Dear Dr. Le Gao,

Thank you for your thorough, rigorous, and constructive review. We sincerely appreciate your careful examination of our manuscript, particularly regarding the exact boundaries of our scientific contribution, our terminology, and the validity of our forecasting setup. Your critical insights have highlighted areas where our manuscript requires refinement to ensure scientific rigor and clarity.

Below, we provide a point-by-point response to your major concerns, detailing the revisions we will implement in the manuscript.

1. Scientific Contribution and Scope of the Study

We appreciate your assessment regarding the specific scope of this study. The primary focus is not on uncovering novel physical or ecological mechanisms driving marine heatwaves (MHWs) in the East China Sea. The mechanistic analysis presented serves primarily as a diagnostic validation to ensure the model's predictions are physically consistent, rather than as an independent discovery of new dynamics.

Our core contribution is strictly methodological. The primary goal is to address a critical bottleneck in data-driven ocean forecasting: the "regression-to-the-mean" effect, which causes neural networks optimized on global errors to systematically underestimate the extreme values defining MHWs. Since reliable high-resolution SST prediction is the foundational prerequisite for MHW early warning, we focused on architectural and loss-function optimizations (e.g., Curriculum Loss, OHEM) to improve extreme-value fidelity. In the revised manuscript, we will clarify this scope by explicitly positioning the work as a methodological advancement in predictive modeling for marine extremes, while framing the diagnostic analysis as essential support for model interpretability.

2. Precision of "Physics-Informed" Terminology

Thank you for the constructive feedback on terminology. Given that "physics-informed" is typically reserved for models explicitly embedding strict governing equations (such as heat conservation or advection-diffusion equations), we will implement a comprehensive terminology revision throughout the text to ensure precision and avoid ambiguity.

All instances of "physics-informed" will be replaced with more accurate descriptors such as "physics-guided feature engineering" or "physics-aware framework". We will clearly articulate that physical variables and structural gradient penalties serve as auxiliary constraints to guide spatial patterns and improve generalization, distinguishing this approach from models that strictly enforce dynamical conservation laws.

3. Interpretation of Learned Physical Processes

We appreciate your caution regarding the strength of our mechanistic claims. Without explicit feature attribution diagnostics, describing the network as having "learned" specific processes like Ekman transport or vertical mixing implies a level of certainty that exceeds current evidence. Accordingly, we will recalibrate these interpretations in the revision to reflect statistical associations rather than proven mechanistic learning.

We will state that incorporating relevant atmospheric boundary conditions ($u_{10}, v_{10}, \Delta T$) provides a necessary physical context that enables the model to statistically approximate SST variations associated with wind and heat fluxes. This adjustment ensures our claims remain fully supported by the presented evidence while maintaining the physical rationale for our feature selection.

4. Experimental Setup and Information Leakage

We sincerely appreciate your careful examination of our forecasting setup, particularly regarding the potential use of future atmospheric forcing—a critical issue for the validity of any operational SST prediction system. We fully recognize that this point must be explicitly clarified.

We confirm unequivocally that NO future atmospheric data were used during training or inference. The model operates in a strictly causal, autoregressive manner: at each forecast initiation time *t*, only observational/reanalysis data from the preceding 14 days (i.e., *t*-13 to *t*) are used as input to predict SST for the subsequent 7 days (*t*+1 to *t*+7). This includes SST, 10-m wind vectors (u_{10}, v_{10}), air–sea temperature difference (ΔT), and their 3-day moving averages—all sourced exclusively from past ERA5 and OSTIA records. Atmospheric fields for lead days *t*+1 to *t*+7 are never accessed by the model at any stage.

This design is explicitly stated in Section 2.6.1 (Lines 207–209) and reinforced in Section 2.6.2 (Lines 245–247). Furthermore, the phrase “accumulation of uncertainty in atmospheric forcing and boundary conditions” in Section 4.2 (Line 540) was indeed imprecise. We will revise it to read: *“…error growth arises from the compounding of initial-condition uncertainties and limitations in representing unresolved oceanic processes, in the absence of future atmospheric forcing.”*

To further ensure reproducibility and transparency, our code enforces this causal constraint programmatically: the data loader explicitly masks all timestamps beyond the input window, and no future-dated files are ever loaded during training or evaluation. The code and data are available in the "Code and Data Availability" section of the manuscript, and reviewers are welcome to verify this implementation via the archived repository.

In summary, our framework represents a true hindcast/forecast experiment without information leakage. The 7-day predictive skill demonstrated in Figures 9–13 is achieved solely from past observations, making it directly relevant to operational early-warning applications where future atmospheric states are unknown.

Thank you again for raising these essential points. We believe these clarifications and forthcoming revisions resolve the concerns and significantly strengthen the methodological rigor of our study.

Best regards,

Zefang Ma (on behalf of all authors)

Marine Science and Technology College, Zhejiang Ocean University