https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


# Deep Learning-Based Prediction of Marine Heatwaves in the East China Sea

Zefang Ma[1], Hui Chen[1], Qiyan Ji[1], Lifang Jiang[2,3], Cui Shen[1], Xiayan Lin[1]

1 Marine Science and Technology College, Zhejiang Ocean University, Zhoushan, 316022, China
2 Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, Guangzhou 510310, China
3 South China Sea Marine Forecast and Hazard Mitigation Center, Ministry of Natural Resources, Guangzhou 510310, China

Correspondence to: Hui Chen (chenhui@zjou.edu.cn) and Qiyan Ji (jiqiyan@zjou.edu.cn)

**Abstract.** Accurate sea surface temperature (SST) prediction in the East China Sea remains challenging because of its highly dynamic oceanic and atmospheric conditions, yet it is essential for regional fisheries management and marine hazard early warning. Here, we propose SwinTrans-ConvLSTM, a spatiotemporal deep-learning framework tailored for SST forecasting in the East China Sea. The model couples the global representation capability of the Swin Transformer with the local temporal-evolution modeling strength of ConvLSTM, incorporates air–sea temperature contrast and vector wind-field features, and is optimized using a curriculum-learning strategy constrained by a physics-informed gradient loss. Experiments show that SwinTrans-ConvLSTM achieves a mean absolute error of only 0.071 °C and a root mean square error of 0.156 °C on the test set, reducing prediction errors by approximately 30 % relative to state-of-the-art baselines. Crucially, hindcasts of the extreme 2022 marine heatwave event further demonstrate that the model can reproduce heatwave occurrence frequency and duration while substantially mitigating the systematic underestimation of extreme SST peaks inherent in purely data-driven models. These results highlight the critical role of thermodynamic-variable reconstruction and training-strategy optimization in improving the robustness of marine extreme-event prediction, and provide a promising technical pathway for high-resolution operational forecasting under complex ocean conditions.

## 1 Introduction

Marine heatwaves (MHWs) are extreme ocean warming events, which are characterized by sea surface temperatures (SSTs) exceeding the local climatological 90th percentile for five or more consecutive days (Hobday et al., 2016). With continued global warming, MHWs have become more frequent, longer-lasting, and more intense (Oliver et al., 2018). MHWs have caused widespread ecological disruptions, including coral bleaching, benthic habitat degradation, and species redistribution, while also amplifying the vulnerability of coastal fisheries and associated economies (Smale et al., 2019).

The East China Sea (ECS) is a marginal shelf sea along the northwestern Pacific continental margin, characterized by complex hydrographic and dynamical conditions. This region is influenced by multiple water masses and circulation systems, including Changjiang Diluted Water, the Zhejiang–Fujian Coastal Current, the Yellow Sea Warm Current, the Taiwan Warm Current, and the Kuroshio (Hickox et al., 2000). Interactions among the water mass and currents, together with the East Asian monsoon, generate pronounced temperature–salinity fronts and active sub-mesoscale processes (Chelton and Xie, 2010).

Long-term observations and reanalysis datasets indicate a significant warming trend in SSTs across the ECS and adjacent coastal waters over recent decades. The regional warming rate exceeds the global ocean average, making the ECS a hotspot of MHWs occurrence (Sasaki and Umeda, 2021; Xu et al., 2025). However, the mechanisms underlying extreme SST events in the ECS are highly complex. In addition to local air–sea interaction processes, they are modulated by multiple large-scale climate modes. For example, the westward extension of the western Pacific subtropical high and the negative phase of the Indian Ocean Dipole suppress surface winds, enhance downward shortwave radiation, and increase net air–sea heat flux, thereby accelerating SST warming (Tan et al., 2023). These coupled atmospheric–oceanic processes, together with the pronounced spatiotemporal variability of the ECS, pose significant challenges for predicting SST and associated MHW characteristics. In practice, most forecasting systems first predict SST fields, from which MHW events are subsequently identified based on climatological thresholds. SST forecasts therefore serve as the basis for MHW detection.

To obtain high-resolution SST and MHW forecasts, various modeling approaches have been developed. Traditional forecasting has mainly relied on high-resolution numerical ocean models, which simulate SST by solving the ocean dynamical equations within the Navier–Stokes framework. These models represent key physical processes but are limited by uncertainties in forcing and initial conditions, as well as by parameterizations of subgrid-scale processes (Zhao et al., 2026). Although physically comprehensive, high-resolution simulations are computationally expensive. Statistical downscaling methods have been widely used as an alternative to numerical models, offering improved computational efficiency but struggling to represent the non-stationarity and abrupt transitions characteristic of extreme events (Flint and Flint, 2012). More recently, data-driven approaches have emerged as a promising alternative, providing higher computational efficiency and improved multi-step forecasting capability, while partially improving the representation of the spatial structure of MHWs (Welandawe et al., 2025).

Among data-driven approaches, deep learning has achieved significant success in SST forecasting (Welandawe et al., 2025; Zhao et al., 2024). Early studies mainly focused on single-point time-series prediction, using methods such as wavelet analysis for multiscale decomposition (Mallat, 1989) and recurrent neural networks (RNN), including long short-term memory networks (LSTM), to construct autoregressive models (Zhang et al., 2017). However, these approaches largely neglect spatial correlations and advective processes, limiting their ability to reproduce realistic SST spatial structures. To address this limitation, ConvLSTM integrates convolutional operations with temporal sequence modeling, enabling joint learning of spatial–temporal dependencies and achieving promising short-term forecasting performance in the ECS and the South China Sea (Hao et al., 2023; Shi et al., 2015; Xiao et al., 2019). Subsequent studies have introduced deformable attention mechanisms to better capture spatial heterogeneity and sharp temperature gradients (Shi et al., 2024). In parallel, traditional machine learning methods such as random forests (Giamalaki et al., 2022) and three-dimensional U-Net models

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


integrating multiple physical variables have been applied to MHW intensity estimation and probabilistic forecasting (Bonino et al., 2024; Parasyris et al., 2025; Sun et al., 2023; Xie et al., 2024).

Nevertheless, the applicability of existing purely data-driven models remains limited for long-duration, large-domain forecasting of extreme SST events. From a spatial modeling perspective, convolution-based architectures are inherently constrained by local receptive fields, requiring multiple layers to propagate long-range dependencies, which weakens their sensitivity to basin-scale thermal forcing. Transformer-based models have therefore been widely adopted to enhance global dependency modeling (Vaswani et al., 2017). However, their computational complexity scales quadratically with grid size, while their ability to resolve fine-scale local structures remains limited (Dai et al., 2024; Song et al., 2024). To balance global context modeling and computational efficiency, the Swin Transformer, a hierarchical architecture based on shifted-window attention, has been proposed (Liu et al., 2021) and applied to meteorological downscaling and bias correction tasks (Xiang et al., 2022). Nevertheless, Swin-based models lack an explicit temporal modeling mechanism, limiting their ability to represent the evolution of continuous geophysical fields such as SST. Beyond architectural limitations, most existing data-driven models treat SST primarily as an autoregressive variable and do not explicitly incorporate key physical controls governing air–sea heat exchange, such as surface winds and air–sea temperature contrast. In addition, optimization based on mean squared error (MSE) tends to emphasize samples near the climatological mean, leading to a regression-to-the-mean effect that systematically underestimates the intensity and spatial gradients of extreme MHW events (Sun et al., 2024).

Motivated by the above limitations, this study develops a physics-variable-guided hybrid spatiotemporal deep-learning framework, termed SwinTrans-ConvLSTM, for high-resolution SST forecasting toward improved MHW prediction. The proposed framework combines the complementary strengths of Swin Transformer and ConvLSTM to simultaneously capture long-range spatial dependencies and temporal evolution. To better represent air–sea heat exchange processes, vector wind fields and air–sea temperature differences are incorporated alongside SST as input variables. Furthermore, a gradient-constrained loss function together with online hard example mining (OHEM) and curriculum learning is employed to improve the prediction of spatial structures and extreme thermal anomalies.

The structure of this paper is organized as follows: Section 2 provides a detailed description of the data, model configurations, and evaluation metrics. Section 3 presents an analysis of the predictive results. Section 4 discusses the findings of this study, and Section 5 concludes with a summary of the principal conclusions.

## 2 Data and Methods

### 2.1 Study Area

The study area in this work is mainly located in the ECS, covering the Changjiang Estuary, Hangzhou Bay, the Zhoushan Archipelago, and the coastal waters off Jiangsu and Shanghai (28°N–33°N, 120°E–125°E, Fig. 1). On the edge of the continental shelf in the northwestern Pacific, this region features shallow waters mostly less than 50 meters and relatively complex seabed. The hydrographic conditions of the ECS are jointly governed by multiple water masses and circulation

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


systems, including the shelf-intrusion branch of the Kuroshio, the Taiwan Warm Current, the Zhejiang–Fujian coastal currents, and the Changjiang Diluted Water. Furthermore, monsoon forcing and tidal processes induce strong vertical and horizontal mixing, resulting in pronounced temperature and salinity structures and distinct seasonal variability across the
100 marine environment.

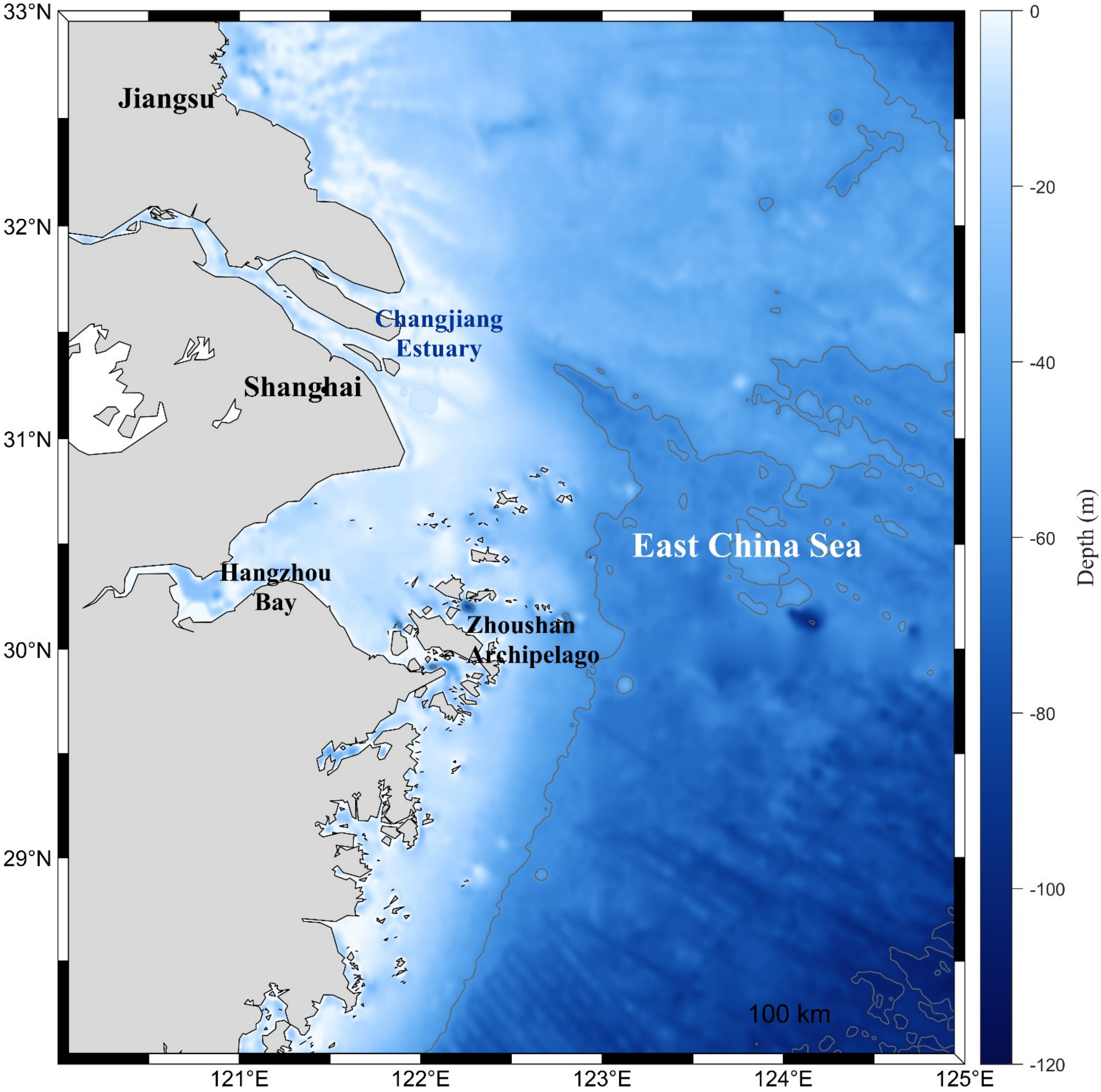



**Figure 1: Bathymetric map of the study area.**

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


### 2.2 Data

This study used the Level-4 reanalysis product from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, provided by the Copernicus Marine Environment Monitoring Service (CMEMS), as the primary sea surface temperature (SST) dataset. The OSTIA product assimilates satellite remote sensing observations with in situ measurements from buoys and ships. It provides gap-free global daily SST fields at a spatial resolution of 0.05° × 0.05°. This resolution allows the representation of fine-scale thermal structures in the ECS and adjacent waters. Owing to its high accuracy and temporal stability, the OSTIA dataset has been widely used in studies of marine environmental variability and has served as the core dataset for constructing and evaluating the MHW prediction model in this study.

To further account for the influence of air–sea interactions on SST variability, atmospheric forcing variables were obtained from the ERA5 global atmospheric reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 is generated using a state-of-the-art data assimilation system and provides spatiotemporally consistent atmospheric fields at a spatial resolution of 0.25° × 0.25°. Key variables affecting SST variability were selected, including 10-m wind vectors and 2-m air temperature. The wind components represent wind-driven ocean dynamical processes, while the 2-m air temperature used to quantify the air–sea thermal contrast describes the heat exchange across the air–sea interface.

To ensure spatial-temporal consistency among the multi-source datasets, the hourly ERA5 data were first aggregated into daily means and then interpolated onto the 0.05° spatial grid as the OSTIA product using bilinear interpolation. This preprocessing procedure ensures that all variables are consistent in both space and time, providing a unified input dataset for subsequent model training. The dataset covering the period from 1982 to 2024 was divided chronologically into three subsets. Data from 1982 to 2015 were used for model training and parameter optimization, data from 2016 to 2019 were used for validation and hyperparameter tuning with early stopping, and data from 2020 to 2024 were reserved for testing and final performance evaluation.

### 2.3 Data Preprocessing with physical constraints

To address the lack of physical constraints in purely data-driven models for ocean environmental prediction, this study incorporates atmospheric variables in addition to SST data to account for air–sea interaction effects different from currently works (Shu et al., 2025; Sun et al., 2024).

First, this study retained the 10-m zonal ($u_{10}$) and meridional ($v_{10}$) wind vectors as the dynamical physical constrain term, rather than using wind speed alone. This enables the model more potentially to capture the response of SST and extreme MHWs caused by wind-driven dynamical processes (Holbrook et al., 2019; Shu et al., 2025).

Second, inspired by the COARE 3.0 flux algorithm (Fairall et al., 2003), this study adopt the air–sea temperature difference ($\Delta T$) as the thermodynamic physical constrain term. This variable effectively reflects the heat transfer tendency across the air–sea interface and serves as an important thermodynamic indicator of SST variability:

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


$$\Delta T(t) = SST(t) - T_{2m}(t) \quad (1)$$

where $SST(t)$ denotes the sea surface temperature at time $t$, and $T_{2m}$ denotes the 2-m air temperature.

Furthermore, 3-day moving averages of the aforementioned physical variables were calculated and used as independent input channels, to partially represent the short-term thermal memory of the ocean system. Considering the large heat capacity of the ocean mixed layer, the SST response to atmospheric forcing generally exhibits a time lag, while MHWs often persist on subseasonal timescales (Sun et al., 2024). This procedure not only smoothed high-frequency stochastic perturbations in the atmospheric forcing but also provided short-term background-state information, helping the model capture temporal dependencies more robustly (Xu et al., 2025).

### 2.4 Marine Heatwave Metrics

MHWs were identified using the widely adopted methods proposed by Hobday (Hobday et al., 2016). MHWs events separated by gaps of less than two days were merged into a single event to ensure temporal continuity. The climatological baseline was defined as 1 January 1982 to 31 December 2011. An 11-day moving window centered on each calendar day was used to estimate the climatological mean and the 90th-percentile threshold, which were subsequently smoothed using a 31-day centered moving average to reduce high-frequency synoptic variability (Hobday et al., 2016).

Following Xu (Xu et al., 2025) and Oliver (Oliver et al., 2018), four metrics were used to characterize MHWs: mean duration, mean intensity, occurrence frequency, and total MHW days (Table 1). Mean duration represents the average event length, mean intensity is the average SST anomaly relative to the climatological mean during each event, occurrence frequency is the number of MHW events within a given period, and total MHW days denote the cumulative duration of all detected events. The predicted SST fields for lead days 1–7 were converted into MHW events using the same detection procedure described above. The predicted events were subsequently compared with those identified from the reanalysis dataset to evaluate the model performance in forecasting MHW occurrence and associated characteristics.

**Table 1: Key metrics of MHWs and their formulations.**

| Metric | Formula | Unit |
|---|---|---|
| Mean duration | $\bar{D} = \frac{1}{N}\sum_{i=1}^{N}(t_{end,i} - t_{start,i} + 1)$ | Days |
| Mean intensity | $I_{mean} = \frac{1}{TD}\sum_{k\in\Omega}(T_k - T_{clim,k})$ | °C |
| Frequency | $\mathrm{F} = \mathrm{N}$ | Times |
| Total MHW days | $TD = \sum_{i=1}^{N} D_i$ | Days |

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


## 2.5 Deep learning models

### 2.5.1 Architecture of SwinTrans-ConvLSTM

This study proposes a deep hybrid spatiotemporal forecasting model, namely SwinTrans-ConvLSTM. The model adopts an asymmetric encoder–decoder architecture (Fig. 2), integrating the local temporal modeling capability of convolutional long short-term memory (ConvLSTM) with the global spatial representation capability of the Swin Transformer.

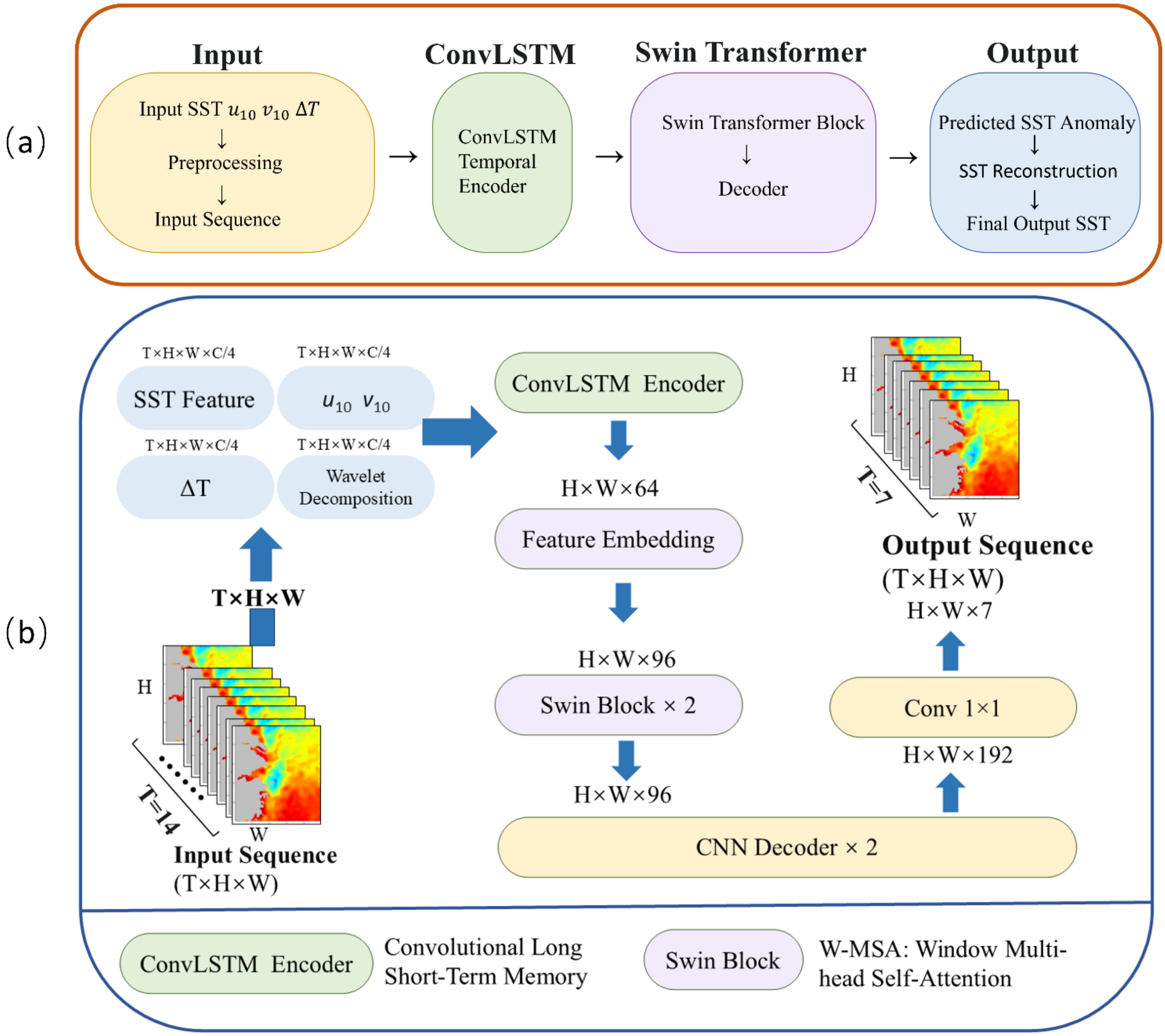



**Figure 2: Architecture of the SwinTrans-ConvLSTM model. (a) Schematic overview of the model workflow, illustrating the complete data flow from multi-source input preprocessing, through temporal feature extraction and spatial feature modeling, to final SST reconstruction. (b) Detailed network architecture, showing the internal interactions among individual modules.**

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


As illustrated in Fig. 2(a), the model ingests a multi-channel input sequence formulated by concatenating historical SST fields with physics-informed variables, namely, 10-m wind vectors ($u_{10}$, $v_{10}$) and air–sea temperature difference ($\Delta T$), along the channel dimension. The encoder then employs ConvLSTM units to extract temporal features from this coupled spatiotemporal tensor. By integrating convolutional operations into recurrent computations, ConvLSTM preserves the spatial structure of the gridded data while capturing local spatiotemporal evolution, making it well suited for representing SST variations driven by tidal processes, coastal currents, and other nearshore dynamics. Previous studies have demonstrated that ConvLSTM can effectively represent the evolution of complex oceanic structures, providing a reliable basis for high-resolution SST prediction (Shi et al., 2024).

To enhance large-scale spatial feature extraction, a Swin Transformer module was incorporated into the deep feature extraction stage. Its shifted-window self-attention mechanism enables efficient information exchange across neighboring regions, allowing the model to capture long-range spatial dependencies while maintaining computational efficiency. This design improves the representation of large-scale oceanic processes, such as Kuroshio intrusion, that influence regional SST variability (Xiao et al., 2019).

At the prediction stage, SST fields exhibit relatively large absolute values but small short-term variations. To facilitate the learning of these subtle changes, a residual learning strategy was adopted (Zhao et al., 2024), in which the model predicts the SST increment relative to the current time step rather than the absolute SST field:

$$\hat{Y}_{t+n} = X_t + \mathcal{F}(X_{t-k:t}) \quad (2)$$

where $X_t$ denotes the observation at the current time step, $X_{t-k:t}$ represents the input historical sequence, $\mathcal{F}$ is the temperature increment predicted by the deep network, and $\hat{Y}_{t+n}$ denotes the final prediction at the target time step $n$.

This formulation enables the model to focus on the nonlinear evolution of SST anomalies, thereby improving prediction efficiency and reducing errors associated with small perturbations. To further preserve fine-scale spatial structures, a gradient-based loss (Mathieu et al., 2016) was incorporated into the objective function to constrain the spatial gradients of the predicted SST fields, thereby improving structural consistency while maintaining prediction accuracy.

#### 2.5.2 Baseline models

To evaluate the effectiveness of the proposed SwinTrans-ConvLSTM model, four representative deep-learning models were selected for comparison: LSTM, ConvLSTM, Transformer, and Swin Transformer.

LSTM models temporal dependence at each grid points independently without explicitly considering spatial correlations. ConvLSTM extends this framework by introducing convolutional operations to jointly model spatial and temporal features, but its receptive field is limited by kernel size, restricting large-scale dependency capture (Xiao et al., 2019). Transformer architectures use global self-attention to model long-range dependencies but suffer from high computational cost and potential loss of local structure in high-resolution gridded data (Zhao et al., 2024). The Swin Transformer reduces computational complexity through shifted-window attention while improving spatial dependency

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


modeling; however, its lack of explicit recurrence limits its ability to represent continuous temporal evolution (Dai et al., 2024).

In contrast, the proposed SwinTrans-ConvLSTM integrates ConvLSTM-based temporal evolution modeling with Swin Transformer-based spatial representation, enabling unified learning of local dynamics and large-scale spatial dependencies.

### 2.6 Experimental Setup

#### 2.6.1 Data preparation and Training strategy

Model construction and training were implemented using the PyTorch deep-learning framework on a high-performance computing platform equipped with an NVIDIA GeForce RTX 3090 GPU. The model was trained using a unified experimental protocol to ensure efficient optimization and reproducible performance. The dataset was divided chronologically into training, validation, and test sets to avoid temporal information leakage. Training samples were generated using a sliding-window strategy, in which multi-source physical variables from the preceding 14 days were used to predict SST fields over the subsequent 7 lead days.

Given the long-tailed distribution of marine heatwave occurrences approximately 5 %–10 % (Sun et al., 2024), standard data-driven models optimized solely on global mean errors frequently underestimate extreme values. To address this and preserve the spatial structural fidelity of SST fields, we designed a custom objective function termed the Physics-Informed Curriculum Loss. The total loss $L_{total}$ at a given training epoch $E$ is formulated as the sum of a base regression loss $L_{base}$ and a spatial gradient constraint $L_{grad}$:

$$L_{total}(E) = L_{base}(E) + \lambda_{grad} L_{grad} \tag{3}$$

where $\lambda_{grad}$ is a scaling hyperparameter (empirically set to 1.0) that balances the physical constraint.

The base regression loss first computes the element-wise Huber loss (Huber, 1964) between the true SST ($y_i$) and the predicted SST ($\hat{y}_i$) at the $i$-th valid grid point. To maintain robustness against extreme outliers while providing stable gradients for small errors, the Huber penalty $l_i$ is defined as:

$$l_i = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2, & \text{if } |y_i - \hat{y}_i| \le \delta \\ \delta|y_i - \hat{y}_i| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases} \tag{4}$$

where $\delta$ is the threshold parameter (set to 1.0). Furthermore, to emphasize the larger SST variability and intensive air-sea heat fluxes during the MHW-prone summer season, a seasonal weighting factor $W_{seasonal}$ is applied:

$$W_{seasonal} = \begin{cases} 2.0, & \text{if the target date falls in Jun, Jul, or Aug} \\ 1.0, & \text{otherwise} \end{cases} \tag{5}$$

The weighted loss for the $i$-th grid point is thus defined as

$$L_i = W_{seasonal} \cdot l_i \tag{6}$$

Deep learning models optimized on global mean errors often suffer from a "regression-to-the-mean" effect, underestimating extreme values. To mitigate this, an epoch-dependent curriculum learning strategy is adopted. Initially ($E <$

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


$E_{start}$, where $E_{start} = 20$), the model learns the global background by minimizing the mean loss across all $N$ valid grid points. Subsequently ($E \geq E_{start}$), the Online Hard Example Mining (OHEM) mechanism is activated (Shrivastava et al., 2016). By backpropagating gradients only from the top $k$ grid points $k = [N \times 20\%]$ with the highest losses, the network focuses exclusively on hard-to-predict MHW anomalies. Thus, the base regression loss $L_{base}(E)$ is formulated as a piecewise function:

$$L_{base}(E) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} L_i\,, & \text{if } E < E_{start} \\ \frac{1}{k}\sum_{j\in\Omega_{hard}} L_j\,, & \text{if } E \geq E_{start} \end{cases} \tag{7}$$

where $L_j$ denotes the point-wise loss at the $j$-th grid point, and $\Omega_{hard}$ represents the index set of the top $k$ grid points with the highest loss values.

Finally, to ensure that the predicted SST fields maintain physically realistic oceanic frontal structures rather than being overly smoothed, an explicit spatial gradient loss $L_{grad}$ is enforced (Mathieu et al., 2016). This constraint calculates the $L_1$-norm penalty on the absolute differences along the longitudinal ($x$) and latitudinal ($y$) directions:

$$L_{grad} = \frac{1}{N}\sum_{i=1}^{N}\left(|\nabla_x y_i - \nabla_x \hat{y}_i| + |\nabla_y y_i - \nabla_y \hat{y}_i|\right) \tag{8}$$

where $\nabla_x$ and $\nabla_y$ denote the spatial derivatives, which are approximated by adjacent grid finite differences (e.g., $\nabla_x y = y_{u,v+1} - y_{u,v}$ for a grid coordinate $(u, v)$). Model optimization was continuously performed using the AdamW optimizer with a Cosine Annealing with Warm Restarts learning rate schedule, which improved training stability and convergence behavior.

### 2.6.2 Forecasting strategy and evaluation setup

A rolling forecasting strategy was applied to the test period with a step size of 1 day. For each forecast instance, the model used the preceding 14-day SST sequence as input to predict the subsequent 7 days. This procedure generated a four-dimensional prediction tensor of shape $[N_{\mathrm{windows}}, 7, H, W]$, where $N_{windows}$ denotes the number of forecast windows and $H$ and $W$ represent spatial dimensions.

This design preserves complete spatiotemporal consistency across lead times and enables independent evaluation at different forecast horizons. Predicted SST fields were converted into MHW events using the definition following Hobday (Hobday et al., 2016). MHW events were identified separately for each lead time and compared with those derived from reanalysis data to evaluate predictive performance.

SwinTrans-ConvLSTM were trained under identical experimental settings to ensure fair comparison. Detailed information on dataset partitioning, model architecture, and training hyperparameters is provided in Table 2.

**Table 2: Model architecture and training configuration of the SwinTrans-ConvLSTM.**

| | | |
|---|---|---|
| Tensor Shape | Input Shape | [$B$,14,25,100,100] |
| | Output Shape | [$B$,7,1,100,100] |

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


| | | |
|---|---|---|
| Model Architecture | Hidden Channels | 64 (ConvLSTM), 96 (Swin) |
| | Kernel Size | 3×3 |
| | Encoder Layers | 2 |
| | Num. Heads | [3,6] |
| | Window Size | 10×10 |
| | Dropout | 0.2 |
| Training Settings | Max Epochs | 120 |
| | Batch Size | 4 |
| | Optimizer | AdamW |
| | Initial LR | $3 \times 10^{-5}$ |
| | LR Scheduler | Cosine Annealing Warm Restarts |
| | Weight Decay | $1 \times 10^{-3}$ |
| | Early Stopping | 30 Epochs |
| | OHEM Ratio | Top 20 % |

**2.7 Evaluation Metrics**

To quantitatively evaluate the prediction performance of the proposed SwinTrans-ConvLSTM model, four statistical metrics commonly used in SST forecasting studies were adopted. All metrics were calculated over valid ocean grid points in the test set.

The mean absolute error (MAE) measures the average absolute difference between predicted and observed SST, providing an intuitive estimate of prediction accuracy.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{9}$$

The root mean square error (RMSE) assigns a squared penalty to large errors and is therefore sensitive to extreme deviations. It is particularly useful for evaluating the model's ability to capture extreme SST variations associated with MHWs (Sun et al., 2024).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{10}$$

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


The Pearson correlation coefficient (PCC) was used to quantify the linear agreement between predicted and observed SST, reflecting phase consistency across temporal scales.

$$\mathrm{PCC} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})\left(\hat{y}_i - \bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{11}$$

The coefficient of determination ($R^2$) quantifies the proportion of variance explained by the model, with values closer to 1 indicating better performance.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{12}$$

where $N$ denotes the total number of samples, $y_i$ and $\hat{y}_i$ represent the observed and predicted values of the $i$-th sample, respectively, and $\bar{y}$ and $\bar{\hat{y}}$ denote the corresponding mean values of the observations and predictions.

To evaluate extreme event detection skill, the Symmetric Extremal Dependence Index (SEDI) was introduced (Shu et al., 2025), which is designed for rare-event verification. SEDI ranges from −1 to 1, with higher values indicating better performance. It is defined as follows:

$$\mathrm{SEDI} = \frac{ln(FAR) - ln(POD) + ln(1 - POD) - ln(1 - FAR)}{ln(FAR) + ln(POD) + ln(1 - POD) + ln(1 - FAR)} \tag{13}$$

where *POD* denotes the probability of detection (hit rate), defined as the fraction of correctly predicted marine heatwave days, and *FAR* denotes the false alarm ratio, defined as the fraction of false alarms among predicted events. The formulations are given as follows:

$$POD = \frac{Hits}{Hits + Misses} \tag{14}$$

$$FAR = \frac{FalseAlarms}{Hits + FalseAlarms} \tag{15}$$

where "*Hits*" denote correctly predicted marine heatwave days, "*Misses*" represent missed observations, and "*FalseAlarms*" correspond to false detections. By using logarithmic transformations of hit and false alarm rates, SEDI provides a robust measure of predictive skill for rare extreme events.

## 3 Results

### 3.1 Metric-based Evaluation of Models' Performance

To evaluate the spatiotemporal SST prediction performance of the SwinTrans-ConvLSTM model, four representative deep learning benchmark models were constructed for comparison (Sect. 2.5.2), including LSTM, ConvLSTM, Transformer, and Swin Transformer. To ensure comparability, all models were trained under a unified strategy. The predictive performance of all models was assessed using multiple evaluation metrics to examine the effectiveness of the proposed approach.

The results in Table 3 show that the proposed SwinTrans-ConvLSTM model achieves the best performance across all metrics. In particular, it yields an MAE of 0.071 °C and an RMSE of 0.156 °C, which are lower than those of all baseline models. In addition, the coefficient of determination ($R^2$) and Pearson correlation coefficient (PCC) reach 0.9994 and 0.9997, respectively, indicating strong agreement between predicted and observed SST fields.

Compared with the Transformer model, the proposed Swin Transformer reduces MAE and RMSE by 19.9 % and 30.2 %, respectively, while increasing $R^2$ from 0.9983 to 0.9992. While the proposed SwinTrans-ConvLSTM reduces MAE and RMSE by 47.8 % and 41.8 %, respectively, and $R^2$ increasing to 0.9994. These results indicate that Transformer-based architectures are effective in capturing spatial dependencies but are limited in modeling temporal dynamics, while the ConvLSTM module enhances spatiotemporal continuity.

Compared with ConvLSTM, the proposed model achieves a 13.6 % reduction in MAE and a 22.4 % reduction in RMSE, demonstrating the benefit of multi-scale spatial feature representation. In contrast, ConvLSTM and LSTM show similar MAE values (0.083 °C and 0.084 °C), suggesting that improved temporal modeling alone provides limited gains in accuracy. The proposed model also achieves the highest PCC (0.9997), indicating improved consistency between predicted and observed SST fields in both magnitude and spatial–temporal patterns. Overall, the SwinTrans-ConvLSTM achieves the best performance by integrating global spatial representation and local spatiotemporal modeling capabilities.

**Table 3: Metric-based evaluation of different models on the test set.**

| Model | MAE (°C) | RMSE (°C) | $R^2$ | PCC |
|---|---|---|---|---|
| SwinTrans-ConvLSTM | 0.071 | 0.156 | 0.9994 | 0.9997 |
| ConvLSTM | 0.083 | 0.201 | 0.9990 | 0.9995 |
| LSTM | 0.084 | 0.208 | 0.9990 | 0.9995 |
| Swin-Transformer | 0.109 | 0.187 | 0.9992 | 0.9996 |
| Transformer | 0.136 | 0.268 | 0.9983 | 0.9992 |

### 3.2 Temporal Evolution of Prediction Errors

In operational ocean forecasting, the growth of prediction errors with increasing lead time is an important indicator for evaluating model robustness and long-range forecasting capability. Due to the nonlinear and multiscale nature of ocean dynamics, initial errors tend to amplify during the forecasting process (Zhang et al., 2017; Zhao et al., 2026). Therefore, this study further analyzes the evolution, spatial distribution, and frequency characteristics of prediction errors across different lead times.

As shown in Fig. 3, both MAE and RMSE increase with forecast lead time across all models, reflecting the general behavior of nonlinear error propagation in ocean systems. Although all models exhibit error growth, the proposed SwinTrans-ConvLSTM consistently shows the slowest increase rate compared with the benchmark models. As summarized

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


in Table 5, from 1-day to 7-day lead time, the MAE of SwinTrans-ConvLSTM increases from 0.022 °C to 0.136 °C, corresponding to a total increase of 0.114 °C. Meanwhile, RMSE increases from 0.082 °C to 0.198 °C, with an increase of 0.116 °C. In contrast, the Transformer model shows larger increases in MAE and RMSE, reaching 0.175 °C and 0.241 °C, respectively. Overall, the proposed model demonstrates a slower error growth rate and better stability for long-lead SST forecasting.

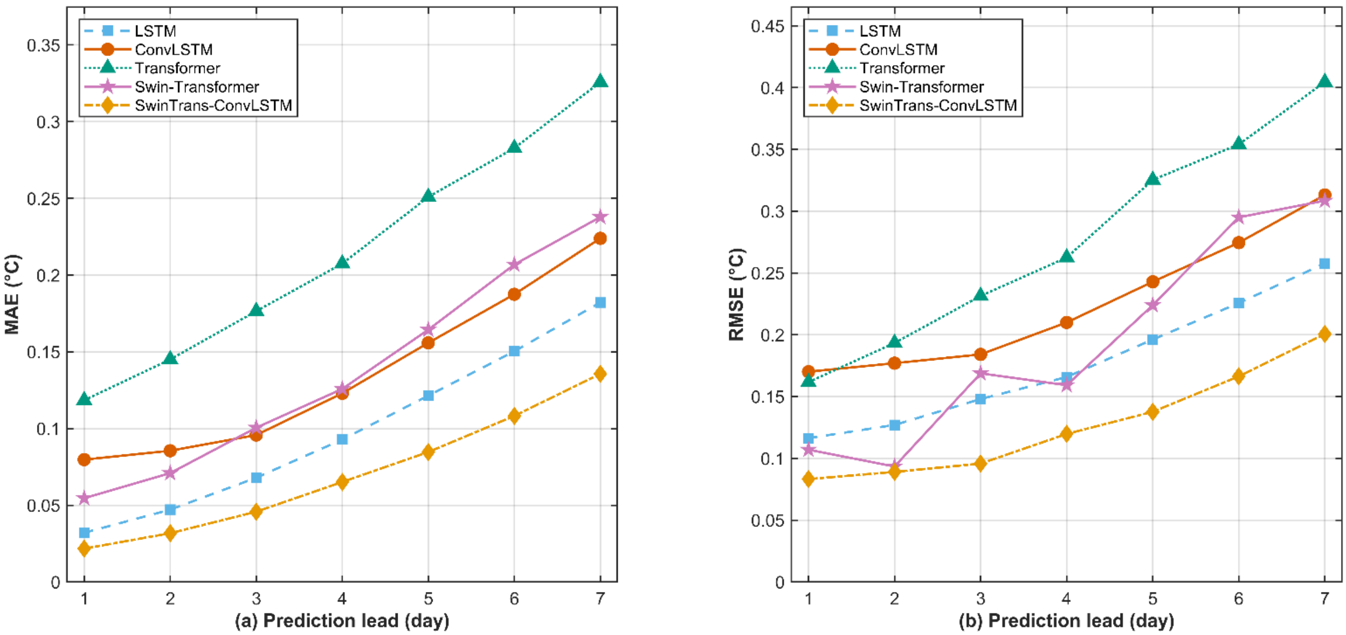



**Figure 3: Comparison of error evolution under different forecast lead times: (a) MAE and (b) RMSE.**

The spatially averaged error distributions under different forecast lead times (Fig. 4) show that the SwinTrans-ConvLSTM model consistently achieves the lowest error across all prediction horizons. Its advantage becomes more evident at medium- to long-term lead times (5–7 days). As shown in Table 4, at the 7-day lead time, the MAE of SwinTrans-ConvLSTM reaches 0.136 °C, representing reductions of 11.7 %, 10.5 %, 35.8 %, and 41.1 % compared with ConvLSTM (0.154 °C), LSTM (0.152 °C), Swin Transformer (0.212 °C), and Transformer (0.231 °C), respectively. The corresponding RMSE is 0.198 °C, which is 36.1 % lower than that of ConvLSTM and 50.7 % lower than Transformer. The proposed SwinTrans-ConvLSTM model exhibits lower error levels and more stable performance across increasing forecast lead times.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026

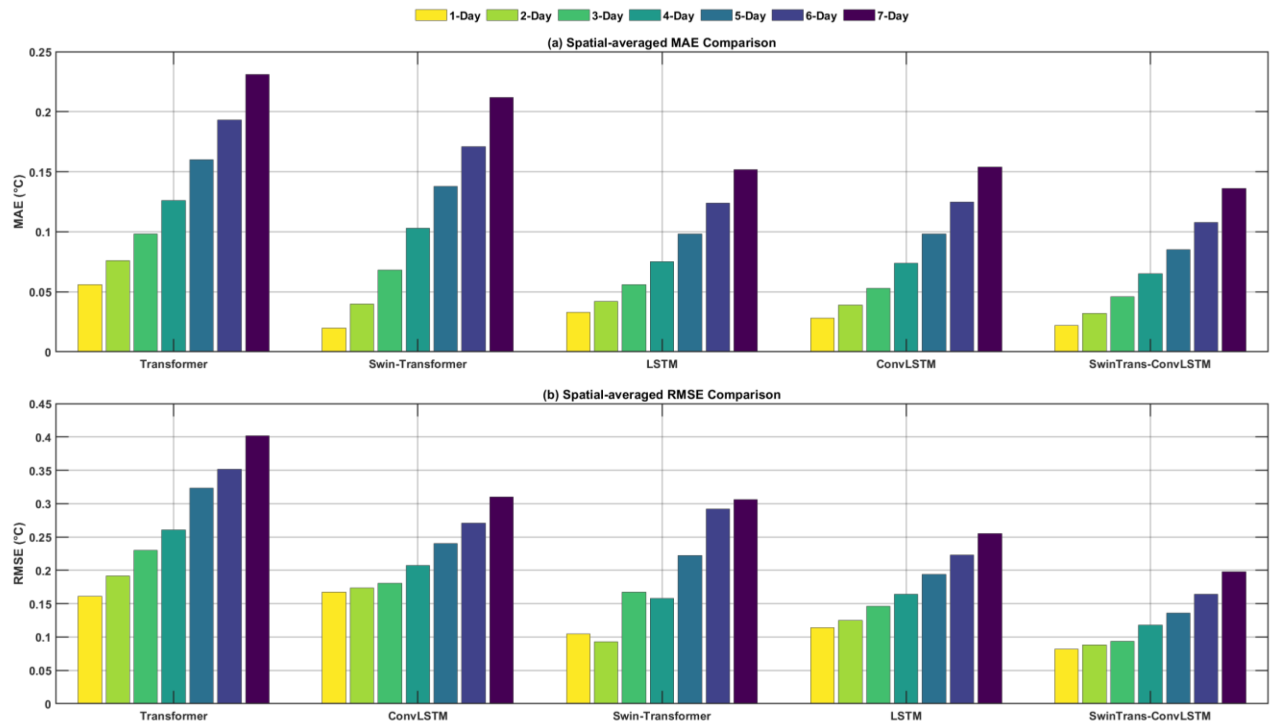




**Figure 4: Comparison of spatially averaged model errors under different forecast lead times: (a) MAE and (b) RMSE.**

**Table 4: Detailed day-by-day performance comparison of different models on the test set. Both MAE and RMSE are in °C.**

| Lead Time | Transformer | | ConvLSTM | | Swin-Transformer | | LSTM | | **SwinTrans-ConvLSTM** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | **MAE** | **RMSE** |
| 1-Day | 0.056 | 0.161 | 0.028 | 0.167 | 0.020 | 0.105 | 0.033 | 0.114 | **0.022** | **0.082** |
| 2-Day | 0.076 | 0.192 | 0.039 | 0.174 | 0.040 | 0.093 | 0.042 | 0.125 | **0.032** | **0.088** |
| 3-Day | 0.098 | 0.23 | 0.053 | 0.181 | 0.068 | 0.167 | 0.056 | 0.146 | **0.046** | **0.094** |
| 4-Day | 0.126 | 0.261 | 0.074 | 0.207 | 0.103 | 0.158 | 0.075 | 0.164 | **0.065** | **0.118** |
| 5-Day | 0.160 | 0.323 | 0.098 | 0.24 | 0.138 | 0.222 | 0.098 | 0.194 | **0.085** | **0.136** |
| 6-Day | 0.193 | 0.352 | 0.125 | 0.271 | 0.171 | 0.292 | 0.124 | 0.223 | **0.108** | **0.164** |
| 7-Day | 0.231 | 0.402 | 0.154 | 0.310 | 0.212 | 0.306 | 0.152 | 0.255 | **0.136** | **0.198** |

Figure 5 illustrates the error frequency distributions of different models under varying forecast lead times. The SwinTrans-ConvLSTM exhibits a highly concentrated error distribution around 0 °C, indicating strong agreement between predictions and observations. At 1-day lead time, 98.6 % of samples fall within the ±0.1 °C error range, outperforming all comparison models. At 7-day lead time, the probabilities $P(\pm0.1\ °C)$, $P(\pm0.3\ °C)$, and $P(\pm0.5\ °C)$ remain 52.9 %, 90.5 %,

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


and 98.1 %, respectively, higher than those of the Transformer model (42.4 %, 75.5 %, and 88.3 %). In contrast, the Transformer and Swin Transformer exhibit broader error distributions and more pronounced long-tail behavior under longer lead times, whereas the proposed model maintains a more compact and symmetric distribution.

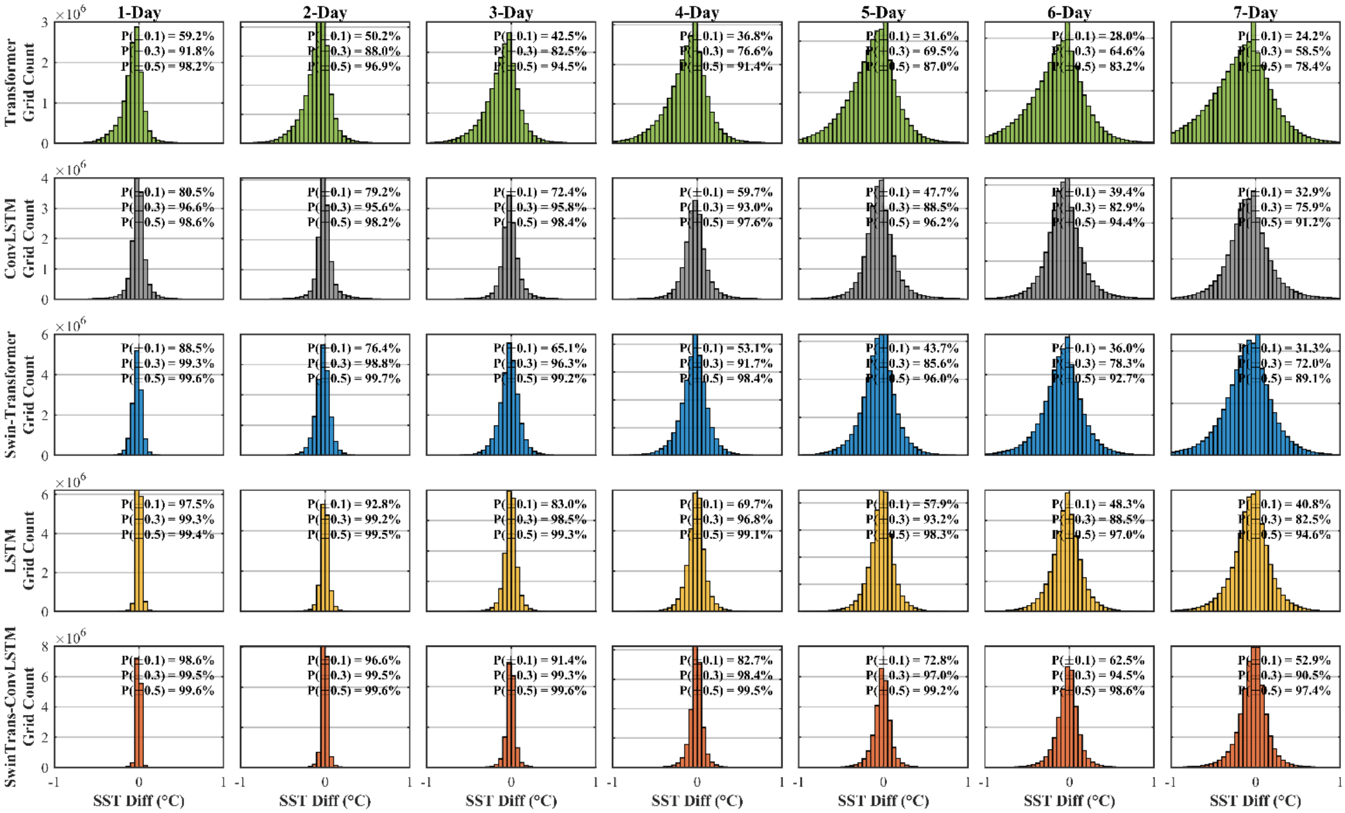



**Figure 5: Histogram of prediction error frequency distributions under different forecast lead times.**

Overall, the results in Fig. 3 to Fig. 5 indicate that the SwinTrans-ConvLSTM model achieves superior predictive performance compared with benchmark models. It also demonstrates improved stability in long-term forecasting, reduced error accumulation, and better control of large prediction deviations. These results highlight the effectiveness of the proposed hybrid spatiotemporal architecture for complex ocean forecasting tasks.

### 3.3 Spatial Distribution of Prediction Errors

To evaluate model performance across different ocean regions and forecast lead times, the spatial distributions of MAE and RMSE from 1 to 7 days are shown in Figs. 6 and 7. Overall, all models exhibit a consistent pattern of higher errors in coastal regions and lower errors in the open ocean, with major error hotspots located in the Yangtze River Estuary, Hangzhou Bay, and the coastal waters of Zhejiang Province.

This spatial distribution is associated with complex shelf dynamics characterized by shallow bathymetry, strong tidal mixing, and freshwater input from the Yangtze River, which lead to pronounced thermohaline fronts (Hickox et al., 2000).

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


The strong gradients and high variability in these frontal zones make accurate representation of nonlinear processes challenging for data-driven models, often resulting in overly smooth predictions and increased errors in these regions (Xiao et al., 2019).

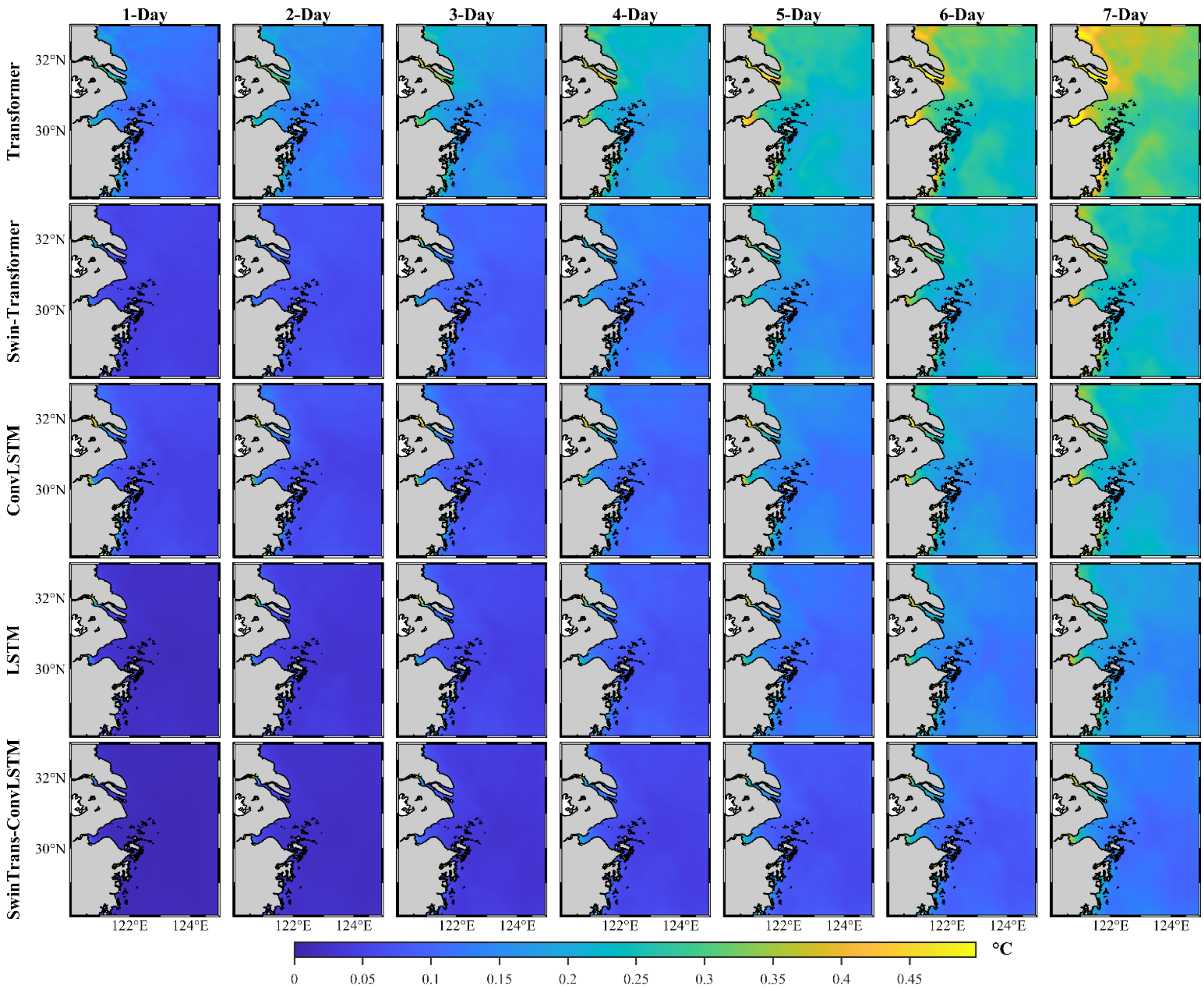



**Figure 6: Spatial distribution of MAE under different forecast lead times.**

A comparison of spatial error distributions across different models under long-term forecasting reveals pronounced differences in their error evolution characteristics, particularly after 5 Days (Figs. 6 and 7). As the forecast lead time increases, both the Transformer and ConvLSTM models exhibit a gradual expansion of high-error regions from nearshore areas toward the open ocean, indicating a reduced capability in representing complex boundary conditions and ocean dynamical processes.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


In contrast, the SwinTrans-ConvLSTM model maintains superior spatial stability throughout the entire forecasting period, with high-error regions consistently confined to relatively limited areas. Notably, in the shelf break zone around 123°E and the Kuroshio branch-influenced region east of 124°E, the proposed model effectively suppresses the diffusion of errors toward deeper waters. Most regions remain characterized by low error levels, demonstrating stronger spatial structure preservation capability and enhanced long-term forecasting stability.

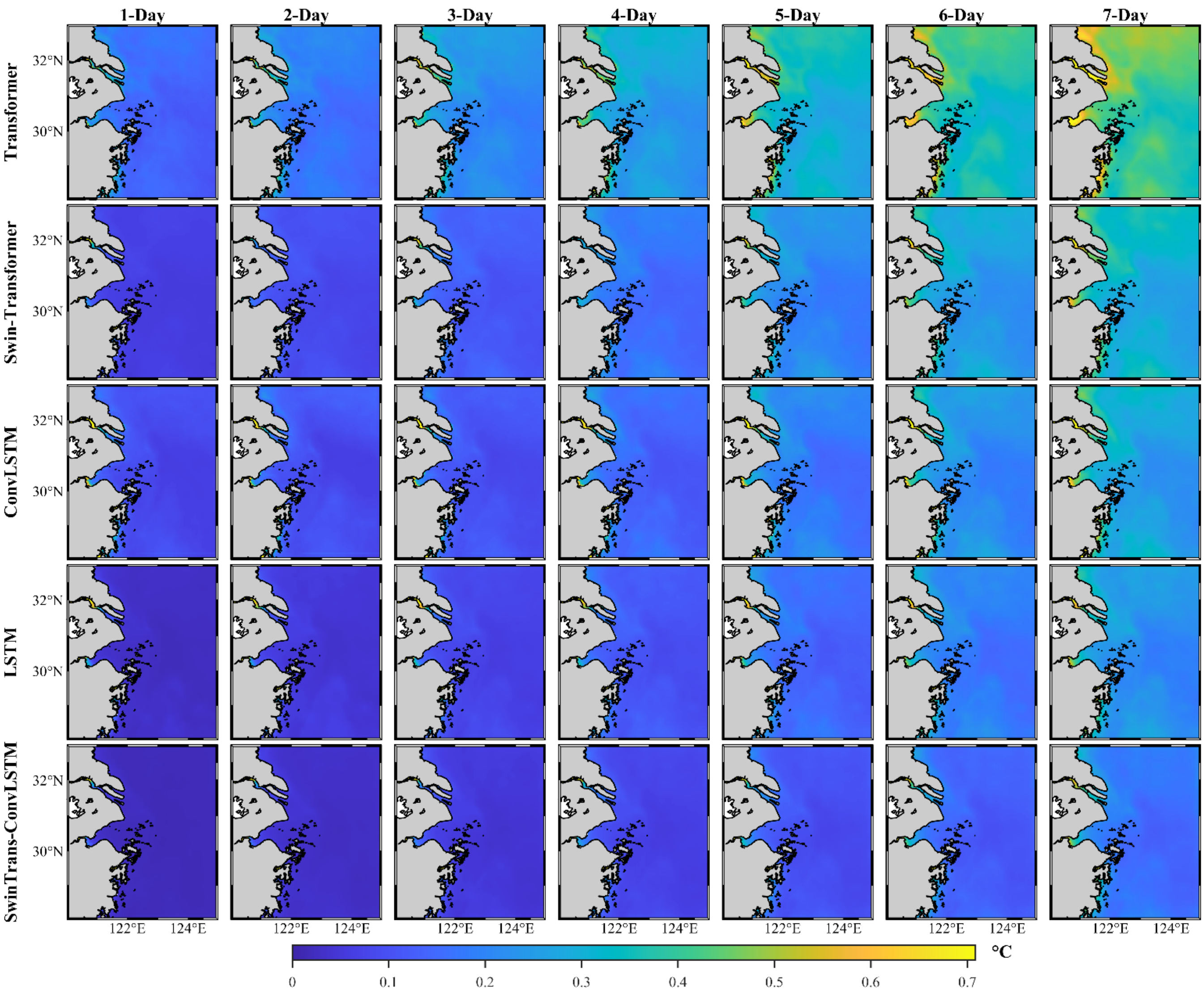



**Figure 7: Spatial distribution of RMSE under different forecast lead times.**

Overall, the results in Figs. 6 and 7 indicate that the SwinTrans-ConvLSTM model achieves better global error performance and maintains improved spatial consistency and long-term forecasting stability in dynamically complex ocean regions.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


The improved performance may be associated with its spatiotemporal feature extraction framework. The inclusion of a gradient constraint helps preserve local temperature gradients and boundary structures, while the incorporation of physical variables such as air–sea temperature difference and vector wind fields provides additional physical information. As a result, the model maintains more coherent SST patterns under long lead times, with high-error regions mainly confined to nearshore areas and reduced offshore error propagation.

### 3.4 Seasonal Performance

To further investigate the seasonal variability of prediction performance, Fig. 8 presents the monthly variations of MAE and RMSE across different models. Overall, all models exhibit clear seasonal patterns in prediction errors, with lower errors during winter (January to March). The SwinTrans-ConvLSTM achieves a minimum MAE of approximately 0.06 °C in this period.

As seasons progresses, prediction errors gradually increase and peak in August and September. Although all models are affected by seasonal variability, the SwinTrans-ConvLSTM consistently maintains the lowest error across all months. In particular, its RMSE peaks at approximately 0.20 °C, which remains lower than that of the benchmark models, indicating improved robustness under seasonal variability.

The seasonal variation in prediction accuracy is closely related to the evolving ocean dynamics in the ECS. In summer, enhanced solar radiation strengthens near-surface stratification, increasing the sensitivity of SST to air–sea heat fluxes and wind forcing (Tan et al., 2023). Meanwhile, the interaction between Changjiang Diluted Water and the Taiwan Warm Current intensifies thermohaline fronts (Hickox et al., 2000), leading to greater SST variability. In contrast, winter monsoonal forcing enhances vertical mixing, and SST variability is mainly governed by large-scale advection, resulting in smoother temporal evolution. These results suggest that the SwinTrans-ConvLSTM model maintains stable performance throughout the year and exhibits strong robustness under complex summer ocean conditions.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026

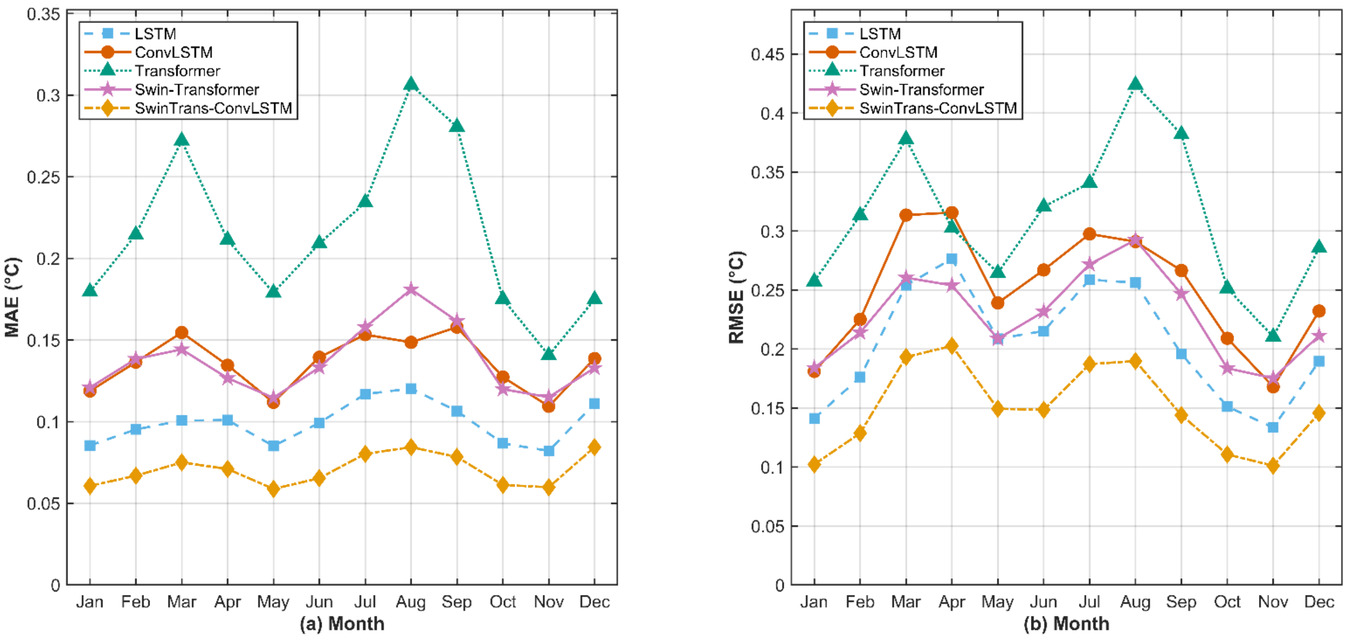




**Figure 8: Monthly variations of prediction errors: (a) MAE and (b) RMSE.**

The SwinTrans-ConvLSTM model consistently maintains the lowest error across all months, with the smallest seasonal variability. Even in August, when errors peak, the RMSE remains approximately 0.25 °C. This improved seasonal performance may be attributed to two key components. First, an air–sea temperature difference variable was constructed based on the COARE framework (Fairall et al., 2003), providing explicit thermodynamic forcing information that helps capture rapid SST variations under strong radiative conditions. Second, an online hard example mining (OHEM) strategy was introduced (Mathieu et al., 2016), which increases the contribution of high-error samples during training and enhances the model's sensitivity to complex summer variability. As a result, these improvements help reduce performance degradation during summer and enhance the model's robustness under strong seasonal forcing.

### 3.5 Marine Heatwave Prediction and Evaluation

An intense and persistent marine heatwave event occurred in the ECS during the summer of 2022, driven by an anomalously strengthened Western Pacific Subtropical High and enhanced air–sea heat exchanges (Tan et al., 2023). The year 2022 is therefore selected as a representative case to evaluate the performance of the SwinTrans-ConvLSTM model in simulating the spatial extent and intensity of MHWs in the ECS. The evaluation is conducted from both temporal evolution and spatial distribution perspectives.

To thoroughly dissect the predictive capability of the proposed approach, we first focus exclusively on the SwinTrans-ConvLSTM model to demonstrate its proficiency in capturing the intricate lifecycle and spatiotemporal dynamics of this extreme event (Figs. 9–12). Following this detailed phenomenological analysis, a systematic evaluation against the four benchmark models is subsequently presented (Fig. 13) to highlight the relative superiority of our approach.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


Figure 9 presents a comparison of mean SST time series in the nearshore region (121°E–122°E, 29°N–30°N) of the ECS during 2022, used to evaluate the model's ability to capture long-duration extreme events. The observed SST persistently exceeds the climatological 90th percentile threshold throughout summer, indicating a marine heatwave lasting approximately two months.

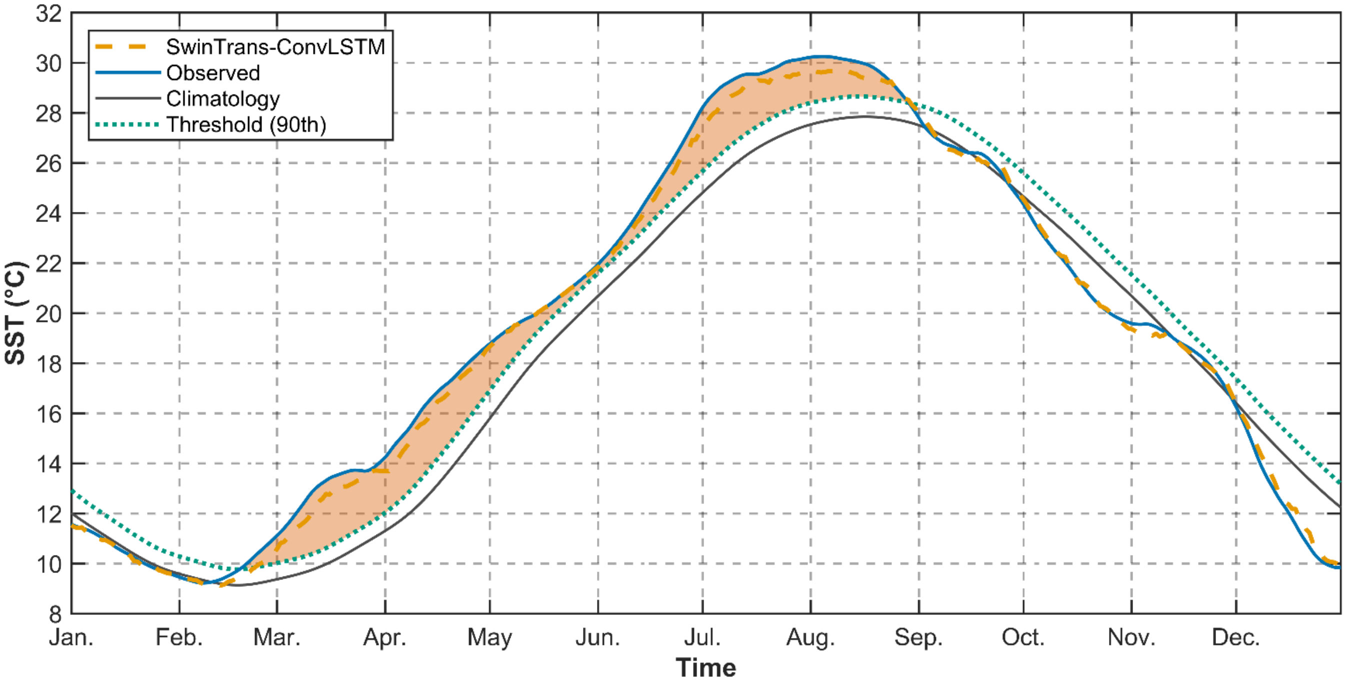



**Figure 9: Lead-day 7 prediction and evaluation of the marine heatwave event in the nearshore waters (121°E–122°E, 29°N–30°N) of the ECS in 2022. The black curve represents the long-term climatological mean, and the green curve denotes the 90th percentile threshold. The blue and orange-yellow curves indicate the observed SST and the SwinTrans-ConvLSTM predicted SST, respectively. The red shaded area highlights the periods identified by the model as extreme marine heatwave conditions.**

The SwinTrans-ConvLSTM predictions show good agreement with observations throughout the full evolution of the event, including onset, peak, and decay phases. In particular, the model captures the extreme temperature peak in mid-August without significant amplitude damping, which is commonly observed in conventional deep learning models. This demonstrates its ability to represent both the intensity and persistence of the marine heatwave.

Figure 10 presents an intensity consistency analysis for different forecast lead times (1, 3, and 7 days). The results show that the SwinTrans-ConvLSTM model maintains good consistency in representing extreme SST values throughout the forecasting period. At 1-day lead time, the coefficient of determination ($R^2$) reaches 0.951 and the Pearson correlation coefficient (PCC) is 0.975, indicating high agreement between predictions and observations.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026

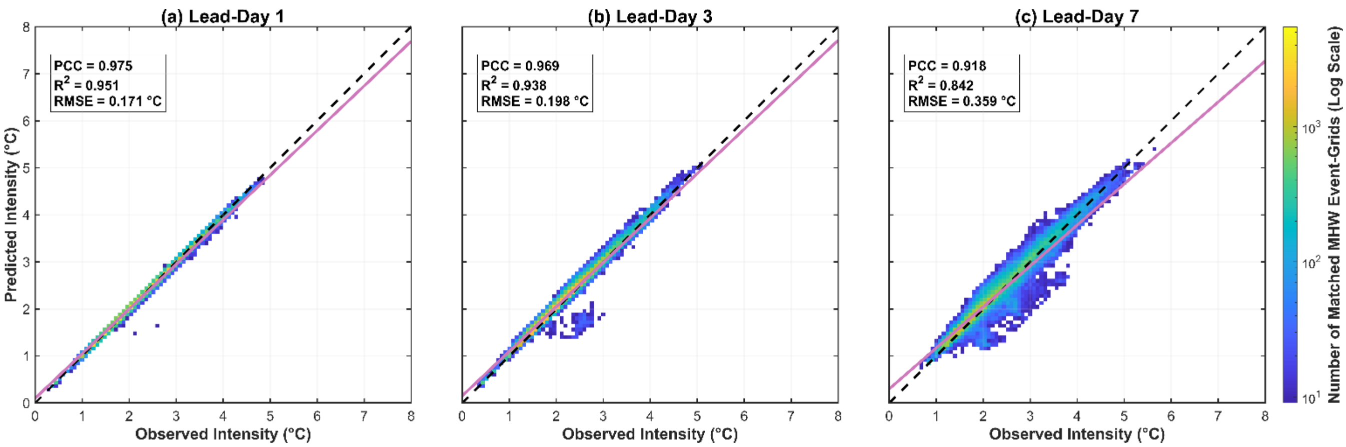




**Figure 10: Consistency analysis of marine heatwave (MHW) intensity across different forecast lead times. Regression distributions between observed and predicted MHW intensities over the entire test set are shown for (a) lead-day 1, (b) lead-day 3, and (c) lead-day 7.**

As lead time increases, model performance gradually decreases, which is consistent with the accumulation of nonlinear forecast errors in ocean dynamics. However, at lead day 7, $R^2$ remains as high as 0.842, and most samples are still concentrated near the reference line, indicating that the model retains reliable predictive skill for marine heatwave intensity at longer lead times.

For higher-intensity MHW events (>4.0 °C), a slight underestimation is observed, reflected by a downward spread of scatter points. This is probably related to increased uncertainty in atmospheric forcing during the forecasting process (Lafferty, 2023). Overall, the proposed model maintains good performance under high-intensity scenarios, suggesting that the incorporation of physical constraints and hard-example mining helps improve extreme-event representation and reduce regression-to-mean effects.

In addition, summer (August) and winter (January) 2022 are selected as representative periods to analyze the spatial distribution of 7-day mean forecasted MHW intensity (lead days 1-7), and the results are compared with observations to further assess the model’s ability to reproduce spatial structures of MHWs.

Figure 11 compares the simulated and observed spatial distributions of mean marine heatwave intensity for 1-7 August 2022, corresponding to a 7-day forecast window. The results show that the SwinTrans-ConvLSTM model successfully reproduces the main spatial patterns of marine heatwave intensity in the ECS at the 7-day lead time.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026

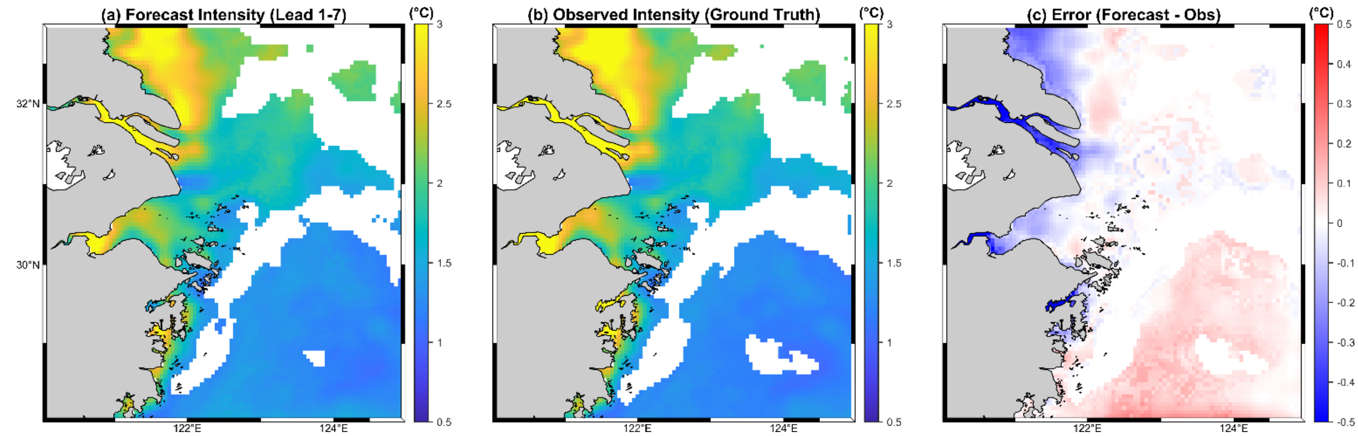




**Figure 11: Spatial comparison of mean intensity of MHWs from 1 to 7 August 2022. (a) Forecast mean intensity from the proposed SwinTrans-ConvLSTM model; (b) observed mean intensity; and (c) error distribution, calculated as predicted intensity minus observed intensity.**

The observed field indicates that the high-intensity core region is mainly located offshore of the Changjiang Estuary and along the Jiangsu coast, where local mean intensity exceeds 2.5 °C. The model captures this spatial structure well and reproduces the overall distribution pattern during this period.

This performance is suggested to be partly associated with the inclusion of a physics-informed gradient loss, which encourages the preservation of spatial gradient information in the SST field and helps reduce excessive smoothing commonly observed in convolution-based models at longer lead times (Mathieu et al., 2016).

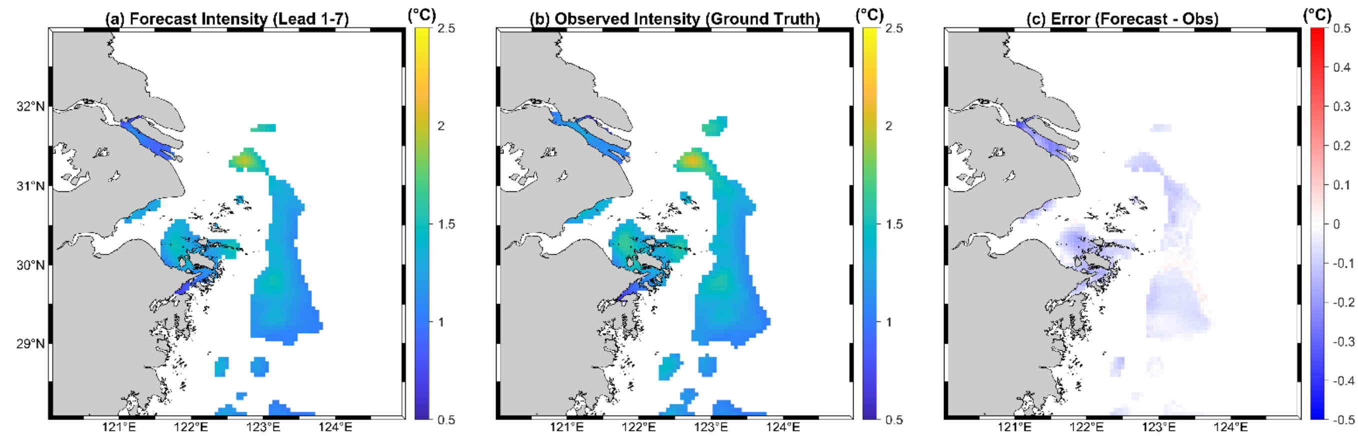



**Figure 12: Spatial comparison of mean intensity of MHWs from 1 to 7 January 2022. (a) Forecast mean intensity from the proposed SwinTrans-ConvLSTM model; (b) observed mean intensity; and (c) error distribution, calculated as predicted intensity minus observed intensity.**

Meanwhile, the marine heatwave event from 1 to 7 January 2022 was selected to further evaluate the model's performance under winter conditions (Fig. 12). Compared with the large-scale and persistent summer events, winter MHWs

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


in the ECS exhibit a more fragmented spatial distribution. The observed high-intensity regions are mainly concentrated along winter thermal frontal zones and near the mouth of Hangzhou Bay around the Zhoushan Archipelago.

For the 1-7 day mean forecast, the SwinTrans-ConvLSTM model captures the general location of the event and reproduces the main spatial pattern of winter MHW intensity. Error analysis indicates relatively better performance under winter conditions than in summer, suggesting reduced dependence on persistent thermodynamic backgrounds.

Winter MHWs in this region are mainly influenced by Kuroshio advection and variations in the winter monsoon (Shu et al., 2025). By incorporating vector wind fields ($u_{10}$ and $v_{10}$), the model can effectively capture changes in wind-driven transport and associated warm-water intrusion pathways, enabling a reasonable reproduction of the dynamical structure of winter MHW events.

Figure 13 presents a comparison between the observed spatial distribution of mean marine heatwave intensity and predictions from five models during 1-7 August 2022. Observations indicate that the event was primarily concentrated in the northern nearshore waters and the Changjiang Estuary, with intensity gradually decreasing toward the southeastern offshore region. None of the baseline models fully captured the observed spatiotemporal structure of the event. The Transformer model significantly underestimated intensity in the northern core region, indicating limited capability in resolving extreme spatial maxima. Although ConvLSTM and Swin Transformer partially reproduced the overall spatial pattern, both exhibited either boundary blurring or local underestimation in high-intensity areas. In contrast, the proposed SwinTrans-ConvLSTM model shows the highest consistency with observations, accurately reproducing the core intensity region as well as the offshore decay gradient. It also achieves the lowest prediction error among all models.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026

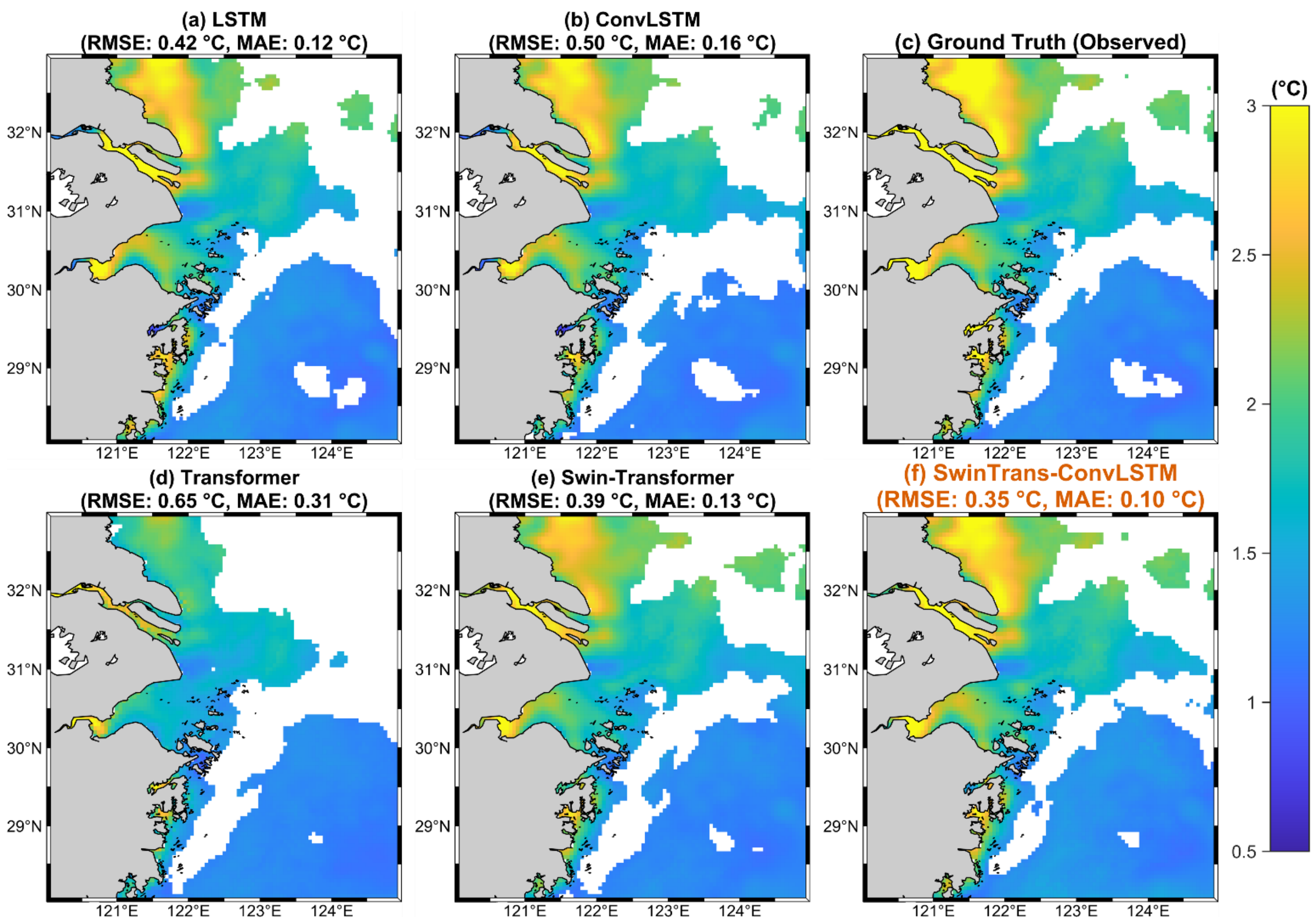




**Figure 13: Spatial comparison of mean MHW intensity predicted by different models from August 1 to 7, 2022: (a) LSTM, (b) ConvLSTM, (c) Gound Truth (Observed), (d) Transformer, (e) Swin-Transformer, and (f) SwinTrans-ConvLSTM.**

These results demonstrate that coupling the global feature extraction capability of the Swin Transformer with the temporal evolution modeling strength of ConvLSTM substantially improves the representation of both spatial distribution and extreme intensity of MHWs. Overall, the proposed model effectively captures the onset, persistence, and decay of extreme events while accurately locating their spatial cores under different oceanic background conditions, confirming its robustness in complex spatiotemporal environments.

## 4 Discussion

The SwinTrans-ConvLSTM model developed in this study demonstrates superior performance in sea surface temperature (SST) prediction and marine heatwave (MHW) early warning in the ECS. Here, we integrate statistical evaluations with case

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


analyses of extreme climate events to examine the model's advantages in MHW prediction, elucidate the mechanisms underlying its performance gains, and discuss current limitations and future directions for improvement.

### 4.1 Predictive Gains from Physics-Informed Variables

As shown in Table 5, an ablation study was systematically conducted to quantify the contributions of different physical variable configurations to the predictive performance of the SwinTrans-ConvLSTM model and to validate the effectiveness of the proposed variable selection strategy. To ensure a fair comparison, all model architectures and training parameters were maintained strictly identical throughout the experiments.

First, inspired by bulk aerodynamic algorithms (Fairall et al., 2003), we utilized the air-sea temperature difference ($\Delta T$) as a robust proxy for the thermodynamic gradients driving sensible and latent heat exchanges. This constraint alone improves MAE by 3.6 % (Control group). Second, incorporating wind vectors ($u_{10}, v_{10}$) introduces critical directional momentum flux, enabling the network to implicitly capture wind-driven kinematic processes such as Ekman transport, advection, and mixing (Chelton and Xie, 2010). This dynamic constraint yields a 5.9 % MAE improvement.

Crucially, the synergistic combination of thermodynamic ($\Delta T$) and dynamic (Wind) variables within the Experimental group yields the most substantial error reduction (12.0 %, 0.0714 °C). Consequently, this final variable combination was adopted as the universal input; all model performances discussed in Sect. 3 are based on this setup.

Furthermore, applying a 3-day lagged moving average to these inputs acts as a mathematical analogue to the upper ocean's thermal inertia, filtering high-frequency atmospheric noise while embedding the mixed layer's short-term memory. These findings confirm that for extreme SST prediction, physics-informed feature design is fundamentally more robust than the indiscriminate use of high-dimensional reanalysis datasets.

**Table 5: Model experimental results under different input variable combinations.**

| Group | Input variables | MAE (°C) | RMSE (°C) | MAE improvement |
|---|---|---|---|---|
| Baseline | SST | 0.0811 | 0.1590 | - |
| Control | SST+$\Delta T$ | 0.0782 | 0.1593 | +3.6% |
| Control | SST+ Wind | 0.0763 | 0.1660 | +5.9% |
| Experimental | SST+$\boldsymbol{\Delta T}$+ Wind | 0.0714 | 0.1555 | +12.0% |

### 4.2 Predictive Performance for Marine Heatwave Events

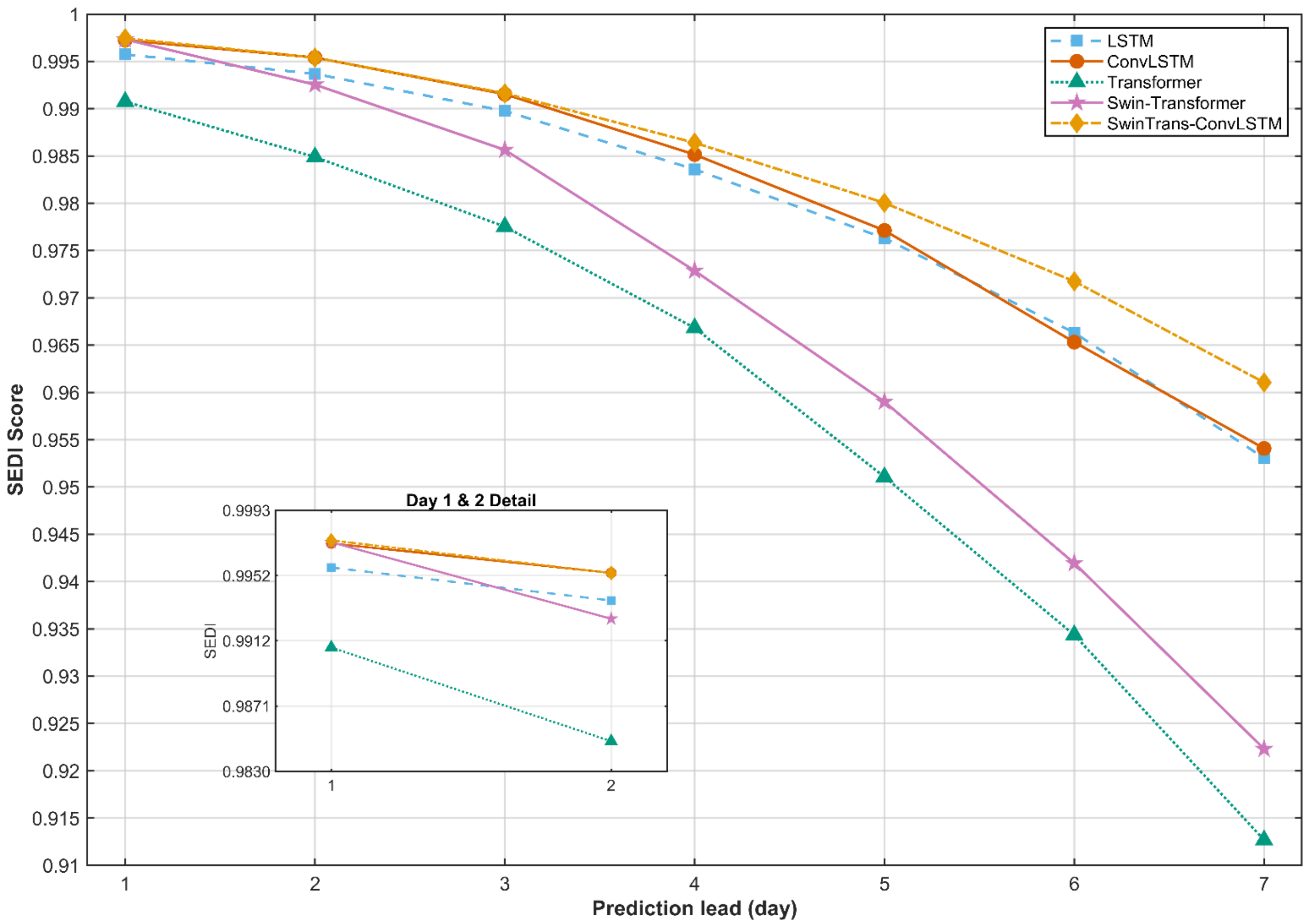
As forecast lead time increases, data-driven models commonly suffer from a structural limitation known as regression to the mean, particularly when predicting extreme events. MHWs are inherently imbalanced within climatological distributions; therefore, models optimized using global mean squared error (MSE) tend to be dominated by abundant non-extreme samples.

This bias leads to systematic underestimation and excessive smoothing when encountering high-temperature anomalies that exceed historical thresholds.

To quantitatively assess the model's ability to capture extreme events, this study employs the symmetric extremal dependence index (SEDI) (Welandawe et al., 2025). The evolution of SEDI across the 1- to 7-day forecast window is shown in Fig. 14.

The temporal evolution of SEDI over the 1-7 day forecast horizon is presented in Fig. 14. All models exhibit strong initial performance at short lead times, benefiting from the persistence of oceanic thermal inertia within the upper mixed layer. However, predictive skill deteriorates as lead time increases due to the accumulation of uncertainty in atmospheric forcing and boundary conditions. Among the compared architectures, the pure self-attention model shows the most rapid degradation, with a sharp decline in SEDI after day 5. This indicates that global attention mechanisms, without explicit physical constraints, are susceptible to error accumulation and instability in longer-term forecasts (Dai et al., 2024). In contrast, the proposed SwinTrans-ConvLSTM model demonstrates consistently superior performance, maintaining a relatively slow decay in SEDI and remaining above 0.95 at day 7. This suggests enhanced robustness in preserving extreme-event predictability over extended forecast horizons.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


535

**Figure 14: Comparison of SEDI variations in marine heatwave forecasts across different lead times. The test period begins on 1 January 2020, with forecast lead times ranging from 1 to 7 days. The inset shows an enlarged view of lead days 1-2.**

The evolution of SEDI shown in Fig. 14 reveals fundamental differences among network architectures in terms of physical-memory decay. During the initial forecast stage, all models exhibited a high level of extreme-event detection skill, benefiting from the strong thermal inertia of the oceanic upper mixed layer. As the forecast lead time increased, uncertainties associated with external forcing and boundary conditions accumulated rapidly. The pure self-attention architecture showed the most pronounced degradation, with its SEDI curve dropping sharply after day 5. This result confirms that a global attention mechanism alone, in the absence of local hydrodynamic constraints, is prone to cascading error amplification during longer-term forecasting (Dai et al., 2024).

In contrast, the proposed SwinTrans-ConvLSTM model demonstrates consistently superior performance, maintaining a relatively slow decay in SEDI and remaining above 0.95 at day 7. This suggests enhanced robustness in preserving extreme-event predictability over extended forecast horizons. The improved performance can be attributed to the incorporation of physics-informed variables and training strategy. The air-sea temperature difference and vector wind components introduce

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


explicit constraints related to surface energy exchange and wind-driven transport, thereby stabilizing the representation of thermodynamic forcing. In addition, the online hard example mining (OHEM) strategy dynamically reweights training toward rare and high-gradient samples, improving the model's sensitivity to extreme conditions (Shrivastava et al., 2016). Together, these mechanisms effectively mitigate the tendency toward extreme-value underestimation commonly observed in standard data-driven models, enhancing their reliability for extreme SST prediction.

### 4.3 Limitations and Future Perspectives

Although the SwinTrans-ConvLSTM model achieves substantial improvements across multiple evaluation metrics, several limitations remain in its application to complex coastal hydrographic environments, providing important directions for future work.

First, relatively large prediction errors are observed in the extremely shallow nearshore waters of the Changjiang Estuary, where water depths are less than 10 m. These errors are associated with strong land contamination signals, highly variable river discharge, and intense tidal mixing. However, the current regular-grid architecture cannot fully conform to irregular coastal geometries, limiting its ability to represent complex boundary-driven dynamics. Future studies may therefore consider unstructured-grid representations, such as graph neural networks, to better capture irregular coastal boundaries and bathymetry-induced perturbations in water-mass transport (Liang et al., 2023).

Second, although ERA5 atmospheric forcing fields are spatially aligned with ocean variables through preprocessing, bilinear interpolation may smooth sub-grid-scale atmospheric variability. This introduces a scale mismatch between atmospheric forcing and high-resolution ocean responses, which may constrain predictive skill in nearshore regions.

Finally, the current framework primarily relies on two-dimensional sea surface information and does not explicitly incorporate subsurface ocean structure. MHWs are inherently three-dimensional processes, where vertical mixing and thermocline variability play a critical role in heat storage and release (Feng et al., 2025). The absence of subsurface dynamical information may therefore limit predictive performance under conditions of strong vertical exchange, such as typhoon forcing or upwelling events.

Future work will focus on integrating multi-source satellite observations with Argo float profiles to construct a quasi-three-dimensional representation of the ocean state, thereby improving the model's ability to capture vertical thermodynamic processes and extreme-event evolution.

## 5 Conclusions

This study proposed a hybrid spatiotemporal deep learning framework, SwinTrans-ConvLSTM, which integrates physics-informed variable reconstruction with a curriculum-learning-based training strategy. Based on multi-source data integration and comparative experiments, the main findings are summarized as follows:

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


The proposed SwinTrans-ConvLSTM model achieves highly accurate SST prediction on the test dataset, with an MAE of 0.0714 °C and an RMSE of 0.1555 °C, corresponding to a 15.5 %–47.8 % reduction in error compared with LSTM and Transformer baselines. The use of reconstructed physics-informed variables, together with a 3-day lagged moving average representing ocean thermal inertia, effectively reduces the influence of high-frequency noise from raw reanalysis inputs and improves robustness under complex coastal dynamics.

The hybrid architecture adopted in the SwinTrans-ConvLSTM model enabled deep coupling between the global representation capability of the Swin Transformer and the local temporal memory of ConvLSTM. Combined with gradient-based constraints and curriculum learning, the model mitigates the tendency of standard deep learning approaches to regress toward mean states and underestimate extremes. In the 2022 summer marine heatwave case, the model accurately reproduced the temporal evolution, intensity, duration, and spatial extent of extreme SST events. Furthermore, the online hard example mining (OHEM) strategy enhanced the learning of rare extreme samples, and the symmetric extremal dependence index (SEDI) remained consistently above 0.95 over a 7-day forecast horizon, demonstrating strong predictive stability for extreme events.

In summary, this study provides a high-accuracy SST prediction framework for the ECS and highlights the importance of physics-guided variable design in improving deep learning performance for marine heatwave prediction. The proposed approach offers a promising tool for supporting marine disaster early warning, fisheries management, and coastal ecosystem protection.

**Code and data availability**

All datasets used in this study are publicly available and properly cited in the text. The OSTIA SST dataset is available from the Copernicus Marine Service (CMEMS), and the ERA5 atmospheric reanalysis data is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The source code and minimal reproducible examples for the SwinTrans-ConvLSTM model are archived on GitHub (https://github.com/ZefangMa/MHW-Prediction-SwinTrans).

**Author contributions**

ZM developed the project conceptualization and methodology, wrote the software, curated the dataset, conducted the formal analysis and validation, and produced the visualizations. HC and QJ provided supervision, managed project administration, and contributed to the formal analysis. ZM prepared the original manuscript. HC, QJ, LJ, SC, and XL contributed to the review and editing of the manuscript.

## Competing interests

The contact author has declared that none of the authors has any competing interests.

## Disclaimer

Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

## Acknowledgements

The authors would like to acknowledge the Copernicus Marine Environment Monitoring Service (CMEMS) for providing the OSTIA sea surface temperature dataset, and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA5 atmospheric reanalysis data used in this study.

## Financial support

This research has been supported by the National Key Research and Development Program of China (Grant No. 2023YFD2401904).

## References

Bonino, G., Galimberti, G., Masina, S., McAdam, R., and Clementi, E.: Machine learning methods to predict sea surface temperature and marine heatwave occurrence: a case study of the mediterranean sea, Ocean Sci., 20, 417–432, https://doi.org/10.5194/os-20-417-2024, 2024.

Chelton, D. and Xie, S.-P.: Coupled ocean-atmosphere interaction at oceanic mesoscales, Oceanog., 23, 52–69, https://doi.org/10.5670/oceanog.2010.05, 2010.

Dai, H., He, Z., Wei, G., Lei, F., Zhang, X., Zhang, W., and Shang, S.: Long-term prediction of sea surface temperature by temporal embedding transformer with attention distilling and partial stacked connection, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 17, 4280–4293, https://doi.org/10.1109/JSTARS.2024.3357191, 2024.

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., and Edson, J. B.: Bulk parameterization of air–sea fluxes: updates and verification for the COARE algorithm, J. Climate, 16, 571–591, https://doi.org/10.1175/1520-0442(2003)016%3C0571:BPOASF%3E2.0.CO;2, 2003.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


Feng, Z., Qi, J., Li, D., Xie, B., Sun, G., Yin, B., and Yang, S.: Attention-enhanced deep learning model for reconstruction and downscaling of thermocline depth in the tropical indian ocean, Ocean Modelling, 196, 102537, https://doi.org/10.1016/j.ocemod.2025.102537, 2025.

Flint, L. E. and Flint, A. L.: Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis, Ecol Process, 1, 2, https://doi.org/10.1186/2192-1709-1-2, 2012.

Giamalaki, K., Beaulieu, C., and Prochaska, J. X.: Assessing predictability of marine heatwaves with random forests, Geophysical Research Letters, 49, e2022GL099069, https://doi.org/10.1029/2022GL099069, 2022.

Hao, P., Li, S., Song, J., and Gao, Y.: Prediction of sea surface temperature in the south china sea based on deep learning, Remote Sensing, 15, 1656, https://doi.org/10.3390/rs15061656, 2023.

Hickox, R., Belkin, I., Cornillon, P., and Shan, Z.: Climatology and seasonal variability of ocean fronts in the east china, yellow and bohai seas from satellite SST data, Geophysical Research Letters, 27, 2945–2948, https://doi.org/10.1029/1999GL011223, 2000.

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Feng, M., Holbrook, N. J., Moore, P. J., Scannell, H. A., Sen Gupta, A., and Wernberg, T.: A hierarchical approach to defining marine heatwaves, Progress in Oceanography, 141, 227–238, https://doi.org/10.1016/j.pocean.2015.12.014, 2016.

Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Alexander, L. V., Burrows, M. T., Donat, M. G., Hobday, A. J., Moore, P. J., Perkins-Kirkpatrick, S. E., Smale, D. A., Straub, S. C., and Wernberg, T.: A global assessment of marine heatwaves and their drivers, Nat Commun, 10, 2624, https://doi.org/10.1038/s41467-019-10206-z, 2019.

Huber, P. J.: Robust Estimation of a Location Parameter, The Annals of Mathematical Statistics, 35, 73–101, https://doi.org/10.1214/aoms/1177703732, 1964.

Lafferty, D. C.: Downscaling and bias-correction contribute considerable uncertainty to local climate projections in CMIP6, 2023.

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, https://doi.org/10.3390/rs15143539, 2023.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B.: Swin transformer: hierarchical vision transformer using shifted windows, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9992–10002, https://doi.org/10.1109/ICCV48922.2021.00986, 2021.

Mallat, S. G.: A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Machine Intell., 11, 674–693, https://doi.org/10.1109/34.192463, 1989.

Mathieu, M., Couprie, C., and LeCun, Y.: Deep multi-scale video prediction beyond mean square error, https://doi.org/10.48550/arXiv.1511.05440, 26 February 2016.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., Benthuysen, J. A., Feng, M., Sen Gupta, A., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Straub, S. C., and Wernberg, T.: Longer and more frequent marine heatwaves over the past century, Nat Commun, 9, 1324, https://doi.org/10.1038/s41467-018-03732-9, 2018.

Parasyris, A., Metheniti, V., Kampanis, N., and Darmaraki, S.: Marine heatwaves in the mediterranean sea: a convolutional neural network study for extreme event prediction, https://doi.org/10.5194/egusphere-2024-4003, 9 January 2025.

Sasaki, Y. N. and Umeda, C.: Rapid warming of sea surface temperature along the kuroshio and the china coast in the east china sea during the twentieth century, Journal of Climate, 34, 4803–4815, https://doi.org/10.1175/JCLI-D-20-0421.1, 2021.

Shi, B., Ge, C., Lin, H., Xu, Y., Tan, Q., Peng, Y., and He, H.: Sea surface temperature prediction using ConvLSTM-based model with deformable attention, Remote Sensing, 16, 4126, https://doi.org/10.3390/rs16224126, 2024.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., and Woo, W.: Convolutional LSTM network: a machine learning approach for precipitation nowcasting, https://doi.org/10.48550/arXiv.1506.04214, 19 September 2015.

Shrivastava, A., Gupta, A., and Girshick, R.: Training region-based object detectors with online hard example mining, https://doi.org/10.48550/arXiv.1604.03540, 12 April 2016.

Shu, R., Wu, H., Gao, Y., Xu, F., Gou, R., Xiong, W., and Huang, X.: Advanced forecasts of global extreme marine heatwaves through a physics-guided data-driven approach, Environ. Res. Lett., 20, 044030, https://doi.org/10.1088/1748-9326/adbddd, 2025.

Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., Burrows, M. T., Alexander, L. V., Benthuysen, J. A., Donat, M. G., Feng, M., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Sen Gupta, A., Payne, B. L., and Moore, P. J.: Marine heatwaves threaten global biodiversity and the provision of ecosystem services, Nat. Clim. Chang., 9, 306–312, https://doi.org/10.1038/s41558-019-0412-1, 2019.

Song, J., Zhang, X., and Jiang, W.: Sea surface temperature prediction in china sea based on SAM-LSTM approach, AJESE, 8, 14–22, https://doi.org/10.11648/j.ajese.20240802.11, 2024.

Sun, D., Jing, Z., and Liu, H.: Deep learning improves sub-seasonal marine heatwave forecast, Environ. Res. Lett., 19, 064035, https://doi.org/10.1088/1748-9326/ad4616, 2024.

Sun, W., Zhou, S., Yang, J., Gao, X., Ji, J., and Dong, C.: Artificial intelligence forecasting of marine heatwaves in the south china sea using a combined U-net and ConvLSTM system, Remote Sensing, 15, 4068, https://doi.org/10.3390/rs15164068, 2023.

Tan, H.-J., Cai, R.-S., Bai, D.-P., Hilmi, K., and Tonbol, K.: Causes of 2022 summer marine heatwave in the east china seas, Advances in Climate Change Research, 14, 633–641, https://doi.org/10.1016/j.accre.2023.08.010, 2023.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: Attention is all you need, https://doi.org/10.48550/arXiv.1706.03762, 12 June 2017.

https://doi.org/10.5194/egusphere-2026-4434
Preprint. Discussion started: 30 July 2026


Welandawe, S., Priyadarshana, Y. H. P. P., Senanayake, N., and Silva, E. N. S.: Machine learning techniques for marine heatwave prediction: a comprehensive review, Intell. Mar. Technol. Syst., 3, 28, https://doi.org/10.1007/s44295-025-00076-1, 2025.

Xiang, L., Guan, J., Xiang, J., Zhang, L., and Zhang, F.: Spatiotemporal model based on transformer for bias correction and temporal downscaling of forecasts, Front. Environ. Sci., 10, 1039764, https://doi.org/10.3389/fenvs.2022.1039764, 2022.

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., Xu, L., Chen, Z., and Gong, J.: 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, 2019.

Xie, B., Qi, J., Yang, S., Sun, G., Feng, Z., Yin, B., and Wang, W.: Sea surface temperature and marine heat wave predictions in the south china sea: a 3D U-net deep learning model integrating multi-source data, Atmosphere, 15, 86, https://doi.org/10.3390/atmos15010086, 2024.

Xu, W., Guo, B., Mantravadi, V. S., Xu, Z., Wan, C., Sabuyi, J. S., and Xu, Z.: Spatial–temporal characteristics and trend analysis of marine heatwaves in the east china sea based on deep learning, Water, 17, 3076, https://doi.org/10.3390/w17213076, 2025.

Zhang, Q., Wang, H., Dong, J., Zhong, G., and Sun, X.: Prediction of sea surface temperature using long short-term memory, IEEE Geoscience and Remote Sensing Letters, 14, 1745–1749, https://doi.org/10.1109/LGRS.2017.2733548, 2017.

Zhao, B., Zhang, T., Wang, Y., Lin, P., Liu, H., Huang, P., Huang, W., Wang, P., and Li, Y.: Combined LFS and ConvLSTM to forecast marine heatwaves: a case study, Atmospheric and Oceanic Science Letters, 19, 100690, https://doi.org/10.1016/j.aosl.2025.100690, 2026.

Zhao, X., Qi, J., Yu, Y., and Zhou, L.: Deep learning for ocean temperature forecasting: a survey, Intell. Mar. Technol. Syst., 2, 28, https://doi.org/10.1007/s44295-024-00042-3, 2024.