Dear Reviewer,

Thank you for taking the time to review our manuscript for the second time, and for providing your detailed and constructive feedback. We sincerely apologize for failing to fully grasp your core concerns in our first response, which led to misunderstandings on both sides.

Over the past two weeks, we have carefully studied your feedback, conducted an extensive review of relevant SST prediction literature, and thoroughly re-evaluated our work. We realize that our unclear explanations in the first response may have caused you to question the prediction accuracy of our model. We are grateful for your current comments, which have helped us identify the root cause of this misunderstanding. We are now revising the manuscript point by point in light of your suggestions, while also optimizing our algorithm accordingly. Before submitting the revised version, we would like to clarify the key issues as follows:

1. Response to the Concern about Prediction Accuracy

We understand that you consider our model's prediction accuracy unacceptably low, citing the ±2.5°C error in the South China Sea (Study Area I) for January 2021 shown in Figure 12, and noting that this performance is inferior to the <1°C error of the climatological forecast. To clarify this misunderstanding, we emphasize the following three key points:

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

Our model is designed to achieve 120-month-ahead SST prediction based on monthly historical data across three study areas. The results displayed in Figure 12—the extremely long figure you pointed out—are intentionally selected as the worst-case scenario: specifically, the predictions for the least accurate year within the 120-month horizon, derived from Study Area I (the poorest-performing region among our three study areas).

While it is true that some pixels exhibit relatively large errors, 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.1). These metrics reflect the minimum performance level of our model, not its overall capability across the full 120-month prediction horizon.

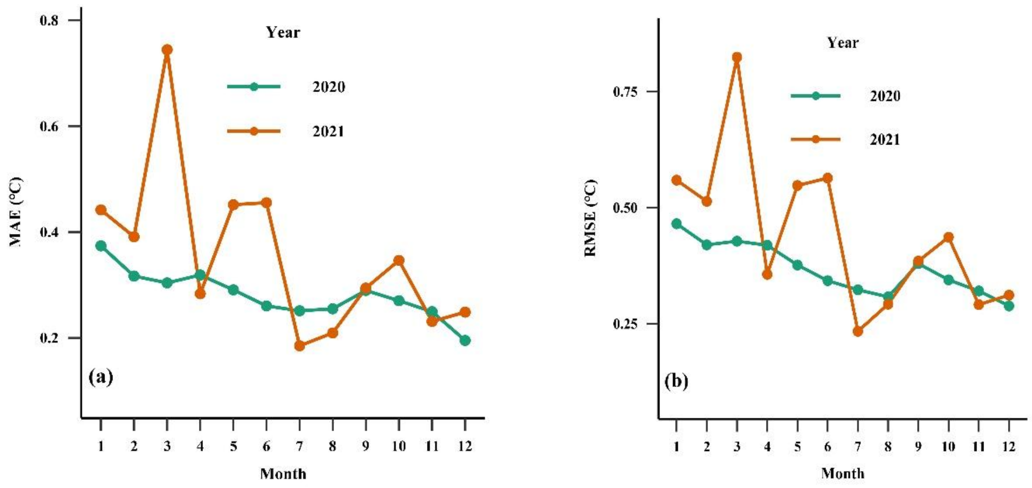

Fig.1 Accuracy Performance of the Two Years with the Lowest Prediction Accuracy in the Worst-Performing Study Area Among Three Regions for 10-Year-Ahead (120-Month) Prediction: (a) MAE Values; (b) RMSE Values

1.2 Comparison with Relevant Studies in the Same Region

Through extensive literature review over the past two weeks, we confirm that no existing study has achieved such an extended lead time of 120 months for SST prediction. For instance, one relevant study focusing on the South China Sea (consistent with our Study Area I, the region shown in Figure 12) only achieved a maximum lead time of 12 months, with an RMSE ranging from 0.464 to 0.530℃ (Figs.2-3). Additionally, this study reported degraded accuracy when applied to other regions, and due to atmospheric circulation anomalies, it also produced numerous high-error pixels in the Beibu Gulf area. In contrast, our model achieves a 120-month-ahead prediction horizon for the same South China Sea region, which is the least accurate region among our three research areas. 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.

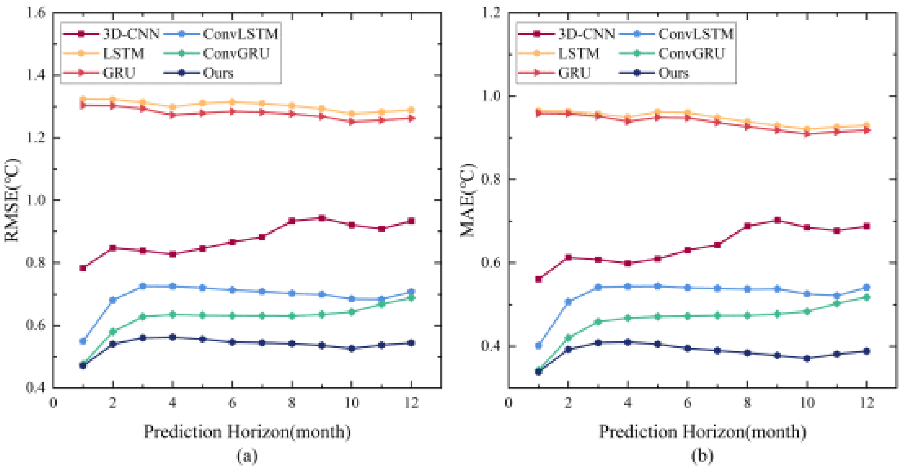

Fig.2 Changes in errors of different models in monthly predictions. (a) and (b) RMSE and MAE variations of 3D-CNN, LSTM, GRU, ConvLSTM, ConvGRU with Ours in monthly predicting, where Ours represents the MSIFM-SMAM-ConvGRU model, respectively. (A Spatio-Temporal Information Fusion Deep Learning Model for Daily and Monthly Sea Surface Temperature Prediction Based on ConvGRU | IEEE Journals & Magazine | IEEE Xplore)

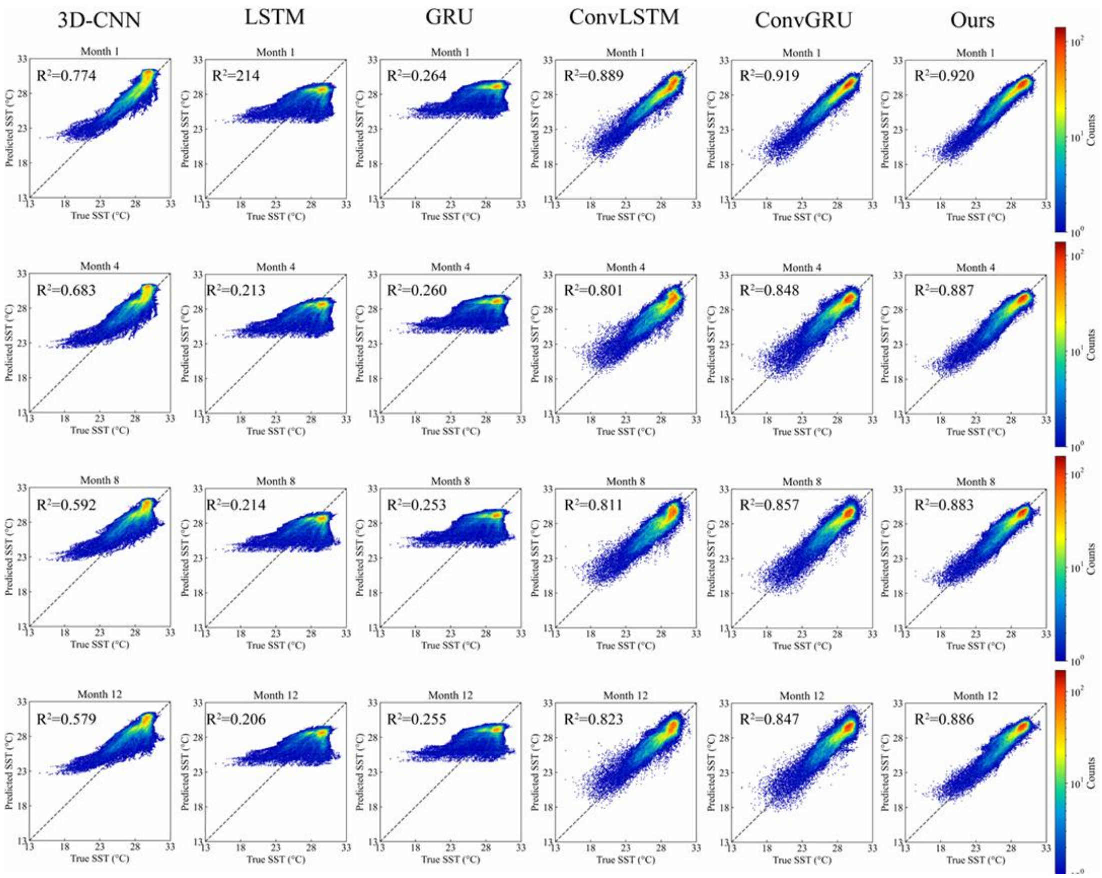

Fig.3 Spatial Distribution of Predicted vs. Observed SST and Corresponding R^2 Values for Different Models at Months 1, 4, 8, and 12(A Spatio-Temporal Information Fusion Deep Learning Model for Daily and Monthly Sea Surface Temperature Prediction Based on

ConvGRU | IEEE Journals & Magazine | IEEE Xplore)

In fact, many other related studies also report the presence of high-error pixels, with prediction errors increasing as the lead time extends (as shown in figs.4-6). This further validates that the existence of a small proportion of high-error pixels is a common challenge in the field of SST prediction, rather than a flaw unique to our model.

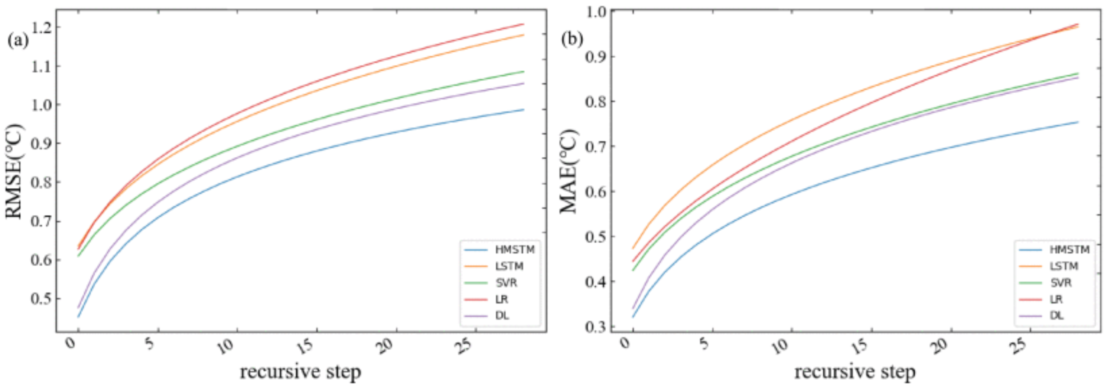

Fig.4 Trends of RMSE and MAE of all models implemented recursively concerning the number of recursive steps. (a) Comparison of RMSE. (b) Comparison of MAE.(Forecasting of Sea Surface Temperature in Eastern Tropical Pacific by a Hybrid Multiscale Spatial–Temporal Model Combining Error Correction Map | IEEE Journals & Magazine | IEEE Xplore)

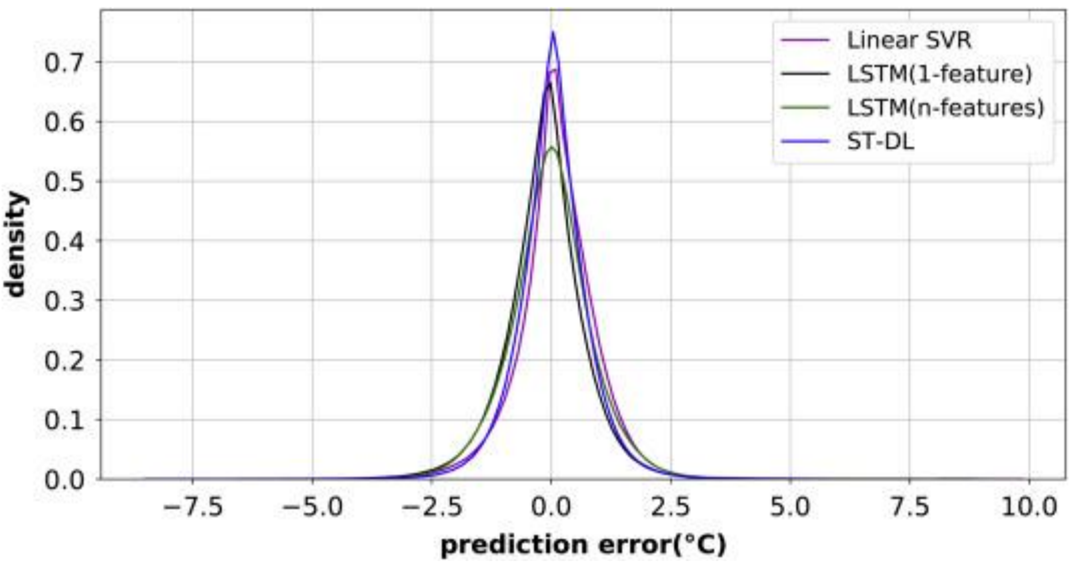

Fig.5 Gaussian kernel density estimation of 1–10 days ahead SST field prediction errors holistically.(https://www.sciencedirect.com/science/article/pii/S1364815218312295)

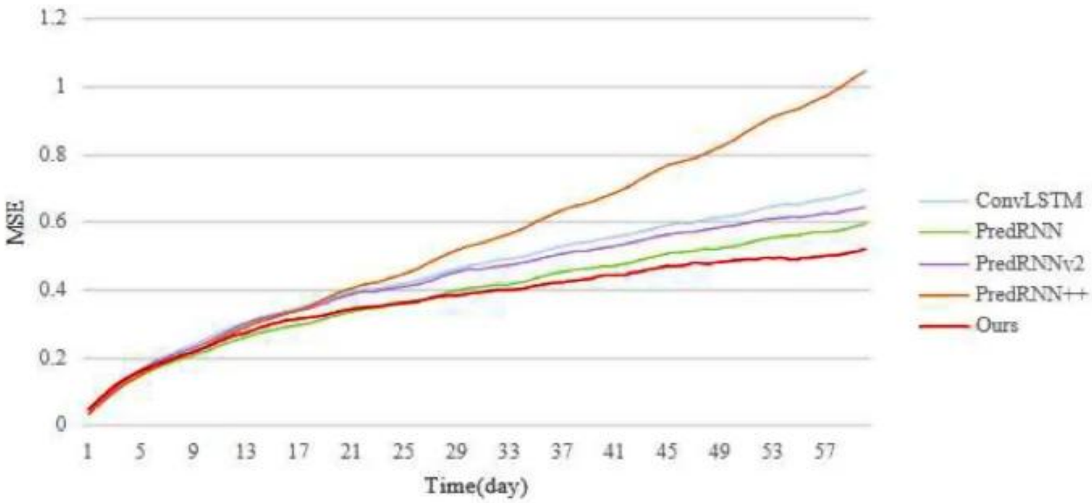

Fig.6 Comparison of prediction error (mean squared error) for different models over a period of 60 days ([A spatio-temporal predictive learning model for efficient sea surface temperature forecasting | Climate Dynamics])

1.3 Superior 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. Response to the Concern about Presentation Quality

We fully agree with your comment on the need to improve result presentation. We acknowledge that some figures (e.g., Figures 11 and 12) are excessively long. To address this concern, we will revise the format of these figures by simplifying redundant subplots, adding clear annotations, and restructuring the layout to enhance readability, strictly following the presentation standards of high-impact journals in the field.

3. Model Optimization Inspired by Your Suggestion

Inspired by your comments, we have been optimizing our model since receiving your feedback. Specifically, we have attempted to incorporate SSTA as an input variable for SST prediction, while also integrating certain oceanographic mechanisms into the loss function. Unfortunately, the results obtained today indicate that the incorporation of oceanographic mechanisms did not lead to a significant improvement in prediction accuracy, whereas the inclusion of SSTA has demonstrated a enhancement in precision. We are currently conducting validation experiments across other study areas to further verify this finding. Once all results are available, we plan to redesign the result visualization format in accordance with your suggestions.

Importantly, even our original model already outperforms many state-of-the-art methods in the literature—especially considering we only presented the worst-case segment of our 120-month predictions to ensure research rigor. Moreover, our original model does not directly take raw SST data as input or predict raw SST values; instead, it operates on feature variables extracted from SST data, and the actual SST values are derived by de-granulating these feature variables. This means that if our model were trained on raw SST data, it could indeed learn the seasonal variation patterns of SST. However, the feature vectors used in our manuscript exhibit almost no seasonal patterns; instead, they largely capture the interannual variation information of SST. Thus, we firmly believe our model's accuracy is not inferior to existing methods.

Thank you again for your invaluable suggestions. We will make every effort to optimize our algorithm and revise the manuscript thoroughly in line with your comments. We hope these clarifications can address your concerns, and we look forward to your further guidance on our work.

Sincerely,

Mengmeng Cao