\text{Niño}} = \frac{1}{N}\sum_{t=1}^{N}(1+\alpha|y_t|)(\hat{y}_t - y_t)^2, \qquad (2)$$

where y_t and $\hat{y}_t$ denote the ground-truth and predicted Niño 3.4 index at time t, respectively. By directly incorporating the absolute magnitude of the true index ($|y_t|$) as a dynamic scaling factor, the loss proportionally increases with the severity of the anomaly. The hyperparameter α is determined empirically on the validation set.

3.3 Incremental Learning

A key challenge facing long-term ocean heat warning systems is the non-stationary nature of the Earth's climate [Ryu *et al.*, 2025; Singh *et al.*, 2021]. As global warming intensifies, ocean temperatures once considered normal gradually shift, and the characteristics of extreme events evolve accordingly. Static models trained once on historical data gradually lose accuracy, while retraining them from scratch is costly.

To address this issue, we introduce an incremental learning mechanism for the core physical model (Figure 3). Instead of fixing model parameters after initial training, the system is regularly self-updated based on the acquisition of new observational data. We can employ a sliding window update strategy that when new satellite and Argo observational data are collected, the model performs lightweight and efficient gradient parameter updates using recent data. This allows the system to rapidly adapt to emerging warming trends and new extreme patterns not present in earlier data. Moreover, we should strictly control the update process to avoid catastrophic forgetting, where new information overwrites previously learned knowledge. We can employ an experience replay strategy that retains the representative historical extreme records and replays them with new data during each update. In addition, physical constraints such as the heat budget and

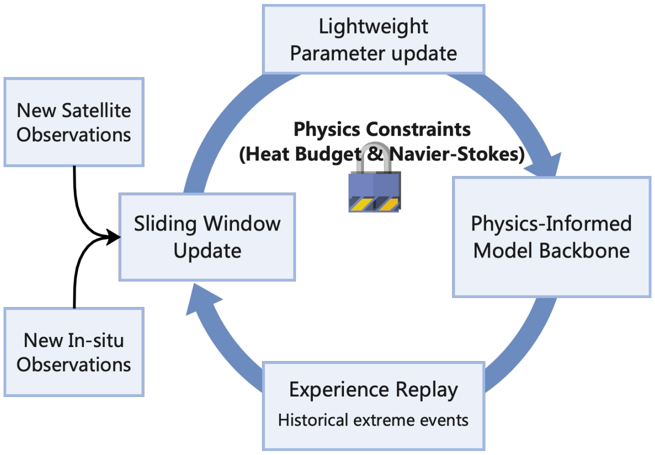

Figure 3: Schematic of the incremental learning loop. The system employs a sliding window approach to integrate incoming observational streams, dynamically updating model weights to adapt to recent ocean heat extremes while preserving historical extreme records via an experience replay buffer.

Navier-Stokes equations are always enforced during entire updates, ensuring that the model can adapt to new conditions while remaining physically consistent and stable over time.

3.4 System Tasks

The proposed system is designed to support two core tasks for monitoring and understanding ocean heat extremes: ocean heat extremes prediction and physical driver attribution.

Ocean Heat Extremes Prediction The primary objective of this task is to forecast the spatiotemporal evolution of 3D ocean temperature fields to detect impending ocean heat extremes. In addition, the system is also tasked with identifying potential extreme events and quantifying their key characteristics, including onset, intensity, duration and spatial extent, across varying lead times. The significance lies in its ability to provide reliable early warnings, enabling proactive risk assessment for marine ecosystems and policy management before an event reaches its peak.

Physical Driver Attribution This task aims to identify the physical processes that drive predicted ocean heat extremes. It does so by decomposing temperature anomalies into contributions from key mechanisms, including horizontal and vertical advection, diffusion, and air–sea heat exchange. By analyzing the relative importance of these processes, the system reveals how each extreme event develops and intensifies. This process explanation provides physically interpretable predictions and supports scientific understanding and decision-making.

3.5 Evaluation Benchmark and Criteria

To provide a holistic assessment of the proposed system, we establish an ocean heat extremes benchmark to enable fair and reproducible comparison across models. Building on this benchmark, we evaluate the system along four perspectives: predictive skill, physical consistency, the reliability of physical driver attribution, and risk assessment capability.

Ocean Heat Extremes Benchmark
We construct a specialized ocean heat extremes benchmark that covers short-term MHWs and long-term ENSO climate patterns. The benchmark is built upon the curated multi-source ocean temperature datasets in Section 3.1. Evaluating such extremes requires more than assessing pointwise temperature prediction error. We therefore implement a standardized labeling pipeline to generate unified, event-level ground truth annotations across all data sources. Specifically, we derive discrete event labels by applying widely accepted oceanographic definitions to temperature anomalies:

- **Marine Heatwave**: Note that the absence of a universally accepted definition for what constitutes a heatwave [Capotondi *et al.*, 2024]. Following the widely used definitions [Hobday *et al.*, 2016], an MHW occurs at a given location when daily SST anomalies exceed the seasonally varying 90th percentile climatology for five days or more. We compute these thresholds for every grid point, creating a binary mask that serves as the classification target.
- **ENSO**: We label temporal windows based on the Oceanic Niño Index (ONI), where the 3-month running mean of SST anomalies in the Niño 3.4 region exceeds $\pm 0.5°$C for five consecutive months as El Niño (warm).

Based on these definitions, we label each ocean heat extreme by its onset time, intensity, duration, and spatial extent. The benchmark supports the assessment of the key prediction tasks defined above. The data are strictly partitioned with a temporal gap: the period 1982–2015 is used for training, while the most recent period (2020–2025) is reserved for testing. To prevent data leakage stemming from potential "long-term ocean memory" effects, a 5-year gap is deliberately introduced between the two splits [Ham *et al.*, 2019]. The proposed system is evaluated against both numerical ocean models and purely data-driven deep learning methods to quantify the specific benefits of physics-informed learning.

Evaluation Criteria
The evaluation criteria below are used to assess prediction skill, physical consistency, attribution Reliability, and risk assessment capability for operational early warning.

- **Prediction Skill.** For global temperature field reconstruction, we report the root mean square error (RMSE) and Anomaly Correlation Coefficient (ACC). For extreme event detection, where strong class imbalance makes standard accuracy unreliable, we therefore use the symmetric extremal dependence index (SEDI) and the F1 score to assess detection skill.
- **Physical Consistency.** The central goal of the proposed system is to ensure physically plausible predictions for extreme heat events. To quantify the physical consistency, we evaluate the degree to which model forecasts satisfy conservation laws by computing the heat budget residual on the test set.
- **Attribution Reliability.** To assess the reliability of driver attribution, we conduct qualitative case studies on historical high-impact events. We will verify whether

the model's internal attribution aligns with post-hoc scientific literature. In particular, we select landmark historical extreme events, e.g., the 1997-1998 and 2015-2016 ENSO, as specific test cases for detailed analysis. We will also quantitatively validate the attribution by correlating the model's learned components with physical heat budget terms.

- **Risk Assessment Capability.** We assess risk assessment capability by measuring the model's lead time in predicting extreme events. Specifically, we evaluate how early (in days or months) the probabilistic prediction confidence exceeds an 80%–90% threshold. This empirically chosen threshold will be rigorously justified using Precision-Recall curves and F1-score trade-offs across various lead times.

4 Expected Results and Social Impact

This project is expected to deliver an early warning system for ocean heat extremes through AI-Ocean Dynamics synergy, which supports forecasting and analyzing these extremes across multiple spatial and temporal scales, including MHWs and large-scale climate modes ENSO. Compared with purely data-driven models, the proposed system is expected to improve the prediction of extreme event onset, intensity, duration and spatial extent. It also provides transparent attribution of driving factors based on explicit heat budget decomposition, contributing to a deeper scientific understanding of the underlying mechanisms. Furthermore, we aim to construct and release a benchmark of ocean heat extremes encompassing surface and subsurface observational data and standardized extremes definitions. This benchmark provides a general evaluation platform for future research on ocean heat extremes and physics-based learning methods. Finally, the deployment of the system incorporating incremental learning promises to achieve long-term, sustainable adaptation to climatic and environmental evolutions, with its deployment serving as a demonstrative application for other researchers and monitoring managers.

5 Challenges and Limitations

Despite the significant advantages of the proposed system, several challenges and limitations remain:

- **Data Bias.** While this project intentionally balanced the spatial distribution of multi-source data, the scarcity of observations in the southern hemisphere and coastal areas in developing regions may lead to less stable and reliable predictions compared to densely observed areas, such as the North Atlantic. Therefore, these results require careful interpretation and continuous supplementation through future observations.
- **Interpretability.** Attribution results based on heat budget decomposition provide physically consistent explanations but do not constitute rigorous causal inferences. We can identify which physical processes are related to the development of extreme events, but we cannot fully explain why these processes change. This limitation can be addressed by exploring the application of causal AI techniques in ocean heat extremes monitoring.
- **Computational Complexity.** The computational cost of physics-informed data-driven models remains relatively high when processing high-resolution 3D temperature fields. This study does not adopt the highest available spatial resolution, but instead selects a resolution that balances physical fidelity and computational feasibility.
- **Real-time Applications.** The current system prioritizes accuracy of ocean heat extremes and requires further optimization for operational early warning with high real-time requirements, such as through model compression or distillation to reduce inference overhead.
- **Model Stability and Drift.** While incremental learning allows the system to adapt to climatic and environmental evolutions, inappropriate updates can lead to model drift or degradation of physical consistency. For example, learning from short-term anomalies may distort long-term dynamics.

6 Collaboration Opportunities and Project Roadmap

This project is naturally well-suited for interdisciplinary collaboration across machine learning, physical oceanography, and climate science. We seek partnerships with marine meteorological agencies and high-performance computing centers to acquire reanalysis products and real-time observational data, while contributing standardized benchmark datasets to the industry. Collaboration with physical oceanographers will ensure the physical consistency of the model and support the interpretation of attribution results. Furthermore, collaboration with regional marine forecasting centers and stakeholders will provide actionable feedback to improve early warning indicators and system design.

The project will be implemented in four successive phases over an expected two-year period. Phase I focuses on data integration and standardization, including building a benchmark dataset for ocean heat extremes and evaluating numerical models and data-driven models. The main deliverable is a publicly released benchmark dataset with standardized evaluation protocols. Phase II aims to build the core physics-informed models and validate it using historical ocean heatwave events. The deliverable is a physics-informed ocean heat extremes prediction model with attribution capabilities for driving factors. Phase III introduces incremental learning, enabling the model to adapt to climatic and environmental evolutions and recent warming trends. The deliverable is an adaptive heat warning system that maintains performance without catastrophic forgetting. Phase IV emphasizes pilot deployment and demonstration in representative regions, including the Northwest Pacific, the China Seas and the Equatorial Pacific ENSO zone, supplemented by case studies and visualization tools. The deliverables are an operational demonstration report and a roadmap for broader application.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under grants 62076232 and 62172049, Fundamental Research Funds for the Beijing University of Posts and Telecommunications under Grant 2025AI4S06.

References

[Aouni *et al.*, 2025] Anass El Aouni, Quentin Gaudel, Juan Emmanuel Johnson, REGNIER Charly, Julien Le Sommer, van Gennip, Ronan Fablet, Marie Drevillon, Yann DRILLET, and Pierre Yves Le Traon. Oceanbench: A benchmark for data-driven global ocean forecasting systems. In *Proceedings of the 39th International Conference on Neural Information Processing Systems*, 2025.

[Bian *et al.*, 2024] Ce Bian, Zhao Jing, Hong Wang, and Lixin Wu. Scale-dependent drivers of marine heatwaves globally. *Geophysical Research Letters*, 51(3):e2023GL107306, 2024.

[Camps-Valls *et al.*, 2025] Gustau Camps-Valls, Miguel-Ángel Fernández-Torres, Kai-Hendrik Cohrs, Adrian Höhl, Andrea Castelletti, Aytac Pacal, Claire Robin, Francesco Martinuzzi, Ioannis Papoutsis, Ioannis Prapas, et al. Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, 16(1):1919, 2025.

[Capotondi *et al.*, 2024] Antonietta Capotondi, Regina R Rodrigues, Alex Sen Gupta, Jessica A Benthuysen, Clara Deser, Thomas L Frölicher, Nicole S Lovenduski, Dillon J Amaya, Natacha Le Grix, Tongtong Xu, et al. A global overview of marine heatwaves in a changing climate. *Communications Earth & Environment*, 5(1):701, 2024.

[Chassignet *et al.*, 2007] 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(1-4):60–83, 2007.

[Chen *et al.*, 2024] Zhiping Chen, Li Li, Bingkun Wang, Jiao Fan, Tieding Lu, and Kaiyun Lv. The impact of global warming on enso from the perspective of objective signals. *Atmospheric Research*, 299:107176, 2024.

[Cui *et al.*, 2025] Yingzhe Cui, Ruohan Wu, Xiang Zhang, Ziqi Zhu, Bo Liu, Jun Shi, Junshi Chen, Hailong Liu, Shenghui Zhou, Liang Su, et al. Forecasting the eddying ocean with a deep neural network. *Nature Communications*, 16(1):2268, 2025.

[Dong *et al.*, 2025] Tianyun Dong, Zhenzhong Zeng, Ming Pan, Dashan Wang, Yuntian Chen, Lili Liang, Shuai Yang, Yubin Jin, Shuxin Luo, Shijing Liang, et al. Record-breaking 2023 marine heatwaves. *Science*, 389(6758):369–374, 2025.

[Gao *et al.*, 2026] Yuan Gao, Hao Wu, Fan Xu, Yanfei Xiang, Ruijian Gou, Ruiqi Shu, Qingsong Wen, Xian Wu, Kun Wang, and Xiaomeng Huang. Neuralom: Neural ocean model for subseasonal-to-seasonal simulation. In *Proceedings of the 40th AAAI Conference on Artificial Intelligence*, volume 40, pages 14756–14764, Mar. 2026.

[Gregory *et al.*, 2024] Catherine H Gregory, Camila Artana, Skylar Lama, Dalena León-FonFay, Jacopo Sala, Fuan Xiao, Tongtong Xu, Antonietta Capotondi, Cristian Martinez-Villalobos, and Neil J Holbrook. Global marine heatwaves under different flavors of enso. *Geophysical Research Letters*, 51(20):e2024GL110399, 2024.

[Ham *et al.*, 2019] Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. Deep learning for multi-year enso forecasts. *Nature*, 573(7775):568–572, 2019.

[Hersbach *et al.*, 2020] 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.

[Hobday *et al.*, 2016] Alistair J. Hobday, Lisa V. Alexander, Sarah E. Perkins, Dan A. Smale, Sandra C. Straub, Eric C.J. Oliver, Jessica A. Benthuysen, Michael T. Burrows, Markus G. Donat, Ming Feng, Neil J. Holbrook, Pippa J. Moore, Hillary A. Scannell, Alex Sen Gupta, and Thomas Wernberg. A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, 141:227–238, 2016.

[Hu *et al.*, 2025] Shuang Hu, Meiqin Liu, Senlin Zhang, Shanling Dong, and Ronghao Zheng. Physics-informed neural networks for modeling ocean dynamics and parameter estimation: Leveraging ocean reanalysis data. *IEEE Journal of Oceanic Engineering*, 50(3):2248–2260, 2025.

[Huang *et al.*, 2025] 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. *arXiv preprint arXiv:2506.03210*, 2025.

[Jean-Michel *et al.*, 2021] Lellouche Jean-Michel, Greiner Eric, Bourdallé-Badie Romain, Garric Gilles, Melet Angélique, Drévillon Marie, Bricaud Clément, Hamon Mathieu, Le Galloudec Olivier, Regnier Charly, et al. The copernicus global 1/12 oceanic and sea ice glorys12 reanalysis. *Frontiers in Earth Science*, 9:698876, 2021.

[Jiang *et al.*, 2026] Zheng Jiang, Wei Wang, Gaowei Zhang, and Yi Wang. Sstode: Ocean-atmosphere physics-informed neural odes for sea surface temperature prediction. In *Proceedings of the 40th AAAI Conference on Artificial Intelligence*, volume 40, pages 38661–38669, 2026.

[Kashinath *et al.*, 2021] Karthik Kashinath, M Mustafa, Adrian Albert, JL Wu, C Jiang, Soheil Esmaeilzadeh, Kamyar Azizzadenesheli, R Wang, Ashesh Chattopadhyay, Aakanksha Singh, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, 379(2194):20200093, 2021.

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

[Li *et al.*, 2020] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[Nairn and Mason, 2025] John Nairn and Simon J Mason. Extreme heat and heatwaves: hazard awareness and impact mitigation. *The Lancet Planetary Health*, 9(7), 2025.

[Nathaniel *et al.*, 2024] Juan Nathaniel, Yongquan Qu, Tung Nguyen, Sungduk Yu, Julius Busecke, Aditya Grover, and Pierre Gentine. Chaosbench: a multi-channel, physics-based benchmark for subseasonal-to-seasonal climate prediction. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, pages 43715–43729, 2024.

[Prochaska *et al.*, 2021] J Xavier Prochaska, Peter C Cornillon, and David M Reiman. Deep learning of sea surface temperature patterns to identify ocean extremes. *Remote Sensing*, 13(4):744, 2021.

[Qiao *et al.*, 2025] Baiyou Qiao, Yanze Yang, Zhong Tang, Donghong Han, and Gang Wu. Cnn-attention combined with improved transformer model for medium-and long-term sst prediction. *Ocean Engineering*, 340:122315, 2025.

[Reynolds *et al.*, 2007] Richard W Reynolds, Thomas M Smith, Chunying Liu, Dudley B Chelton, Kenneth S Casey, and Michael G Schlax. Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, 20(22):5473–5496, 2007.

[Roemmich *et al.*, 2009] Dean Roemmich, Gregory C Johnson, Stephen Riser, Russ Davis, John Gilson, W Brechner Owens, Silvia L Garzoli, Claudia Schmid, and Mark Ignaszewski. The argo program: Observing the global ocean with profiling floats. *Oceanography*, 22(2):34–43, 2009.

[Ryu *et al.*, 2025] Jung-Hee Ryu, Jae-Heon Lee, and Song-Lak Kang. Non-stationary temperature extremes in south korea: An extreme value analysis of global warming impacts. *Atmospheric Research*, page 108388, 2025.

[Sen Gupta *et al.*, 2020] Alex Sen Gupta, Mads Thomsen, Jessica A Benthuysen, Alistair J Hobday, Eric Oliver, Lisa V Alexander, Michael T Burrows, Markus G Donat, Ming Feng, Neil J Holbrook, et al. Drivers and impacts of the most extreme marine heatwave events. *Scientific reports*, 10(1):19359, 2020.

[Shu *et al.*, 2025] Ruiqi Shu, Yuan Gao, Hao Wu, Ruijian Gou, Kun Wang, Yanfei Xiang, Fan Xu, Qingsong Wen, and Xiaomeng Huang. Ocean-e2e: Hybrid physics-based and data-driven global forecasting of extreme marine heatwaves with end-to-end neural assimilation. *arXiv preprint arXiv:2505.22071*, 2025.

[Singh *et al.*, 2021] Harsimrenjit Singh, Mohammad Reza Najafi, and Alex J Cannon. Characterizing non-stationary compound extreme events in a changing climate based on large-ensemble climate simulations. *Climate Dynamics*, 56(5):1389–1405, 2021.

[Sun *et al.*, 2023] Di Sun, Furong Li, Zhao Jing, Shijian Hu, and Bohai Zhang. Frequent marine heatwaves hidden below the surface of the global ocean. *Nature Geoscience*, 16(12):1099–1104, 2023.

[Wang *et al.*, 2024] Xiang Wang, Renzhi Wang, Ningzi Hu, Pinqiang Wang, Peng Huo, Guihua Wang, Huizan Wang, Senzhang Wang, Junxing Zhu, Jianbo Xu, et al. Xihe: A data-driven model for global ocean eddy-resolving forecasting. *arXiv preprint arXiv:2402.02995*, 2024.

[Wernberg *et al.*, 2025] Thomas Wernberg, Mads S Thomsen, Michael T Burrows, Karen Filbee-Dexter, Alistair J Hobday, Neil J Holbrook, Shinae Montie, Pippa J Moore, Eric CJ Oliver, Alex Sen Gupta, et al. Marine heatwaves as hot spots of climate change and impacts on biodiversity and ecosystem services. *Nature Reviews Biodiversity*, 1(7):461–479, 2025.

[Wu *et al.*, 2024a] Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

[Wu *et al.*, 2024b] Lichuan Wu, Jinbao Song, Leonie Esters, and Victor Alari. Air-sea interaction and oceanic extremes. *Frontiers in Marine Science*, Volume 11 - 2024, 2024.

[Xu *et al.*, 2025] Daosheng Xu, Zebin Lu, Jeremy Cheuk-Hin Leung, Dingchi Zhao, Yi Li, Yang Shi, Bin Chen, Gaozhen Nie, Naigeng Wu, Xiangjun Tian, et al. Ai models still lag behind traditional numerical models in predicting sudden-turning typhoons. *arXiv preprint arXiv:2502.16036*, 2025.

[Yang *et al.*, 2023] Hanchen Yang, Wengen Li, Siyun Hou, Jihong Guan, and Shuigeng Zhou. Higrn: a hierarchical graph recurrent network for global sea surface temperature prediction. *ACM Transactions on Intelligent Systems and Technology*, 14(4):1–19, 2023.

[Zhao *et al.*, 2025] Shan Zhao, Zhitong Xiong, Jie Zhao, and Xiao Xiang Zhu. Exebench: Benchmarking foundation models on extreme earth events. *arXiv preprint arXiv:2505.08529*, 2025.

[Zhou and Zhang, 2023] 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.

[Zhu *et al.*, 2026] Xiao Xiang Zhu, Zhitong Xiong, Yi Wang, Adam J. Stewart, Konrad Heidler, Yuanyuan Wang, Zhenghang Yuan, Thomas Dujardin, Qingsong Xu, and Yilei Shi. On the foundations of earth foundation models. *Communications Earth & Environment*, 7(1):103, 2026.