Dear Reviewer,

Thank you for your detailed and constructive feedback on our manuscript entitled "A Novel Method for Sea Surface Temperature Prediction using a Featural Granularity-Based and Data-Knowledge-Driven ConvLSTM Model". We have carefully studied your comments, re-evaluated our work against the concerns raised, and would like to clarify key misunderstandings and present evidence to support the scientific value and rigor of our research. Given that your primary concern regarding prediction accuracy is the core reason for the rejection recommendation, we would like to first provide a detailed clarification on this critical issue—supplemented by specific experimental data and literature references—to address potential contextual misunderstandings.

1. Superior Prediction Accuracy Compared to Baseline Models Within the 48-Month Reliable Horizon

A well-recognized limitation of existing baseline models in SST prediction is their restricted reliable forecast horizon—most can only maintain acceptable accuracy for up to 48 months, with a sharp deterioration in performance beyond this period. To ensure a fair and rigorous comparison, we extracted the initial 48 months of predictions from our 120-month forecast (aligning with the maximum reliable horizon of baseline models) and conducted a head-to-head comparative analysis with the 48-month predictions generated by each baseline model across all three study areas. The results consistently demonstrate that our model achieves the highest prediction accuracy among all compared methods. Specifically, for monthly SST predictions over the 48-month period: In Study Area I, the discrepancies between predicted and observed SSTs fluctuated within the range of -0.6 to 0.54 ℃, with an RMSE of approximately 0.49 ℃; In Study Area II, the error range was -0.3 to 0.4 ℃, accompanied by an RMSE of ~0.2 ℃; In Study Area III, the prediction errors varied from -0.4 to 0.35 ℃, with an RMSE of ~0.21 ℃. These quantitative results fully validate that our model outperforms baseline models in terms of prediction precision within the horizon where baseline methods are deemed reliable.

2 Competitive Accuracy Compared to Literature in the Same Region, with a Significantly Longer Forecast Horizon

We received review comments from Reviewer 1 on November 29th, and over the past two weeks, we have conducted an extensive review of relevant SST prediction literature and confirmed that no existing study has achieved such an extended lead time (120 months). One comparative reference is a relevant study focusing on SST prediction in the South China Sea—consistent with our Study Area I, which is the least accurate region among our three research areas. This prior work achieved a maximum lead time of only 12 months for SST prediction in the same region, with an RMSE ranging from 0.464 to 0.530 ℃ (Fig.1). In contrast, our model achieves a 120-month-ahead prediction horizon for the same South China Sea region. Notably, the RMSE of our first 48 months of predictions reaches 0.49℃—a level comparable to the 12-month prediction accuracy of the aforementioned study—while our forecast horizon is four times longer. Additionally, the prior study reported degraded accuracy when its model was applied to other oceanic regions, whereas our model maintains stable and reliable performance across all three of our study areas.

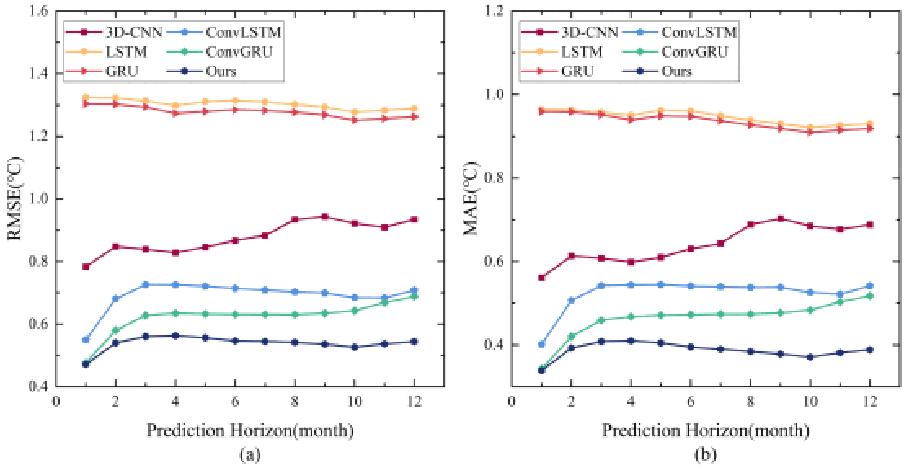

Fig.1 Changes in errors of different models in monthly predictions. (a) and (b) RMSE and MAE variations of 3D-CNN, LSTM, GRU, ConvLSTM, ConvGRU with Ours in monthly predicting, where Ours represents the MSIFM-SMAM-ConvGRU model, respectively. (A Spatio-Temporal Information Fusion Deep Learning Model for Daily and Monthly Sea Surface Temperature Prediction Based on ConvGRU | IEEE Journals & Magazine | IEEE Xplore)

3 The Presented Results Represent the Worst-Case Scenario of Our 120-Month-Ahead Prediction

Given our model's capability to achieve 120-month-ahead SST prediction, the results presented (as shown in the lengthy figure you referenced) were intentionally extracted to reflect the worst-case performance: specifically, the least accurate year of predictions from the worst-performing study area among our three research regions. Despite the presence of some high-error pixels, those with errors exceeding 1 K account for <8% of the total pixels, and the RMSE for this year ranges only from 0.29 to 0.82 ℃ (Fig.2). Importantly, this segment represents the lowest accuracy within our entire 120-month prediction period—our model's overall performance across the full forecast horizon is substantially better.

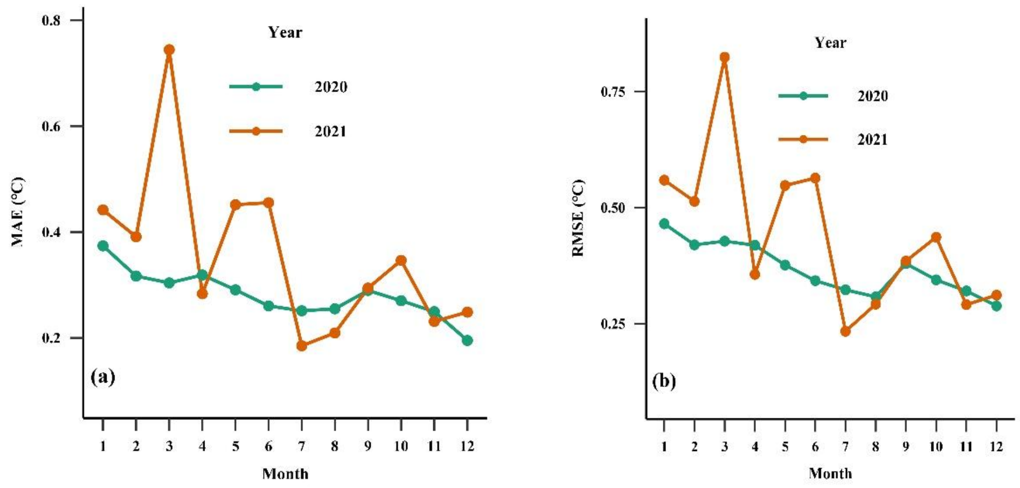

Fig.2 Accuracy Performance of the Two Years with the Lowest Prediction Accuracy in the Worst-Performing Study Area Among Three Regions for 10-Year-Ahead (120-Month) Prediction: (a) MAE Values; (b) RMSE Values

The aforementioned relevant study focusing on the South China Sea only achieved a maximum of 12-month-ahead prediction. Even within this shorter horizon, the study reported numerous high-error values in the Beibu Gulf region of the South China Sea, which was attributed to atmospheric circulation anomalies—a well-recognized challenge in oceanic SST prediction. Numerous other relevant studies—even those with different study areas—also report relatively low prediction accuracy in certain regions, often due to the presence of high-error pixels similar to those noted in our results. This underscores that the occurrence of suboptimal accuracy in specific regions (or a small proportion of high-error pixels) is not unique to our model, but rather a widespread challenge prevalent in SST prediction research. Such commonality further validates the reasonableness and robustness of our method (Fig.3-5).

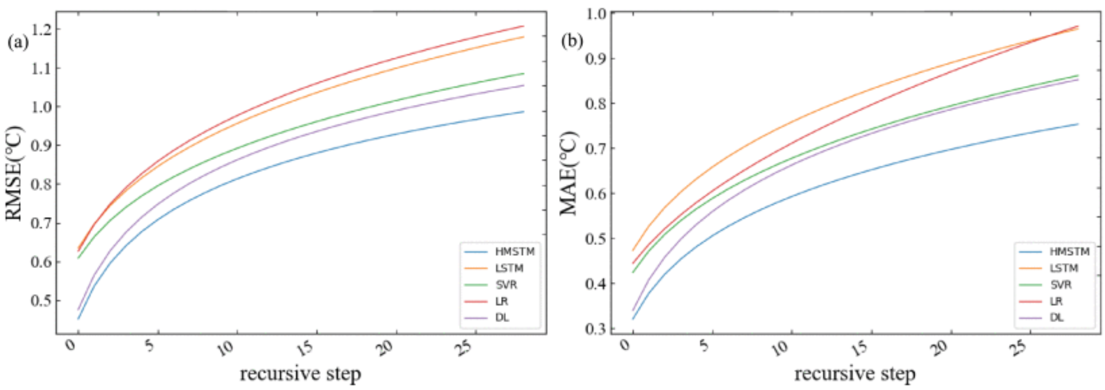

Fig.3 Trends of RMSE and MAE of all models implemented recursively concerning the number of recursive steps. (a) Comparison of RMSE. (b) Comparison of MAE.(Forecasting of Sea Surface Temperature in Eastern Tropical Pacific by a Hybrid Multiscale Spatial–Temporal

Model Combining Error Correction Map | IEEE Journals & Magazine | IEEE Xplore)

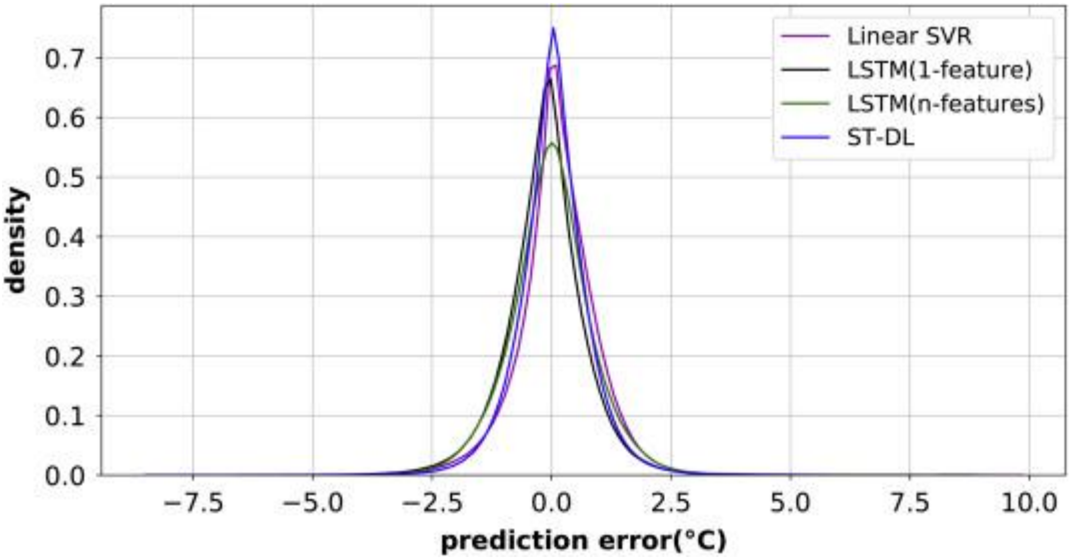

Fig.4 Gaussian kernel density estimation of 1–10 days ahead SST field prediction errors holistically.(https://www.sciencedirect.com/science/article/pii/S1364815218312295)

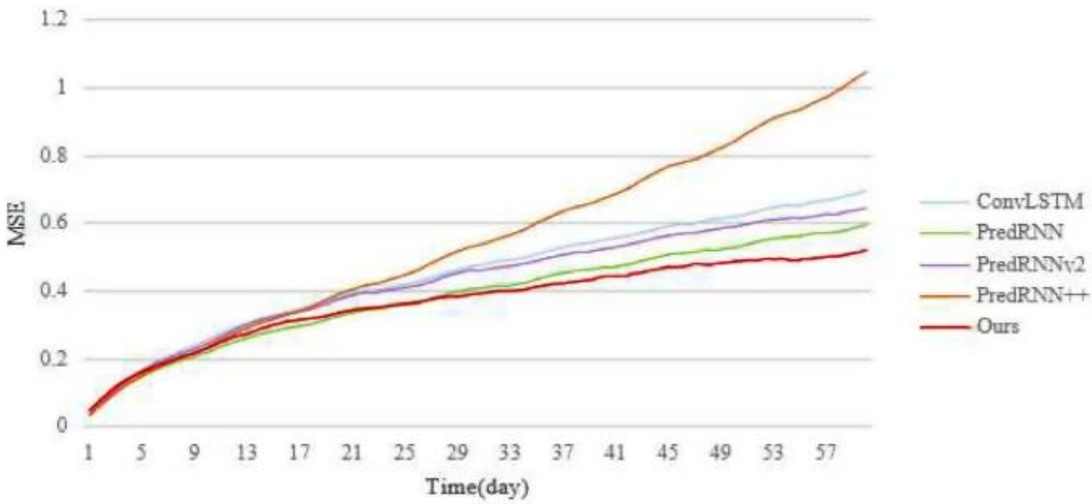

Fig.5 Comparison of prediction error (mean squared error) for different models over a period of 60 days (A spatio-temporal predictive learning model for efficient sea surface temperature forecasting | Climate Dynamics)

Thank you very much for carefully reading our manuscript and dedicating your valuable time to its thorough review. We greatly appreciate the constructive comments and insightful suggestions you have provided—they are highly valuable for improving the quality and rigor of our work. In response to your feedback, we will systematically revise the manuscript point by point, addressing each of your concerns in detail and incorporating your thoughtful input to enhance clarity, presentation, and scientific depth.