TPRformer: A Sea Surface Temperature Prediction Method Based on Temporal Periodic Feature Mining and Reconstruction Transformer

1st Jiabao Zhang
College of Meteorology and Oceanology
National University of Defense Technology
ChangSha, China
zhangjiabao23@nudt.edu.cn

2nd Qian Li
College of Meteorology and Oceanology
National University of Defense Technology
ChangSha, China
public_liqian@163.com

3rd Bing Sui
Institute of Meteorological Sciences
Hunan Province
ChangSha, China
sui_bing@163.com

4th Yanwen Wang
College of Meteorology and Oceanology
National University of Defense Technology
ChangSha, China
wangyw@nudt.edu.cn

5th Sheng Li
College of Meteorology and Oceanology
National University of Defense Technology
ChangSha, China
lisheng24@nudt.edu.cn

***Abstract*—Sea Surface Temperature (SST) is a key factor in ocean-atmosphere interactions, which exerts significant influence on climate, ecology and maritime activities. Existing methods are limited to the modeling and optimization of long-term temporal dependence, and fail to adequately capture the complex temporal patterns in SST sequences, which are formed by the combined effects of long-term periodic trends and short-term non-stationary variations, resulting in suboptimal performance for long-term SST prediction tasks. To address this issue, we propose a temporal periodic feature mining and reconstruction transformer (TPRformer). In TPRformer, the encoder employs a temporal decomposition block to separate the historical sequence into periodic and non-stationary components, with a focus on mining long-term temporal dependencies within the periodic component to construct periodic features. The decoder performs the temporal decomposition on the SST sequence, employs temporal cross-attention to dynamically align historical periodic features, and ultimately achieves SST prediction by reconstructing the two decomposed components. Experimental results on multiple SST datasets demonstrate that TPRformer is superior to state-of-the-art baseline methods across multiple forecasting horizons.**

***Index Terms*—SST Prediction, Transformer Networks, Temporal Decomposition, Long-Term Time Series Forecasting, Physics-Inspired Neural Networks**

I. Introduction

The variations of Sea Surface Temperature (SST) can significantly drive adjustments in atmospheric circulation and weather patterns [1], playing an irreplaceable and central role in the accurate prediction of key climate phenomena such as El Niño and La Niña [2]. However, SST prediction is regulated by multiple factors including solar radiation, land-sea thermal contrast, and ocean currents [3]–[5]. The complex coupling and feedback mechanisms among these factors pose substantial challenges to the accurate long-term forecasting of SST.

Existing SST prediction methods can be categorized into physically-driven and data-driven models. Physically-driven models are primarily based on principles of fluid dynamics and thermodynamics [6], achieving SST prediction by simulating ocean physical processes. For example, the Regional Ocean Modeling System (ROMS) [7] employs finite difference or finite volume methods for spatial discretization, combined with explicit time integration, to simulate coastal dynamics and SST variations. Although physically-driven models offer strong interpretability in simulating ocean physics, they suffer from limitations such as high computational complexity, strong parameter sensitivity, and a reliance on extensive prior knowledge [8], [9].

In contrast, data-driven models do not require explicit modeling of complex physical processes; instead, they learn the spatio-temporal patterns of SST directly from observational data, offering advantages such as lower computational cost and greater flexibility. This category encompasses both traditional machine learning and deep learning techniques. Traditional machine learning methods model time series data using statistical models to extract temporal dependencies from historical data for the prediction of SST. For example, Mahongo et al. [10] employed linear regression to model the temporal dependencies of SST, while Lins et al. [11] used Support Vector Machines (SVM) to capture the temporal relationships of SST. However, these methods generally lack a direct mechanism for modeling long-term temporal dependencies, often rely on manually designed temporal features, and struggle to adapt to the nonlinearity and long-memory characteristics of long SST sequences, which consequently limits their accuracy in long-term forecasting [12].

Deep learning has overcome the limitations of traditional machine learning through gating mechanisms (e.g., LSTM [13], GRU [14]) and self-attention mechanisms (e.g., Transformer [15]). These approaches can directly model long-term temporal dependencies without manually engineered temporal features, effectively mitigating the issue of inadequate long-term dependency capture. For example, Zhang et al. [16] employed fully connected LSTMs to capture temporal dependencies in SST data, while Yang et al. [17] proposed the CFCC-LSTM model to extract spatiotemporal features for the prediction of SST. Gao et al. [18] proposed a Transformer-based correction method for error correction in the daily SST forecast products of ocean numerical prediction systems, which significantly reduced the root mean square error of 2-5-day SST forecasts in the Northwest Pacific region.

However, existing methods primarily rely on modeling single long-term temporal dependencies and fail to adequately account for the complex temporal patterns in SST sequences, which are formed by the synergistic interaction of periodic and non-stationary trends. Specifically, the periodic trend refers to the long-term evolving pattern of SST corresponding to seasonal changes, typically with a one-year cycle, driven primarily by seasonal variations in solar radiation distribution and seasonal adjustments in atmospheric circulation. This trend exhibits stability and repeatability, manifesting itself as periodic variations with higher SST in summer and lower SST in winter, which directly reflects the seasonal interactions within the ocean-atmosphere system.

The non-stationary trend, in contrast, refers to the short-term fluctuating characteristics where the statistical properties of SST change continuously over time, making them impossible to describe with a fixed statistical model in the long term. It arises from the combined effects of long-term trends, such as global warming, and various fluctuating anomalies, including localized small-scale and large-scale anomalies. This is reflected not only in the dynamic changes of mean and variance, but also in the interactive amplification between trends and anomalies. In summary, we characterize SST as having a complex trend comprising both periodic and non-stationary components.

Based on the above analysis, we propose a sea surface temperature (SST) forecasting model called the Temporal Periodic feature mining and Reconstruction Transformer (TPRformer). Guided by the aforementioned physical mechanisms, this model decomposes the input sequence into periodic and non-stationary components, captures long-term temporal dependencies within the periodic component, and reconstructs the prediction by fusing both components to obtain accurate SST forecasts. Specifically, the encoder first processes the SST sequence through a temporal enhancement block, which supplements positional information and captures global temporal correlations. It then employs a temporal decomposition mechanism to decompose the input historical SST sequence into a periodic component and a non-stationary component, focusing on mining the long-term temporal dependencies within the periodic component to construct a structured periodic feature. The decoder takes the historical SST sequence and the output characteristics of the encoder as input, decomposes the non-stationary components of the sequence, fuses the historical periodic features output by the encoder with the help of the cross-attention mechanism, and finally integrates the periodic features and the non-stationary features to complete the sequence reconstruction and realize the SST prediction.

The contributions of this paper are as follows.

- We propose a temporal decomposition block which separates the input SST sequence into periodic and non-stationary components, corresponding to the long-term temporal dependency features and short-term non-stationary dependency features of the sequence, respectively.
- We construct an SST time series prediction model named TPRformer. The encoder of TPRformer extracts core periodic features from historical SST sequences, and the decoder incorporates non-stationary features from the target sequence while fusing the historical periodic features output by the encoder, ultimately accomplishing SST prediction through feature reconstruction.
- Experiments conducted on observed datasets from multiple sea areas demonstrate that the proposed SST forecasting method outperforms existing state-of-the-art approaches.

II. Related Work

Deep learning has achieved successes in addressing a series of spatio-temporal challenges [19]–[22], and its application in SST prediction can be categorized into two groups: models that focus on temporal features and models that focus on spatial features.

The first category of methods primarily emphasizes feature extraction within the temporal domain. Hao et al. [23] proposed TransDtSt-Part, a Transformer-based model which incorporates temporal embedding, attention mechanisms, and stacking connections to improve SST prediction accuracy. Hou et al. [24] explored the spatio-temporal properties within SST data and employed an inter-scale fusion technique to integrate correlations across the temporal dimension. Agarwal et al. [25] utilized wavelet-based climate networks to characterize correlations among SST time series at different scales, thereby facilitating a deeper understanding of the related patterns.

The second category of methods focuses on intra-spatial feature extraction. Sun et al. [26] employed Graph Convolutional Networks (GCNs) to identify spatial dependencies in SST data, combined with LSTMs for time series analysis. Liang et al. [27] designed a graph-based representation for SST and integrated it with LSTMs, proposing a graph memory neural network approach. Yang et al. [28] combined adaptive node embedding with GCNs to capture local SST patterns and global spatial dependencies, introducing a novel perspective for SST prediction. Ren et al. [29] developed a UNet-based model which leverages parallel convolutional layers with different kernel sizes for spatial feature extraction and employs LSTMs for temporal modeling.

Although deep learning models have demonstrated remarkable capability in modeling underlying patterns and relationships to predict future SST variations, a significant gap remains in adequately modeling the temporal dependencies of SST within these models.

III. Preliminaries

We formulate the Sea Surface Temperature (SST) prediction problem as a spatio-temporal sequence forecasting task. Given a historical SST sequence $X_{\tau:t}^{(N)} = \{x_{\tau+1}^{(N)}, x_{\tau+2}^{(N)}, x_{\tau+3}^{(N)}, \ldots, x_{t}^{(N)}\}$, where X denotes the input sequence, $\tau : t$ represents the input time span $t - \tau$, and N indicates the number of SST observation stations, the objective of the model F is to predict the future SST sequence for the next p time steps, i.e., $\hat{X}_{t:t+p}^{(N)} = \{\hat{x}_{t+1}^{(N)}, \hat{x}_{t+2}^{(N)}, \hat{x}_{t+3}^{(N)}, \ldots, \hat{x}_{t+p}^{(N)}\}$, where $\hat{X}$ denotes the output sequence and $t : t+p$ represents the prediction horizon P. This process can be formally expressed as:

$$\hat{X}_{t:t+p}^{(N)} = f(X_{\tau:t}^{(N)}; \theta) \in \mathbb{R}^{P\times N\times D} \tag{1}$$

where the output $\hat{X}$ has a dimension of $P \times N \times D$, in which D corresponds to the feature dimension of the nodes. In NN models, the time ranges for both $X_{\tau:t}^{(N)}$ and $\hat{X}_{t:t+p}^{(N)}$ are temporally resolved with "day" as the unit. For example, the historical period can be set to the past 30 days, and the forecast period to the next 7 days. Furthermore, the temporal resolution of $X_{\tau:t}^{(N)}$ and $\hat{X}_{t:t+p}^{(N)}$ remains consistent.

IV. METHOD

A. Overview

To address the challenge of accurately capturing complex temporal dependencies in SST prediction, we propose a temporal periodic feature mining and reconstruction transformer (TPRformer). As illustrated in Fig. 1, TPRformer consists of encoder and decoder, the primary blocks of which are the temporal enhancement block (TEB), temporal decomposition (Temporal Decomp) block and temporal cross-attention block (TCA).

The encoder takes the historical SST sequence $X \in \mathbb{R}^{(L\times D)}$ as input, where L denotes the time step length and D denotes the feature dimension, and adopts a multi-layer architecture. In the $l-1$-th encoder, the TEB first receives X_{en}^{l-1}, performs position embedding and temporal autocorrelation computation, and outputs enhanced features $\hat{X}_{en}^{l-1}$, which are used to supplement temporal position information and capture global dependencies. Subsequently, the Temporal Decomp block performs multi-scale filtering and adaptive fusion on $\hat{X}_{en}^{l-1}$, separating it into the non-stationary component and the periodic component $\hat{X}_{en}^{l-1,1}$, thereby achieving trend decomposition and focusing on periodic information. Finally, the Temporal Decomp block refines $\hat{X}_{en}^{l-1,1}$ processed by the feedforward neural network(FFD), generating the high-order periodic feature $\hat{X}_{en}^{l-1,2}$, which is used to preserve the core periodic patterns.

The decoder takes the historical sequence $X \in \mathbb{R}^{(L\times D)}$ and the encoder output X_{en}^{l} as inputs, and adopts a multi-layer architecture. In the $l-1$-th decoder, the TEB and Temporal Decomp block process the input data X_{de}^{l-1}, decomposing it into a periodic component $\hat{X}_{de}^{l-1,1}$ and a non-stationary component. The non-stationary component is retained for the final sequence reconstruction. The TCA block then takes the periodic component $\hat{X}_{de}^{l-1,1}$ and the output of the $l-1$-th encoder $\hat{X}_{en}^{l}TEB$ as input, performs a cross-attention computation to associate historical periodic information, and outputs the fused features. The second Temporal Decomposition block refines the fused features to further enhance the core periodic characteristics while retaining the non-stationary component for sequence reconstruction. Subsequently, the Temporal Decomposition block processes the FFD-refined output $\hat{X}_{de}^{l-1,2}$, producing the refined periodic feature $\hat{X}_{de}^{l-1,3}$ and continuing to preserve the non-stationary component. Finally, the output periodic feature $\hat{X}_{de}^{l-1,3}$ is integrated with the aforementioned non-stationary component to generate the SST prediction.

B. Temporal Enhancement Block

The temporal enhancement block (TEB) adopts a temporal autocorrelation attention structure, which aims to explicitly encode periodic position information and capture global temporal dependencies for the historical input SST sequence $X \in \mathbb{R}^{L\times D}$, ultimately generating enhanced features for subsequent use by the Temporal Decomp block.

TEB first processes the input sequence X_{en}^{l-1}using a shared rotary position encoding (RoPE) function [30], a process that is formalized as

$$\tilde{Q} = \mathrm{RoPE}(X_{en/de}^{l-1})W_Q, \quad \tilde{K} = \mathrm{RoPE}(X_{en/de}^{l-1})W_K \tag{2}$$

where $W_Q, W_K \in \mathbb{R}^{d_{\text{model}}\times d_k}$ are learnable projection matrices. The operation $RoPE$ is a vector transformation method: for a given component $x_{t,d}$ of vector x at position t and dimension d, the transformation is defined as $x_{t,d} \mapsto x_{t,d}\cos(\lambda_d t) + x_{t,\pi(d)}\sin(\lambda_d t)$. Here, λ_d is a frequency factor related to the preset period length and $\pi(d)$ is a predefined dimension index mapping function used to achieve rotational interaction between different dimensions. This encoding mechanism ensures that the inner product of the transformed vectors can simultaneously reflect the similarity of the original content and the phase relationships with periodicity.

Subsequently, based on the rotation-encoded sequences $\tilde{Q}$ and $\tilde{K}$, TEB efficiently computes their temporal autocorrelation sequence using the Fast Fourier Transform(FFT) [31], which is formalized as:

$$\hat{X}_{en/de}^{l-1} = F^{-1}\left(F(\tilde{Q}) \cdot F^{*}(\tilde{K})\right) \tag{3}$$

where the symbol F denotes FFT, F^{*} denotes its complex conjugate. The computed autocorrelation sequence $\hat{X}_{en/de}^{l-1}$ reflects the similarity between the sequence and its delayed copies, which is then used to construct attention weights that highlight key periodic patterns within the sequence.

C. Temporal Decomposition Block

The Temporal Decomp block receives the enhanced feature output $\hat{X}_{en/de}^{l-1}$ by the TEB, aiming to decompose the input into components with different temporal characteristics through multi-scale filtering and adaptive fusion, thereby preliminarily separating non-stationary features and focusing on periodic patterns. Temporal Decomp builds on the autocorrelation

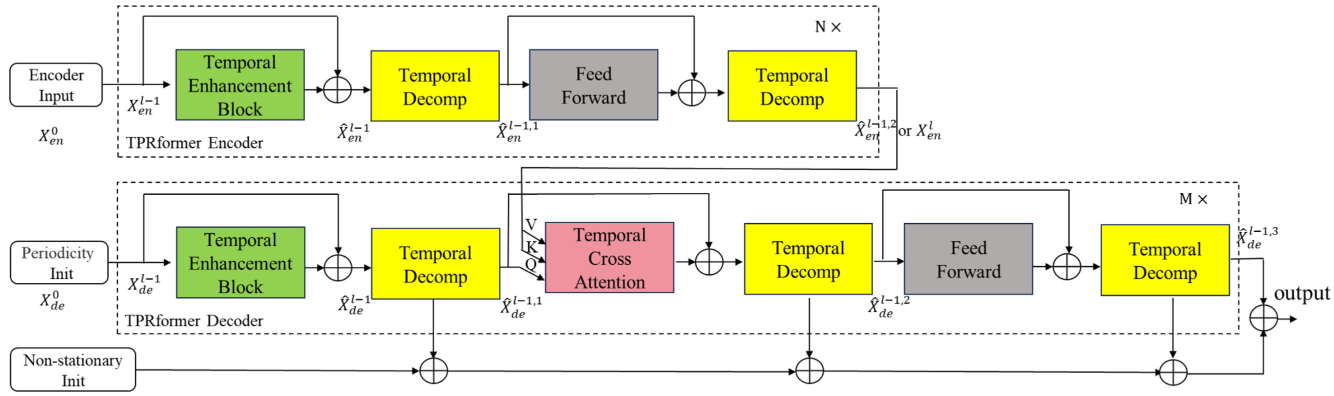

Fig. 1. TPRformer structure. TPRformer consists of N encoders and M decoders. Temporal enhancement blocks (TEB, green blocks) and temporal cross-attention (TCA, pink squares) are used for temporal representation learning. The temporal decomposition block (Temporal Decomp, yellow block) is used to extract seasonal trend patterns from input data.

sequence $\hat{X}^{l-1}_{en/dec}$ computed by the preceding TEB, from which the top k time lags with the highest autocorrelation values are selected to form the key time lag set $\mathcal{T}_{\text{topk}}$.

Subsequently, through cyclic shifting and weighted aggregation operations, the block fuses pattern information from historically similar time periods into the current moment. Its core computational process is formalized as

$$\hat{X}^{l-1,1}_{en/de} = \sum_{\tau\in\mathcal{T}_{\text{topk}}} \text{Roll}(V_T,\tau)\cdot\text{Softmax}(\hat{X}^{l-1,\tau}_{en/de}) \tag{4}$$

where V_T denotes the vector sequence of values associated with the characteristic F_{att}, the symbol $\text{Roll}(V_T,\tau)$ represents the cyclical shift of the sequence V_T along the time axis by steps τ, $\hat{X}^{l-1,\tau}_{en/de}$ is the autocorrelation value corresponding to time lag τ, and the function Softmax normalizes these autocorrelation values to generate weights aggregation. This operation realizes information aggregation based on periodic similarity, which ultimately outputting the initially separated periodic component $\hat{X}^{l-1,1}_{en/de}$.

D. Temporal Cross-attention Block

The Temporal Cross-Attention (TCA) block takes the temporal features of decoder $\hat{X}^{l-1,1}_{de}\in\mathbb{R}^{L\times d_{\text{model}}}$ and the encoder output features $X^{l}_{en}\in\mathbb{R}^{L\times d_{\text{model}}}$ as input. TCA achieves the fusion between decoder features and historical encoder features through a cross-attention mechanism. First, absolute temporal and periodic phase information are respectively injected into the decoder and encoder features using ROPE to generate period-aware query and key representations. Specifically, for the $l-1$-th decoder characteristic $atX^{l-1,1}_{de}$ and encoder features X^{l}_{en}, a shared function RoPE is applied, formalized as follows:

$$\tilde{Q} = \text{RoPE}(\hat{X}^{l-1,1}_{de})W_Q, \quad \tilde{K} = \text{RoPE}(\hat{X}^{l}_{en})W_K \tag{5}$$

where $W_Q, W_K\in\mathbb{R}^{d_{\text{model}}\times d_k}$ are learnable projection matrices. The operation of rotary position encoding RoPE is consistent with that described in the aforementioned TEB, thus embedding periodic phase information in the features.

The value sequence is obtained via a linear transformation: $V = \hat{X}^{l}_{en}W_V$, where $W_V\in\mathbb{R}^{d_{\text{model}}\times d_v}$. Subsequently, the cross-attention score matrix is calculated and dynamic alignment is introduced along with a causal mask. The dynamic alignment layer maps future decoder timesteps to phase reference points within the historical cycle according to the known period length T of the sequence, which is achieved through temporal index offset: for a decoder timestep i, its corresponding historical phase reference point is defined as $\text{ref}(i) = i \mod T$. The mask matrix $M\in\mathbb{R}^{L\times L}$ is constructed accordingly, with its elements given by:

$$M_{i,j} = \begin{cases} 0, & \text{if encoder timestep } j \le \text{ref}(i) \\ -\infty, & \text{otherwise} \end{cases} \tag{6}$$

This mask ensures that each position in the decoder attends only to positions in the encoder that are not later than its historical phase reference point, thereby maintaining prediction causality. The attention weights are calculated as follows:

$$\text{Attention}(\tilde{Q},\tilde{K},V) = \text{Softmax}\left(\frac{\tilde{Q}\tilde{K}^T}{\sqrt{d_k}} + M\right)V \tag{7}$$

where d_k is the dimensionality of the key vectors used for scaling. Finally, the block outputs the fused feature, which integrates periodic pattern information from the historical sequence of the encoder, thereby providing an enhanced contextual representation for subsequent prediction.

V. EXPERIMENTS

A. Datasets

To evaluate the predictive performance of the TPRformer model for sea surface temperature (SST), we employs the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) [32] dataset released by the UK Met Office. The OSTIA dataset generates a daily SST product at a spatial resolution of 0.05° by assimilating in situ observations with satellite infrared and microwave radiometer measurements. This work focuses on two representative oceanic regions: the Yellow and Bohai Sea (30°N–40°N, 120°E–130°E) and the South China Sea (12°N–20°N, 112°E–120°E), as illustrated in Fig. 2, and the original spatial resolution of 0.05° is maintained

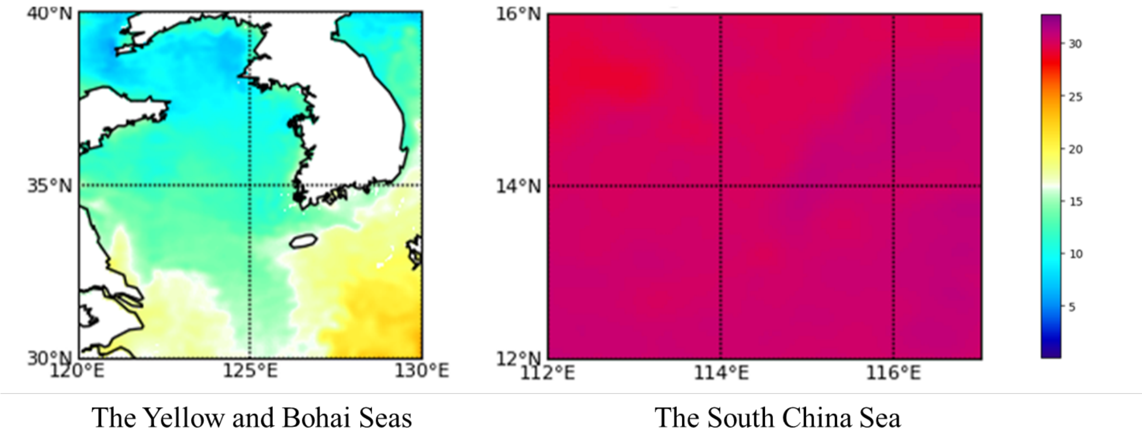

Fig. 2. The two images represent the Bohai Sea and the Yellow Sea and the South China Sea, respectively.

within the study areas. The data spans the period from January 1, 1992, to December 31, 2020, comprising a total of 10,585 daily sea surface temperature sequences.

B. Experimental Setup

All experiments were performed on a single NVIDIA GeForce RTX 4090 GPU using the PyTorch framework. A learning rate of 0.001, a batch size of 4, and 200 training epochs were used, with early stopping implemented with patience of 20. To capture the short-period non-stationary fluctuation characteristics of the SST series, we define a single time step as 1-day observed SST data, generate multiple training samples for model training via the sliding window approach, and set the sliding window length to 10 time steps. And the dataset was partitioned into 80% for training, 10% for testing and 10% for validation in chronological order. The Adam optimizer was used to minimize the mean squared error (MSE) loss function, which is defined as:

$$\text{Loss} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

where y_i denotes the true value, $\hat{y}_i$ represents the predicted value, and n is the number of samples.

The performance of the model was evaluated using the mean absolute error (MAE) and the root mean squared error (RMSE), which are defined as follows.

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

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

To validate the performance advantages of the TPRformer in SST forecasting, we selects two categories of classical baseline models for comparative experiments. The category for multi-scale temporal feature modeling includes CFCC-LSTM [17], Scaleformer [33], and MIMO [24], while the category for multi-scale spatial feature modeling and extraction includes MGCN [34], OVPGCN [35], HIGRN [28], and ST-Unet [29].

To evaluate predictive performance of the model across different time scales, this study establishes three typical forecasting scenarios:

- **Daily**: 7-day history → 1-day prediction
- **Weekly**: 30-day history → 7-day prediction
- **Monthly**: 120-day history → 30-day prediction

C. Experimental Results

As shown in Tables I and II, the proposed TPRformer achieves state-of-the-art performance on both datasets, surpassing all baseline methods.

On the Yellow and Bohai Sea dataset, TPRformer demonstrates superior results across all baselines. Compared to multi-scale temporal feature modeling methods, it improves MAE by 5.5% and RMSE by 6.25% over ScaleFormer, by 17.0% in MAE and 22.4% in RMSE over MIMO, and by 26.0% in MAE and 17.9% in RMSE over CFCC-LSTM. Against multi-scale spatial feature modeling methods, it achieves MAE and RMSE improvements of 8.1% and 10.0% over OVPGCN, 12.8% and 13.4% over ST-Unet, 12.8% and 16.7% over HIGRN, and 19.5% and 25.0% over MGCN.

Similarly, on the South China Sea dataset, TPRformer consistently outperforms all benchmarks. It reduces MAE by 10.3% and RMSE by 5.4% compared to ScaleFormer, by 13.8% in MAE and 25.5% in RMSE compared to MIMO, and by 28.3% in MAE and 31.3% in RMSE compared to CFCC-LSTM. Relative to spatial modeling baselines, it attains MAE and RMSE improvements of 16.1% and 10.2% over OVPGCN, 16.1% and 14.6% over ST-Unet, 21.2% and 20.4% over HIGRN, and 25.7% and 25.5% over MGCN.

The prediction results are visualized to facilitate model performance comparison. Fig. 3 shows the comparison between the predicted value and the measured value of the monthly prediction task in the typical area of the Yellow Sea and the Bohai Sea from January 10 to February 9, 2020. Data for the

TABLE I
PREDICTION RESULT OF SST IN THE YELLOW AND BOHAI SEA

Model	Typical area of Yellow and Bohai Sea					
	Daily		Weekly		Monthly	
	MAE(°C)	RMSE(°C)	MAE(°C)	RMSE(°C)	MAE(°C)	RMSE(°C)
CFCC-LSTM	0.46	0.67	0.60	0.87	0.81	1.18
MGCN	0.42	0.60	0.58	0.83	0.79	1.07
MIMO	0.41	0.58	0.55	0.81	0.72	1.01
HiGRN	0.39	0. 54	0.52	0.76	0.67	0.91
ST-UNet	0.39	0.52	0.48	0.74	0.64	0.86
OVPGCN	0.37	0.50	0.47	0.69	0.59	0.82
Scaleformer	0.36	0.48	0.45	0.63	0.56	0.80
TPRformer(ours)	**0.34**	**0.45**	**0.43**	**0.62**	**0.52**	**0.70**

TABLE II
PREDICTION RESULT OF SST IN SOUTH CHINA SEA

Model	South China Sea					
	Daily		Weekly		Monthly	
	MAE(°C)	RMSE(°C)	MAE(°C)	RMSE(°C)	MAE(°C)	RMSE(°C)
CFCC-LSTM	0.38	0.51	0.53	0.72	0.74	1.01
MGCN	0.35	0.47	0.48	0.69	0.71	0.94
MIMO	0.35	0.46	0.47	0.70	0.69	0.86
HiGRN	0.33	0.44	0.45	0.66	0.63	0.90
ST-UNet	0.31	0.41	0.44	0.64	0.60	0.84
OVPGCN	0.31	0.39	0.43	0.65	0.55	0.78
Scaleformer	0.29	0.37	0.43	0.64	0.52	0.73
TPRformer(ours)	**0.26**	**0.35**	**0.38**	**0.57**	**0.48**	**0.70**

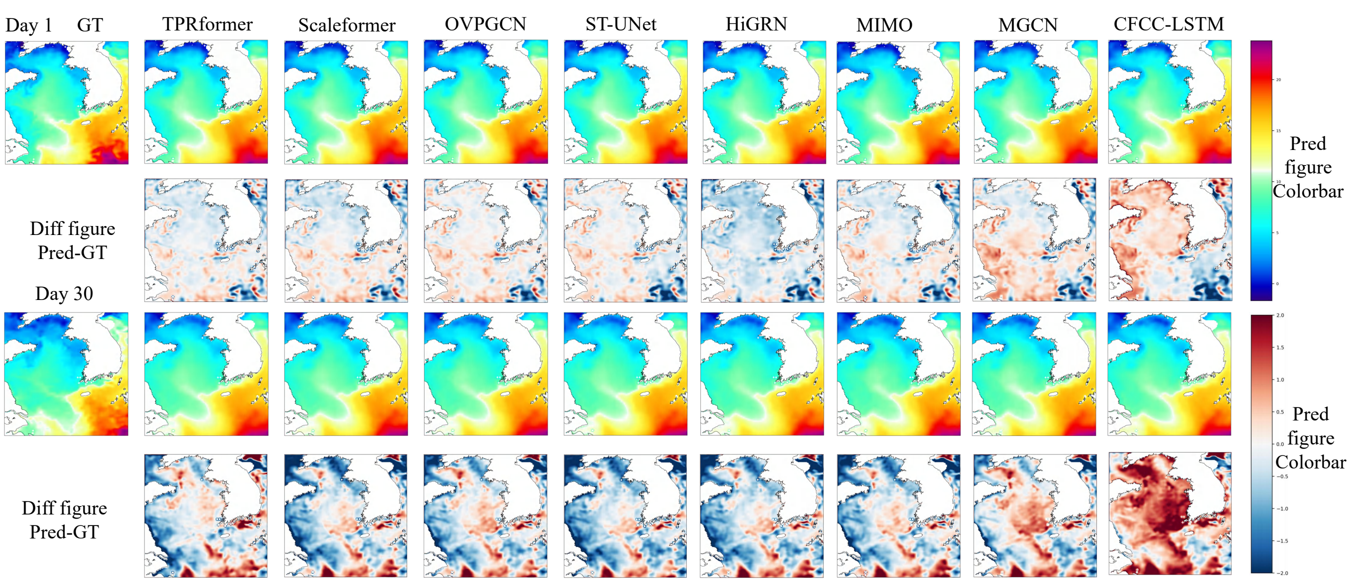

Fig. 3. Visualization of results of different models in monthly prediction

TABLE III
ABLATION EXPERIMENT OF TEB AND TEMPORAL DECOMP BLOCK IN THE DAILY,WEEKLY AND MONTHLY PREDICTION OF THE YELLOW AND BOHAI SEA

Model	MAE		
	Daily	Weekly	Monthly
TPRformer w/o TEB	0.39	0.56	0.66
TPRformer w/o Temporal Decomp	0.37	0.52	0.64
TPRformer w/o secondary Temporal Decomp	0.35	0.47	0.57
TPRformer	**0.34**	**0.43**	**0.52**

first day, the 15th day and the 30th day of the period were selected for comparative analysis. It can be clearly seen that the prediction effect of TPRformer at each time point is better than all baseline methods.

D. Ablation Study

To validate the necessity of each core innovative block in the model, we designed ablation studies by constructing and evaluating multiple model variants in which specific blocks were either removed or replaced.

1) TPRformer w/o the temporal enhancement block (TEB): This variant completely replaces the temporal enhancement block described in the paper—which includes rotary position encoding (RoPE) and temporal autocorrelation computation—with standard sinusoidal position encoding. This aims to verify the necessity of the explicit periodic encoding and global dependency capturing mechanism.

2) 2) TPRformer w/o temporal decomposition (Temporal Decomp): In this variant, the multi-scale filtering and adaptive fusion mechanism in the temporal decomposition block is replaced with a simple moving average. This is designed to validate the effectiveness of the adaptive trend separation method.

3) TPRformer w/o secondary Temporal Decomp: This variant removes the second temporal decomposition and refinement step in the decoder to verify the contribution of this design in eliminating redundant noise from the prediction features.

As shown in Table III, on the Yellow and Bohai Sea dataset, TPRformer achieves the best MAE performance across multiple forecasting time scales.

E. Predictive stability analysis

To further illustrate the performance of the proposed model, a comparison of the daily average MAE between TPRformer and seven baseline models for SST prediction in the typical regions of the Yellow-Bohai Sea and the South China Sea during October 2020 is presented in Fig. 4. A general decline in prediction accuracy is observed as the forecast horizon increases. Although long-term predictions exhibit higher absolute errors, the error distribution of TPRformer is tighter, indicating greater stability in long-term forecasting when the proposed model is employed.

VI. CONCLUSION

We propose TPRformer, which is the SST prediction model based on temporal decomposition and reconstruction. The encoder of TPRformer incorporates the temporal enhancement block to supplement positional information and capture global temporal dependencies, while the temporal decomposition block extracts long-term periodic dependencies and constructs a structured periodic feature library. The decoder first applies temporal enhancement and decomposition to the target SST sequence to obtain initial features, then utilizes a temporal cross-attention module to perform cross-computation between the historical periodic features and the current prediction state for feature fusion. Finally, feature reconstruction is performed to accomplish SST prediction, addressing the challenges faced by existing methods in capturing the complex temporal dependencies of SST. Experimental results demonstrate that the TPRformer model outperforms current state-of-the-art baseline methods. This research is expected to provide robust support for more accurate SST forecasting, with significant implications for various application domains reliant on SST data, such as disaster prevention and marine ecosystem conservation.

REFERENCES

[1] Johnstone, James A., and Nathan J. Mantua. ”Atmospheric controls on northeast Pacific temperature variability and change, 1900–2012.” Proceedings of the National Academy of Sciences, vol. 111, no. 40, pp. 14360-14365, 2014.

[2] S. J. Sohn, C. Y. Tam, and H. I. Jeong, “How do the strength and type of ENSO affect SST predictability in coupled models,” Scientific Reports, vol. 6, no. 1, p. 33790, 2016.

[3] J. A. Brotzge, D. Berchoff, D. L. Carlis, F. H. Carr, R. Hogan, J. J. Gerth, B. D. Gross, T. M. Hamill, S. E. Haupt, N. Jacobs, et al., “Challenges and opportunities in numerical weather prediction,” Bulletin of the American Meteorological Society, vol. 104, no. 3, p. E698–E705, 2023.

[4] Y. Kushnir, A. A. Scaife, R. Arritt, G. Balsamo, G. Boer, F. Doblas-Reyes, E. Hawkins, M. Kimoto, R. K. Kolli, A. Kumar, et al., “Towards operational predictions of the near-term climate,” Nature Climate Change, vol. 9, no. 2, p. 94–101, 2019.

[5] Y. Meng, F. Gao, E. Rigall, R. Dong, J. Dong, and Q. Du, “Physical knowledge-enhanced deep neural network for sea surface temperature prediction,” IEEE Transactions on Geoscience and Remote Sensing, vol. 61, p. 1–13, 2023.

[6] A. F. Shchepetkin and J. C. McWilliams, “The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model,” Ocean modelling, vol. 9, no. 4, p. 347–404, 2005.

[7] P. Costa, B. Gómez, A. Venâncio, E. Pérez, and V. Pérez-Muñuzuri, “Using the Regional Ocean Modelling System (ROMS) to improve the sea surface temperature predictions of the MERCATOR Ocean System,” Scientia Marina, vol. 76, no. S1, p. 165–175, 2012.

[8] C. C. Mutai, M. N. Ward, and A. W. Colman, “Towards the prediction of the East Africa short rains based on sea-surface temperature–atmosphere coupling,” International Journal of Climatology: A Journal of the Royal Meteorological Society, vol. 18, no. 9, p. 975–997, 1998.

[9] A. J. Miller, M. Collins, S. Gualdi, T. G. Jensen, V. Misra, L. P. Pezzi, D. W. Pierce, D. Putrasahan, H. Seo, and Y. H. Tseng, “Coupled ocean–atmosphere modeling and predictions,” 2017.

[10] S. B. Mahongo and M. C. Deo, “Using artificial neural networks to forecast monthly and seasonal sea surface temperature anomalies in the Western Indian Ocean,” The International Journal of Ocean and Climate Systems, vol. 4, no. 2, p. 133–150, 2013.

[11] I. D. Lins, M. Araujo, M. das Chagas Moura, M. A. Silva, and E. L. Droguett, “Prediction of sea surface temperature in the tropical Atlantic by support vector machines,” Computational Statistics Data Analysis, vol. 61, p. 187–198, 2013.

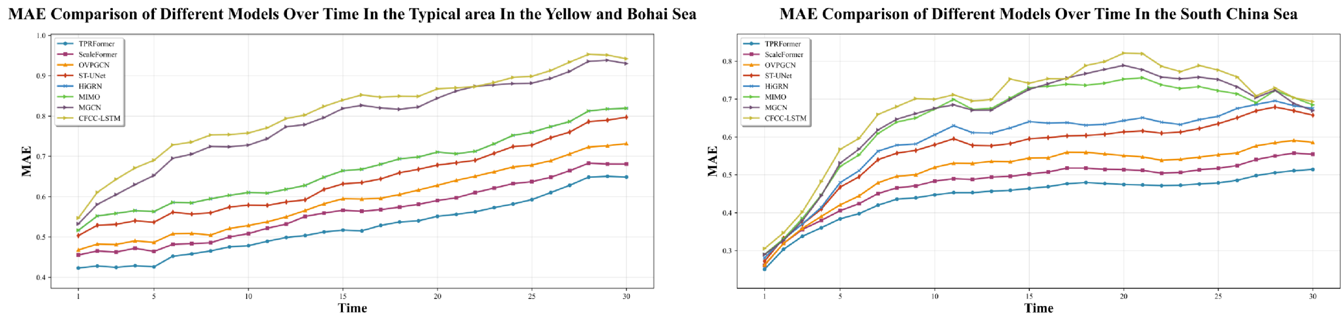

Fig. 4. The SST prediction error of each model in the typical area of the Yellow Sea and Bohai Sea and the South China Sea is analyzed.

[12] S. Kartal, "Assessment of the spatiotemporal prediction capabilities of machine learning algorithms on sea surface temperature data: A comprehensive study," Engineering Applications of Artificial Intelligence, vol. 118, p. 105675, 2023.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, p. 1735–1780, 1997.

[14] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, , , 2014.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, , 2017.

[16] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, p. 1745–1749, 2017.

[17] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, "A CFCC-LSTM model for sea surface temperature prediction," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 2, p. 207–211, 2017.

[18] G. Zhang, X. Kang, Y. Luo, Q. Wang, H. Song, and X. Yin, "A transformer-based method for correcting daily SST numerical forecasting products," Frontiers in Earth Science, vol. 13, p. 1530475, 2025.

[19] X. Xu, C. Chen, Y. Liang, B. Huang, G. Bai, L. Zhao, and K. Shu, "SST: Multi-Scale Hybrid Mamba-Transformer Experts for Time Series Forecasting," Proceedings of the 34th ACM International Conference on Information and Knowledge Management, vol. , p. 3655–3665, 2025.

[20] B. Zhenxiang, Z. Sun, B. Fan, A.-A. Liu, Z. Wei, and B. Yin, "Multiscale spatio-temporal attention network for sea surface temperature prediction," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, p. 5866–5877, 2025.

[21] Z. Ma, H. Zhang, and J. Liu, "MS-LSTM: Exploring spatiotemporal multiscale representations in video prediction domain," Applied Soft Computing, vol. 147, p. 110731, 2023.

[22] S. Li, Q. Li, Z. Du, Z. Zhou, M. Wang, L. Zhang, and W. Li, "Heterogeneous Spatiotemporal Graph Learning for Localized Sparse Meteorological Forecasting," IEEE Transactions on Geoscience and Remote Sensing, , vol. 63, p. 1–18, 2025.

[23] H. Dai, Z. He, G. Wei, F. Lei, X. Zhang, W. Zhang, and S. Shang, "Long-term prediction of sea surface temperature by temporal embedding transformer with attention distilling and partial stacked connection," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, p. 4280–4293, 2024.

[24] S. Hou, W. Li, T. Liu, S. Zhou, J. Guan, R. Qin, and Z. Wang, "MIMO: A unified spatio-temporal model for multi-scale sea surface temperature prediction," Remote Sensing, vol. 14, no. 10, p. 2371, 2022.

[25] A. Agarwal, L. Caesar, N. Marwan, R. Maheswaran, B. Merz, and J. Jürgen, "Network-based identification and characterization of teleconnections on different scales," Scientific Reports, vol. 9, no. 1, p. 8808, 2019.

[26] Y. Sun, X. Yao, X. Bi, X. Huang, X. Zhao, and B. Qiao, "Time-series graph network for sea surface temperature prediction," Big Data Research, vol. 25, p. 100237, 2021.

[27] S. Liang, A. Zhao, M. Qin, L. Hu, S. Wu, Z. Du, and R. Liu, "A graph memory neural network for sea surface temperature prediction," Remote Sensing, vol. 15, no. 14, p. 3539, 2023.

[28] H. Yang, W. Li, S. Hou, J. Guan, and S. Zhou, "HiGRN: a hierarchical graph recurrent network for global sea surface temperature prediction," ACM Transactions on Intelligent Systems and Technology, vol. 14, no. 4, p. 1–19, 2023.

[29] J. Ren, C. Wang, L. Sun, B. Huang, D. Zhang, J. Mu, and J. Wu, "Prediction of sea surface temperature using U-Net based model," Remote Sensing, vol. 16, no. 7, p. 1205, 2024.

[30] J. Su, M. Ahmed, Y. Lu, S. Pan, W. Bo, and Y. Liu, "RoFormer: Enhanced transformer with rotary position embedding," Neurocomputing, vol. 568, p. 127063, 2024.

[31] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," International Conference on Machine Learning, p. 27268–27286, 2022.

[32] M. Worsfold, S. Good, C. Atkinson, and O. Embury, "Presenting a long-term, reprocessed dataset of global sea surface temperature produced using the OSTIA system," Remote Sensing, vol. 16, no. 18, p. 3358, 2024.

[33] A. Shabani, A. Abdi, L. Meng, and T. Sylvain, "Scaleformer: Iterative multi-scale refining transformers for time series forecasting," arXiv preprint arXiv:2206.04038, 2022.

[34] X. Zhang, Y. Li, A. C. Frery, and P. Ren, "Sea surface temperature prediction with memory graph convolutional networks," IEEE Geoscience and Remote Sensing Letters, vol. 19, p. 1–5, 2021.

[35] M. Ye, B. Li, J. Nie, Q. Wen, Z. Wei, and L.-L. Yang, "Graph convolutional network-assisted SST and Chl-a prediction with multicharacteristics modeling of spatio-temporal evolution," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, p. 1–14, 2023.