

Article

Not peer-reviewed version

A New Transformer Network for Short-Term Global Sea Surface Temperature Forecasting: Importance of Eddies

[Tao Zhang](#) , [Pengfei Lin](#) ^{*} , [Hailong Liu](#) , [Pengfei Wang](#) , Ya Wang , [Weipeng Zheng](#) , Zipeng Yu , [Jinrong Jiang](#) , [Yiwen Li](#) , [Hailun He](#)

Posted Date: 3 March 2025

doi: 10.20944/preprints202503.0067.v1

Keywords: global sea surface temperature; mesoscale eddies; deep learning; forecast



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

A New Transformer Network for Short-Term Global Sea Surface Temperature Forecasting: Importance of Eddies

Tao Zhang ^{1,5}, Pengfei Lin ^{1,5*}, Hailong Liu ², Pengfei Wang ¹, Ya Wang ¹, Weipeng Zheng ¹, Zipeng Yu ¹, Jinrong Jiang ³, Yiwen Li ⁴ and Hailun He ⁶

¹ State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

² Laoshan Laboratory, Qingdao 266000, China

³ Computer Network Information Center, Chinese Academy of Sciences, Beijing, 100190, China

⁴ School of Ocean Sciences, China University of Geosciences, Beijing 100083, China

⁵ College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

⁶ State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, 310012, Zhejiang, China

* Correspondence: linpf@mail.iap.ac.cn

Abstract: Short-term sea surface temperature (SST) forecasts are crucial for operational oceanology. This study introduced a specialized Transformer model (U-Transformer) to forecast global short-term SST variability and to compare with those from Convolutional Long Short-Term Memory (ConvLSTM) and Residual Neural Network (ResNet) models. The U-Transformer model achieved SST root mean square errors (RMSEs) of 0.2–0.54 °C for lead times of 1–10 days during 2020–2022, with anomaly correlation coefficients (ACCs) from 0.97 to 0.79. In regions characterized by active mesoscale eddies, RMSEs from the U-Transformer model exceeded the global averages by at least 40%, with increases exceeding 100% for the Gulf Stream region. Additionally, ACC values in active mesoscale eddy regions declined more sharply with forecast lead time compared to the global averages, decreasing from approximately 0.96 to 0.73. Specifically, in the Gulf Stream region, the ACC value dropped to 0.89 at a 3-day lead time, while the value can maintain 0.92 globally. Compared with the ConvLSTM and ResNet models, the U-Transformer model consistently delivered smaller RMSEs and larger ACCs, especially in regions with active mesoscale eddies. These findings imply the importance of advanced approaches to enhance SST forecast accuracy in regions with active mesoscale eddies.

Keywords: global sea surface temperature; mesoscale eddies; deep learning; forecast

1. Introduction

Sea surface temperature (SST) plays a crucial role in air–sea interactions and is a key climate factor of global change. Variation in SST has substantial impact on regional climate variability, influencing global precipitation patterns and potentially leading to extreme events such as droughts and floods [1–4]. SST short-term fluctuations can be indicative of marine heatwaves, which can have severe impact on marine ecosystems globally [5–9]. Short-term SST forecasting is influenced by numerous factors, among which oceanic mesoscale eddies are particularly important. Due to their strong dynamical effects, these eddies can induce extreme short-term SST anomalies. Therefore, understanding and addressing the influence of mesoscale eddies is essential for improving short-term SST forecasting.

In recent years, data-driven deep learning (DL) methods have gained widespread application in ocean and atmospheric sciences due to their ability to learn complex nonlinear relationships, such as

eddy identification, downscaling, SST reconstruction, and parameterization of physical processes [10–15]. Various types of DL models have been widely explored in SST forecasting. Recurrent neural networks (RNNs), like long short-term memory networks (LSTMs) and gated recurrent units (GRUs), primarily focus on the temporal evolution of SST at individual locations and have been applied to SST forecasting in specific regions [16–19]. However, these models struggle to capture complex spatial correlations across areas. Researchers have adopted convolutional neural networks (CNNs) to address this limitation, which leverages their strengths in spatial feature extraction to significantly enhance SST prediction performance [20]. Residual neural networks (ResNets) further improve model depth by introducing a residual learning framework, effectively mitigating the gradient vanishing problem in deep networks [21]. Therefore, different types of CNN networks and their variants have also been applied to short-term SST forecasting [22–26]. A typical network is Convolutional Long Short-Term Memory (ConvLSTM), which combines the advantages of CNN and RNN, can effectively extract temporal and spatial information, and improve the model's forecast accuracy [27,28]. These models have shown significant progress in improving prediction accuracy and reliability.

With the rapid advancement of DL, the Transformer architecture [29,30] (see Appendix A for details) has emerged as a powerful tool across various fields due to its capability to capture long-range dependencies and model complex spatiotemporal relationships. While it has been successfully applied to SST super-resolution tasks [31], its potential for short-term SST forecasting remains relatively unexplored.

Mesoscale eddies are widely distributed in the ocean and they are the primary drivers of mesoscale SST variability [32]. These eddies significantly impact SST forecasts, with dynamical forecast models often exhibiting notable errors in eddy-active regions due to their complex dynamics and temperature structures [33,34]. These errors are also particularly pronounced over eddy-active areas [35–37] in the simulations of the High-Resolution Ocean Model Intercomparison Project. Studies on quantifying short-term SST forecast errors and their spatial distribution in eddy-active regions remain limited. Furthermore, forecasting short-term SST anomalies caused by eddies has received little attention as previous DL researches have concentrated on forecasting sea surface height associated with eddies [38–40].

This study introduces an innovative Transformer-based variant, the U-Transformer model, to improve short-term global SST forecasting. The U-Transformer model is designed to capture spatial and temporal features simultaneously, enabling more accurate multi-step forecasts for the coming days. In this study, we compare the performance of the U-Transformer model with two classic CNN-based model types—ConvLSTM and ResNet—across global areas and regions with active mesoscale eddies. The paper is organized in the following. Section 1 provides an introduction. Section 2 describes the data and methods. Section 3 presents the results. Section 4 contains discussion and conclusions.

2. Data and Methods

2.1. Data

The data used in this study is derived from the NOAA/NESDIS/NCEI Daily Optimum Interpolation Sea Surface Temperature (OISST), version 2.1, dataset, as detailed in [41,42]. This dataset provides global SST observations with a spatial resolution of 0.25° and a daily temporal resolution, covering the period from January 1, 1982, to December 31, 2022. The global coverage spans from 89.975°S to 89.875°N and from 0.125°E to 359.875°E .

OISST v2.1 integrates SST measurements from satellite observations (e.g., AVHRR), in-situ measurements from ships, drifting buoys, and Argo floats. These data sources are blended using an optimum interpolation algorithm, which ensures consistency and accuracy through bias adjustments based on in-situ observations. The interpolation leverages spatial autocorrelation and temporal consistency to generate a high-resolution, gridded SST product.

While OISST v2.1 provides high global accuracy, regions with sparse in-situ observations—such as the Indian, South Pacific, and South Atlantic Oceans—rely more heavily on satellite data and interpolation, which may lead to higher uncertainties. However, v2.1 improvements, including bias corrections, incorporation of Argo data above 5m depth, and enhanced Arctic SST estimates, have significantly reduced these biases and improved overall accuracy.

To evaluate the SST forecasts, we employed data from three prominent oceanographic research programs: the Tropical Atmosphere Ocean (TAO) project, the Research Moored Array for African–Asian–Australian Monsoon Analysis and Prediction (RAMA) project, and the Prediction and Research Moored Array in the Tropical Atlantic (PIRATA) project. These programs utilize moored buoys that provide essential real-time data on oceanic and atmospheric conditions.

All three arrays provide data daily, measuring key parameters such as SST, air temperature, wind stress, 10m wind speed, and longwave radiation. The spatial resolution of these arrays is approximately 2° latitude by 10° longitude, with the TAO array covering the equatorial Pacific, RAMA covering the tropical Indian Ocean, and PIRATA covering the tropical Atlantic Ocean.

This study employed the spatial filtering methods of [43] and [44] to extract the ocean mesoscale signal. A filter box with dimensions of 3° in both longitude and latitude was used to calculate the mean value within the box, which comprised the low-pass filtered value representing the large-scale signal. The difference between the original SST and the low-pass filtered value was then utilized to isolate and reflect the role of the mesoscale signal.

2.2. Model

The proposed U-Transformer architecture eliminates convolutional and recursive operations, replacing them with a self-attention mechanism to extract multivariate relationships in parallel, regardless of spatial and temporal distance. The U-Transformer, as shown in Figure 1a, comprises an encoder, decoder, and skip connections [45], and was built on the Swin Transformer module [46]. The Swin Transformer module employs self-attention within nonoverlapping local windows to reduce network complexity and build hierarchies for multiscale feature extraction.

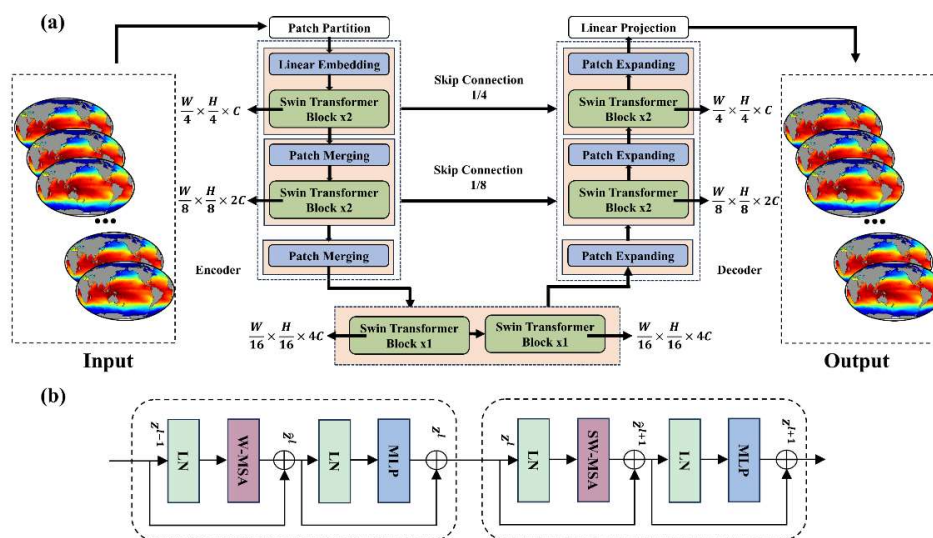


Figure 1. (a) Architecture of the U-Transformer model and (b) two successive Swin Transformer Blocks.

The encoder starts by dividing the input SST field into 4×4 non-overlapping patches, each with a feature dimension $T(10)$. These patches are then projected to an arbitrary dimension C (e.g., $C=96$) through a linear embedding layer, reducing both the spatiotemporal dimensions and memory usage. The resulting matrix has dimensions $(\frac{H}{4} \times \frac{W}{4}) \times C$, where H and W represent the height, and width dimensions of the input.

These encoded patches pass through a series of Swin Transformer Blocks and a patch merge layer. The patch merge layer reduces the spatial dimensions by half while doubling the feature dimension, enabling hierarchical feature representation. For instance, after the first patch merge layer, the matrix dimensions become $(\frac{H}{8} \times \frac{W}{8}) \times 2C$.

Similarly, the decoder employs Swin Transformer Blocks and a patch expand layer. The patch expand layer upsamples the feature mappings to restore the spatial dimensions progressively. For example, the dimensions change from $(\frac{H}{8} \times \frac{W}{8}) \times 2C$ back to $(\frac{H}{4} \times \frac{W}{4}) \times C$. Skip connections from the encoder provide contextual features to the decoder, mitigating spatial information loss. Finally, a linear projection layer converts the output of the decoder into the desired shape, $(T \times H \times W)$, to generate the future SST field forecast.

A key innovation is the shifted window-based multi-head self-attention (SW-MSA) module in the Swin Transformer, which addresses the lack of cross-window connectivity in a standard window-based MSA. The SW-MSA alternates between two partitioning configurations, with each Swin Transformer Block comprising an SW-MSA, a 2-layer multilayer perceptron (MLP) with Gaussian Error Linear Unit activation, Layer Normalization (LN), and residual connections (Figure 1b). This process can be formulated as follows:

$$\hat{z}^l = W\text{-MSA}(LN(z^{l-1})) + z^{l-1} \quad (1)$$

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l \quad (2)$$

$$\hat{z}^{l+1} = SW\text{-MSA}(LN(z^l)) + z^l \quad (3)$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (4)$$

where \hat{z}^l and z^l denote the output features of the (S)W-MSA module and the MLP module for block l , respectively. The self-attention is computed as follows:

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (5)$$

where $Q, K, V \in R^{M^2 \times d}$ are the query, key, and value matrices, respectively, d is the query/key dimension, and M^2 is the number of patches in a window.

2.3. Implementation Details

After the data is preprocessed through normalization, land locations are assigned a value of 0 to exclude them from the model's predictions. Subsequently, the generated input-output pairs are segmented along the time axis, with each input sample comprising the SST data from the previous ten days and the corresponding output representing the SST data for the next ten days. This allows the model to capture the temporal dependencies within SST patterns and predict future values. The dataset is then split into training, validation, and test sets. Data from 1982 to 2019 are used for training and validation, with 90% allocated to training and 10% reserved for validation. The model is trained using this split, with the validation set utilized for hyperparameter tuning to prevent overfitting. Finally, data from 2020 to 2022 are retained as an independent test set, enabling a robust evaluation of the model's performance on unseen data. During model evaluation, a spatial filtering method is applied to isolate mesoscale and large-scale signals, ensuring a comprehensive assessment of the model's forecasting ability.

The model was trained using a latitude-weighted L2 loss function, which accounts for the area of grid points across different latitudes. To improve training efficiency and model stability, inputs and outputs were normalized using zero-mean normalization.

Our weighted L2 loss function can be expressed as:

$$L2 = \frac{1}{N} \sum_{i=1}^N \cos(\theta_i) (y_i - \hat{y}_i)^2 \quad (6)$$

where y_i is the true value, \hat{y}_i is the predicted value, θ_i is the latitude of the corresponding point, and $\cos(\theta_i)$ is the latitude-based weight.

Our zero-mean normalization function can be expressed as:

$$X' = \frac{X - \mu}{\sigma} \quad (7)$$

where X is the original data, μ is the mean of the feature, σ is the standard deviation of the feature, X' is the normalized data. The mean and standard deviation values are computed based on the historical dataset spanning from 1982 to 2019.

All models were implemented using the PyTorch framework and trained on a cluster of 16 nodes, each with two accelerator cards (16 GB memory). The training process ran for 100 epochs, with a batch size 2 per card. We used the AdamW [47,48] optimizer, as it is known for its effectiveness in stabilizing training in deep learning models. The optimizer parameters, $\beta_1 = 0.9$ and $\beta_2 = 0.95$ were chosen based on their widespread use in similar tasks, which balances gradient momentum and stability. An initial learning rate of 10^{-3} was applied, as it provided a good balance between convergence speed and performance during preliminary tests. Additionally, we used a weight decay of 0.1 to mitigate overfitting, ensuring generalization. All training parameters were kept consistent across models to maintain fairness and comparability.

We compared the proposed U-Transformer model with ConvLSTM and ResNet models, using the same input-output structures and preprocessing across all models. The ConvLSTM architecture was adapted from [49], while the ResNet model was based on ResNet-18 [21]. The parameter details of ConvLSTM and ResNet can be seen in Appendix B. All forecasts were derived from the test set, ensuring consistent methodology across models.

2.4. Evaluation Methods

We evaluated the forecast performance using the area-weighted RMSE, Bias, and anomaly correlation coefficient (ACC), which were calculated as follows:

$$RMSE_t = \frac{1}{|D|} \sum_{t_0 \in D} \sqrt{\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \alpha_i (\hat{Y}_{i,j} - Y_{i,j})^2} \quad (8)$$

$$Bias_t = \frac{1}{|D|} \sum_{t_0 \in D} \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \alpha_i (\hat{Y}_{i,j} - Y_{i,j}) \quad (9)$$

$$ACC_t = \frac{1}{|D|} \sum_{t_0 \in D} \frac{\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \alpha_i (\hat{Y}_{i,j} - C_{i,j})(Y_{i,j} - C_{i,j})}{\sqrt{\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \alpha_i (\hat{Y}_{i,j} - C_{i,j})^2} \sqrt{\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \alpha_i (Y_{i,j} - C_{i,j})^2}} \quad (10)$$

where t_0 is the forecast initialization time in testing set D and t is the forecast lead time step added to t_0 ; H , and W are the number of time steps and the grid points in the latitude and longitude directions, respectively; α_i represents the weights of different latitudes; $\hat{Y}_{i,j}$ and $Y_{i,j}$ are the forecast field and the true field at time t , respectively; and C represents the climatological mean calculated using data from 2000–2010.

3. Results

The U-Transformer model has good ability to forecast short-term global SST. The global average RMSEs produced by the U-Transformer model are in the range 0.2–0.54 °C from 1- to 10-day lead times during 2020–2022 (Figure 2a), demonstrating its consistent accuracy. At the 5-day lead time, the U-Transformer model achieves a global RMSE of 0.42 °C, slightly outperforming the ConvLSTM and ResNet models that have RMSEs of 0.43 and 0.44 °C, respectively. Although the RMSEs increase with increasing lead time for all three models, the U-Transformer model maintains the lowest RMSE values consistently for all (from 1- to 10-day) lead times. Meanwhile, the smaller RMSEs for all grid points (spread of RMSEs) are achieved in the ConvLSTM and U-Transformer models than those in

the ResNet model from 4- to 10-day lead times. According to previous studies, dynamical forecast models have RMSEs of forecasted SSTs of approximately 0.35–1.1 °C at the 1-day lead time [50] with the smallest forecast SST from Met Office Forecast Ocean Assimilation Model (FOAM) dynamical forecast system [51,52]. For example, in the LICOM Forecast System v1.0 [53], the global averaged RMSE of forecasted SST is 0.45–0.55 °C at the 1-day lead time. Compared with the RMSEs of SST forecasted by dynamical models, the U-Transformer model can substantially reduce the global SST forecast errors. The RMSE of forecasted SST produced by the model presented by [27] at the 1-day lead time is 0.27 °C, which is larger than that of the three models built in this study, particularly the U-Transformer model. The above analysis indicates that the U-Transformer model is a good model for forecasting short-term SST globally because it produces the smallest global averaged RMSEs.

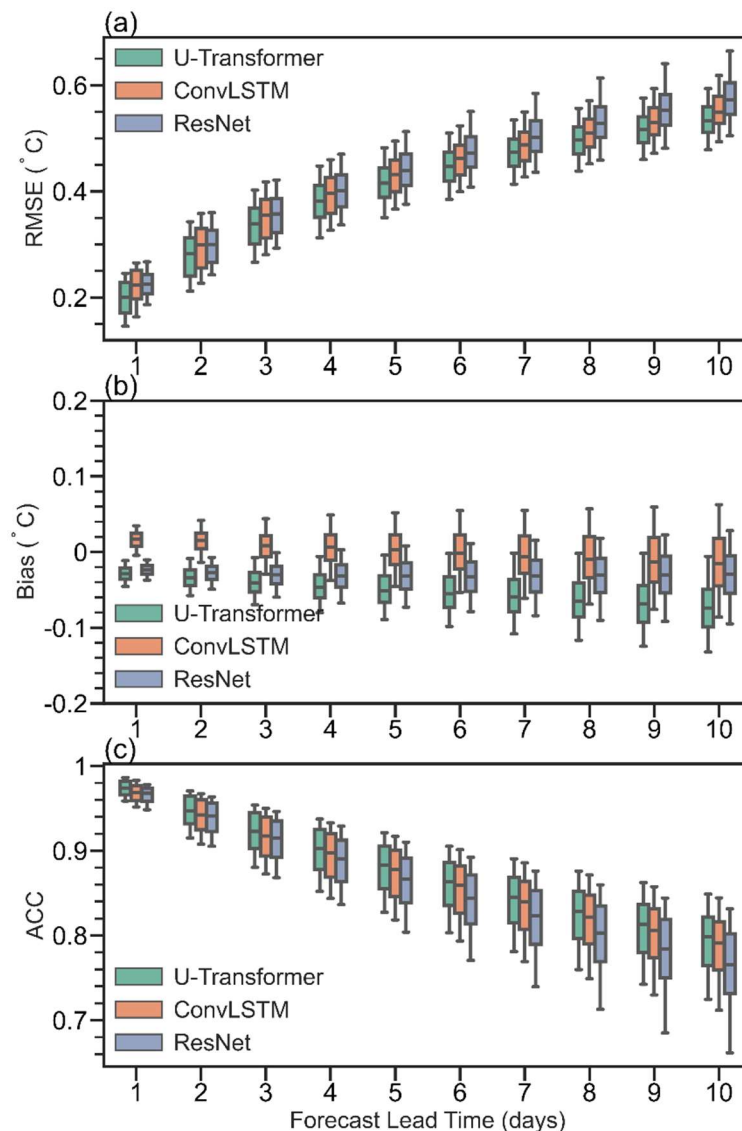


Figure 2. Variation of different evaluation metrics with lead time (1- to 10-day) for the three DL models (i.e., the U-Transformer, ConvLSTM, and ResNet models): (a) RMSEs, (b) Bias, (c) ACC during 2020–2022. Metrics were obtained through global weighted averaging.

The good ability of the U-Transformer model to forecast short-term global SST is also represented in the small biases and large ACC (Figure 2b and 2c). Generally, all three models exhibit small biases (<0.1 °C) in global averaged SST, although the biases among the three models differ slightly. The U-Transformer model exhibits a cold bias of 0.03–0.06 °C at 1- to 10-day lead times. The ResNet model shows a relatively stable cold bias of approximately 0.02 °C, whereas the ConvLSTM model transitions from a slight warm bias to a cold bias over the same period but with a larger spread

of values. The consistency in the sign of the global averaged SST bias for all lead times implies that systematic bias exists for the U-Transformer and ResNet models, which requires further study. The U-Transformer model consistently achieves the largest ACC values, starting at approximately 0.97 at a 1-day lead time and gradually decreasing to 0.79 at the 10-day lead time. Meanwhile, the spreads are smaller for the U-Transformer and ConvLSTM models, implying consistently large ACCs and thus, high forecast skill.

The above analysis highlights the reasonable ability of the U-Transformer model to forecast short-term SST globally, outperforming both the ConvLSTM and the ResNet models in terms of forecast error and forecast skill from the perspective of global statistics and the 1- to 10-day lead times.

Statistical analysis was also performed for the tropics and the regions with active mesoscale eddies. In the tropical and subtropical oceans, the RMSEs are relatively low and generally do not exceed 0.2 °C, except in the eastern equatorial Pacific (Figure 3a), which is a region characterized by the Tropical Instability Wave (TIW). In this area, RMSEs exceed 0.3 °C, consistent with the findings of previous studies [23,54], and highlighting the challenges in forecasting daily SST changes in the TIW region. In the Atlantic and Indian oceans, higher RMSEs (calculated using observed data from moored stations) are observed in the northeast, while smaller RMSEs (<0.2 °C) are found in the western Atlantic (Figure 3c). The Indian Ocean exhibits a relatively uniform distribution of RMSEs, with values generally exceeding 0.25 °C (Figure 3e). In the Pacific Ocean, RMSEs are distributed unevenly, with higher values in the eastern equatorial region and lower values in the western parts (Figure 3a). Overall, the anomaly correlation coefficient (ACC) decreases as the forecast lead time increases, with the U-Transformer model demonstrating the largest correlation forecast in the Pacific and Atlantic regions (Figure 3b, d). On the first forecast day, the ACC is approximately 0.9, with the Atlantic region showing the best correlation performance. Interestingly, the ACC does not exhibit a strictly monotonic decrease as the forecast lead time extends, which may be attributed to the limited number of observed samples available for evaluation. It is evident from Figures 2 and 3 that the U-Transformer model can substantially reduce forecast errors relative to those of the ConvLSTM and ResNet models, particularly in the TIW region, which is a region with complex oceanic activities and notable regional interactions. The ConvLSTM and ResNet models both struggle to capture these varied regional dynamics because of the use of convolutional networks, whereas the self-attention mechanism in the U-Transformer model might effectively capture the remote dependencies and intricate spatial patterns, making it particularly suited for SST forecasts in such complex regions.

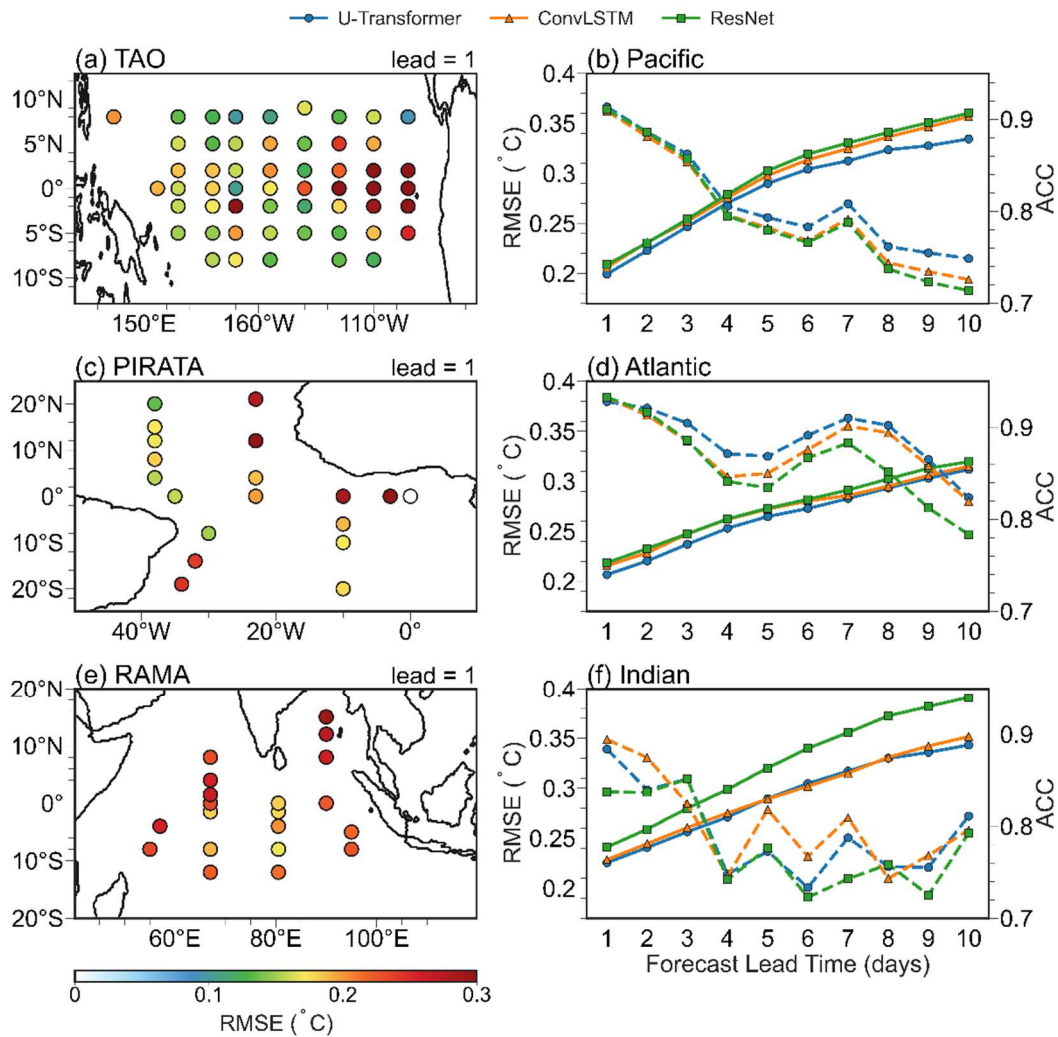


Figure 3. Comparison of different model forecasts with observations from TAO/PIRATA/RAMA buoys in different oceans. (a, b) Pacific; (c, d) Atlantic; (e, f) Indian. The a, c, and e represent the distribution and RMSE of spatial points across various oceans, as forecasted 1-day lead time by the U-Transformer. Panels (b), (d), and (f) display the RMSE and ACC of different models at varying lead times, where the solid line represents the RMSE, and the dashed line represents the ACC.

Figure 4 illustrates the spatial RMSE distribution for forecast SST at different lead times (days) in the test set. At the 1-day lead time, the U-Transformer model performs exceptionally well with a global area average RMSE of 0.2 °C (Figure 4a). In the U-Transformer model results, small RMSEs are found mainly in tropical or subtropical ocean areas, whereas large RMSEs are found in regions with active mesoscale eddies. The ConvLSTM and ResNet models can also reproduce the observed SST distribution in the 1-day lead time forecast with RMSE values (0.22–0.23 °C) slightly larger than those of the U-Transformer model (Figure 4d, g). As the forecast lead time increases, the forecast error also grows. When forecasting 10 days, the U-Transformer achieves a global average RMSE of 0.54°C, compared to 0.55°C for ConvLSTM and 0.58°C for ResNet (Figure 4c, f, i). Notably, forecast errors are more pronounced in regions with active mesoscale eddies than in other areas. The observed large-scale features of the SST distribution can also be well reproduced on January 1, 2022 at 1- or 5-day lead times by the three DL models (Figure 5a, d, g, j), such as warm SST in the tropics, cold SST at high latitudes (Arctic Ocean and Southern Ocean), the Indo-Pacific warm pool, and the cold tongue in the equatorial eastern Pacific (Figure 5b, e, h, k). It is evident that while different models successfully capture the general characteristics and spatial morphology of mesoscale signals, their errors remain significant. The forecast error for mesoscale processes accounts for approximately 70% of the total error, highlighting the complexity of mesoscale activities as a primary contributor to

inaccuracies in SST forecasts. Among the evaluated models, the RMSE is consistently around 0.25; however, the U-Transformer model demonstrates superior correlation performance, achieving the largest value of 0.92 (Figure 5c, f, i, l).

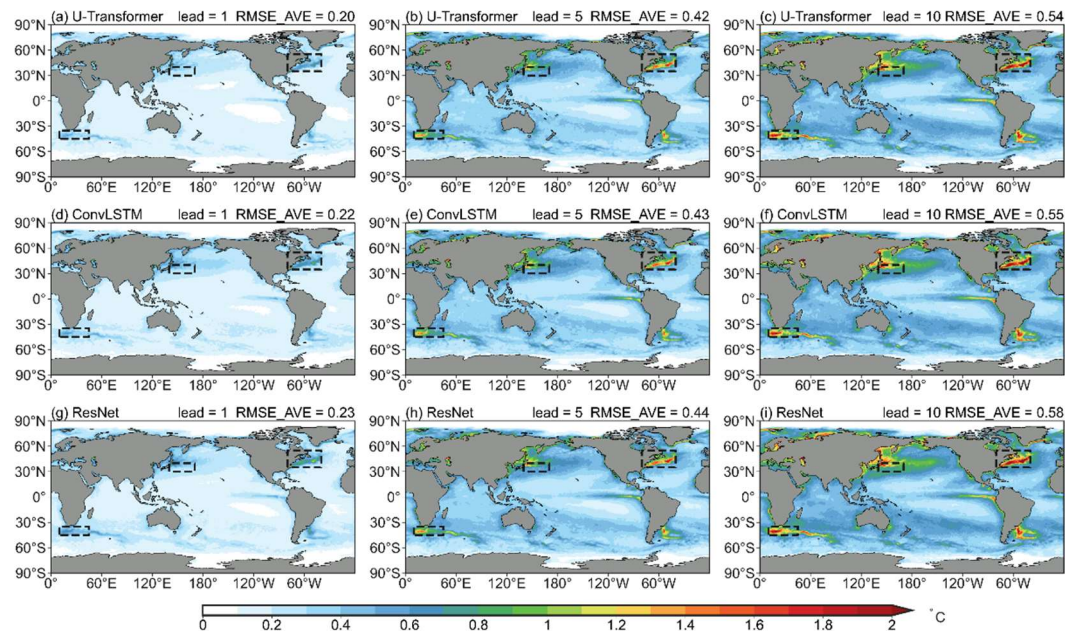


Figure 4. Spatial distribution of RMSEs for 1-day (a, d, g), 5-day (b, e, h), and 10-day (c, f, i) lead times by the U-Transformer model during 2020–2022 (a–c), by the ConvLSTM model (d–f), and by the ResNet model (g–i). Global average RMSE values are displayed in the upper-right corner of each panel. Dashed boxes indicate the locations of selected regions with active mesoscale eddies: the Kuroshio Extension (30°–40°N, 140°–170°E), Gulf Stream (35°–55°N, 40°–80°W), and the oceans around Southern Africa (35°–45°S, 10°–45°E).

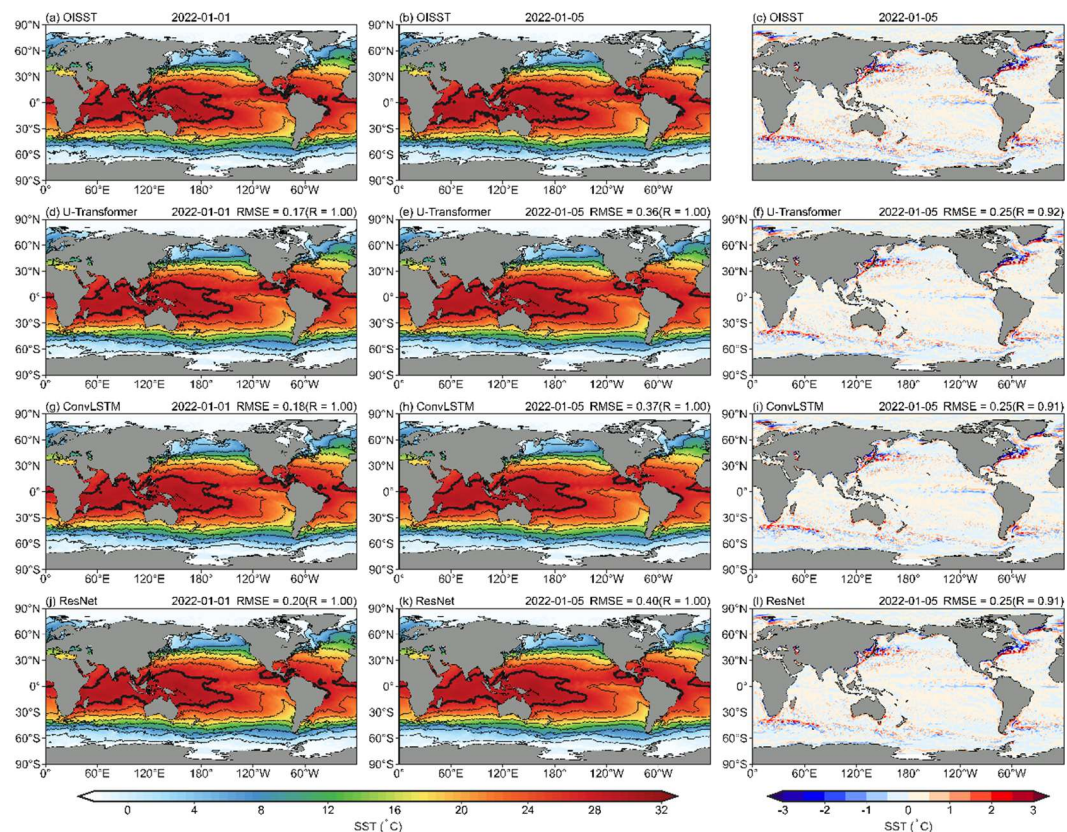


Figure 5. Global SST from observation and forecasts from three DL models, at 1-day leading (January 1, 2022), 5-days leading (January 5, 2022) starting from January 1, 2022. (a–c) OISST; (d–f) U-Transformer; (g–i) ConvLSTM; (j–l) ResNet.

(j-l) ResNet. RMSEs and pattern correlation coefficients (R) of forecast SST from models and observations in the upper right corner. The first and second columns display the raw SST values, with thin black contour intervals representing 4°C isotherms and thick black lines denoting 28°C isotherms. The third column shows the filtered mesoscale signal obtained by subtracting the low-pass filtered SST ($3^{\circ}\times 3^{\circ}$) from the raw SST values.

In regions with active mesoscale eddies, the RMSEs are large and the behavior of the forecasted local SSTs requires investigation. The Kuroshio Extension (KE), Gulf Stream (GS), and the oceans around Southern Africa (OSA) are regions chosen to characterize these areas with active mesoscale eddies. In these regions, all three DL models exhibit large RMSEs compared with the global average, particularly within the black dashed boxes shown in Figure 4 (details as Figure A1), with errors exceeding 0.6°C for 1-day lead time forecasts. Additionally, the mesoscale pattern correlation coefficients of the forecasted SSTs are notably lower than those for large-scale patterns, highlighting the challenges in forecasting SST changes associated with mesoscale eddies (Figure 5).

The forecast of a specific day further reflects the ability of SST to forecast in complex ocean regions. The KE was chosen to display the forecast SST evolution in the eddy-active areas. Figure 6 presents the observed OISST SST and daily SST forecast biases in this region from July 14 to July 20, 2022, based on initial conditions from OISST on July 14, 2022. The results demonstrate that DL models effectively capture the overall spatial distribution of SST. The U-Transformer model exhibits smaller biases at the 1-day lead time (July 14, 2022) than the ConvLSTM and ResNet models. South of 36°N , the absolute SST biases are less than 0.3°C , while between 36°N and 39°N , biases exceed 0.5°C in all models at the 1-day lead time. The U-Transformer achieves an RMSE of 0.3°C , outperforming the ConvLSTM (0.4°C) and ResNet (0.5°C) models. The forecast biases become large as the lead times increase. Since the lead times of 4 days, biases south of 36°N become comparable to those north of 36°N in the U-Transformer model. The RMSEs of forecast SST at all lead times are smaller using the U-Transformer model than those using the ConvLSTM and ResNet models, with a slight difference compared to the ConvLSTM model and a larger difference compared to the ResNet model. All models exhibit common biases around finer-scale SST features, particularly near extreme local high or low SST values linked to mesoscale and submesoscale eddies. From 1-day to 3-day lead times, the RMSEs increase significantly, from 0.3 to 0.6°C using the U-Transformer model. Similar evolution behaviors are found in the other two models. This may be caused by the eddy movement and their nonlinear behaviors. Similar patterns of large forecast SST biases and their evolution are evident in the GS and the OSA, as shown in Figures A2 and A3. Therefore, the advanced DL models exhibit better capabilities for forecasting SST in eddy-active regions. Further optimization is required to enhance their accuracy and reliability when addressing complex ocean processes, like eddy-rich regions.

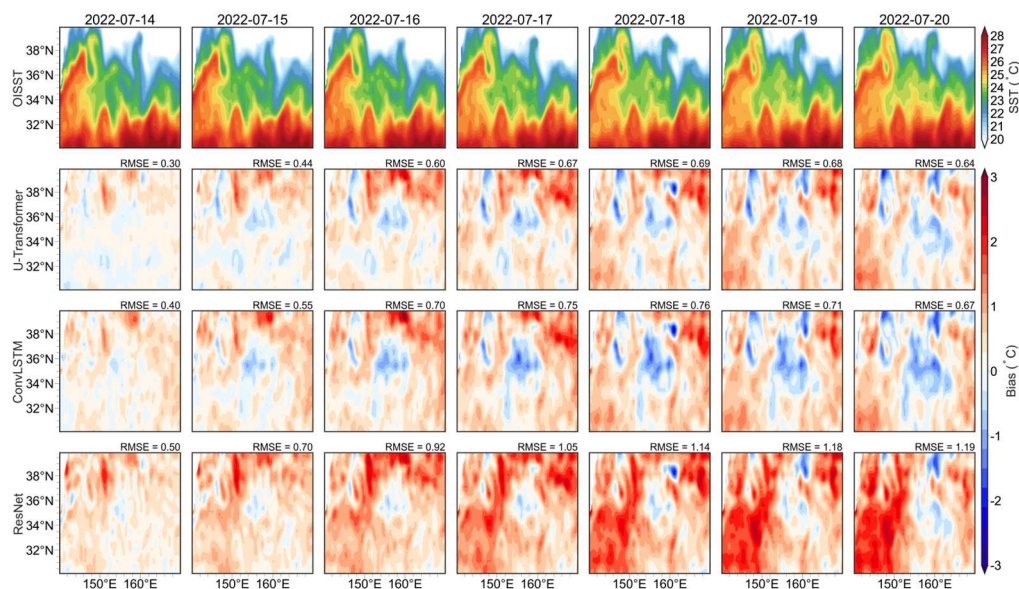


Figure 6. Comparison of OISST and SST forecasts by three deep learning models in the Kuroshio Extension region from July 14, 2022, to July 20, 2022. The first row represents OISST, while the second, third, and fourth rows show forecasts biases from the U-Transformer, ConvLSTM, and ResNet models.

Figures 7 and 8 illustrate the spatial distributions of forecast SST biases for the U-Transformer model, selected based on the 10th percentile (lower RMSE) and 90th percentile (higher RMSE) of sorted RMSE values in ascending order. The error distributions across the three DL models are generally consistent under various forecast initial conditions, reflecting the inherent physical characteristics of SST variations in eddy-active regions. However, notable differences in bias magnitudes are observed among the models. In the KE region, the SST bias from the U-Transformer model is predominantly below 0.2°C (Figure 7a), which is 10-30% smaller than those produced by the ConvLSTM and ResNet models. In contrast, the GS region exhibits significantly more irregular SST bias structures (Figure 7d-f), resembling features associated with eddies. In this region, the RMSEs are 40-50% larger than those in the KE region for the same model. Similarly, eddy-related bias patterns are evident in the OSA (Figure 7g-i) but much more obvious in the ConvLSTM and ResNet models. In this area, the RMSEs of the ConvLSTM and ResNet models are 22-61% larger than those of the U-Transformer. While biases in this region are larger than those in the KE, they remain smaller than those in the GS for the same model. These comparisons across different areas indicate the challenges posed by active eddies, which can induce significant forecast SST biases due to their complex capture dynamics.

In the cases of larger RMSEs (Figure 8), larger biases exist for the same region and the same DL model compared with those in the case of smaller RMSEs (Figure 7). Even under these cases, the U-Transformer model demonstrates smaller SST biases than the ConvLSTM and ResNet models, but the differences vary by region and model. In the KE region, the U-Transformer achieves RMSE reductions of 17% and 3% compared to ConvLSTM and ResNet, respectively. In the GS region, the RMSEs for the U-Transformer are 11-14% smaller than those of the other two models. In the OSA, the U-Transformer reduces RMSEs by 7% compared to ConvLSTM and by 20% compared to ResNet. While the U-Transformer consistently outperforms the other models, the relative improvements are smaller in the larger RMSE case. The above results may be related to much more apparent eddy structures in Figure 8 than in Figure 7. These intensified mesoscale eddies significantly influence SST forecasts, highlighting the challenges of accurately capturing such complex dynamics.

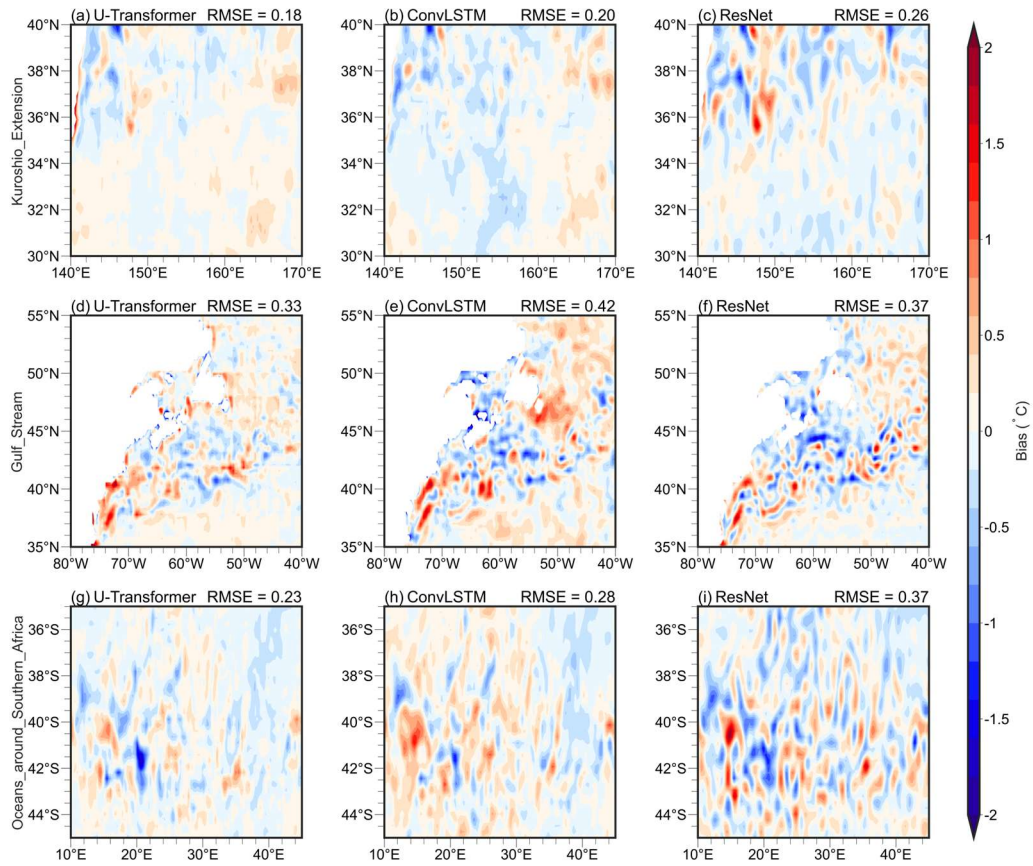


Figure 7. Forecast SST from forecast cases from the U-Transformer, ConvLSTM, and ResNet at the 1-day lead time using the same forecast initial value. These cases are selected according to the 10th percentile (smaller RMSE) of the sorted RMSE values by ascending order for the U-Transformer in three eddy-active regions (Kuroshio Extension, Gulf Stream, and the oceans around Southern Africa). The average RMSE values for each area are displayed in the upper-right corner of each panel. Panels (a-c) correspond to forecasts initialized on December 19, 2021; panels (d-f) forecasts initialized on October 6, 2022; and panels (g-i) forecasts initialized on January 1, 2022.

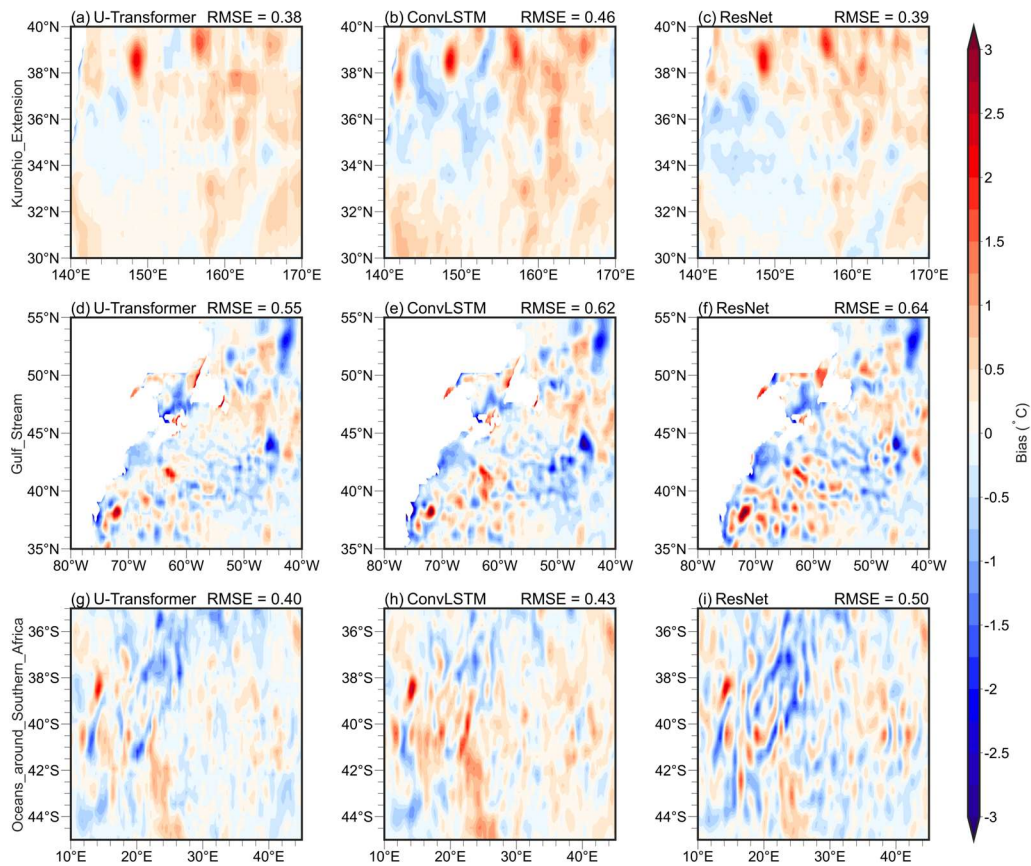


Figure 8. Forecast SST from forecast cases from the U-Transformer, ConvLSTM, and ResNet at the 1-day lead time using the same forecast initial value. These cases are selected according to the 90th percentile (larger RMSE) of the sorted RMSE values by ascending order for the U-Transformer in three eddy-active regions (Kuroshio Extension, Gulf Stream, and the oceans around Southern Africa). The average RMSE values for each area are displayed in the upper-right corner of each panel. Panels (a-c) correspond to forecasts on August 30, 2020; panels (d-f) forecasts on June 3, 2020; and panels (g-i) forecasts on February 13, 2021.

Statistical analysis of these ACC and RMSEs in the three regions further underscores the existence of large SST error associated active mesoscale eddies. In the selected active eddy regions, local forecast SST errors are notably larger than those of the global SST forecasts, with RMSEs rising from 0.2–0.6 °C globally to 0.28–1.2 °C in regions with active mesoscale eddies. The RMSEs of forecasted SSTs in regions with active mesoscale eddies are 40%–130% greater than the global average (Figure 9b, 9d, and 9f). Among the selected regions, relative to the global average RMSEs, the RMSEs in the KE (GS) region increase by 42%–60% (>100%), suggesting that the errors are relatively small in the KE region but large in the GS region. This further indicates the obvious difficulty in forecasting short-term SST over different regions with active mesoscale eddies.

Similar to the statistics of forecasted short-term SST in the global domain, as the forecast lead time extends, the RMSEs increase and the ACC values decline markedly from approximately 0.96 to 0.73 in regions with active mesoscale eddies (Figure 9a, 9c, and 9e). The ACC values in the three regions with active mesoscale eddies are consistently lower than the global averages analyzed at the same lead time. Notably, in the GS region, ACC values for all models drop below 0.9 at the 3-day lead time; however, on the global scale, the U-Transformer model maintains an ACC value of >0.9 at the 4-day lead time. This suggests that forecasting short-term SST has lower skill in the regions with active mesoscale eddies. From 1- to 3-day lead times, the ACC values in the KE and OSA regions are slightly higher than those in the GS region. The U-Transformer model has superior forecasting skill among all three models across all areas. However, the decline in the ACC values is sharper in the regions with active mesoscale eddies than that observed in relation to the global forecast, with the ACC value decreasing by approximately 0.18 from the 1- to 10-day lead time forecast (Figure 2c), and

by approximately 0.24 in the regions with active mesoscale eddies. This indicates that forecast skill is lower in the regions with active mesoscale eddies, and that it is more difficult to forecast SSTs in these regions than to forecast SST globally. The presence of mesoscale eddies causes these regions to be dynamically complex. Meanwhile, the forecast skill declines more sharply as the forecast lead time increases, which implies that it is more difficult to forecast SSTs for the regions with active mesoscale eddies as the lead time extends.

Regions such as the KE, GS, and OSA exhibit larger SST forecast errors, primarily due to their distinct dynamic characteristics. These factors make forecasting SST in these regions more challenging than in other oceanic areas. Active mesoscale eddies in these regions play a significant role in SST variability through their movements and nonlinear behaviors. The frequent formation and dissipation of these eddies introduce additional uncertainties, as their small spatial scales (ranging from tens to hundreds of kilometers) often approach the resolution limits of the models [55]. These regions also experience intense air-sea interactions, which are not adequately accounted for by the DL models, leading to substantial forecast discrepancies [56,57]. The GS is a high-speed western boundary current, posing unique challenges. Compared to the KE, the GS exhibits stronger mass transport, heat, and salt transport, which can easily lead to flow instabilities [58–60]. Moreover, the GS's stronger current is confined within the narrower Atlantic Ocean Basin than the Pacific Ocean Basin. These combined factors make SST forecasting in the GS region even more difficult than in other eddy-active regions.

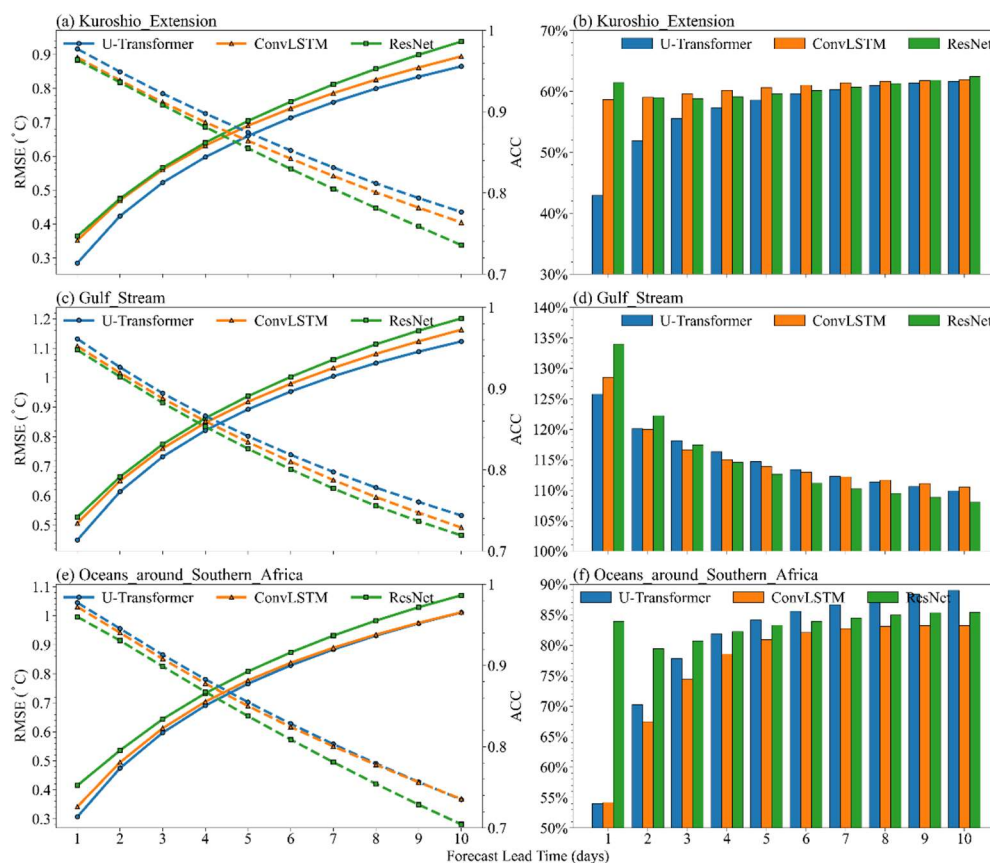


Figure 9. Comparison of RMSEs and ACC values across three regions with active mesoscale eddies for different models at various lead times (a, c, e). Solid lines represent RMSEs and dashed lines represent ACC values. Panels (b, d, f) show the percentage increase in RMSEs within the selected regions (denoted RMSEc) compared with the global average RMSEs (denoted RMSEd), calculated as $((\text{RMSEc} - \text{RMSEd})/\text{RMSEd}) \times 100\%$.

The following quantifies the RMSE difference (denoted as the RMSE reduction percentage) of the forecasted SSTs when using the U-Transformer model compared with the SSTs forecasted using the ConvLSTM and ResNet models (Table 1), particularly in the regions with active mesoscale eddies.

At the 1-day lead time, the U-Transformer model achieves RMSE reductions of 20.42% and 21.32% in the KE region, 12.19% and 16.82% in the GS region, and 10.41% and 25.79% in the OSA region relative to the ConvLSTM and ResNet models, respectively. Globally, the U-Transformer model reduces the RMSEs by 10.13% compared with the ConvLSTM model and by 11.66% compared with the ResNet model. As the lead time increases, the RMSE reduction percentages also decrease. For example, at the 10-day lead time, compared with the ConvLSTM and ResNet models, the RMSE reductions for the U-Transformer model decrease to 3.73% and 9.08% in the KE region, 4.09% and 8.05% in the GS region, and 0.08% and 6.36% in the OSA region, respectively. This comparison demonstrates that the U-Transformer model consistently outperforms the other two models, not only in terms of forecast SSTs globally but also for forecast SSTs in the regions with active mesoscale eddies. This is important when forecasting SSTs induced by mesoscale eddies.

Table 1. Percentage reduction in RMSE ($\frac{|RMSE_a - RMSE_b|}{RMSE_b} \times 100\%$) of the U-Transformer model (denoted $RMSE_a$) compared with the different models (denoted $RMSE_b$) across various regions, including the global region and the three regions with active mesoscale eddies: the Kuroshio Extension (KE), Gulf Stream (GS), and oceans around Southern Africa (OSA).

Model	Lead times Region	1-day	2-day	3-day	4-day	5-day	6-day	7-day	8-day	9-day	10-day
ConvLSTM	Global	10.13	5.65	4.37	3.67	3.17	2.88	2.76	2.76	2.91	3.15
	KE	20.42	10.60	7.45	5.96	4.80	4.10	3.72	3.51	3.52	3.73
	GS	12.19	5.96	4.13	3.53	3.25	3.15	3.20	3.40	3.70	4.09
	OSA	10.41	4.19	2.51	1.81	1.29	0.87	0.50	0.25	0.12	0.08
ResNet	Global	11.66	6.66	5.75	5.59	5.68	5.94	6.25	6.58	6.95	7.33
	KE	21.32	10.94	8.24	7.41	7.17	7.23	7.52	7.88	8.41	9.08
	GS	16.82	8.03	5.80	5.28	5.37	5.73	6.29	6.90	7.52	8.05
	OSA	25.79	11.47	7.45	6.08	5.59	5.51	5.64	5.87	6.13	6.36

4. Discussion and Conclusions

This study used the U-Transformer model to forecast global short-term SST, and compared its performance with that of the ConvLSTM and ResNet models. The U-Transformer model consistently outperformed the other two models, achieving the lowest RMSEs globally, ranging from 0.2–0.54 °C for 1- to 10-day lead times, with larger ACC of 0.97–0.79. Notably, the RMSEs produced by the U-Transformer model at the 1-day lead time were more than 10% smaller than those of the ConvLSTM and ResNet models.

In regions with active mesoscale eddies, such as the KE, GS, and OSA, the U-Transformer model also produced smaller RMSEs and higher ACC values. It reduced the RMSEs by 20.42% and 21.32% in the KE region, 12.19% and 16.82% in the GS region, and 10.41% and 25.79% in the OSA region, compared with those of the ConvLSTM and ResNet models, respectively. However, the RMSEs in these regions were 40%–130% above those of the global average, reflecting the difficulty in forecasting short-term SSTs in regions with active mesoscale eddies. The richness of mesoscale eddies or swift perturbation processes leads to strong nonlinearity because of their large energy and motion. The RMSE is generally low (<0.2 °C) in the tropical and subtropical oceans, except the eastern equatorial Pacific, where the influence of the TIW results in significantly higher prediction errors (>0.3 °C). In TIW regions, strong shear instability are dominated. It is very obvious mesoscale perturbation process. This underscores the challenges of accurately forecasting SST in regions dominated by strong

nonlinear dynamics and complex interactions. Such traits can lead to a more rapid loss of skills related to short-term forecasting of SSTs in such areas.

The good performance demonstrated by the U-Transformer model implies that the use of the self-attention mechanism can recognize pattern connections and nonlinear (eddies having strong nonlinear features) temporal relationships, which can enhance the ability for short-term SST forecasting in the regions with active mesoscale eddies.

These findings of this study emphasize the importance of selecting appropriate DL models for accurate SST forecasting, especially in regions characterized by complex physical processes. Future research should explore strategies intended to improve model performance in relation to forecasting in such challenging regions. Key directions for improving SST forecasting accuracy include incorporating physical constraints into neural networks, integrating physical prior knowledge [14,15], and designing more diverse input variables.

Author Contributions: Conceptualization, P.L. and T.Z.; methodology, T.Z.; software, T.Z.; validation, P.L., H.L., P.W., and Y.W.; formal analysis, T.Z.; investigation, T.Z.; resources, P.L.; data curation, T.Z.; writing—original draft preparation, T.Z.; writing—review and editing, P.L., H.L., P.W., Y.W., W.Z., Z.Y., J.J., Y.L., and H.H.; visualization, T.Z.; supervision, P.L.; project administration, H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Program for Developing Basic Sciences (Grant No. 2022YFC3104802), National Natural Science Foundations of China (Grant Nos. 41931183). The Strategic Priority Research Program of Chinese Academy of Sciences (XDB0500303).

Data Availability Statement: All data sets used in this study are publicly available. The OISST datasets is available and can be accessed at <https://psl.noaa.gov/thredds/catalog/Datasets/noaa.oisst.v2.highres/catalog.html>. The buoy observation datasets can be accessed at <https://www.pmel.noaa.gov/tao/drupal/disdel/>. The code for this study was developed using PyTorch. The Swin Transformer code can be found at <https://github.com/microsoft/Swin-Transformer>. The ConvLSTM code is available at <https://github.com/jhhuang96/ConvLSTM-PyTorch>. The ResNet code can be found at <https://github.com/weiaicunzai/pytorch-cifar100/blob/master/models/resnet.py>.

Acknowledgments: We thank for the technical support of the National Large Scientific and Technological Infrastructure “Earth System Numerical Simulation Facility” (<https://cstr.cn/31134.02.EL>).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Overview of Transformers

Transformer was initially introduced by Vaswani et al. (2017) [29] for natural language processing (NLP). Its core innovation lies in modeling long-range dependencies through a self-attention mechanism. This mechanism calculates the correlation between each input element and all other elements in the sequence, enabling the model to focus on globally significant features. To retain the sequential information of input data, the Transformer incorporates position encoding that introduces relative or absolute positional information. The typical Transformer architecture consists of multiple self-attention layers stacked together, each followed by a feedforward neural network, layer normalization, and residual connections. These components ensure training stability and enhance the model's expressiveness.

Although initially designed for sequential data, Transformers' capability to model long-range dependencies has led to widespread adoption in computer vision and remote sensing fields. For image data processing, Dosovitskiy et al. (2020) [30] introduced a Vision Transformer (ViT) that divides images into fixed-size patches (e.g., 16×16 pixels) and treats these patches as input tokens analogous to words in NLP. Each patch is embedded into a vector, and position encoding is added to preserve the spatial structure of the image. These embedded vectors are then processed through a series of Transformer layers, enabling the model to capture global dependencies across all patches. The convolutional neural networks (CNNs) excel at extracting local features, and Transformers demonstrate significant advantages in learning global long-range dependencies.

The introduction of ViT has made Transformers highly adaptable to remote sensing tasks with spatiotemporal attributes, particularly in scenarios requiring capturing large-scale spatiotemporal dependencies. In this study, inspired by ViT, we propose the U-Transformer architecture, which leverages the spatiotemporal characteristics of SST data (time, latitude, and longitude) to model the inherent relationships in SST effectively. This design facilitates more accurate short-term SST forecasts.

Appendix B. Model Architectures

Appendix B.1 ConvLSTM Architecture

The ConvLSTM model uses an encoder-decoder architecture to capture spatial and temporal input data features. The encoder begins with two convolutional layers: the first layer has 16 output channels, uses a kernel size of 3, a stride of 2, and padding of 1. The second convolutional layer also has 16 output channels and applies similar kernel size and padding. These convolutional layers help downsample the input data while preserving key spatial features. Following these layers, two ConvLSTM cells process the data. The first ConvLSTM cell operates on a spatial resolution of 360x720, while the second works on a reduced resolution of 180x360. Each ConvLSTM cell retains 16 feature maps per location, learning spatial and temporal dependencies across the data.

In the decoder, the model uses deconvolutional layers to restore the input data's spatial resolution progressively. The first deconvolutional layer has 16 output channels, a kernel size of 4, and a stride of 2, followed by a second deconvolutional layer with 32 output channels. Finally, a 1x1 convolutional layer changes the shape of the output. The architecture effectively bridges spatial and temporal dependencies through the ConvLSTM cells, making it suitable for spatiotemporal forecasting tasks.

Appendix B.2 ResNet Architecture

The ResNet model is a deep convolutional neural network adapted for spatiotemporal forecasting. It starts with an initial convolutional layer that processes the input data with 10 channels, applying a convolution operation with 64 output channels, a kernel size of 3, and padding of 1. The output is then passed through a ReLU activation function to introduce non-linearity. The core of ResNet consists of several residual blocks, where each block learns hierarchical features from the input data. The first block extracts features with 64 output channels, followed by blocks that progressively increase the output channels to 128, 256, and 512. These residual blocks utilize skip connections, which allow the model to learn residual mappings, mitigating the vanishing gradient problem and enabling the training of deeper networks.

Following the residual blocks, the network performs a convolution to produce the desired output shape, with additional convolutional layers to refine the predictions. Instead of using global pooling and fully connected layers, the model uses deconvolutional layers to upsample the output back to its original spatial resolution gradually. The first deconvolutional layer has 32 output channels, followed by another layer increasing the output channels to 64. A final deconvolutional layer refines the prediction, bringing the output to the required shape. The ResNet-18 architecture's use of residual connections helps it effectively capture complex spatial features while maintaining efficient training, making it a suitable model for spatiotemporal forecasting tasks.

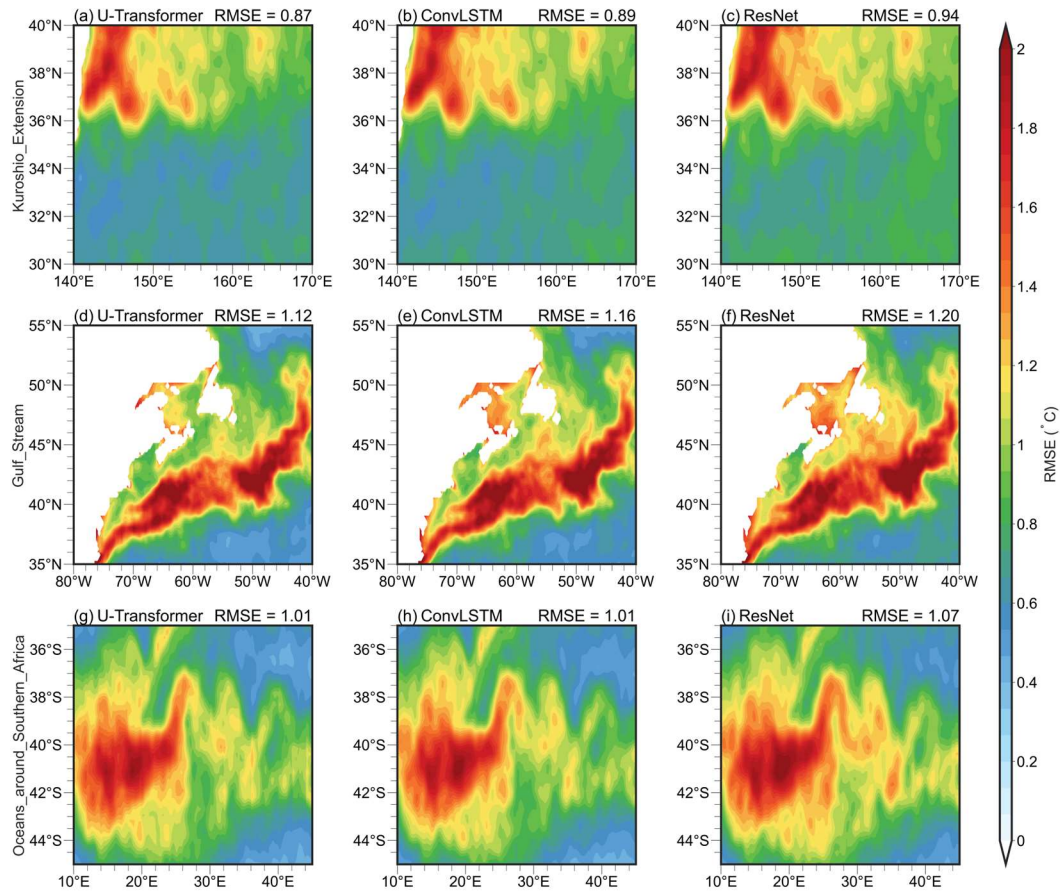


Figure A1. Spatial distribution of RMSE for forecast 10 days ahead by the U-Transformer model (a, d, g), ConvLSTM model (b, e, h), and ResNet model (c, f, i) during 2020–2022 in three eddy-active regions. The regional average RMSE values are displayed in the upper-right corner of each panel. Panels (a–c) correspond to the Kuroshio Extension (30°–40°N, 140°–170°E), panels (d–f) to the Gulf Stream (35°–55°N, 40°–80°W), and panels (g–i) to the oceans surrounding Southern Africa (35°–45°S, 10°–45°E).

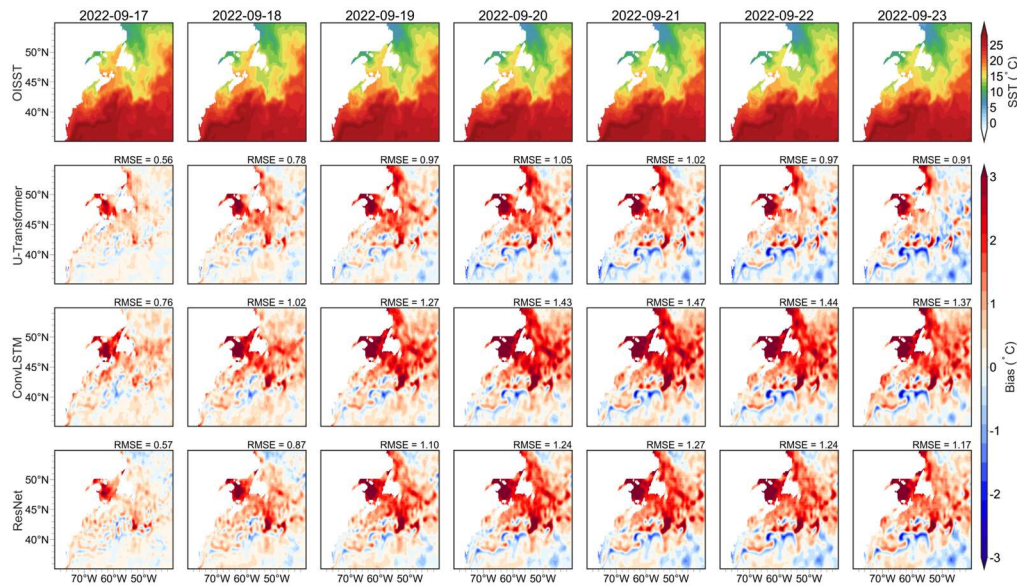


Figure A2. Comparison of OISST and SST forecasts by three deep learning models in the Gulf Stream region from September 17, 2020, to September 23, 2022. The first row represents OISST, while the second, third, and fourth rows show forecasts biases from the U-Transformer, ConvLSTM, and ResNet models.

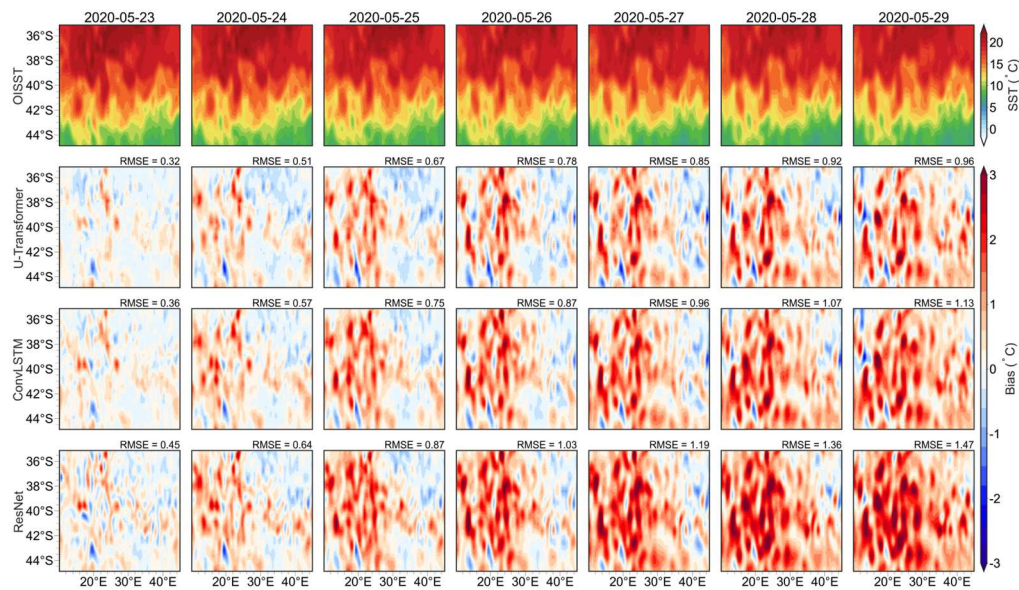


Figure A3. Comparison of OISST and SST forecasts by three deep learning models in the oceans around Southern Africa region from May 23, 2020, to May 29, 2022. The first row represents OISST, while the second, third, and fourth rows show forecasts biases from the U-Transformer, ConvLSTM, and ResNet models.

References

1. Behera, S. K., Luo, J.-J., Masson, S., Delecluse, P., Gualdi, S., Navarra, A., & Yamagata, T. (2005). Paramount Impact of the Indian Ocean Dipole on the East African Short Rains: A CGCM Study. *Journal of Climate*, 18(21), 4514–4530. <https://doi.org/10.1175/JCLI3541.1>
2. Zhou, L.-T., Tam, C.-Y., Zhou, W., & Chan, J. C. L. (2010). Influence of South China Sea SST and the ENSO on winter rainfall over South China. *Advances in Atmospheric Sciences*, 27(4), 832–844. <https://doi.org/10.1007/s00376-009-9102-7>
3. Rauscher, S. A., Jiang, X., Steiner, A., Williams, A. P., Cai, D. M., & McDowell, N. G. (2015). Sea Surface Temperature Warming Patterns and Future Vegetation Change. *Journal of Climate*, 28(20), 7943–7961. <https://doi.org/10.1175/JCLI-D-14-00528.1>
4. Salles, R., Mattos, P., Iorgulescu, A.-M. D., Bezerra, E., Lima, L., & Ogasawara, E. (2016). Evaluating temporal aggregation for predicting the sea surface temperature of the Atlantic Ocean. *Ecological Informatics*, 36, 94–105. <https://doi.org/10.1016/j.ecoinf.2016.10.004>
5. Cane, M. A., Clement, A. C., Kaplan, A., Kushnir, Y., Pozdnyakov, D., Seager, R., Zebiak, S. E., & Murtugudde, R. (1997). Twentieth-Century Sea Surface Temperature Trends. *Science*, 275(5302), 957–960. <https://doi.org/10.1126/science.275.5302.957>
6. Friedel, M. J. (2012). Data-driven modeling of surface temperature anomaly and solar activity trends. *Environmental Modelling & Software*, 37, 217–232. <https://doi.org/10.1016/j.envsoft.2012.04.016>
7. Castro, S. L., Wick, G. A., & Steele, M. (2016). Validation of satellite sea surface temperature analyses in the Beaufort Sea using UpTempO buoys. *Remote Sensing of Environment*, 187, 458–475. <https://doi.org/10.1016/j.rse.2016.10.035>
8. Bouali, M., Sato, O. T., & Polito, P. S. (2017). Temporal trends in sea surface temperature gradients in the South Atlantic Ocean. *Remote Sensing of Environment*, 194, 100–114. <https://doi.org/10.1016/j.rse.2017.03.008>
9. Chaidez, V., Dreano, D., Agusti, S., Duarte, C. M., & Hoteit, I. (2017). Decadal trends in Red Sea maximum surface temperature. *Scientific Reports*, 7(1), 8144. <https://doi.org/10.1038/s41598-017-08146-z>
10. Su, H., Wang, A., Zhang, T., Qin, T., Du, X., & Yan, X.-H. (2021). Super-resolution of subsurface temperature field from remote sensing observations based on machine learning. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102440. <https://doi.org/10.1016/j.jag.2021.102440>
11. Xu, G., Xie, W., Lin, X., Liu, Y., Hang, R., Sun, W., Liu, D., & Dong, C. (2024). Detection of three-dimensional structures of oceanic eddies using artificial intelligence. *Ocean Modelling*, 190, 102385. <https://doi.org/10.1016/j.ocemod.2024.102385>
12. Zhu, Y., Zhang, R.-H., Moum, J. N., Wang, F., Li, X., & Li, D. (2022). Physics-informed deep-learning parameterization of ocean vertical mixing improves climate simulations. *National Science Review*, 9(8), nwac044. <https://doi.org/10.1093/nsr/nwac044>

13. Qi, J., Xie, B., Li, D., Chi, J., Yin, B., & Sun, G. (2023). Estimating thermohaline structures in the tropical Indian Ocean from surface parameters using an improved CNN model. *Frontiers in Marine Science*, 10, 1181182. <https://doi.org/10.3389/fmars.2023.1181182>
14. Putra, D. P., & Hsu, P.-C. (2024). Leveraging Transfer Learning and U-Nets Method for Improved Gap Filling in Himawari Sea Surface Temperature Data Adjacent to Taiwan. *ISPRS International Journal of Geo-Information*, 13(5), 162. <https://doi.org/10.3390/ijgi13050162>
15. Young, C.-C., Cheng, Y.-C., Lee, M.-A., & Wu, J.-H. (2024). Accurate reconstruction of satellite-derived SST under cloud and cloud-free areas using a physically-informed machine learning approach. *Remote Sensing of Environment*, 313, 114339. <https://doi.org/10.1016/j.rse.2024.114339>
16. Zhang, Q., Wang, H., Dong, J., Zhong, G., & Sun, X. (2017). Prediction of Sea Surface Temperature Using Long Short-Term Memory. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1745–1749. <https://doi.org/10.1109/LGRS.2017.2733548>
17. Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., Xu, L., Chen, Z., & Gong, J. (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling & Software*, 120, 104502. <https://doi.org/10.1016/j.envsoft.2019.104502>
18. Sarkar, P. P., Janardhan, P., & Roy, P. (2020). Prediction of sea surface temperatures using deep learning neural networks. *SN Applied Sciences*, 2(8), 1458. <https://doi.org/10.1007/s42452-020-03239-3>
19. Jia, X., Ji, Q., Han, L., Liu, Y., Han, G., & Lin, X. (2022). Prediction of Sea Surface Temperature in the East China Sea Based on LSTM Neural Network. *Remote Sensing*, 14(14), 3300. <https://doi.org/10.3390/rs14143300>
20. Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., & Chen, Z. (2019). Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sensing of Environment*, 233, 111358. <https://doi.org/10.1016/j.rse.2019.111358>
21. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition (arXiv:1512.03385). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90
22. Shi, B., Hao, Y., Feng, L., Ge, C., Peng, Y., & He, H. (2024) An Attention-Based Context Fusion Network for Spatiotemporal Prediction of Sea Surface Temperature. *IEEE Geoscience and Remote Sensing Letters*, 21, 1504405. <https://doi.org/10.1109/LGRS.2024.3431586>
23. Zheng, G., Li, X., Zhang, R.-H., & Liu, B. (2020). Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Science Advances*, 6(29), ea-ba1482. <https://doi.org/10.1126/sciadv.aba1482>
24. Shi, B.; Ge, C.; Lin, H.; Xu, Y.; Tan, Q.; Peng, Y.; He, H. Sea Surface Temperature Prediction Using ConvLSTM-Based Model with Deformable Attention. *Remote Sens.* 2024, 16, 4126. doi: 10.3390/rs16224126
25. He, H. L., Shi, B. Y., Hao, Y. J., Feng, L., Lyu, X., & Ling, Z. (2024) Forecasting sea surface temperature during typhoon events in the Bohai Sea using spatiotemporal neural networks, *Atmospheric Research*, 309, 107578. <https://doi.org/10.1016/j.atmosres.2024.107578>
26. Xu, S., Dai, D., Cui, X., Yin, X., Jiang, S., Pan, H., & Wang, G. (2023). A deep learning approach to predict sea surface temperature based on multiple modes. *Ocean Modelling*, 181, 102158. <https://doi.org/10.1016/j.ocemod.2022.102158>
27. Xu, T., Zhou, Z., Li, Y., Wang, C., Liu, Y., & Rong, T. (2023). Short-Term Prediction of Global Sea Surface Temperature Using Deep Learning Networks. *Journal of Marine Science and Engineering*, 11(7), 1352. <https://doi.org/10.3390/jmse11071352>
28. Pan, X., Jiang, T., Sun, W., Xie, J., Wu, P., Zhang, Z., & Cui, T. (2024). Effective attention model for global sea surface temperature prediction. *Expert Systems with Applications*, 254, 124411. <https://doi.org/10.1016/j.eswa.2024.124411>
29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, 6000–6010. Curran Associates Inc., Red Hook, NY, USA.
30. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint*. <https://arxiv.org/abs/2010.11929>
31. Zou, R.; Wei, L.; Guan, L. Super Resolution of Satellite-Derived Sea Surface Temperature Using a Transformer-Based Model. *Remote Sens.* 2023, 15, 5376. doi: 10.3390/rs15225376
32. Lv, M., Wang, F., Li, Y., Zhang, Z., & Zhu, Y. (2022). Structure of Sea Surface Temperature Anomaly Induced by Mesoscale Eddies in the North Pacific Ocean. *Journal of Geo-physical Research: Oceans*, 127(3), e2021JC017581. <https://doi.org/10.1029/2021JC017581>
33. Carneiro, D.M., King, R., Martin, M. & Aguiar, A. (2021) Short-range ocean forecast error characteristics in high resolution assimilative systems. *Forecasting Research Technical Report 645*, Met Office. https://digital.nmla.metoffice.gov.uk/IO_e084c2c3-dc73-4cf3-acc1-44091ce6ef32

34. Lea, D.J., While, J., Martin, M.J., Weaver, A., Storto, A. & Chrust, M. (2022) A new global ocean ensemble system at the met Office: assessing the impact of hybrid data assimilation and inflation settings. *Quarterly Journal of the Royal Meteorological Society*, 148, 1996–2030. <https://doi.org/10.1002/qj.4292>
35. Chassignet, E. P., Yeager, S. G., Fox-Kemper, B., Bozec, A., Castruccio, F., Danabasoglu, G., Horvat, C., Kim, W. M., Koldunov, N., Li, Y., Lin, P.F., Liu, H., Sein, D. V., Sidorenko, D., Wang, Q., and Xu, X., (2020). Impact of horizontal resolution on global ocean–sea ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2), *Geosci. Model Dev.*, 13, 4595–4637, doi: 10.5194/gmd-13-4595-2020
36. Li, Y., Liu, H., Ding, M., Lin, P., Yu, Z., Yu, Y., Meng, Y., Li, Y., Jian, X., Jiang, J., Chen, K., Yang, Q., Wang, Y., Zhao, B., Wei, J., Ma, J., Zheng, W., & Wang, P. (2020). Eddy-resolving Simulation of CAS-LICOM3 for Phase 2 of the Ocean Model Intercomparison Project. *Advances in Atmospheric Sciences*, 37(10), 1067–1080. <https://doi.org/10.1007/s00376-020-0057-z>
37. Ding, M., Liu, H., Lin, P., Hu, A., Meng, Y., Li, Y., & Liu, K. (2022). Overestimated eddy kinetic energy in the eddy-rich regions simulated by eddy-resolving global ocean–sea ice models. *Geophysical Research Letters*, 49, e2022GL098370. <https://doi.org/10.1029/2022GL098370>
38. Nian, R., Cai, Y., Zhang, Z., He, H., Wu, J., Yuan, Q., Geng, X., Qian, Y., Yang, H., & He, B. (2021). The Identification and Prediction of Mesoscale Eddy Variation via Memory in Memory With Scheduled Sampling for Sea Level Anomaly. *Frontiers in Marine Science*, 8, 753942. <https://doi.org/10.3389/fmars.2021.753942>
39. Zhu, R., Song, B., Qiu, Z., & Tian, Y. (2024). A Metadata-Enhanced Deep Learning Method for Sea Surface Height and Mesoscale Eddy Prediction. *Remote Sensing*, 16(8), 1466. <https://doi.org/10.3390/rs16081466>
40. Wang, X., Li, C., Wang, X., Tan, L., & Wu, J. (2022). Spatio-Temporal Attention-Based Deep Learning Framework for Mesoscale Eddy Trajectory Prediction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 3853–3867. <https://doi.org/10.1109/JSTARS.2022.3174239>
41. Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Journal of Climate*, 20(22), 5473–5496. <https://doi.org/10.1175/2007JCLI1824.1>
42. Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., & Zhang, H.-M. (2021). Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *Journal of Climate*, 34(8), 2923–2939. <https://doi.org/10.1175/JCLI-D-20-0166.1>
43. Bryan, F. O., Tomas, R., Dennis, J. M., Chelton, D. B., Loeb, N. G., & McClean, J. L. (2010). Frontal Scale Air–Sea Interaction in High-Resolution Coupled Climate Models. *Journal of Climate*, 23(23), 6277–6291. <https://doi.org/10.1175/2010JCLI3665.1>
44. Lin, P., Liu, H., Ma, J., & Li, Y. (2019). Ocean mesoscale structure–induced air–sea interaction in a high-resolution coupled model. *Atmospheric and Oceanic Science Letters*, 12(2), 98–106. <https://doi.org/10.1080/16742834.2019.1569454>
45. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2021). Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation (arXiv:2105.05537). arXiv. <http://arxiv.org/abs/2105.05537>
46. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986. 0
47. Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization (arXiv:1412.6980). arXiv. <http://arxiv.org/abs/1412.6980>
48. Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization (arXiv:1711.05101). arXiv. <http://arxiv.org/abs/1711.05101>
49. Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. 2015. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'15)*. MIT Press, Cambridge, MA, USA, 802–810.
50. Zheng, W., Lin, P., Liu, H., Luan, Y., Ma, J., Mo, H., & Liu, J. (2023). An assessment of the LICOM Forecast System under the IVTT class4 framework. *Frontiers in Marine Science*, 10, 1112025. <https://doi.org/10.3389/fmars.2023.1112025>
51. Barbosa Aguiar, A., Bell, M. J., Blockley, E., Calvert, D., Crocker, R., Inverarity, G., King, R., Lea, D. J., Maksymczuk, J., Martin, M. J., Price, M. R., Siddorn, J., Smout-Day, K., Waters, J., & While, J. (2024). The Met Office Forecast Ocean Assimilation Model (FOAM) using a 1/12-degree grid for global forecasts. *Quarterly Journal of the Royal Meteorological Society*, qj.4798. <https://doi.org/10.1002/qj.4798>
52. Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A., & Storkey, D. (2014). Recent development of the Met Office operational ocean forecasting system: An overview and assessment of the new Global FOAM forecasts. *Geoscientific Model Development*, 7(6), 2613–2638. <https://doi.org/10.5194/gmd-7-2613-2014>

53. Liu, H., Lin, P., Zheng, W., Luan, Y., Ma, J., Ding, M., Mo, H., Wan, L., & Ling, T. (2023). A global eddy-resolving ocean forecast system in China – LICOM Forecast System (LFS). *Journal of Operational Oceanography*, 16(1), 15–27. <https://doi.org/10.1080/1755876X.2021.1902680>
54. Zhang, T., Lin P., Liu H., Zheng, W., Wang, P., Xu, T., Li, Y., Liu, J., & Chen, C. (2024). Short-Term Sea Surface Temperature Forecasts for the Equatorial Pacific Based on Long Short-Term Memory Network [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 48(2): 745–754. 10.3878/j.issn.1006-9895.2302.22128
55. Fu, L.-L., Chelton, D., Le Traon, P.-Y., & Morrow, R. (2010). Eddy Dynamics From Satellite Altimetry. *Oceanography*, 23(4), 14–25. <https://doi.org/10.5670/oceanog.2010.02>
56. Kwon, Y.-O., Alexander, M. A., Bond, N. A., Frankignoul, C., Nakamura, H., Qiu, B., & Thompson, L. A. (2010). Role of the Gulf Stream and Kuroshio–Oyashio Systems in Large-Scale Atmosphere–Ocean Interaction: A Review. *Journal of Climate*, 23(12), 3249–3281. <https://doi.org/10.1175/2010JCLI3343.1>
57. Ni, X., Zhang, Y., & Wang, W. (2025). Hurricane influence on the oceanic eddies in the Gulf Stream region. *Nature Communications*, 16(1), 583. <https://doi.org/10.1038/s41467-025-55927-6>
58. Chelton, D. B., Schlax, M. G., & Samelson, R. M. (2011). Global observations of nonlinear mesoscale eddies. *Progress in Oceanography*, 91(2), 167–216. <https://doi.org/10.1016/j.pocean.2011.01.002>
59. Kang, D., & Curchitser, E. N. (2013). Gulf Stream eddy characteristics in a high-resolution ocean model. *Journal of Geophysical Research: Oceans*, 118(9), 4474–4487. <https://doi.org/10.1002/jgrc.20318>
60. Kang, D., & Curchitser, E. N. (2015). Energetics of Eddy–Mean Flow Interactions in the Gulf Stream Region. *Journal of Physical Oceanography*, 45(4), 1103–1120. <https://doi.org/10.1175/JPO-D-14-0200.1>

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.