Ms. Ref. No.: EGUSPHERE-2025-239

Title: A Novel Method for Sea Surface Temperature Prediction using a Featural Granularity-Based ConvLSTM Model of Data-Knowledge-Driven

Journal: GMD

Responses to RC #3:

I challenge the value of proposed SST forecast. In the ocean, the monthly SST, as the predicted variable of this study, has strong seasonality. If we use the SST of, for example, Jan of 2008, as the 'forecast' of Jan of 2009, we will find unexpectedly small differences between them, generally <2 degree. This is called 'persistence' forecast, but using the seasonality information. We may call it 'climatological forecast'. Fig. 1 shows an example of such a climatology forecast. My point is the proposed method does not exceed such a simple 'forecast', with higher forecast errors, making the forecast operationally useless. When given SST for a deep learning to learn, it only learns the seasonal pattern, not subtle year-to-year SST changes. That is the reason why many studies predict SST anomaly, not SST itself.

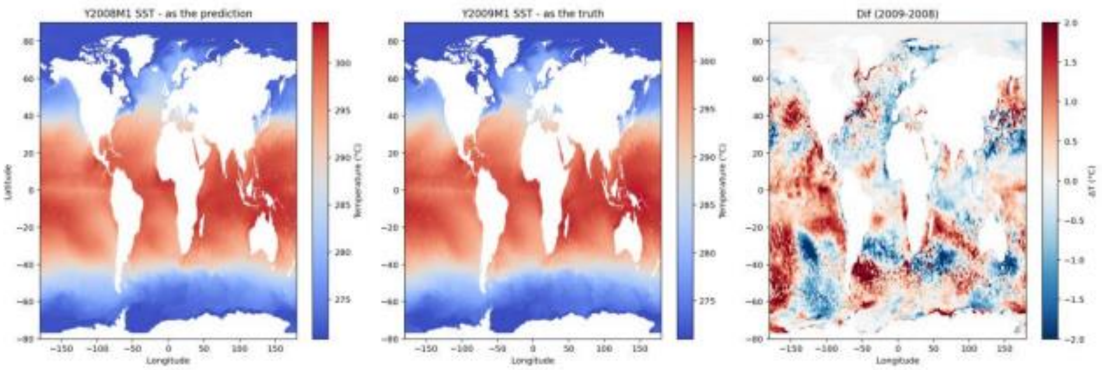


Fig. 1 Example of SST climatological forecast

Reply: Thank you sincerely for your valuable perspective on SST forecasting, which has helped us clarify the positioning and significance of our work. We fully agree with your emphasis on the importance of SST anomaly (SSTA) forecasting, as anomalies effectively isolate interannual variability and are critical for understanding climate dynamics. However, we respectfully argue that monthly absolute SST forecasting retains distinct value, and our work aims to contribute to this domain by addressing specific gaps.

First, monthly SST prediction remains a focus of numerous studies, including work by Mehmet Bilgili and colleagues (e.g., their 2025 study on Global monthly SST forecasting), where absolute temperatures are directly relevant to applications such as marine ecosystem management, fisheries planning, and coastal engineering—fields where threshold-based decisions (e.g., thermal tolerance of species) depend on absolute values rather than anomalies alone. Our motivation aligns with this practical need: to develop a systematic framework that captures the complex spatiotemporal dependencies of SST and extends prediction horizons, which remains underexplored in long-term (multi-year) monthly SST forecasting.

To achieve this, our approach was designed to go beyond replicating seasonal patterns:

We first analyzed the multi-faceted drivers of SST variability (e.g., ocean currents, wind stress, solar radiation) and curated a comprehensive database integrating universal oceanic and meteorological variables. Recognizing regional differences in these drivers, we partitioned the study area using pixel-wise similarity analysis, then employed random forest to identify and validate region-specific key factors (e.g., upwelling intensity in coastal zones vs. atmospheric forcing in open oceans). Building on this, we developed a ConvLSTM model enhanced with granularity information, leveraging these identified predictors to capture both large-scale seasonal cycles and fine-scale interannual variations.

To rigorously test the model's capability, we selected three representative regions differing in distance from land (coastal vs. open ocean), latitude (tropical vs. temperate), and baseline temperature regimes. Results confirm that our model achieves reliable 120-month (10-year) SST predictions across all regions, with errors consistently lower than those from simple climatological persistence.

To further validate its effectiveness, we compared our model with five widely used machine learning baselines for SST forecasting, and—responding directly to your concern—incorporated the Graph Memory Neural Network (GMNN) proposed by Liang et al. (2023), a state-of-the-art spatiotemporal graph model for SST prediction. After adjusting GMNN's hyperparameters (e.g., hidden layer dimensions) to align with our experimental setup, our ConvLSTM model retained the highest precision across all metrics and regions, demonstrating its ability to outperform both traditional and advanced baselines (Figs.2-4).

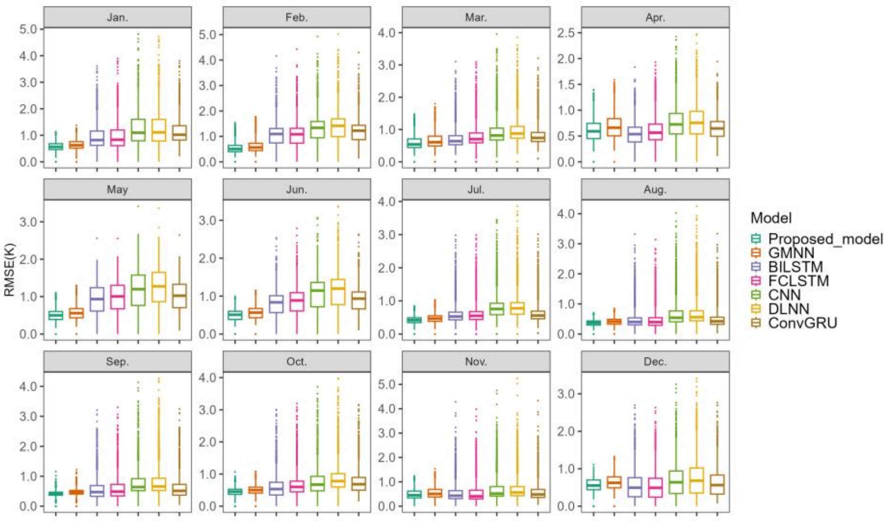


Fig.2 RMSE of SST predictions for the years 2012-2021 within study area I, obtained using the baseline models and the proposed model.

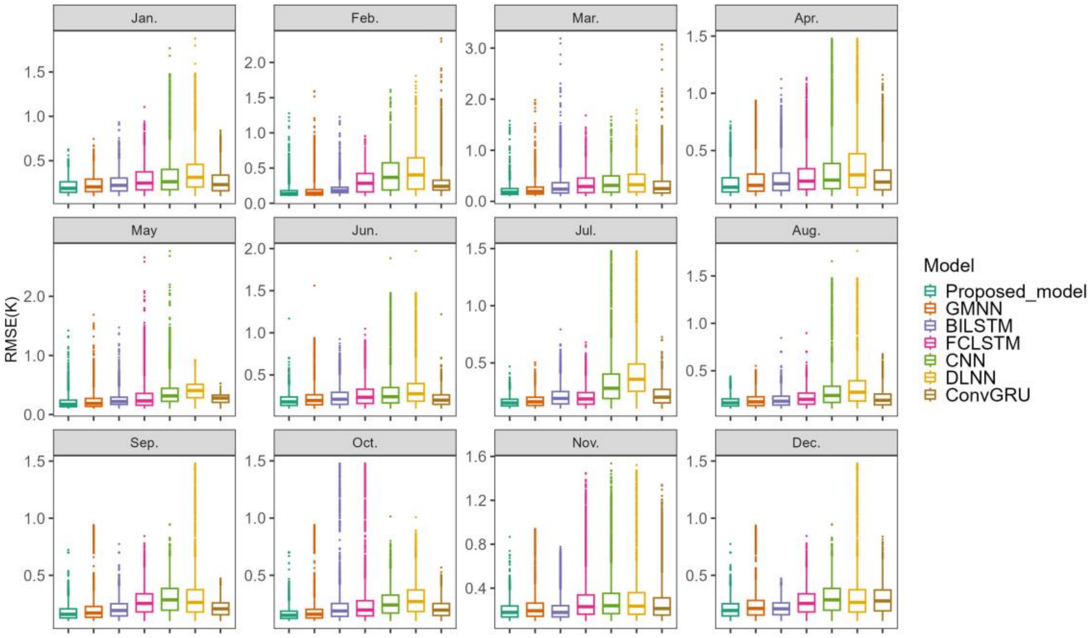


Fig.3 RMSE of SST predictions for the years 2012-2021 within study area II, obtained using the baseline models and the proposed model.

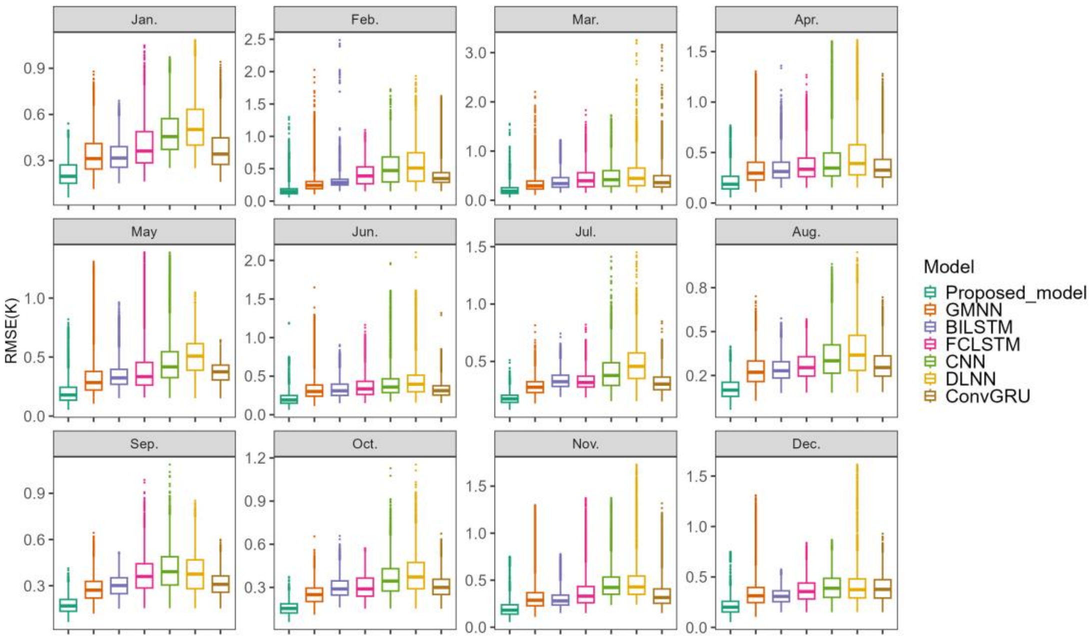


Fig.4 RMSE of SST predictions for the years 2012-2021 within study area III, obtained using the baseline models and the proposed model.

In summary, while we acknowledge the central role of SSTA forecasting in climate research, our work focuses on a complementary goal: providing a robust framework for long-term absolute SST prediction that serves practical applications and advances understanding of spatiotemporal SST dynamics. The comparative results confirm that our model goes beyond replicating seasonality, capturing meaningful interannual variations and outperforming established methods—thus justifying its value in the broader landscape of SST forecasting. Building on this constructive input, we have also modified our ongoing research on daily SST prediction to focus on SSTA forecasting.

Reference：

Bilgili, M., Pinar, E., and Durhasan, T.: Global monthly sea surface temperature forecasting using the SARIMA, LSTM, and GRU models, Earth Science Informatics, 18, 10, 10.1007/s12145-024-01585-z, 2024.

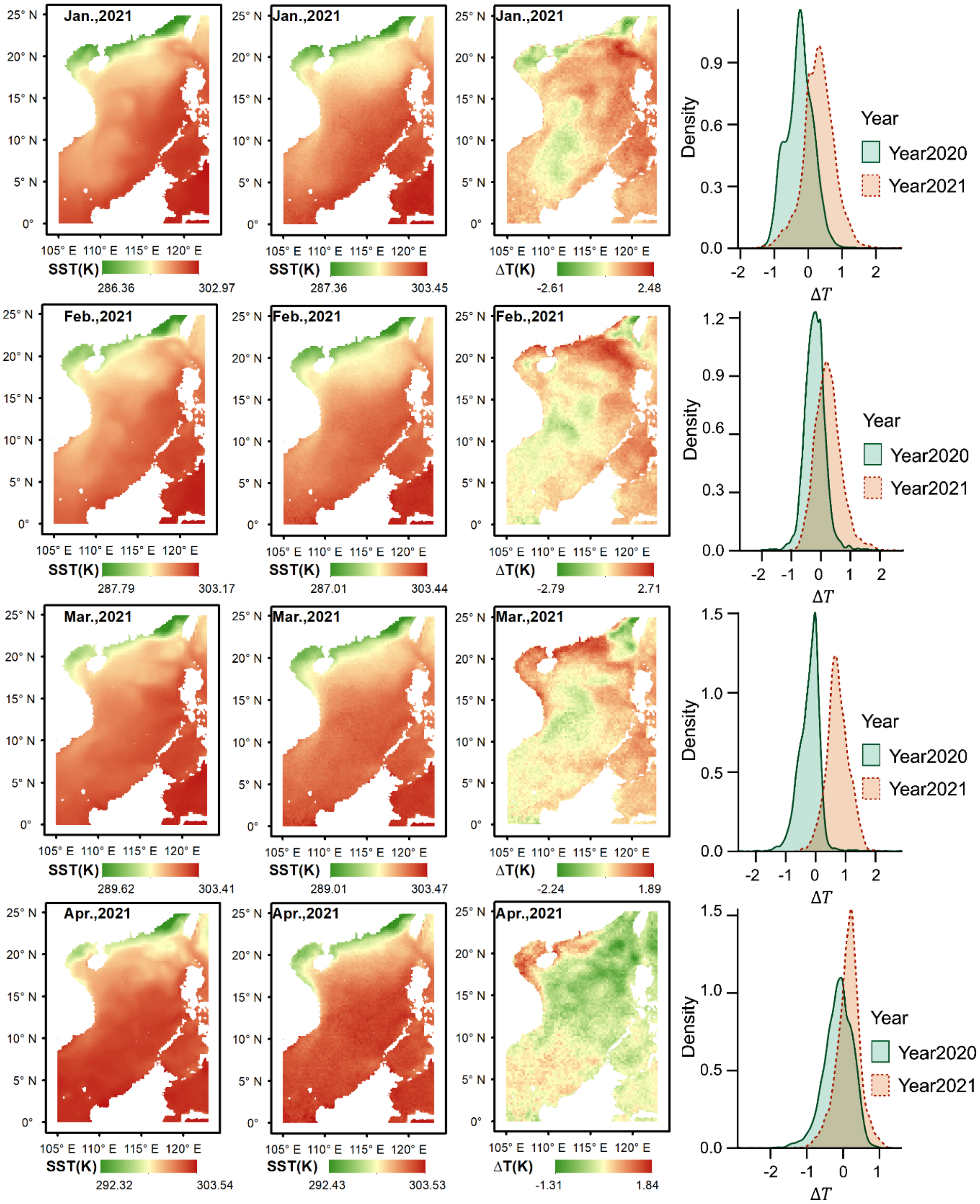
Liang, S., Zhao, A., Qin, M., Hu, L., Wu, S., Du, Z., and Liu, R.: A Graph Memory Neural Network for Sea Surface Temperature Prediction, Remote Sensing, 15,3539, 10.3390/rs15143539, 2023.

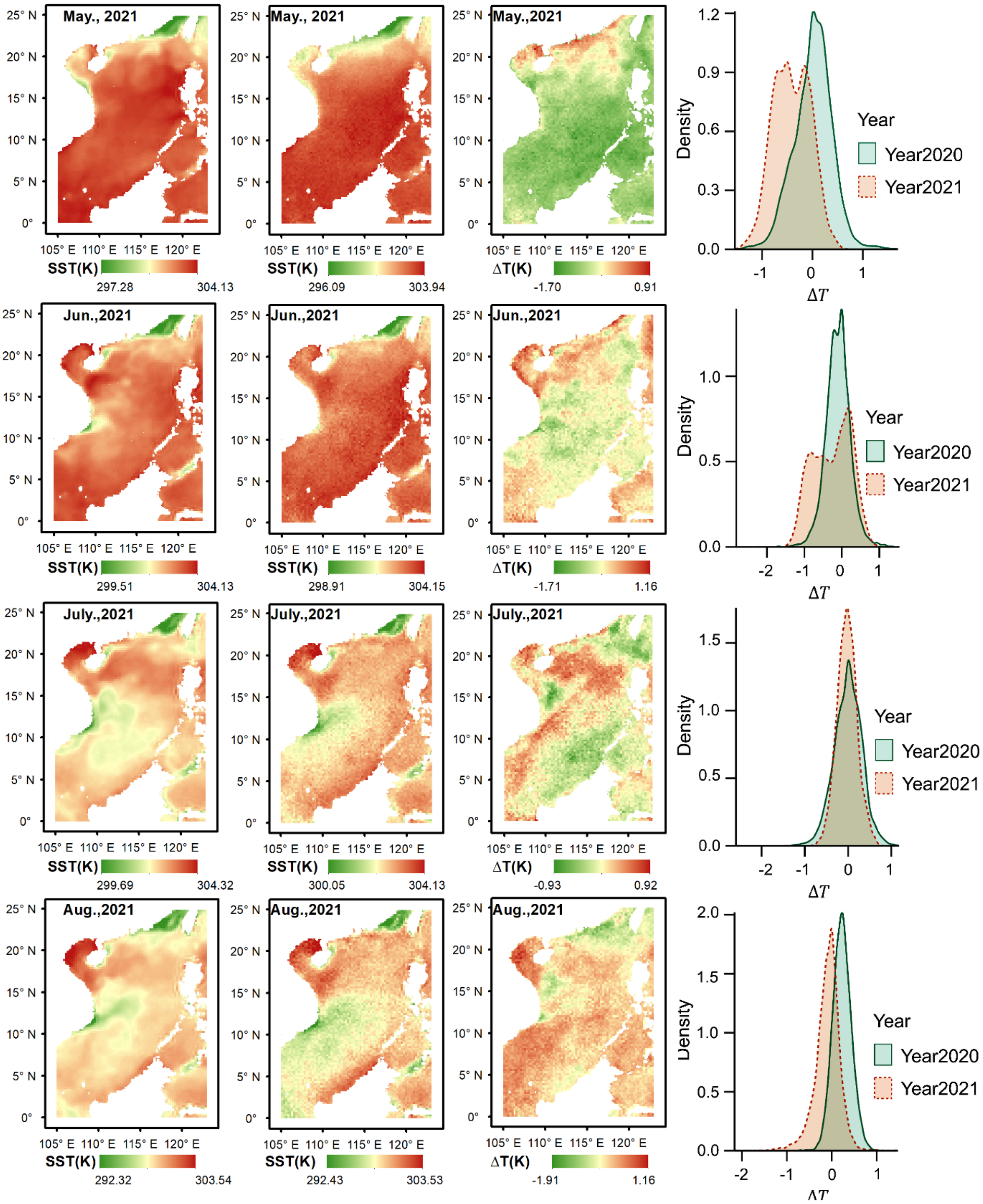
Besides, the manuscript is poorly written. (1) Although in the abstract, the authors claim that 120-month forecast is obtained, but no results were shown in the manuscript. (2) Section 3.3.4 is written in a non-informative and verbose way.

Reply: Thank you for your feedback. We note that the two points raised are related to the same content—specifically, the presentation of the 120-month SST forecast results, which is addressed in Section 3.3.4. We appreciate the opportunity to clarify and strengthen this section.

Section 3.3.4 is explicitly dedicated to elaborating on the 120-month forecast results. Given the impracticability of presenting all 120 months of data due to space constraints, we strategically selected two representative years: 2021 (the year with the lowest overall forecast accuracy) and 2020 (the second-lowest). This choice was intentional: by focusing on the two years with the weakest performance, we aimed to demonstrate the model's reliability even in suboptimal scenarios, thereby supporting the robustness of predictions across the entire 10-year horizon. If there is high prediction accuracy in 2021, it proves that the predictions of other years are also reliable. Fig. 5 shows the predicted SSTs for each month of the year 2021 within study area I. The first column represents the observed SST images, the second column is the predicted SST images, the third column is the difference between the observed and predicted SST images, and the fourth column exhibits the density plot depicting the disparities between the predicted and observed values for the respective months in 2021 and 2020. Our proposed methodology demonstrates the ability to forecast SST distribution for the year 2021, with discrepancies between predicted and observed values typically falling within the range of -1 to 1 K. The forecast accuracy for 2021 is slightly lower than that for 2020, with February, March, May, June, October, and December having lower forecast accuracy for the SST in 2021 than that for 2020. In contrast, the forecast accuracy for the other months is similar to that of 2020, indicating that the forecasts for other years are reliable.

By including both 2020 and 2021, we sought to provide a more holistic view of the model's long-term stability: if the model performs reliably in the two years with the lowest accuracy, it logically follows that it maintains or improves in years with higher accuracy across the 120-month horizon. This approach, we believe, better supports the robustness of the full 10-year predictions than showcasing only the best-performing years. We have added this explanation to Section 3.3.4 in the revised manuscript to enhance clarity. Thank you again for prompting us to elaborate on this rationale. Furthermore, to address the concern that the section was "non-informative and verbose," we have streamlined the content in the revised manuscript.


Jan.,2021
Feb.,2021
Mar.,2021
Apr.,2021
25° N
20° N
15° N
10° N
5° N
0°
105° E 110° E 115° E 120° E
SST(K)
ΔT(K)
286.36 302.97
287.36 303.45
-2.61 2.48
287.79 303.17
287.01 303.44
-2.79 2.71
289.62 303.41
289.01 303.47
-2.24 1.89
292.32 303.54
292.43 303.53
-1.31 1.84
Density
ΔT
Year
Year2020
Year2021


May., 2021
May., 2021
May.,2021
SST(K)
297.28
304.13
SST(K)
296.09
303.94
ΔT(K)
-1.70
0.91
Jun.,2021
Jun.,2021
Jun.,2021
SST(K)
299.51
304.13
SST(K)
298.91
304.15
ΔT(K)
-1.71
1.16
July.,2021
July.,2021
July.,2021
SST(K)
299.69
304.32
SST(K)
300.05
304.13
ΔT(K)
-0.93
0.92
Aug.,2021
Aug.,2021
Aug.,2021
SST(K)
292.32
303.54
SST(K)
292.43
303.53
ΔT(K)
-1.91
1.16
25° N
20° N
15° N
10° N
5° N
0°
105° E 110° E 115° E 120° E
Density
ΔT
Year
Year2020
Year2021

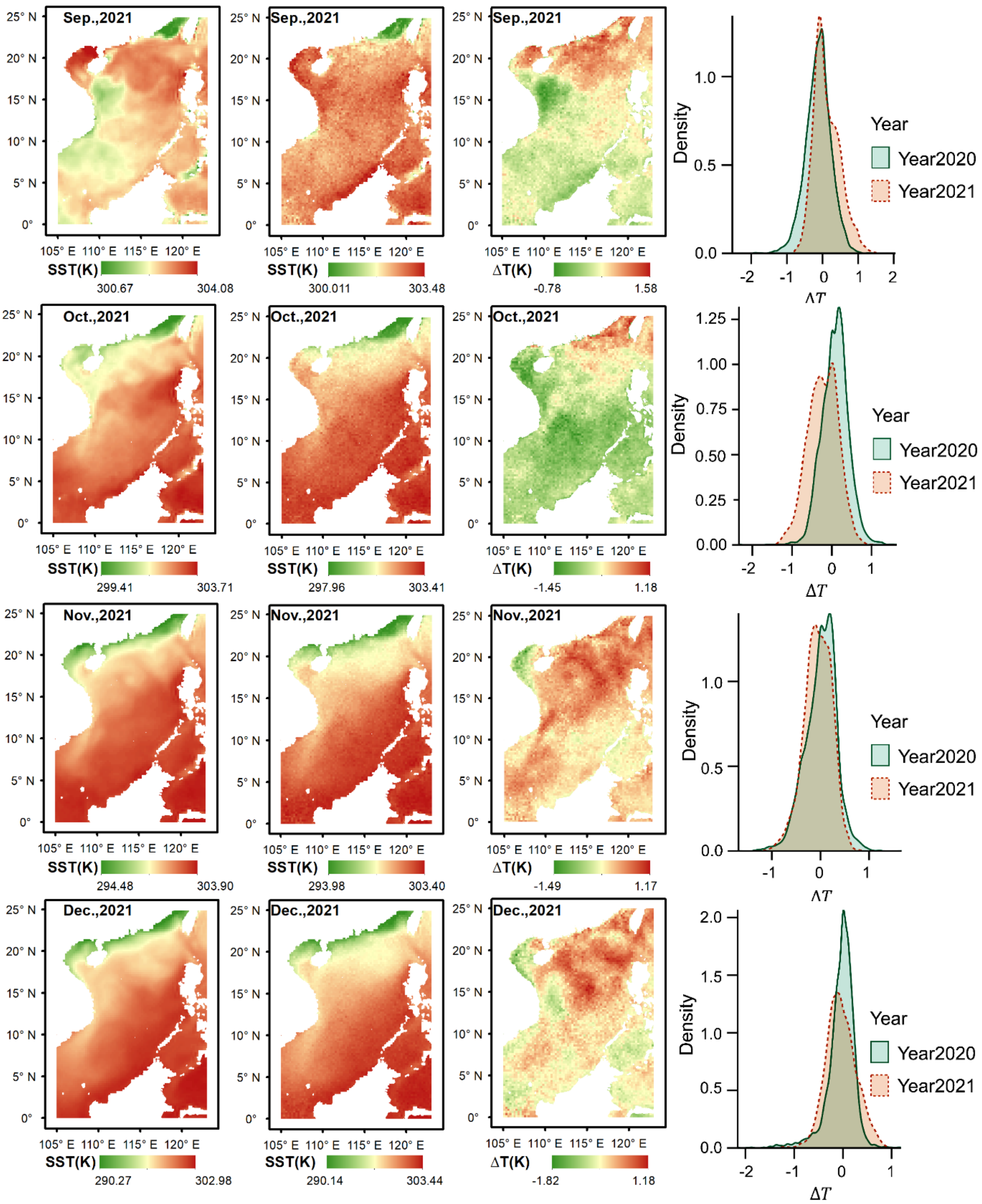


Fig.5 The predicted SSTs (first column), observed SSTs (second column), and spatial distribution (third column) and statistics (fourth column) of prediction errors for the study area I in 2021

(3) https://doi.org/10.1080/17538947.2023.2260779 this DOI in the Code and Data Availability is in-accessible.

Reply: Thank you for your valuable feedback. We deeply apologize for any concerns caused and are truly grateful for your dedication to transparent scholarship.

We immediately downloaded the original Zenodo attachment (https://doi.org/10.5281/zenodo.14759549) and verified that it contains code files when accessed via our database interface, as shown in the screenshot Fig. 6. To ensure universal accessibility, we proactively uploaded a new ZIP-format code package (https://doi.org/10.5281/zenodo.15714288) with identical content and will update the manuscript to prioritize this direct-access option.

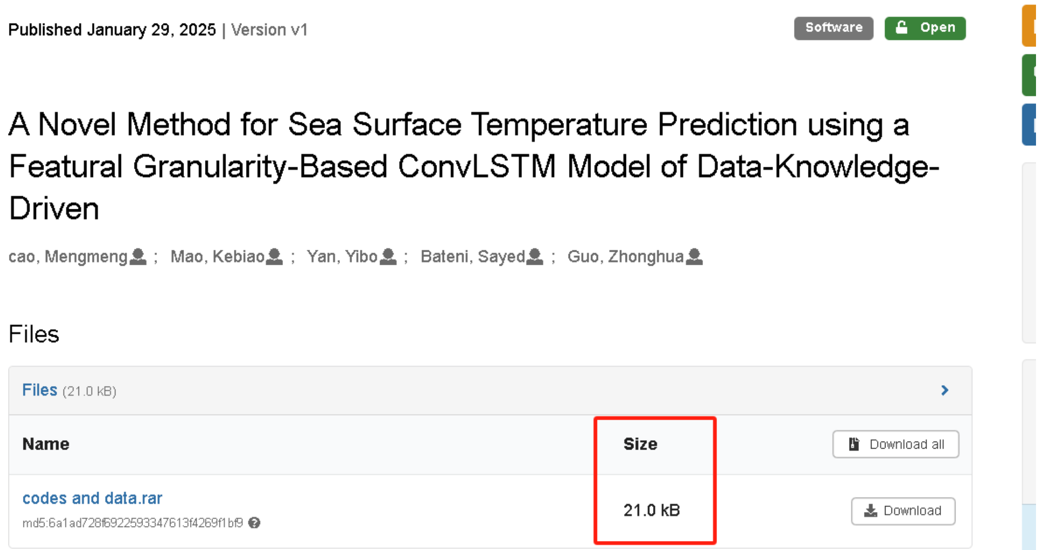


Fig.6 Original Zenodo file opened via database interface, showing active code files

Regarding data availability, we have transferred all public dataset links (originally in Section 3.1) into the “Code and Data Availability” section, adding explicit URLs for all datasets used. The revised section now states:

Code and Data availability: The codes for conducting the analyses can be downloaded from https://zenodo.org/records/15714288. The data on which this article is based are available in https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means?tab=overview (previous interface: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means) and https://cds.climate.copernicus.eu/datasets/projections-cmip5-monthly-single-levels?tab=overview (previous interface: https://cds.climate.copernicus.eu/cdsapp#!/dataset/projections-cmip5-monthly-single-levels).

(4) The analysis in Section 3 is superficial and pointless.

Reply: Thank you for your critical feedback on Section 3. We appreciate this opportunity to clarify the purpose and depth of the analyses presented, as they are foundational to validating our study’s core contributions.

The primary goal of this work is to develop a systematic, multi-module SST forecasting model—comprising a featural granularity sub-model and a data-knowledge-driven ConvLSTM sub-model—designed to capture spatiotemporal dependencies, integrate multi-variable influences, and extend prediction horizons. To rigorously demonstrate that this model achieves its intended goals, Section 3 is structured to validate each component’s necessity and the overall model’s superiority through targeted, incremental analyses. This design is deliberate: complex multi-module models require stepwise validation to isolate the impact of each innovation, ensuring conclusions about the model's effectiveness are robust.

Specifically, Section 3 includes three interconnected analyses:

Effect of study area segmentation: We designed controlled experiments to test whether partitioning the study area via pixel-wise similarity improves prediction performance. Validation confirms that regional customization—accounting for spatially heterogeneous SST drivers—enhances accuracy, justifying the inclusion of this module.

Effect of knowledge-driven parameters: We analyzed how integrating oceanographic/thermodynamic variables (e.g., currents, wind stress) into the ConvLSTM affects results. This step isolates the value of incorporating domain knowledge, distinguishing our model from data-only baselines and validating its "data-knowledge-driven" design.

Comparison with baseline models: By contrasting our model with five established machine learning methods and the SOTA GMNN model, we quantify its overall superiority. This not only demonstrates practical value but also situates our work within the broader literature, showing that our innovations (segmentation, knowledge integration) collectively drive improved performance.

These analyses are tightly linked to our study's objective: they do not merely report "accuracy" but explain why the model performs well—by validating which components contribute to its success. This stepwise validation avoids black-box conclusions, ensuring the model's design choices are supported by evidence.

We believe these analyses are neither superficial nor pointless—they are essential to demonstrating that our model's innovations are not arbitrary, but rather collectively enable its ability to achieve long-term, accurate SST forecasts. Furthermore, we have made further revisions to this section to enhance its clarity. Thank you again for pushing us to enhance their clarity and depth; we are confident the revised section better conveys their significance.

Additional remark:

<u>Lines 519-520 in the revised manuscript (Manuscript with author details):</u> "The authors are grateful to three reviewers and the editor for their constructive comments and suggestions on this paper." has been added to Acknowledgments.

Special thanks are extended to you for your valuable comments.

We have tried our best to improve the manuscript and made substantial changes to the manuscript to correct certain shortcomings.

We greatly appreciate your help and hope that the corrections will meet with approval.

Once again, we would like to extend our sincere gratitude and appreciation for the valuable comments and suggestions.